IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 April 2023, accepted 5 May 2023, date of publication 9 May 2023, date of current version 15 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274691

RESEARCH ARTICLE

Cybersecurity-Enhanced Encrypted Control System Using Keyed-Homomorphic Public Key Encryption

MASAKI MIYAMOTO[1], (Graduate Student Member, IEEE), KAORU TERANISHI[1,2], (Graduate Student Member, IEEE), KEITA EMURA[3], AND KIMINAO KOGISO[1], (Member, IEEE)

[1]Department of Mechanical and Intelligent Systems Engineering, The University of Electro-Communications (UEC), Chofu, Tokyo 1828585, Japan
[2]Japan Society for the Promotion of Science (JSPS), Chiyoda-ku, Tokyo 1020083, Japan
[3]Cybersecurity Research Institute, National Institute of Information and Communications Technology (NICT), Koganei, Tokyo 1848795, Japan

Corresponding author: Kiminao Kogiso (kogiso@uec.ac.jp)

This work was supported by the Japan Society for the Promotion of Science KAKENHI under Grant JP22H01509 and Grant JP21K11897.

ABSTRACT Encrypted control systems are secure control methods that use the cryptographic properties of a specific homomorphic encryption scheme. This study proposes a cyberattack-detectable encrypted control system and validates its effectiveness using a proportional integration derivative (PID) position-control system for an industrial motor. The proposed encrypted control system uses a keyed-homomorphic public-key encryption scheme for real-time detection of cyberattacks, such as signal and control parameter falsification. Additionally, a novel quantizer is presented to reduce the computation cost and quantization-error effects on control performance. The quantizer demonstrated a significant improvement, reducing the computation time by 47.3 % compared to using our previous quantizer, and decreasing the quantization-error effect by 30.6 % compared to a widely-used gain-multiplying quantizer. Moreover, this study establishes conditions through a theorem to avoid an overflow in the proposed control system. Experimental validation confirms that the proposed control system effectively conceals the control operation, and the presented theorem aids in designing the quantization gains to prevent overflows. Notably, the results of falsification attack tests highlight that the proposed control system enables real-time detection of attacked components within control parameters and signals, representing a significant advantage of this study.

INDEX TERMS Cybersecurity, encrypted control, keyed-homomorphic public key encryption, quantization, experimental validation.

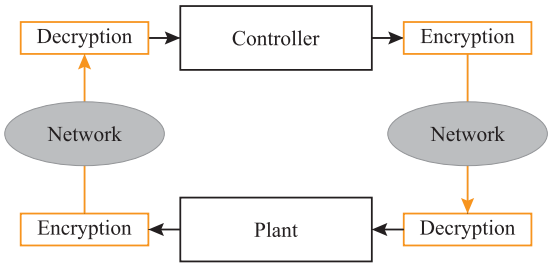
I. INTRODUCTION

Cybersecurity is important in networked control systems. Networked control systems are connected to information networks used in factory automation and power grids for supporting modern life. However, being connected to an information network entails the risk of a cyberattack on the control system. Furthermore, unlike in the case of conventional information technology systems, cyberattacks on control systems can cause physical damage. Stuxnet destroyed centrifuges at an Iranian nuclear facility [1], and Industroyer caused massive power outages in Ukraine [2].

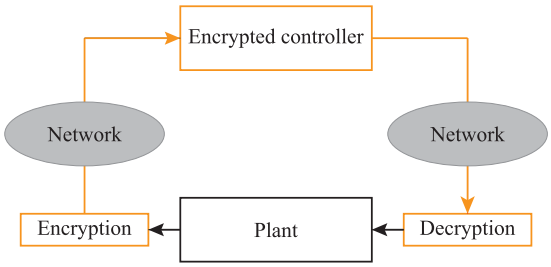
The associate editor coordinating the review of this manuscript and approving it for publication was Hosam El-Ocla.

Various techniques that are used to attack control systems and several main attacks [3], [4], [5], [6] are classified based on the impact and adversary's knowledge of the control system [7]. Eavesdropping attacks are the easiest to execute because they do not require any model knowledge of the target control system, but they lead to more sophisticated attacks [8]. However, cryptography can prevent eavesdropping attacks; thus, it increases the security of control systems.

From both control-theoretic and cryptographic viewpoints, a multidisciplinary method can develop a cyber-secure automatic control technology. Homomorphic encryption (HE) [9] enables arithmetic operations on encrypted data. The method of incorporating homomorphic encryption into a control system is known as encrypted control [10], [11],



(a) Traditional (unencrypted) control system.



(b) Encrypted control system.

FIGURE 1. Conceptual configuration of encrypted control systems.

[12], [13]. As shown in Fig. 1(a), networked control systems communicate control-system signals over a network. For cloud-based control systems, which have attracted attention in recent years, computations of the controller are performed in the cloud. In contrast, as shown in Fig. 1(b), an encrypted control system directly determines the control inputs in ciphertext without decrypting the signals. Hence, it has attracted considerable attention because it can reduce the risk of raw data leakage even if an attacker accesses the control network through cyberspace.

A. OUR CONTRIBUTIONS

The objective of this study is to propose an encrypted control system based on a keyed-homomorphic public-key encryption (KH-PKE) scheme that enables the detection of the malleability-based falsification of signals and control parameters. To enhance security, this study considers controller encryption using the KH-PKE scheme [14]. The falsification attack detection is inherited from the feature that allows the decryption and evaluation algorithms of the underlying KH-PKE scheme to output an error symbol when tampering occurs. The proposed encrypted control system benefits from the feature that enables the identification of attacked components within signals and control parameters, which constitutes a significant advantage over the conventional studies [10], [11], [12], [13], [15], [16], [17], [18], [19], [20], [21]. Moreover, a novel efficient quantizer is presented for constructing the encrypted control system, which is developed by modifying a conventional quantizer [16]. The efficient quantizer introduced in this study significantly improves the performance by reducing computation time, as compared to conventional quantizers [16]. Furthermore, this study investigates the conditions for avoiding overflows caused by quantization processes, which is summarized in a theorem. In addition, an experimental validation was conducted to confirm the effectiveness of the proposed encrypted control system using an industrial linear stage. First, this study evaluates three encrypted control systems with a modified quantizer and two conventional quantizers [10], [16] in terms of computation time and quantization-error effects on control performance, and the encryption scheme used is KH-PKE and common. Subsequently, this study verifies the theorem that provides a design policy for quantization gains, compared with the situation where quantization gains do not hold for the theorem. The final validation demonstrates that the proposed encrypted control system enables the real-time detection of falsification attacks. These experimental results indicate that the proposed encrypted control system is adequate and appropriate for developing secure control technologies.

The contributions of this study are threefold. i) This study developed a more secure encrypted control system that enables the real-time detection of cyberattacks compared with conventional encrypted control systems. This implies that there is expertise and knowledge in cryptography that helps enhance the security of control systems. ii) The developed linear-stage control system is a secure automatic control technology, and the proposed method can be implemented and can run in real-time with a certain key length. iii) This study provides the possibility and relevance of appropriate security concepts for real-time control systems in terms of provable and computational security. The findings of this study will lead to the creation of a new fusion area for cryptography and control engineering.

B. ORGANIZATION OF THE PAPER

The remainder of this paper is organized as follows: Section II introduces the notations and syntax of the KH-PKE scheme. Section III presents the proposed encrypted control system that involves a novel quantizer. Section IV introduces a practical testbed control system and determines the parameters for implementing an encrypted controller. Section V validates the proposed encrypted control system in terms of the control performance effect, overflow avoidance, and cyberattack detection. Section VI discusses the security of the proposed encrypted control system. Finally, Section VII concludes the paper.

C. RELATED WORK

Encrypted control systems using Paillier encryption [22], which is an additive homomorphic encryption (AHE) that enables the addition of encrypted data, have been studied [11], [17], [18], [19]. Encrypted control systems using ElGamal encryption [23], which is a multiplicative homomorphic encryption (MHE) that enables the multiplication of encrypted data, have also been studied [10], [15], [16]. Furthermore, encrypted control systems using fully homomorphic encryption (FHE) [24], which can perform both addition and multiplication, have been proposed [12]. Some recent studies have considered encrypted control systems using AHE or leveled FHE based on learning with errors [20], [21]. Only the signal of the control system is encrypted when using AHE, whereas MHE and FHE enable the encryption of signals and controller parameters.

Countermeasures against cyberattacks such as tampering, are necessary to secure control systems. Although eavesdropping attacks can be prevented, the direct manipulation of signals or controller parameters enables the degradation of control performance or compromises it to break in the worst case. Encrypted control systems are vulnerable to attacks that use the malleability of the homomorphic encryption scheme [25], which means an attacker can manipulate encrypted signals and parameters to adjust controlled outputs without decrypting them. To reduce the vulnerability to attacks, there are related studies that consider countermeasures such as homomorphic authentication [26], obfuscation of controller parameters [25], and cancellation and detection by a modified somewhat homomorphic encryption [27], [28], which uses the malleability of a homomorphic encryption scheme. However, these studies could hardly identify an attacked component within signals and control parameters. The enhancement of cyberattack detection motivated us to develop a secure control technology for a quick response to cyberattack incidents.

The concept of KH-PKE, as proposed in [14] and [29], introduces another private key specifically dedicated to performing homomorphic operations. This approach aims to achieve indistinguishability under an adaptive chosen ciphertext attack (IND-CCA2) against adversaries who do not possess a homomorphic operation key. The IND-CCA2 security property enables the detection of attacks based on malleability, and various configurations of KH-PKE have been proposed in [14], [29], [30], [31], [32], [33], and [34]. Furthermore, a study has also proposed a KH-PKE scheme that supports multiplicative homomorphic operations [14]. Notably, the KH-PKE scheme is secure under the decisional Diffie-Hellman (DDH) assumption, which is commonly used to prove the security of ElGamal encryption. As a result, the use of DDH-based KH-PKE scheme is expected to enhance the security of encrypted control systems, making them more resillient against potentail cyberattacks.

II. PRELIMINARIES

This section provides notations of variables and functions and introduces the KH-PKE as preliminaries for constructing encrypted control systems.

A. NOTATIONS

Sets of real numbers, integers, plaintext spaces, and ciphertext spaces are denoted by $\mathbb{R}$, $\mathbb{Z}$, $\mathcal{M}$, $\mathcal{C}$, respectively. We define $\mathbb{R}^+ := \{x \in \mathbb{R} \mid 0 < x\}$, $\mathbb{Z}^+ := \{z \in \mathbb{Z} \mid 0 < z\}$, $\mathbb{Z}_n := \{z \in \mathbb{Z} \mid 0 \leq z < n\}$, $\mathbb{Z}_n^+ := \{z \in \mathbb{Z} \mid 0 < z < n\}$, and $\mathfrak{P}_a^b := \{a^i \bmod b \mid i \in \mathbb{Z}_b\}$. A set of vectors of size n is denoted by $\mathbb{R}^n$. The jth element of vector v is denoted by v_j. ℓ_2 norm and infinity norm v are denoted by $||v||$ and $||v||_\infty$, respectively. The set of matrices of size $m \times n$ is denoted by $\mathbb{R}^{m\times n}$. (i, j) entry of matrix M is denoted by M_{ij}. The induced 2-norm and maximum norm of M are denoted by $||M||$ and $||M||_{\max}$, respectively. The greatest common divisor of the two positive integers $a, b \in \mathbb{Z}^+$ is denoted by $\gcd(a, b)$.

Definition 2.1: The minimal residue of integer $a \in \mathbb{Z}$ modulo $m \in \mathbb{Z}^+$ is defined as

$$a \operatorname{Mod} m := \begin{cases} b & \text{if } b < |b - m|, \\ b - m & \text{otherwise,} \end{cases}$$

where $b = a \bmod m$. For example, let $m = 10$, $a_1 = 3$, and $a_2 = 7$, then $a_1 \operatorname{Mod} m = 3$ and $a_2 \operatorname{Mod} m = -3$, where $a_1 \bmod m = 3$ and $a_2 \bmod m = 7$.

Definition 2.2: Let p be an odd prime number and z be an integer satisfying $\gcd(z, p) = 1$. If there exists integer b such that $b^2 = z \bmod p$, then integer z is the quadratic residue of modulo p. If integer b does not exist, then integer z is a quadratic nonresidue of modulo p. This can be expressed using the Legendre symbol $(\cdot/\cdot)_L$ as follows:

$$\left(\frac{z}{p}\right)_L := z^{\frac{p-1}{2}} \operatorname{Mod} p = \begin{cases} 1 & \text{if } z \text{ is a quadratic residue,} \\ -1 & \text{if } z \text{ is a quadratic nonresidue.} \end{cases}$$

Definition 2.3: The rounding function $\lceil \cdot \rfloor$ of $\sigma \in \mathbb{R}^+$ to the nearest positive integer is defined as

$$\lceil \sigma \rfloor = \begin{cases} \lfloor \sigma + 0.5 \rfloor & \text{if } \sigma \geq 0.5, \\ 1 & \text{otherwise,} \end{cases}$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

B. KEYED-HOMOMORPHIC PUBLIC KEY ENCRYPTION

The syntax of KH-PKE for homomorphic operations [14] is introduced as follows:

Definition 2.4 (KH-PKE): Let $\odot$ be a binary operation over $\mathcal{M}$. The KH-PKE scheme $\mathcal{E} = (\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec}, \mathbf{Eval})$ for homomorphic operation $\odot$ consists of the following four algorithms:

- **Gen**: This key-generation algorithm takes a security parameter $\kappa \in \mathbb{R}^+$ as the input and returns public key $\mathbf{pk}$, private key $\mathbf{sk_d}$, and homomorphic operation key $\mathbf{sk_h}$;
- **Enc**: This encryption algorithm takes $\mathbf{pk}$ and plaintext $m \in \mathcal{M}$ as the input and returns ciphertext $c \in \mathcal{C}$;
- **Dec**: This decryption algorithm takes $\mathbf{sk_d}$ and c as inputs and returns m or $\perp$.
- **Eval**: This evaluation algorithm takes $\mathbf{sk_h}$, two ciphertexts c_1 and c_2 as inputs and returns ciphertexts c or $\perp$,

where $\perp$ denotes the error symbol.

Definition 2.5 (Correctness): A KH-PKE scheme for homomorphic operation $\odot$ is correct if, for all $(\mathbf{pk}, \mathbf{sk_d}, \mathbf{sk_h}) \leftarrow \mathbf{Gen}(\mathbf{1}^\kappa)$, the following two conditions are satisfied:

1) For all $m \in \mathcal{M}$ and $c \in \mathcal{C}_{\mathbf{pk},m}$, it holds that $\mathbf{Dec}(\mathbf{sk_d}, c) = m$;
2) For all $m_1, m_2 \in \mathcal{M}$, $c_1 \in \mathcal{C}_{\mathbf{pk},m_1}$, and $c_2 \in \mathcal{C}_{\mathbf{pk},m_2}$, it holds that $\mathbf{Eval}(\mathbf{sk_h}, c_1, c_2) \in \mathcal{C}_{\mathbf{pk},m_1 \odot m_2}$,

where $\mathcal{C}_{\mathbf{pk},m}$ denotes the set of all ciphertexts of $m \in \mathcal{M}$ under the public key $\mathbf{pk}$. For simplicity, the arguments $\mathbf{pk}$, $\mathbf{sk_d}$, and $\mathbf{sk_h}$ will be omitted henceforth.

In this study, the multiplicative KH-PKE scheme proposed in [14] is used to encrypt communication signals and

control parameters. This security is provided by the DDH assumption. In the case of multiplicative DDH-based KH-PKE, $\odot$ is replaced with a multiplicative homomorphic operation, and the plaintext space is a multiplicative cyclic group, defined as $\mathbb{G} := \{g^i \bmod p \mid i \in \mathbb{Z}_q\}$ such that $g^q \bmod p = 1$ and $p-1 \bmod q = 0$ with generator g of cyclic group $\mathbb{G}$, which is a set of positive integers with discrete values. The four algorithms can be specified as **Appendix A**. Moreover, the DDH-based KH-PKE scheme enables the error symbol to be returned when processing ill-formed ciphertexts in the **Dec** and **Eval** algorithms. Therefore, the DDH-based KH-PKE scheme helps us realize encrypted control systems with real-time detection of tampered communication signals and/or control parameters.

Remark 2.6: The original definition of the KH-PKE scheme states that if tampering occurs, then the error symbol is output to terminate the algorithm. However, because control systems require availability, this study modifies the algorithm such that it outputs the corresponding signals and never terminates them even if tampering occurs.

III. ENCRYPTED CONTROL SYSTEM

This section presents an appropriate quantizer for the proposed encrypted control systems, introduces the controller encryption technique, and explains the quantizer design policy to avoid overflows.

A. QUANTIZER

A quantizer is required to construct the encryption control system because the plaintexts and ciphertexts in the encryption scheme are integers, and the processes at the controller are reconstructed using the encryption scheme. Hence, this study presents a novel quantizer that maps $x \in \mathbb{R}$ onto $\bar{x} := (\bar{x}^1, \bar{x}^2) \in \mathbb{G}^2$, an encoding map $\mathsf{Ecd}_\gamma := \mathscr{C} \circ \mathscr{A}_\gamma$, and a decoding map $\mathsf{Dcd}_\gamma := \mathscr{B}_\gamma \circ \mathscr{D}$ with

$$
\begin{aligned}
\mathscr{A}_\gamma &: \mathbb{R} \to \mathfrak{P}_2^q \times \mathbb{Z}_q^+, \\
&: x \mapsto \begin{cases} (1, \lceil \gamma |x| \rfloor \bmod q) & \text{if } x \geq 0, \\ (2, \lceil \gamma |x| \rfloor \bmod q) & \text{if } x < 0, \end{cases} \\
\mathscr{B}_\gamma &: \mathfrak{P}_2^q \times \mathbb{Z}_q^+ \to \mathbb{R}, \\
&: (\zeta, z) \mapsto \left(\frac{\zeta}{3}\right)_L \frac{z}{\gamma} := \check{x}, \\
\mathscr{C} &: \mathfrak{P}_2^q \times \mathbb{Z}_q^+ \to \mathbb{G}^2, \\
&: (\zeta, z) \mapsto \left(\left(\frac{\zeta}{p}\right)_L \zeta, \left(\frac{z}{p}\right)_L z\right) \bmod p := (\bar{x}^1, \bar{x}^2), \\
\mathscr{D} &: \mathbb{G}^2 \to \mathfrak{P}_2^q \times \mathbb{Z}_q^+, \\
&: (\bar{x}^1, \bar{x}^2) \mapsto (|\bar{x}^1 \operatorname{Mod} p|, |\bar{x}^2 \operatorname{Mod} p|),
\end{aligned}
$$

where $\gamma \in \mathbb{R}^+$ is the quantization gain, $\zeta \in \{1, 2\}$, and $z := \lceil \gamma |x| \rfloor \bmod q$. The relationship between the maps is shown in Fig. 2.

The presented quantizer comprising Ecd_γ and Dec_γ is a modified version of the conventional quantizer [16] and has the advantage of reducing computation time and resource consumption compared to the conventional quantizer. The conventional quantizer uses plaintext space $\mathbb{G}^3$ that assigns

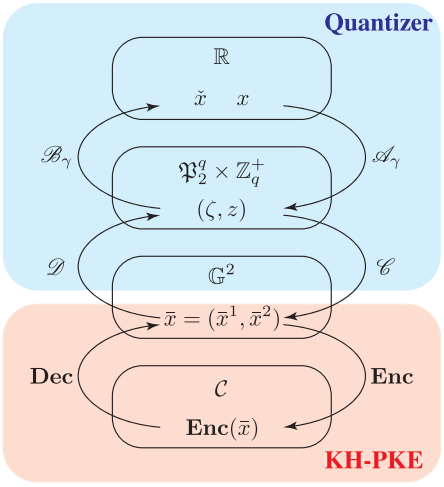


FIGURE 2. Relationship of the maps between the real and ciphertext spaces, realized by the presented $\mathbb{G}^2$-based quantizer.

$\mathbb{G}$ to a zero component. The assignment is inefficient in computing; therefore, the presented quantizer removes the zero component to obtain the plaintext space $\mathbb{G}^2$. In this study, the conventional quantizer is called a $\mathbb{G}^3$-based quantizer. In addition, the effects of the presented quantizer on the computation time are presented in Section IV-B. The proposed encrypted control system requires the plaintext space $\mathbb{G}^2$; therefore, the KH-PKE scheme must be conducted twice to encrypt data, as indicated in Fig. 2.

Remark 3.1: IND-CCA2 security holds against a KH-PKE ciphertext, and it does not imply IND-CCA2 security against two ciphertexts in a strict manner. In other words, the ciphertexts of $\bar{x}^1$ and $\bar{x}^2$ are non-malleable. However, if we consider that $(\mathbf{Enc}(\bar{x}^1), \mathbf{Enc}(\bar{x}^2))$ is a ciphertext, then it is malleable; for example, one can replace a component $\mathbf{Enc}(\bar{x}^1)$ (or $\mathbf{Enc}(\bar{x}^2)$) with other KH-PKE ciphertexts. In our system, this study does not consider such a replacement as tampering but considers element-wise tampering.

B. CONTROLLER ENCRYPTION

Let us consider a linear controller in a discrete-time state-space representation $f : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n \times \mathbb{R}^m$,

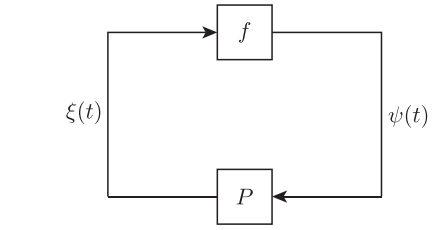
$$
f : \begin{cases} x_c(t+1) = A_c x_c(t) + B_c v_c(t), \\ u(t) = C_c x_c(t) + D_c v_c(t), \end{cases} \tag{1}
$$

where $t \in \mathbb{Z}^+$ is the time step, $u \in \mathbb{R}^m$ is the control input, $v_c \in \mathbb{R}^l$ is the measured output, $x_c \in \mathbb{R}^n$ is a controller state, and A_c, B_c, C_c, and D_c are controller parameters. Controller (1) can be rewritten as follows:

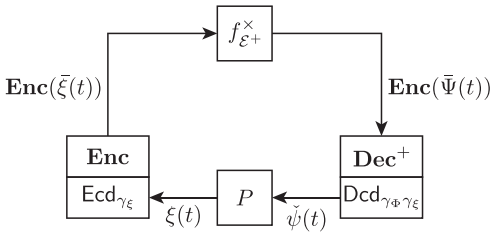
$$
\psi(t) = \Phi \xi(t) =: f(\Phi, \xi(t)), \tag{2}
$$

where $\Phi \in \mathbb{R}^{\alpha \times \beta}$ denotes a matrix that collects the control parameters, and $\psi \in \mathbb{R}^\alpha$ and $\xi \in \mathbb{R}^\beta$ denote a vector gathering the arguments and computed variables in the controller, respectively, which are written as follows:

$$
\Phi := \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix}, \psi(t) := \begin{bmatrix} x_c(t+1) \\ u(t) \end{bmatrix}, \xi(t) := \begin{bmatrix} x_c(t) \\ v_c(t) \end{bmatrix}, \tag{3}
$$



(a) Unencrypted control system.



(b) Encrypted control system using the KH-PKE scheme.

FIGURE 3. Block diagrams of feedback control systems before and after the controller encryption.

where $\alpha = n+m$ and $\beta = n+l$. The control system is shown in Fig. 3(a).

Because f is a composition product of multiplication $f^{\times}$ and addition f^{+}, the decryption algorithm is modified to yield $\mathbf{Dec}^{+} = f^{+} \circ \mathbf{Dec}$ [10]. Using the modified homomorphic encryption scheme $\mathcal{E}^{+} = (\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec}^{+}, \mathbf{Eval})$, the linear controller (1) can be encrypted, that is, it can be reconstructed into a different representation consisting of only encrypted parameters and signals. Then, the resulting encrypted controller when using $\mathcal{E}^{+}$ forms, $\forall t \in \mathbb{Z}^{+}$,

$$f^{\times}_{\mathcal{E}^+} : (\mathbf{Enc}(\bar{\Phi}), \mathbf{Enc}(\bar{\xi}(t))) \mapsto \mathbf{Enc}(\bar{\Psi}(t)), \tag{4}$$

where $\bar{\Phi} = \mathsf{Ecd}_{\gamma_\Phi}(\Phi)$, $\bar{\xi} = \mathsf{Ecd}_{\gamma_\xi}(\xi)$, $\bar{\Psi} = \mathsf{Ecd}_{\gamma_\Phi\gamma_\xi}(f^{\times}(\Phi, \xi))$, γ_Φ and γ_ξ are quantization gains regarding Φ and ξ, respectively, and $\mathbf{Enc}(\bar{\Psi}(t))$ is calculated using **Eval** as follows, $\forall t \in \mathbb{Z}^{+}$,

$$\mathbf{Enc}(\bar{\Psi}^{\theta}_{ij}(t)) = \mathbf{Eval}(\mathbf{Enc}(\bar{\Phi}^{\theta}_{ij}), \mathbf{Enc}(\bar{\xi}^{\theta}_{j}(t)), \\ \forall \theta \in \{1, 2\}, \forall i \in \mathbb{Z}^{+}_{\alpha+1}, \forall j \in \mathbb{Z}^{+}_{\beta+1}.$$

The process (4) is a ciphertext version of (1); therefore, function f running in the controller is replaced by $f^{\times}_{\mathcal{E}^+}$, and the controller output $\mathbf{Enc}(\bar{\Psi}(t))$ is decrypted and decoded at the plant side to extract control input u via a signal $\check{\psi} = \mathsf{Dcd}_{\gamma_\Phi\gamma_\xi}(\mathbf{Dec}^{+}(\mathbf{Enc}(\bar{\Psi}(t))))$. The resulting encrypted control system is illustrated in Fig. 3(b).

The merits of using the KH-PKE scheme are as follows. The proposed encrypted control system can operate using encrypted Φ, ξ, and ψ. An index of vectors or matrices falsified by attackers can be identified because **Eval** and **Dec** are performed element-wise. Thus, controller encryption can protect the controller device and communication from cyberattacks, such as eavesdropping, and can also detect the falsification of signals and control parameters.

C. DESIGN POLICY OF QUANTIZATION GAIN

The plaintext space is finite; thus, overflows may occur with inappropriate gains. This study then provides the design policy of quantization gain as a theorem.

Definition 3.2: An overflow occurs when quantizing Φ, ξ, and a homomorphic operation if $\lceil \gamma_\Phi |\Phi_{ij}| \rfloor \geq q$, $\lceil \gamma_\xi |\xi_j| \rfloor \geq q$, and $\lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\Phi |\xi_j| \rfloor \geq q$ hold, respectively.

Theorem 3.3: Consider a discrete-time linear controller with input $\xi \in \mathbb{R}^{\beta}$ and coefficient matrix $\Phi \in \mathbb{R}^{\alpha \times \beta}$, where $\Phi_{\max} := ||\Phi||_{\max}$ is nonzero. For a given positive integer q, if there exists a quantization gain $\gamma_\Phi \in \mathbb{R}^{+}$ such that the following inequality condition:

$$\gamma_\Phi < \frac{1}{\Phi_{\max}} \left(q - \frac{1}{2} \right), \tag{5}$$

holds, and if there exists a quantization gain $\gamma_\xi \in \mathbb{R}^{+}$ and $||\xi(t)||_\infty < \infty, \forall t \in \mathbb{Z}^{+}$, such that the following inequality condition:

$$\gamma_\xi < \frac{1}{\gamma_\Phi \Phi_{\max} ||\xi(t)||_\infty} \left(q - \frac{1}{2} \right) \ \forall t \in \mathbb{Z}^{+}, \tag{6}$$

holds, then an overflow never occurs in the control operation using quantization gains γ_Φ and γ_ξ.

Proof: The inequality (5) is transformed into

$$q - \frac{1}{2} > \gamma_\Phi \Phi_{\max} \geq \gamma_\Phi |\Phi_{ij}|, \ \forall i \in \mathbb{Z}^{+}_{\alpha+1}, \forall j \in \mathbb{Z}^{+}_{\beta+1}, \\ \Rightarrow \left\lceil q - \frac{1}{2} \right\rfloor = q > \lfloor \gamma_\Phi |\Phi_{ij}| \rfloor = \lceil \gamma_\Phi |\Phi_{ij}| \rfloor, \forall i, j, \tag{7}$$

which is the condition in which an overflow never occurs when quantizing Φ. Next, we define $\xi_{\max} := ||\xi(t)||_\infty$, $\forall t \in \mathbb{Z}^{+}$. The inequality (6) is transformed into

$$q - \frac{1}{2} > \gamma_\Phi \Phi_{\max} \gamma_\xi \xi_{\max} \\ \geq \gamma_\Phi |\Phi_{ij}| \gamma_\xi \xi_{\max}, \ \forall i \in \mathbb{Z}^{+}_{\alpha+1}, \forall j \in \mathbb{Z}^{+}_{\beta+1}, \\ \Rightarrow \left\lceil q - \frac{1}{2} \right\rfloor = q > \lfloor \gamma_\Phi |\Phi_{ij}| \gamma_\xi \xi_{\max} \rfloor \\ = \lceil \gamma_\Phi |\Phi_{ij}| \gamma_\xi \xi_{\max} \rfloor, \forall i, j. \tag{8}$$

When we choose $\xi_{\max}$ such that the following inequality $\lceil \gamma_\Phi |\Phi_{ij}| \gamma_\xi \xi_{\max} \rfloor \geq \lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\xi \xi_{\max} \rfloor$ is satisfied, the following inequality holds,

$$q > \lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\xi \xi_{\max} \rfloor, \forall i, j, \\ \Rightarrow q > \lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\xi \xi_{\max} \rfloor \geq \lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\xi |\xi_j| \rfloor, \forall i, j, \\ \Rightarrow q > \lceil \gamma_\Phi |\Phi_{ij}| \rfloor \lceil \gamma_\xi |\xi_j| \rfloor, \forall i, j, \tag{9}$$

which implies that an overflow never occurs when quantizing the homomorphic operation. Furthermore, the inequalities (7) and (9) imply that an overflow never occurs when quantizing ξ. Therefore, the overflow avoidance conditions shown in **Definition 3.2** are derived. ■

Remark 3.4: The systematic computation of $\xi_{\max}$, which is needed to confirm the overflow avoidance conditions, is difficult. However, there is a situation where we can estimate it to some extent using the specifications of

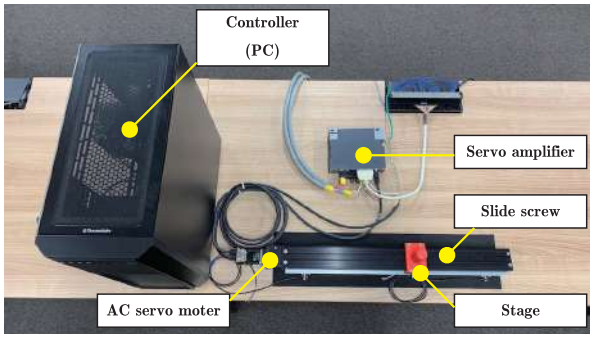


FIGURE 4. Experimental equipment.

TABLE 1. Experimental apparatus.

Servo amplifier	MITSUBISHI MR-J5-10A
Main circuit power supply	1/3-phase 200-240 VAC 50/60 Hz
AC servo motor	MITSUBISHI HK-KT13W
Rated power	0.1 kW
Rated torque	0.32 Nm
Rated speed	3000 rpm
Rated current	1.2 A
Pulse per rotation	67108864 ppr
Slide screw	MiSUMi LX3010CP-MX
Length	1250 mm
Lend	10 mm
PC	
CPU	Intel Core i7-10700K 3.80 GHz
Memory	64 GB
OS	CentOS Linux 8
Language	C++17
DA/AD board	Interface PEX-340216 (16-bit resolution)
Counter board	Interface PEX-632104 (32-bit resolution)

the control systems, such as the allowable position range of a linear stage. Section IV-B explains a method for estimating $\xi_{\max}$.

IV. IMPLEMENTATION

This section introduces a practical testbed control system and determines the parameters for implementing an encrypted controller in the control system.

A. PID POSITION CONTROL SYSTEM

We constructed a position-control system for the linear stage. Fig. 4 presents an overview of the stage position-control system. The actuator used to drive the stage via a slide screw (MiSUMi LX3010CP-MX) is an industrial AC servo motor (MITSUBISHI HK-KT13W) with a servo amplifier (MITSUBISHI MR-J5-10A). The controller device is a PC (Intel Core i7 and CentOS Linux 8), where a robot-control development tool, Advanced Robot Control System V6 (ARCS6) [35], was used for real-time control. The PC outputs a control input to the servo amplifier to actuate the AC servomotor. The position of the stage is measured by a rotary encoder installed in the motor unit and fed back to the PC via a counter board to update the control input. The processes run in the control algorithm are written as C++17. The apparatus and their specifications are listed in TABLE 1.

Throughout the experiments of the position control, we used a PID controller with proportional, integral, and derivative gains, K_p, K_i, and K_d, respectively. Defining the state and input of the controller as $x_c := [e\ w]^{\mathrm{T}}$ and $v_c := [r\ y]^{\mathrm{T}}$, respectively, the discrete-time state-space representation of the PID controller in (1) has the following matrices:

$$A_c = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, B_c = \begin{bmatrix} 1 & -1 \\ T_s & -T_s \end{bmatrix}, C_c = \begin{bmatrix} -\frac{K_d}{T_s} & K_i \end{bmatrix},$$
$$D_c = \begin{bmatrix} K_p + K_i T_s + \frac{K_d}{T_s} & -(K_p + K_i T_s + \frac{K_d}{T_s}) \end{bmatrix}, \quad (10)$$

where $r \in \mathbb{R}$ is a reference for the stage position, $y \in \mathbb{R}$ is the measured stage position, $e := r - y$ is the tracking error, and w is an integrated value defined as $w(t+1) := \sum_{\tau=0}^{t} e(\tau)T_s = w(t) + e(t)T_s$ with sampling period T_s. In this experiment, $K_p = 1.465 \times 10^{-2}$, $K_i = 6.000 \times 10^{-3}$, $K_d = 1.500 \times 10^{-4}$, and $T_s = 20$ ms. In this case, Φ is given as follows:

$$\Phi = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 1 & 0.02 & -0.02 \\ -0.0075 & 0.006 & 0.0223 & -0.0223 \end{bmatrix},$$

and ξ is defined by $\xi := [e\ w\ r\ y]^{\mathrm{T}}$, where $\alpha = 3$ and $\beta = 4$.

B. SECURE IMPLEMENTATION OF THE PID CONTROLLER

To encrypt the PID controller, we set the key length to 256 bits, which was determined by evaluating the computation time of the encrypted control processes, including **Enc**, **Eval**, and **Dec**$^+$ over key lengths. The averages of 100 computation times from 256 to 3072 bits for every 256 bits are shown in Fig. 5. Fig. 5(a) shows the total computation time of **Enc**, **Eval**, and **Dec**$^+$ of the encrypted controls using the proposed $\mathbb{G}^2$-based, $\mathbb{G}^3$-based [16], and gain-multiplying [10] quantizers, as shown in Figs. 5(b), (c), and (d), respectively. Table 2 presents the total computation time and its comparison to the proposed quantizer. The two conventional methods employed the DDH-based KH-PKE scheme, which is common in the method proposed in this study. The figure and table confirm that the computation time of the proposed encrypted control is 47.3 % lower than that of the control system with the $\mathbb{G}^3$-based quantizer, as mentioned in Section III-A. Although processing the control computation with the gain-multiplying quantizer is 45.6 % faster than with the proposed control, it tends to degrade the control performance, as shown in Section V-A. Furthermore, the computation at key length $\ell = 256$ must be completed within a sampling period of 20 ms; therefore, the presented faster quantizer is preferred to the $\mathbb{G}^3$-based quantizer.

We determine γ_Φ and γ_ξ using **Theorem 3.3**. First, we set γ_Φ to 1.0×10^{20} from the inequality (5) of $\gamma_\Phi < 4.3 \times 10^{76}$, with $||\Phi||_{\max} = 1$ and

$$q = 436330242283591462479179208760550548662 \times 291077161756781216195899944211526105 93,$$
$$\approx 4.3 \times 10^{76}.$$

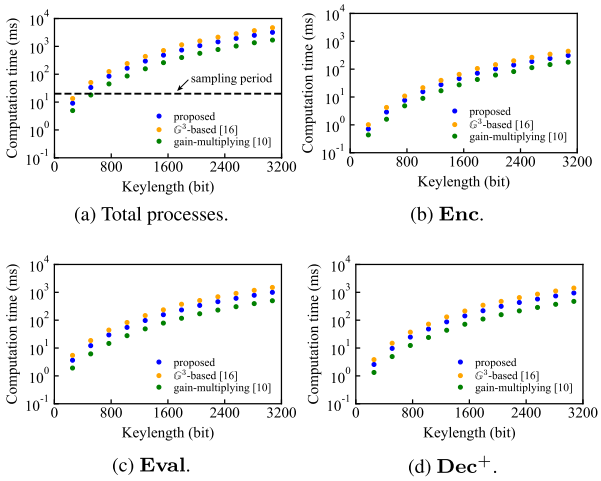


FIGURE 5. Computation time of each process over the key length.

Next, we estimate $||\xi(t)||_\infty$ over a step; that is, we estimate the maximum error e, state w, reference r, and output y. When the length of the linear stage is 1250 mm, the maximum error is 1250 mm because of $e = r - y = 625 - (-625)$. w is given by $w(t+1) := \sum_{\tau=0}^{t} e(\tau)T_s = w(t) + e(t)T_s$. If $T_s = 0.02$ and the duration of this experiment is 10 s, then the maximum state is 1250. Subsequently, $||\xi(t)||_\infty$ can be set to 1250. Therefore, we set γ_ξ to 1.0×10^{53} from inequality (6) of $\gamma_\xi < 3.44 \times 10^{53}$.

Using KH-PKE with the parameters above, the controller gains Φ are encrypted and implemented on the PC. For example, $\mathbf{Enc}(\bar{\Phi}^1_{11})$, $\mathbf{Enc}(\bar{\Phi}^2_{31})$, and $\mathbf{Enc}(\bar{\Phi}^2_{32})$ are (11), as shown at the bottom of the page.

V. EXPERIMENTAL VALIDATION

This section validates the three attributes of the proposed encrypted control system, namely, the control-performance effect, overflow avoidance, and cyberattack detection, compared with the unencrypted control (10) and the two conventional encrypted controls with a $\mathbb{G}^3$-based quantizer [16] and gain-multiplying quantizer [10]. The used encryption scheme, which is a DDH-based PH-PKE, is the same for the three encrypted control methods. In addition, the used control parameters are common, and the other parameters such as γ_Φ, γ_ξ, and ℓ are the same.

A. CONTROL RESULTS

We present the experimental results of position control using the proposed, conventional, and unencrypted controls with a step-like reference, which is given as follows:

$$\begin{cases} 0 & \text{if } 0 \le T_s t < 2 \text{ or } 8 \le T_s t < 10, \\ 50 & \text{if } 2 \le T_s t < 4 \text{ or } 6 \le T_s t < 8, \\ 100 & \text{if } 4 \le T_s t < 6, \end{cases} \tag{12}$$

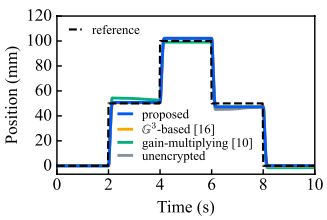
to examine the effects of the proposed secure implementation on a control system.

The control results are presented in Fig. 6. Figs. 6(a) and (b) show the time responses of the stage position and control input, respectively. In the figures, the blue, green, yellow, and gray lines represent the proposed, conventional [10], [16], and unencrypted control methods, respectively, where the broken line represents the reference. Fig. 6(c) shows three stage-position errors between each of the three encrypted controls and the unencrypted control, respectively, and Fig. 6(d) shows the three control input errors between the encrypted control and the unencrypted control. In Figs. 6(c) and (d), the blue line represents the proposed method, and green and yellow lines represent the two conventional methods. Figs. 6(e), (f), and (g) show the quantization error between the quantized and unquantized control inputs, that is, $|\check{u}(t) - u(t)|$, for the proposed method and the two conventional methods, respectively. Table 2 presents ℓ_1-norm values of the quantization-error signals for each quantizer up to 10 s. Figs. 6(h) and (i) show the time responses of the encrypted signals $\mathbf{Enc}(\Psi^1_{33}(t))$ and $\mathbf{Enc}(\xi^1_3(t))$, respectively, which correspond to parts of the output and input signals of the controller.

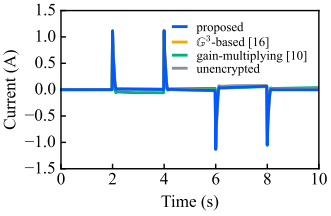
$$\mathbf{Enc}(\bar{\Phi}^1_{11}) = \begin{array}{l}\texttt{(ef99a26e99df01b7c50118dea8b8826fa169177f3c94333f6de844b7faa738a5,}\\ \texttt{f5e78a92d80b0a6ad503a8338905d373b6afae3b1615bdc6280bab18cac4571e,}\\ \texttt{da6e18939379d11f1fb6ca2a7350156adade88f55fac80ecad62b142fe34bd72,}\\ \texttt{4fe15d197c003ee86ee3a35404a8c7cde1e1c1f12a43d963c1d1e2d5f20d5a98,}\\ \texttt{b386194f8a1f016ffc6afdd8469630f7d242db11e5ff6332471048ab8ff7c7f),}\end{array} \tag{11a}$$

$$\mathbf{Enc}(\bar{\Phi}^2_{31}) = \begin{array}{l}\texttt{(c350361fb7d41633662736edb5028dba4cc7ae4d4189d75778aa73c357f2f9d0,}\\ \texttt{73bd85f730994868b02ee2512272c96452aca1575f0317c1ef32de3840c527df,}\\ \texttt{941863c7e67f6ec4f48ba7cdcb04f2da1c1871442da5daaf8e69f88987243546,}\\ \texttt{179a00c5a06fc763972cc5254bdf7beb9d8e4d3057391b679a23dc072c2e3114,}\\ \texttt{112f5a3adc5bdf4d4509195e4c97879328dd158ddee5899c978ba0defb7034b8),}\end{array} \tag{11b}$$

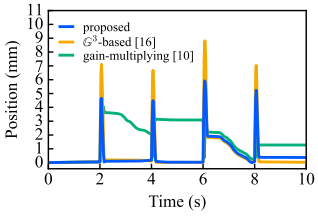
$$\mathbf{Enc}(\bar{\Phi}^2_{32}) = \begin{array}{l}\texttt{(eb6906386c2e66dbfa88403f3297ca0356c28b00e9ac6dbd1db81caffa4a7312,}\\ \texttt{ee15995c854fb987ee47338a31f4dce480635e2e75123d03dc1370b29f204abd,}\\ \texttt{e434b267068ca03fde81d39dd1644f466c8588dc91e1bae366d0591add6d5c09,}\\ \texttt{d9d89949da579ecb3b662c434575425470fffcf1e19c52f34bd9d11b36c35e65,}\\ \texttt{a4cf383732d68e30a05a96398a3dd3fdda246b08a8786c0cd4be46cfdcdcf866).}\end{array} \tag{11c}$$



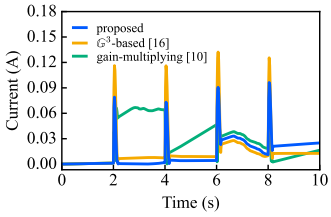
(a) Controlled output: $y(t)$.



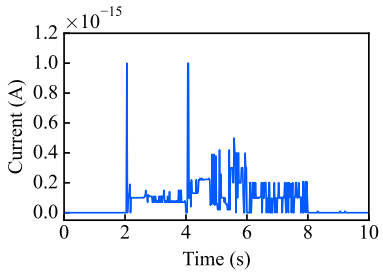
(b) Control input: $u(t)$.



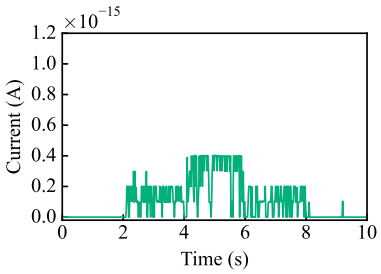
(c) Deviation from the output of the unencrypted control system.



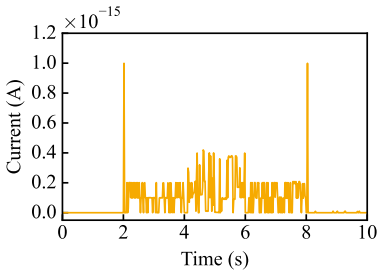
(d) Deviation from the control input of the unencrypted control system.



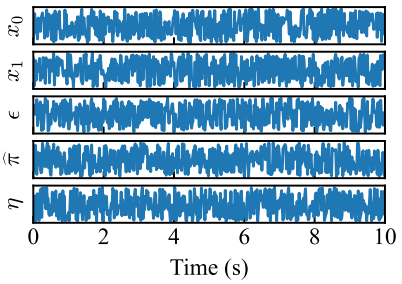
(e) Quantization error in u of the proposed $\mathbb{G}^2$-based quantizer.



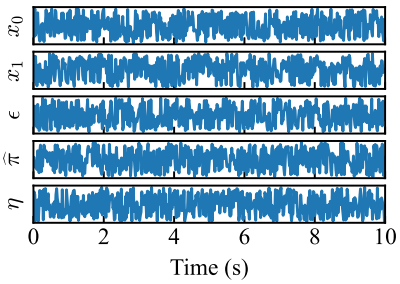
(f) Quantization error in u of the conventional gain-multiplying quantizer.



(g) Quantization error in u of the conventional $\mathbb{G}^3$-based quantizer.



(h) $\mathbf{Enc}(\Psi_{33}^1(t))$.



(i) $\mathbf{Enc}(\xi_3^1(t))$.

FIGURE 6. Experimental results of the four control methods: the KH-PKE-based encrypted control systems with the proposed $\mathbb{G}^2$-based quntizer, the conventional gain-multiplying [10] and $\mathbb{G}^3$-based [16] quantizers, and the unencrypted PID control system (10). The parameters $\ell = 256$, $\gamma_\Phi = 10^{20}$, and $\gamma_\xi = 10^{53}$ are the same for the proposed and conventional encrypted controls.

Figs. 6(c) and (d) confirm that the proposed control system achieves less deviation from the conventional method using the gain-multiplying quantizer. As shown in Table 2, the ℓ_1-norm values of the quantization-error signals in Figs. 6(e), (f), and (g) are 3.880×10^{-14}, 5.074×10^{-14}, and 3.856×10^{-14}, respectively. These scores demonstrate that the modification of the plaintext space from $\mathbb{G}^3$ to $\mathbb{G}^2$ has minimal impact on the quantization error, resulting in only a 0.6 % difference in the resulting control signals. The proposed control system achieves a 30.6 % improvement compared to the gain-multiplying quantizer. Furthermore, the parameters of (11) and signals shown in Figs. 6(h) and (i) were concealed in random numbers. The norm scores imply that the proposed control system is better than conventional systems in terms of control performance degradation and that they are negligibly small from the viewpoint of using variables in the C++ language.

Therefore, the control experimental results confirm that the proposed encrypted control system has a smaller impact on the control performance than the conventional encrypted control systems.

B. OVERFLOW AVOIDANCE

This section shows that **Theorem 3.3** helps us choose the appropriate values of γ_Φ and γ_ξ to avoid overflows in the control operation. Overflow avoidance is validated by showing another control result of the proposed encrypted control with inappropriate values, such as $\gamma_\Phi = 1.0 \times 10^{20}$ and $\gamma_\xi = 1.0 \times 10^{57}$, which do not satisfy the inequality in (6).

The control results for this case are shown in Fig. 7. Figs. 7(a) and (b) show the stage position and the control input, respectively, using the blue line. In this case, an overflow occurred between 2.00 s and 8.16 s, highlighted in yellow. The figures confirm that the control input was too small for the stage to follow the reference. This is because of the large γ_ξ, such that the term $\lceil \gamma |x| \rfloor$ of $\mathscr{A}_\gamma$ is greater than q in encoding, which implies an overflow. Therefore, the control result confirms that **Theorem 3.3** facilitates the desig of parameter values to avoid overflow.

Remark 5.1: If a systematic method of designing parameters is established, $||\xi(t)||_\infty$ must be estimated before starting the control operation or a dynamic quantizer related to γ_ξ using the observation of $\xi(t)$ [36], [37]. The estimation may

TABLE 2. Performance comparison of computation time and quantization error in the experimental results.

Encrypted control system with	Computation time (ms) at a 256-bit key length	Rate to the proposed quantizer (%)	Evaluation of quantization error	Rate to the proposed quantizer (%)
the proposed quantizer	9.12	100	3.880×10^{-14}	100
the $\mathbb{G}^3$-based quantizer [16]	13.43	147.3	3.856×10^{-14}	99.4
the gain-multiplying quantizer [10]	4.96	54.4	5.074×10^{-14}	130.6

be discussed using the equipment properties as stated in Section IV-B, while the dynamic quantizer design requires a different problem setting and further discussion; therefore, the systematic parameter design will be addressed in our future study.

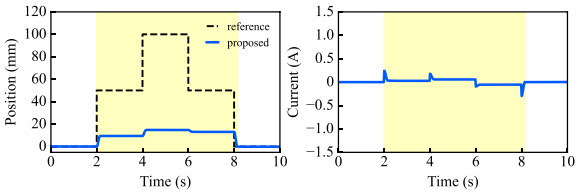
C. ATTACK DETECTION

It is demonstrated that the proposed encrypted control system enables the detection of malleability-based cyberattacks. An attacker does not know the homomorphic operation key $\mathbf{sk_h}$ for **Eval**; therefore, they tamper with the third element of the ciphertext using a constant factor. Because the homomorphic encryption scheme yields malleability, the attacker can obtain the desired decryption result by falsifying the ciphertext [25]. In KH-PKE, the last element of the **Dec** algorithm is $m = \epsilon/\pi \bmod p$, which means, if the third element e of the ciphertext is multiplied by $\lambda \in \mathbb{G}$, the decryption result is λ-fold. Therefore, even if the attackers do not know $\mathbf{sk_h}$, they can tamper with the data.

We consider two cyberattacks in this test: One falsifies an encrypted signal, double $\mathbf{Enc}(\bar{\xi}_4^2(t))$ in 4.0 s to 4.1 s, and the other falsifies two components of the control parameter triple $\mathbf{Enc}(\bar{\Phi}_{31}^2)$ and $\mathbf{Enc}(\bar{\Phi}_{32}^2)$ at 4.0 s, respectively. This means that the cyberattack doubles the signal or triples the two components of the parameter matrix.

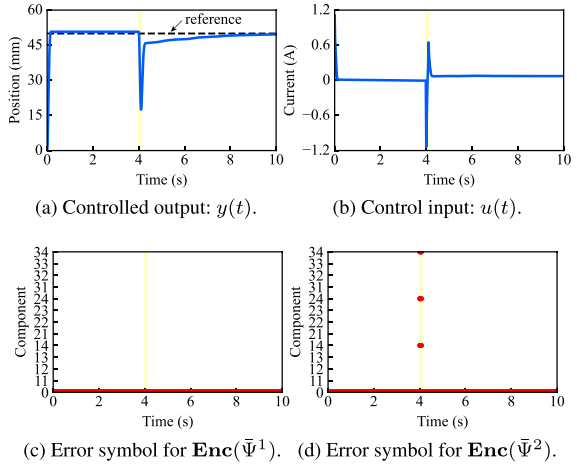
The cyberattack test results for falsifying a signal are shown in Fig. 8. Figs. 8(a) and (b) show the stage position and the control input, respectively. Figs. 8(c) and (d) show the error symbols for $\mathbf{Enc}(\bar{\Psi}^1)$ and $\mathbf{Enc}(\bar{\Psi}^2)$ in **Eval**, respectively. Fig. 8(a) shows that falsification of the sensor values caused spike-like changes in the stage position, indicating that tampering with the sensor values can destroy the control system. For such a falsification attack, the falsification time is confirmed from the error symbol, as shown in Fig. 8(c) and (d). Moreover, the indices of $\mathbf{Enc}(\bar{\Psi}_{14}^2)$, $\mathbf{Enc}(\bar{\Psi}_{24}^2)$, and $\mathbf{Enc}(\bar{\Psi}_{34}^2)$, which were calculated with the attacked signal $\mathbf{Enc}(\bar{\xi}_4^2)$ between 4.0 s and 4.1 s, can be identified.

The cyberattack test results for falsifying the two components in the control parameter are shown in Fig. 9. Figs. 9(a) and (b) show the stage positions of the stage and control input, respectively. Figs. 9(c) and (d) show the error symbols for $\mathbf{Enc}(\bar{\Psi}^1)$ and $\mathbf{Enc}(\bar{\Psi}^2)$ in **Eval**, respectively. Fig. 9(a) shows that the stage position was gradually shifted by falsifying the controller parameters. Fig. 9(b) shows that the falsification effect is not significantly reflected in the control input, which confirms that it is difficult to detect an attack using the threshold method [38]. For the falsification attack, the falsification time was confirmed using the error symbol, as shown in Figs. 9(c) and (d). The indices of the attacked signals $\mathbf{Enc}(\Phi_{31}^2)$ and $\mathbf{Enc}(\Phi_{32}^2)$ could be detected.



(a) Controlled output: $y(t)$. (b) Control input: $u(t)$.

FIGURE 7. Experimental results of the proposed encrypted PID control method. The used parameters are $\gamma_\Phi = 10^{20}$ and $\gamma_\xi = 10^{57}$, which do not satisfy the inequality condition (6).



(a) Controlled output: $y(t)$. (b) Control input: $u(t)$.

(c) Error symbol for $\mathbf{Enc}(\bar{\Psi}^1)$. (d) Error symbol for $\mathbf{Enc}(\bar{\Psi}^2)$.

FIGURE 8. Experimental result of the cyberattack performed by tampering with signals between 4.0 and 4.1 s against the proposed encrypted control system.

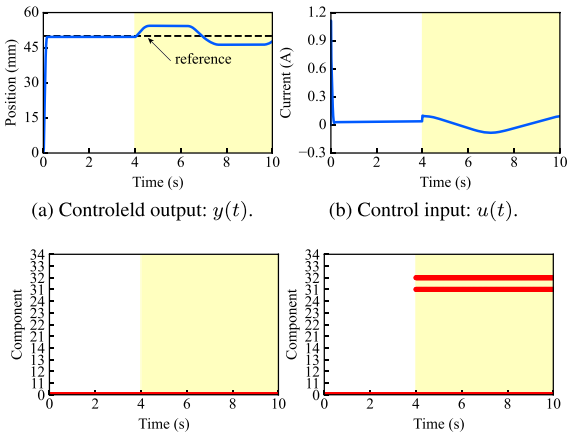
VI. DISCUSSIONS

Based on the results of this study, this section discusses two important issues for future work, to secure control systems.

A. EFFECTS OF A LEAKED HOMOMORPHIC OPERATION KEY

This study considered a situation in which the attacker never has a homomorphic operation key $\mathbf{sk_h}$ in tampering; however, it is also important to consider a situation in which a homomorphic operation key is leaked to a third party. Attackers may use a homomorphic operation to tamper with the ciphertexts in **Eval** to adjust the controlled outputs.

In this case, another detection mechanism is required because the algorithms **Eval** and **Dec** do not output error symbols for detection. One detection approach is to observe unencrypted control inputs from the perspective of a control



(a) Controleld output: $y(t)$. (b) Control input: $u(t)$.

(c) Error symbol for $\mathbf{Enc}(\bar{\Psi}^1)$. (d) Error symbol for $\mathbf{Enc}(\bar{\Psi}^2)$.

FIGURE 9. Experimental results of the cyberattack performed by tampering with control parameters within 4 s against the proposed encrypted control system.

theory, such as threshold-based detection methods [38], [39]. The other is to update the encryption keys in time from the perspective of cryptography, such as storing and replacing the keys periodically [40] and employing an updatable public-key ElGamal encryption scheme [36], [41]. The switching and updatable encryption schemes make it easy to detect falsification and replay attacks, which are difficult to detect in [4]. In this sense, an updatable KH-PKE scheme is expected to enhance the ability to detect cyberattacks on the control systems.

In addition, even if the homomorphic operation key is leaked, the proposed encrypted control system is more secure than an ElGamal-based encryption control system [10]. This is because the KH-PKE scheme retains stronger security than indistinguishability under chosen-ciphertext (IND-CCA1) security even when a key is leaked [14], where IND-CCA1 is the security that homomorphic encryption schemes can achieve [42]. Furthermore, because the controller parameters and signals remain encrypted, the control system is resistant to cyberattacks that require a model of the control system [6], [43].

B. APPROPRIATE SECURITY FOR CONTROL SYSTEMS

Implementing a key of several thousand bits in an encrypted control system with a real-time constraint is challenging, implying the control processes must be completed within a sampling period. The KH-PKE scheme is based on the DDH assumption, and from the viewpoint of cryptology, a key length that can assume the DDH-hardness is needed. Specifically, the NIST document [44] states that a key length of at least 2048 bits is desirable. However, the key length set used in this study was 256 bits to fulfill the real-time constraint under a sampling period of 20 ms, as shown in Fig. 5.

Currently, it is difficult to conclude whether the key length is satisfactory because the answer depends on the security concept we consider. An appropriate security concept exists for control systems, such as indistinguishability against parameter estimation attack (IND-PEA), proposed in [45], which is in provable security and revealed that IND-PEA is equal to indistinguishability under the chosen plaintext attack (IND-CPA). The updatable ElGamal-based encrypted control system [41], [46], which covers computational security, protects the controller parameters from being identified by attackers, even though the key length is shorter than that required by NIST. Furthermore, key updating ideas help solve real-time constraint issues. Such a security concept that is appropriate for control systems has recently been studied; therefore, appropriate security may exist for the 256-bits KH-PKE scheme. Exploring the appropriate security concept is significant for developing the control theory and cryptography fields, which will be explored in our future study. Additionally, in the sense of achieving IND-CCA1/2, the key length is insufficient because it is less than 2048 bits.

VII. CONCLUSION

This study proposed a cyberattack-detectable encrypted control system and validated its effectiveness using an industrial motor PID position-control system. The KH-PKE scheme was employed in the proposed control system for real-time cyberattack detection, and our novel quantizer was used to reduce computation time, as demonstrated by experiments that showed a 47.3 % reduction in computation time while maintaining similar quantization-error impact (with only a 0.6 % difference) compared to our previous quantizer. Furthermore, this study analyzed the conditions for overflow, which are summarized in **Theorem 3.3**. Experimental validations confirmed that the proposed control system concealed the control operation, and the results also confirmed that the theorem helps design quantization gains to avoid overflows. Importantly, this study demonstrated the results of falsification attack tests using homomorphism and confirmed that the proposed control system enables real-time detection of attacked components within signals and control parameters, which is a significant advantage.

Future studies will focus on the following areas. Firstly, the development of countermeasures against leaked homomorphic operation keys to potential attackers, as mentioned in Section VI. The homomorphic operation key is placed in the controller, which poses a risk of leakage to attackers who are interested in breaking into and compromising the control system. Secondly, ensuring the stability of the proposed encrypted control system. The stability of control systems is crucial for safe operation, but the proposed system requires a quantizer, which may destabilize the control system if not stabilized in advance. Finally, exploring security concepts appropriate for control systems and evaluating the security of the encrypted control system developed in this study.

APPENDIX A ALGORITHMS OF DDH-BASED KH-PKE SCHEME

This study uses KH-PKE with multiplicative homomorphism [14] to construct encrypted control systems. The KH-PKE scheme, denoted as $\mathcal{E} = (\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec}, \mathbf{Eval})$, consists of the following four algorithms.

Gen: $\kappa \mapsto (\mathbf{pk}, \mathbf{sk_d}, \mathbf{sk_h})$. The **Gen** algorithm takes security parameter κ and key length ℓ regarding ℓ-bit prime number p and outputs public, private, and homomorphic operation keys, denoted as $\mathbf{pk}$, $\mathbf{sk_d}$, and $\mathbf{sk_h}$, respectively:

$$\begin{aligned}\mathbf{pk} &= (g_0, g_1, s, \widehat{s}, \widetilde{s}_0, \widetilde{s}_1)\,,\\ \mathbf{sk_d} &= \left(k_0, k_1, \widehat{k}_0, \widehat{k}_1, \widetilde{k}_{0,0}, \widetilde{k}_{0,1}, \widetilde{k}_{1,0}, \widetilde{k}_{1,1}\right),\\ \mathbf{sk_h} &= \left(\widetilde{k}_{0,0}, \widetilde{k}_{0,1}, \widetilde{k}_{1,0}, \widetilde{k}_{1,1}\right),\end{aligned}$$

where g_0 and g_1 are randomly chosen from $\mathbb{G}$; $s := g_0^{k_0} g_1^{k_1} \bmod p$; $\widehat{s} := g_0^{\widehat{k}_0} g_1^{\widehat{k}_1} \bmod p$; $\widetilde{s}_0 := g_0^{\widetilde{k}_{0,0}} g_1^{\widetilde{k}_{0,1}} \bmod p$; $\widetilde{s}_1 := g_0^{\widetilde{k}_{1,0}} g_1^{\widetilde{k}_{1,1}} \bmod p$; k_0, k_1, $\widehat{k}_0$, $\widehat{k}_1$, $\widetilde{k}_{0,0}$, $\widetilde{k}_{0,1}$, $\widetilde{k}_{1,0}$, and $\widetilde{k}_{1,1}$ are randomly chosen from $\mathbb{Z}_q$, where $p = 2q + 1$.

Enc: $(\mathbf{pk}, m \in \mathcal{M}) \mapsto c = (x_0, x_1, \epsilon, \widehat{\pi}, \eta) \in \mathcal{C}$. The **Enc** algorithm takes a public key $\mathbf{pk}$ and a plaintext m and outputs a ciphertext c. The components of c are as follows: $x_0 := g_0^{\omega} \bmod p$; $x_1 := g_1^{\omega} \bmod p$; $\epsilon := m\pi \bmod p$; $\widehat{\pi} := \widehat{s}^{\omega} \bmod p$, where $\pi := s^{\omega} \bmod p$ and ω is chosen randomly from $\mathbb{Z}_q$; $\eta := f_{hk}\left(\left(\widetilde{s}_0 \cdot \widetilde{s}_1^{\delta}\right)^{\omega} \bmod p\right)$ with $\delta := \gamma_{hk}(x_0, x_1, \epsilon, \widehat{\pi})$, where γ_{hk} is target collision resistance hash family and f_{hk} is a smooth function [14]. We use SHA-256 to both γ_{hk} and f_{hk}.

Dec: $(\mathbf{sk_d}, c \in \mathcal{C}) \mapsto m \in \mathcal{M} \cup \{\perp\}$. The **Dec** algorithm takes a private key and a ciphertext $c = (x_0, x_1, \epsilon, \widehat{\pi}, \eta)$ and outputs a plaintext m or an error symbol $\perp$. Compute $\widehat{\pi}' := x_0^{\widehat{k}_0} x_1^{\widehat{k}_1} \bmod p$, $\delta := \gamma_{hk}(x_0, x_1, \epsilon, \widehat{\pi})$, and $\eta' := f_{hk}\left(x_0^{\widetilde{k}_{0,0}+\delta\widetilde{k}_{1,0}} x_1^{\widetilde{k}_{0,1}+\delta\widetilde{k}_{1,1}} \bmod p\right)$, where f_{hk} is a smooth function [14]. If either $\widehat{\pi} \neq \widehat{\pi}'$ or $\eta \neq \eta'$, then return an error symbol $\perp$; Otherwise, return $m = \epsilon/\pi \bmod p$, where $\pi := x_0^{k_0} x_1^{k_1} \bmod p$.

Eval: $(\mathbf{sk_h}, c_1, c_2 \in \mathcal{C}) \mapsto c \in \mathcal{C} \cup \{\perp\}$. The **Eval** algorithm takes a homomorphic operation key and two ciphertexts c_i $\forall i \in \{1, 2\}$ and outputs a ciphertext $(x_0, x_1, \epsilon, \widehat{\pi}, \eta)$ or an error symbol $\perp$. The components of the output c are computed as follows: $x_0 := x_{1,0} x_{2,0} g_0^{\omega} \bmod p$, $x_1 := x_{1,1} x_{2,1} g_1^{\omega} \bmod p$, $\epsilon := \epsilon_1 \epsilon_2 s^{\omega} \bmod p$, $\widehat{\pi} := \widehat{\pi}_1 \widehat{\pi}_2 \widehat{s}^{\omega} \bmod p$, and $\eta = f_{hk}\left(x_0^{\widetilde{k}_{0,0}+\delta\widetilde{k}_{1,0}} x_1^{\widetilde{k}_{0,1}+\delta\widetilde{k}_{1,1}} \bmod p\right)$, where $c_i := (x_{i,0}, x_{i,1}, \epsilon_i, \widehat{\pi}_i, \eta_i)$; $\delta := \gamma_{hk}(x_0, x_1, \epsilon, \widehat{\pi})$; $\delta_i := \gamma_{hk}\left(x_{i,0}, x_{i,1}, \epsilon_i, \widehat{\pi}_i\right)$; ω is randomly chosen from $\mathbb{Z}_q$; $\eta_i' := f_{hk}\left(x_{i,0}^{\widetilde{k}_{0,0}+\delta_i\widetilde{k}_{1,0}} x_{i,1}^{\widetilde{k}_{0,1}+\delta_i\widetilde{k}_{1,1}} \bmod p\right)$. If either $\eta_1 \neq \eta_1'$ or $\eta_2 \neq \eta_2'$, then return $\perp$; Otherwise, return c.

REFERENCES

[1] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security Privacy*, vol. 9, no. 3, pp. 49–51, May/Jun. 2011.

[2] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317–3318, Jul. 2017.

[3] A. Cetinkaya, H. Ishii, and T. Hayakawa, "An overview on denial-of-service attacks in control systems: Attack models and security analyses," *Entropy*, vol. 21, no. 2, p. 210, Feb. 2019.

[4] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2009, pp. 911–918.

[5] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Proc. 1st Workshop Secure Control Syst.*, 2010, pp. 1–6.

[6] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[7] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, Jan. 2015.

[8] M. S. Chong, H. Sandberg, and A. M. H. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *Proc. 18th Eur. Control Conf. (ECC)*, Jun. 2019, pp. 968–978.

[9] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Comput. Surveys*, vol. 51, no. 4, pp. 1–35, Jul. 2019.

[10] K. Kogiso and T. Fujita, "Cyber-security enhancement of networked control systems using homomorphic encryption," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Dec. 2015, pp. 6836–6843.

[11] F. Farokhi, I. Shames, and N. Batterham, "Secure and private cloud-based control using semi-homomorphic encryption," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 163–168, 2016.

[12] J. Kim, C. Lee, H. Shim, J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Encrypting controller using fully homomorphic encryption for security of cyber-physical systems," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 175–180, 2016.

[13] M. Schulze Darup, A. B. Alexandru, D. E. Quevedo, and G. J. Pappas, "Encrypted control for networked systems: An illustrative introduction and current challenges," *IEEE Control Syst.*, vol. 41, no. 3, pp. 58–78, Jun. 2021.

[14] K. Emura, G. Hanaoka, K. Nuida, G. Ohtake, T. Matsuda, and S. Yamada, "Chosen ciphertext secure keyed-homomorphic public-key cryptosystems," *Designs, Codes Cryptogr.*, vol. 86, no. 8, pp. 1623–1683, Aug. 2018.

[15] K. Teranishi, N. Shimada, and K. Kogiso, "Stability-guaranteed dynamic ElGamal cryptosystem for encrypted control systems," *IET Control Theory Appl.*, vol. 14, no. 16, pp. 2242–2252, Nov. 2020.

[16] K. Teranishi and K. Kogiso, "ElGamal-type encryption for optimal dynamic quantizer in encrypted control systems," *SICE J. Control, Meas., Syst. Integr.*, vol. 14, no. 1, pp. 59–66, Jan. 2021.

[17] A. B. Alexandru, M. S. Darup, and G. J. Pappas, "Encrypted cooperative control revisited," in *Proc. IEEE Conf. Decis. Control*, Mar. 2019, pp. 7196–7202.

[18] N. Schluter and M. S. Darup, "Encrypted explicit MPC based on two-party computation and convex controller decomposition," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Dec. 2020, pp. 5469–5476.

[19] M. Kishida, "Encrypted control system with quantizer," *IET Control Theory Appl.*, vol. 13, no. 1, pp. 146–151, 2019.

[20] R. Alisic, J. Kim, and H. Sandberg, "Model-free undetectable attacks on linear systems using LWE-based encryption," *IEEE Control Syst. Lett.*, vol. 7, pp. 1249–1254, 2023.

[21] J. Kim, H. Shim, and K. Han, "Dynamic controller that operates over homomorphically encrypted data for infinite time horizon," *IEEE Trans. Autom. Control*, vol. 68, no. 2, pp. 660–672, Feb. 2023.

[22] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, vol. 5, 1999, pp. 223–238.

[23] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 469–472, Jul. 1985.

[24] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2009.

[25] K. Teranishi and K. Kogiso, "Control-theoretic approach to malleability cancellation by attacked signal normalization," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 297–302, 2019.

[26] J. H. Cheon, K. Han, H. Kim, J. Kim, and H. Shim, "Need for controllers having integer coefficients in homomorphically encrypted dynamic system," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 5020–5025.

[27] M. Fauser and P. Zhang, "Resilient homomorphic encryption scheme for cyber-physical systems," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 5634–5639.

[28] M. Fauser and P. Zhang, "Detection of cyber attacks in encrypted control systems," *IEEE Control Syst. Lett.*, vol. 6, pp. 2365–2370, 2022.

[29] K. Emura, G. Hanaoka, G. Ohtake, T. Matsuda, and S. Yamada, "Chosen ciphertext secure keyed-homomorphic public-key encryption," in *Proc. Int. Workshop Public Key Cryptography*. Cham, Switzerland: Springer, 2013, pp. 32–50.

[30] B. Libert, T. Peters, M. Joye, and M. Yung, "Non-malleability from malleability: Simulation-sound quasi-adaptive NIZK proofs and CCA2-secure encryption from homomorphic signatures," in *Proc. Adv. Cryptol. EUROCRYPT*, 2014, pp. 514–532.
[31] C. Jutla and A. Roy, "Dual-system simulation-soundness with applications to UC-PAKE and more," in *Proc. Adv. Cryptol. ASIACRYPT*, 2015, pp. 630–655.
[32] Y. Maeda and K. Nuida, "Chosen ciphertext secure keyed two-level homomorphic encryption," in *Proc. Inf. Secur. Privacy*, 2022, pp. 209–228.
[33] J. Lai, R. H. Deng, C. Ma, K. Sakurai, and J. Weng, "CCA-secure keyed-fully homomorphic encryption," in *Proc. Public-Key Cryptography (PKC)*, 2016, pp. 70–98.
[34] S. Sato, K. Emura, and A. Takayasu, "Keyed-fully homomorphic encryption without indistinguishability obfuscation," in *Applied Cryptography and Network Security* (Lecture Notes in Computer Science), vol. 13269, G. Ateniese and D. Venturi, Eds. Cham, Switzerland: Springer, 2022, pp. 3–23.
[35] Y. Yokokura. *Side Warehouse of Laboratory*. Accessed: Apr. 19, 2023. [Online]. Available: https://www.sidewarehouse.net/arcs6/index.html
[36] K. Teranishi and K. Kogiso, "Dynamic quantizer for encrypted observer-based control," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Dec. 2020, pp. 5477–5482.
[37] H. Kawase, K. Teranishi, and K. Kogiso, "Dynamic quantizer synthesis for encrypted state-feedback control systems with partially homomorphic encryption," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2022, pp. 75–81.
[38] B. Rikuna, K. Kogiso, and M. Kishida, "Detection method of controller falsification attacks against encrypted control system," in *Proc. SICE Annu. Conf.*, 2018, pp. 5032–5037.
[39] D. Martynova and P. Zhang, "An approach to encrypted fault detection of cyber-physical systems," in *Proc. Asian Control Conf.*, 2019, pp. 1501–1506.
[40] K. Kogiso, "Attack detection and prevention for encrypted control systems by application of switching-key management," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 5032–5037.
[41] K. Teranishi, T. Sadamoto, A. Chakrabortty, and K. Kogiso, "Designing optimal key lengths and control laws for encrypted control systems based on sample identifying complexity and deciphering time," *IEEE Trans. Autom. Control*, vol. 68, no. 4, pp. 2183–2198, Apr. 2023.
[42] R. Canetti, S. Raghuraman, S. Richelson, and V. Vaikuntanathan, "Chosen-ciphertext secure fully homomorphic encryption," in *Proc. Public Key Cryptography*. Cham, Switzerland: Springer, 2017, pp. 213–240.
[43] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proc. Volumes*, vol. 44, no. 1, pp. 90–95, Jan. 2011.
[44] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid, "Recommendation for key management part 1: General (revision 5)," Special Publication, NIST, Gaithersburg, MD, USA, Tech. Rep. 800–57, 2020.
[45] K. Teranishi and K. Kogiso, "Towards provably secure encrypted control using homomorphic encryption," in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, Dec. 2022, pp. 7740–7745.
[46] K. Teranishi and K. Kogiso, "Optimal controller and security parameter for encrypted control systems under least squares identification," 2023, *arXiv:2302.12154*.

MASAKI MIYAMOTO (Graduate Student Member, IEEE) received the B.E. and M.E. degrees from The University of Electro Communications, Tokyo, Japan, in 2021 and 2023, respectively. His research interest includes encrypted controls.

KAORU TERANISHI (Graduate Student Member, IEEE) received the B.E. degree in electromechanical engineering from the National Institute of Technology, Ishikawa College, Ishikawa, Japan, in 2019, and the M.E. degree in mechanical and intelligent systems engineering from The University of Electro-Communications, Tokyo, Japan, in 2021, where he is currently pursuing the Ph.D. degree. From October 2019 to September 2020, he was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, GA, USA. Since April 2021, he has been a Research Fellow with the Japan Society for the Promotion of Science. His research interests include control theory and cryptography for the cybersecurity of control systems.

KEITA EMURA received the M.E. degree from Kanazawa University, in 2004, and the Ph.D. degree in information science from the Japan Advanced Institute of Science and Technology (JAIST), in 2010. He was with Fujitsu Hokuriku Systems Ltd., from 2004 to 2006. He was a Postdoctoral Researcher with the Center for Highly Dependable Embedded Systems Technology, JAIST, from 2010 to 2012. He has been a Researcher with the National Institute of Information and Communications Technology (NICT), since 2012. Since 2014, he has been a Senior Researcher with NICT, where he has been a Research Manager, since 2021. His research interests include public-key cryptography and information security. He is a member of IEICE, IPSJ, and IACR. He was a recipient of the SCIS Innovation Paper Award from IEICE, in 2012, the CSS Best Paper Award from IPSJ, in 2016, the IPSJ Yamashita SIG Research Award, in 2017, and the Best Paper Award from ProvSec 2022.

KIMINAO KOGISO (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in mechanical engineering from Osaka University, Japan, in 1999, 2001, and 2004, respectively. He was appointed as a Postdoctoral Fellow with the 21st Century COE Program and as an Assistant Professor with the Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan, in April 2004 and July 2005, respectively. From November 2010 to December 2011, he was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, GA, USA. In March 2014, he was promoted to the position of Associate Professor with the Department of Mechanical and Intelligent Systems Engineering, The University of Electro-Communications, Tokyo, Japan, where he has been a Full Professor, since April 2023. His research interests include cybersecurity of control systems, constrained control, control of decision-makers, and their applications.

QualSec: An Automated Quality-Driven Approach for Security Risk Identification in Cyber-Physical Production Systems

Matthias Eckhart, Andreas Ekelhart, Stefan Biffl, *Member, IEEE*, Arndt Lüder, *Senior Member, IEEE*, and Edgar Weippl, *Senior Member, IEEE*

***Abstract*—As the threat landscape in the industrial domain continually advances, security-by-design is an ever-growing concern in the engineering of cyber-physical production systems (CPPSs). Often, quality aspects are not considered when securing CPPSs, which creates attack vectors that could lead to malicious activity affecting the products' quality. Since quality control systems generally provide inadequate protection against intentionally introduced defects, and can be susceptible to attacks, quality considerations must be integrated into security-aware CPPS engineering. For this purpose, we propose the QualSec method that automatically identifies security risks pertaining to CPPSs, building on the quality characteristics associated with manufacturing operations to determine cascading effects. QualSec is based on a semantic representation of engineering knowledge, allowing to efficiently reuse engineering models from AutomationML artifacts. Moreover, QualSec utilizes Petri nets to facilitate the analysis of security risks and cascading effects. In this way, QualSec informs users about possible attack paths for compromising quality characteristics, how attackers may disguise their malicious actions, and the possible consequences of attacks with respect to product quality. We demonstrate the benefits of QualSec in a case study and analyze its scalability through a rigorous performance evaluation.**

***Index Terms*—AutomationML, cyber-physical production systems (CPPSs), industrial control systems (ICSs), information security, petri net (PN), production systems engineering (PSE).**

Manuscript received 9 March 2022; revised 11 June 2022; accepted 5 July 2022. Date of publication 22 July 2022; date of current version 22 March 2023. This work was supported in part by the Austrian Research Promotion Agency (FFG) through the Austrian Competence Center for Digital Production (CDP) under Grant 881843, and in part by Bridge 1 Program under Grant 880609, in part by the Christian Doppler Research Association, in part by the Austrian Federal Ministry for Digital and Economic Affairs, and in part by the National Foundation for Research, Technology, and Development. The COMET Center SBA Research (SBA-K1) was supported by BMVIT, BMDW, and the Federal State of Vienna, through COMET—Competence Centers for Excellent Technologies, and managed by the FFG. Paper no. TII-22-1013. *(Corresponding author: Matthias Eckhart.)*

Matthias Eckhart, Andreas Ekelhart, and Edgar Weippl are with SBA Research, 1040 Vienna, Austria, and also with the Christian Doppler Laboratory for Security and Quality Improvement in the Production System Lifecycle, University of Vienna, 1090 Vienna, Austria (e-mail: meckhart@sba-research.org; aekelhart@sba-research.org; edgar.weippl@univie.ac.at).

Stefan Biffl is with the Institute of Information Systems Engineering, TU Wien, 1040 Vienna, Austria, and also with CDP, 1220 Vienna, Austria (e-mail: stefan.biffl@tuwien.ac.at).

Arndt Lüder is with CDP, 1220 Vienna, Austria, and also with the Institute of Ergonomics, Manufacturing Systems and Automation, Otto-von-Guericke U., 39106 Magdeburg, Germany (e-mail: arndt.lueder@ovgu.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3193119.

Digital Object Identifier 10.1109/TII.2022.3193119

I. Introduction

SINCE new threats that can compromise the secure and safe operation of cyber-physical production systems (CPPSs) are continuously emerging, managing security risks at the beginning of the systems' lifecycle is paramount. This requires, on the one hand, that the individual engineering activities (e.g., software development and testing [1]), including the exchanged artifacts, are sufficiently protected against adversaries [2]. On the other hand, security must be established as a "first-class citizen" in the engineering process to achieve CPPSs that are secure by design [3]. In the latter case, knowledge from diverse domain experts is essential, given that the engineering of CPPSs is by itself a highly multidisciplinary endeavor. The cyber-physical nature of attacks launched against CPPSs further underlines this need: Attacks executed from cyberspace can lead to physical harm and may endanger human life. Thus, both security and safety concerns need to be considered jointly. In this context, it is worth pointing out that quality is likewise interdependent with security but often not perceived as such. For instance, security risks may manifest themselves as symptoms of a quality control (QC) issue (e.g., data integrity breach due to poor handling of QC logbooks), meaning that addressing this underlying problem could also improve the overall security. Conversely, strengthening the security of a CPPS may also lead to higher quality (e.g., additional sensors put in place to prevent covert product modifications may at the same time unveil defects). Recognizing this interdependence may not only help to promote the fact that information security adds value to an organization (in this case, realized via quality improvements), but also increases the awareness of cyberattacks that focus on the quality of the manufactured products.

The potential severity and multidimensional characteristic of sabotage attacks targeting product quality necessitate a holistic security risk assessment approach that also incorporates quality

considerations. However, current risk assessment workflows defined in leading industrial security standards and guidelines (e.g., IEC 62443-3-2 [4] or VDI/VDE 2182-1 [5]) adopt a rather resource-centric view, neglecting the product and process components. This leads to an incomplete understanding of security risks that is also inconsistent with the Product, Process, and Resource (PPR) concept [6], which plays a predominant role in the engineering of CPPSs. In other words, engineers currently consider quality concerns without assuming intentional wrongdoing (i.e., in isolation from security concerns). This isolated view weakens both QCs and security controls.

Moreover, given the vast complexity of designing secure CPPSs, systems integrators need a highly efficient security risk identification method that leverages the data and models that emerge during the engineering process.

The article at hand aims to remedy these pressing issues. Building upon prior work [7], the QualSec method presented in this article interprets the interlinking of PPR engineering information to automatically identify

(i) critical quality characteristics of products,
(ii) attack steps to compromise them, and
(iii) the resulting consequences on the production process.

Since the method can be seamlessly embedded into existing toolchains and makes direct use of already available engineering knowledge contained in AutomationML artifacts, the effectiveness and efficiency of the security risk identification step can be raised significantly. Further, the method automatically generates Petri nets (PNs) that model the sequence of manufacturing processes in a quality-oriented way, allowing to employ reachability analysis that supports risk identification.

The main contributions of this work are as follows:

1) We propose QualSec, that is, a quality-driven method for the automated identification of security risks sourced from engineering models of CPPSs. QualSec draws upon PPR information, including the sequences of manufacturing steps, to thoroughly inform about security risk sources and consequences.
2) We present a quality ontology that contains the QC domain knowledge available in production systems engineering (PSE). This ontology enriches semantics-based security risk assessments and can be interlinked with other ontologies to build knowledge graphs (KGs) for security applications.
3) We introduce the notion of a quality-oriented Petri net (QOPN) to represent the relationships between manufacturing operations, QC steps, and cyberattacks.
4) We provide an open-source implementation of QualSec, test its practicality by conducting a case study, and analyze its scalability via a rigorous performance evaluation.

To the best of our knowledge, this is the first work that considers the relationship between quality and security in a risk identification context with an emphasis on the PPR concept, making it highly relevant to the industrial informatics and information security communities.

The rest of this article is organized as follows. Section II provides background information and discusses related work. In Section III, we motivate the need for quality-driven security risk identification and define the scope of QualSec. Then, in Section IV, we explain the details of our novel method. Section V demonstrates the benefits and practicality of the introduced method by means of a case study. After that, in Section VI, we discuss the results of our performance evaluation. Finally, in Section VII, we conclude our work and give an outlook on future research.

II. Background and Related Work

In this section, we briefly review background information on AutomationML and QC in the context of cyber-physical systems (CPSs) and discuss related work on supporting security-aware CPPS engineering.

A. AutomationML

The Automation Markup Language (AutomationML, hereafter abbreviated as AML) is an XML-based data format that aims to improve the data exchange among heterogeneous engineering tools [8]. This format harmonizes and unifies data models of different engineering disciplines by integrating the Computer Aided Engineering Exchange (CAEX) data format, COLLADA, and PLCopen XML to enable modeling of the topology, geometry and kinematics, and behavior and sequencing of the CPPS [9]. The reason for utilizing AML artifacts to implement the automated identification of quality-driven security risks in CPPSs is threefold: First, AML has been standardized in the IEC 62714 series and gained wide acceptance within the CP(P)S engineering community, many of whom have joined the AutomationML association[1] to develop the format further. Second, the scope of AML far exceeds the mere exchange of information by enabling a model-based engineering approach [10]. Third, the PPR concept fits naturally into the AML architecture as a way of structuring plant models [6]. Thus, the interlinking of information regarding products (e.g., features, quality requirements), processes (e.g., sequencing of manufacturing steps), and resources (e.g., physical and logical objects, networks) can be directly harnessed for risk assessment purposes.

B. Role of QC in CPS Security

Surprisingly, little scholarly work has focused on QC in the context of CPS security thus far. However, of the few works published in this area, we consider the papers by Elhabashy et al. [11], [12] to be most relevant to the article at hand. Elhabashy et al. [11] proposed a taxonomy of cyber-physical attacks involving QC systems, which is composed of

(i) attack objectives,
(ii) targeted components,
(iii) attack methods, and
(iv) attack locations.

Their subsequent work [12] reveals that QC systems may have numerous potential vulnerabilities and shortcomings that attackers can passively exploit (i.e., without changing the QC

[1][Online]. Available: https://www.automationml.org

systems themselves). The findings presented in [11] and [12] highlight the importance of adopting a QC perspective when assessing security risks and, therefore, strongly motivate the proposed method.

An interesting observation reported by Wells et al. [13] was that there is a significant need to raise awareness about cyberattacks that have an adverse effect on product quality. Their finding suggests that security needs to be firmly established in the engineering and quality improvement process to become a natural part of the engineer's work. For this reason, our method is designed to allow tight integration into the engineering environment.

Other works, such as [14], [15], analyze sabotage attacks in additive manufacturing (AM) processes. Sturm et al. [14] explored different attack vectors in AM that cybercriminals may use to trick systems into producing faulty products. In particular, they conducted a case study to investigate how `STL` files can be manipulated such that voids inside the produced parts are created. Sturm et al. [14] accentuated that void attacks in AM are typically difficult to detect and may cause a loss of structural integrity. Belikovetsky et al. [15] demonstrated a complete attack scenario involving an AM process, targeting the 3D-printed propellers of a quadcopter. This attack is particularly devious, as the introduced defects remain unnoticed by basic quality checks and cause a critical failure after a certain amount of operating time. Both publications simulate realistic threat scenarios that challenge the state of how product quality issues can be mitigated in the event of an attack, thereby motivating a quality-driven consideration of security risks in CPPSs.

C. Model-Based Security Risk Identification in Cyber-Physical Systems

Several model-driven, risk-based approaches have been proposed in the past years that aim to support the engineering of secure CPPSs. In the following, we briefly summarize the most relevant works.

In [16] and [17], Apvrille and Roudier present an extension for the Systems Modeling Language (SysML) named *SysML-Sec*, which facilitates the model-driven design of safe and secure (sub-)systems (e.g., embedded systems). This extension enables users to incorporate security and safety properties into SysML models, which can then be validated by means of formal verification and simulation. Another security extension for SysML was introduced in [18], which focuses primarily on the architectural aspects of industrial control systems (ICSs), such as CPPSs, rather than the design of the systems' individual components (e.g., a controller). Lemaire et al. [19], [20] have further improved the security-aware, model-based engineering of ICSs by utilizing a formal reasoning framework to automate the identification of security risks in SysML models.

Besides SysML, researchers have also investigated AML for the purpose of extracting relevant information from CPPS blueprints to automate security risk assessments. In [21]–[23], a knowledge-based approach was introduced that applies security rules to AML artifacts in order to discover vulnerabilities in engineering models. These rules were created based on security domain knowledge [24] and modeled with the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL). In this context, it is worth noting that their approach directly accesses the engineering data in AML without converting it to OWL.

Recently, Eckhart et al. [7] proposed a new method that further advanced this research area. Their method employs an AML-to-OWL transformation mechanism, enabling semantic interlinking, and the use of semantic technologies (e.g., applying semantic reasoning to infer new knowledge). In this way, the method is able to identify threats, vulnerabilities, and consequences automatically by executing a set of queries and rules, which were written in the SPARQL Protocol and RDF Query Language (SPARQL) and the Shapes Constraint Language (SHACL), respectively. The results of the risk identification then serve as an input for the automated generation of attack graphs, which visualize the most critical paths adversaries may take when launching cyberattacks against CPPSs. In this article at hand, we build upon the approach described in [7] to automate the identification of quality-driven security risks in CPPS engineering models.

Finally, it is worth noting that researchers have also applied Petri nets (PNs) for security analysis purposes [25], [26]. Henry et al. [27], [28] employed PNs for attack analysis in the context of ICSs. The authors of [27], [28] then use coverability analysis to determine the extent to which an adversary can gain unauthorized access to resources. Ten et al. [29] used generalized stochastic Petri nets (GSPNs) as part of a framework that aims to quantify the vulnerability of power systems. In comparison to [27]–[29], our proposed method has a clear focus on the quality aspects of the produced parts and provides a significant level of automation in terms of risk identification.

III. Considered Attack Scenario and Scope of QualSec

The attack model considered in the article at hand assumes resourceful adversaries capable of remaining under the radar until defective products caused by intentional sabotage slip through QC and are shipped to customers. Based on a casual review of past cyberattacks against CPPSs, we sketch a realistic scenario in which threat actors either gain their initial foothold within the business network and then pivot to the control system network or directly gain unauthorized access to control devices via unprotected remote maintenance services. Furthermore, we assume that adversaries attack the CPPS at the weakest point they can find, which commonly coincides with exploiting publicly known vulnerabilities. The objective of attackers is to compromise manufacturing systems during operation in order to cause product quality issues deliberately. From an attacker's perspective, overcoming QC that functions as a defense against such attacks can be achieved in two ways: Either by manipulating the products' quality characteristics selectively, affecting only those which are not subject to quality inspection, or by exploiting QC vulnerabilities [12] to avoid detection of malicious product alterations.

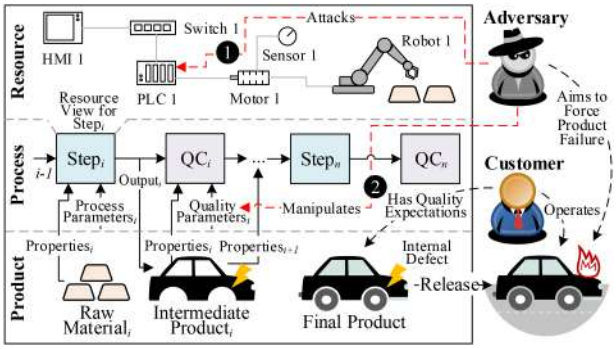


Fig. 1. Attack scenario targeting the products' quality.

Fig. 1 illustrates an example scenario in which an adversary attacks a vulnerable programmable logic controller (PLC) ① in a car manufacturing process. Since the compromised PLC controls a spot welding robot, the adversary can induce subtle changes in the welds, resulting in loss of product integrity (e.g., poor durability of the produced car body) and eventual failure of the vehicle. The consequence of this cyber-physical attack remains undetected throughout the manufacturing process as subsequent inspection for the purpose of QC can be evaded. The reason for this is that QC systems are typically not designed to uncover issues that have been created with malicious intent [13]. Even if the implemented QC checks would detect malicious product changes, an adversary may also exploit the QC systems to manipulate quality parameters (e.g., inspection locations, thresholds) ②, ensuring that any modifications go unnoticed [12]. Furthermore, increasing the defect rate and disrupting production processes constitute additional attack objectives that adversaries may pursue [11].

Our novel method, named QualSec, aims to automate tasks of the risk identification step that are carried out as part of security risk assessments during the engineering of CPPSs. One of its core features is to incorporate the semantics, structure, and sequence of the manufacturing process to identify

(i) product quality characteristics that attackers may compromise, and
(ii) possible propagation effects thereof.

To illustrate the scope and purpose of our contribution, we define the following set of questions.

Q1 *What are the security vulnerabilities in assets of CPPSs that threats may exploit?*
The first question aims to uncover architectural security weaknesses and vulnerabilities in systems that are intended to be integrated into the plant topology. Answers to this question build upon public sources, such as Common Vulnerabilities and Exposures (CVE), security advisories, and industrial security standards and guidelines. We repurpose the method presented in [7] to enable a quality-driven consideration of cyber-physical risk that is realized by answering the next questions.

Q2 *Given a set of vulnerable assets, which quality characteristics of the workpiece or product can attackers deliberately alter, and would these defects remain undetected due to insufficient QC?*
Based on the answer given to Q1, this question aims to inform engineers about potential consequences on product quality that may be caused by an adversary, who exploits vulnerable assets to execute such sabotage attacks. Answers to this question provide engineers guidance on how to prioritize risks.

Q3 *What are the consequences of an attack that targets a certain quality characteristic in terms of cascading effects relating to product quality?*
Similar to the previous question, Q3 focuses on the quality characteristics that adversaries may be able to influence in the course of an attack. However, as the sequence of manufacturing steps can create dependencies among quality attributes (e.g., diameter and location of drilled pilot holes must be correct for subsequent joining), this question places special emphasis on the indirect effects of sabotage attacks. As a result, engineers can quickly spot critical quality characteristics whose malicious alteration would lead to a chain reaction.

Q4 *How can attackers disguise their malicious actions to evade QC?*
Finally, the last question addresses the case, where an adversary might attempt to attack those QC systems that would catch product defects caused by prior manipulations of quality characteristics. Informing engineers about the minimal set of assets needed to be hacked to bypass the QC in place may provide guidance on prioritizing the systems to be hardened.

IV. Method

An overview of our proposed method and its steps is shown in Fig. 2. In the course of engineering CPPSs, professionals from various disciplines design and model systems using specialized tools. The created engineering artifacts are managed in the AML format to facilitate data exchange. In step ①, engineers annotate the plant topology contained in the AML document with security- and quality-relevant information using the AML extension libraries (AMLsec and AMLqual). Step ② transforms both the plant topology and the description of the manufacturing process to OWL. Step ③ builds the Knowledge Base (KB) by connecting the semantic representation of the plant topology and production process with additional know-how from the security ontology [30], the ICS security ontology [7], the quality ontology, and linked open security data. Based on the process description contained in the KB, step ④ generates the quality-oriented Petri net (QOPN). Finally, step ⑤ automatically performs the quality-driven security risk identification by executing rules and queries against the KB and analyzing the QOPN.

Before, we explain each element of QualSec in detail, we state the assumptions that the QualSec method relies on:

1) *Risk Identification at Design Time:* As the purpose of QualSec is to reveal security risks in the CPPS during the engineering process, we only consider what the QC system can check at design time.
2) *Model of the Manufacturing Process:* It is assumed that the manufacturing process is modeled in the sequential

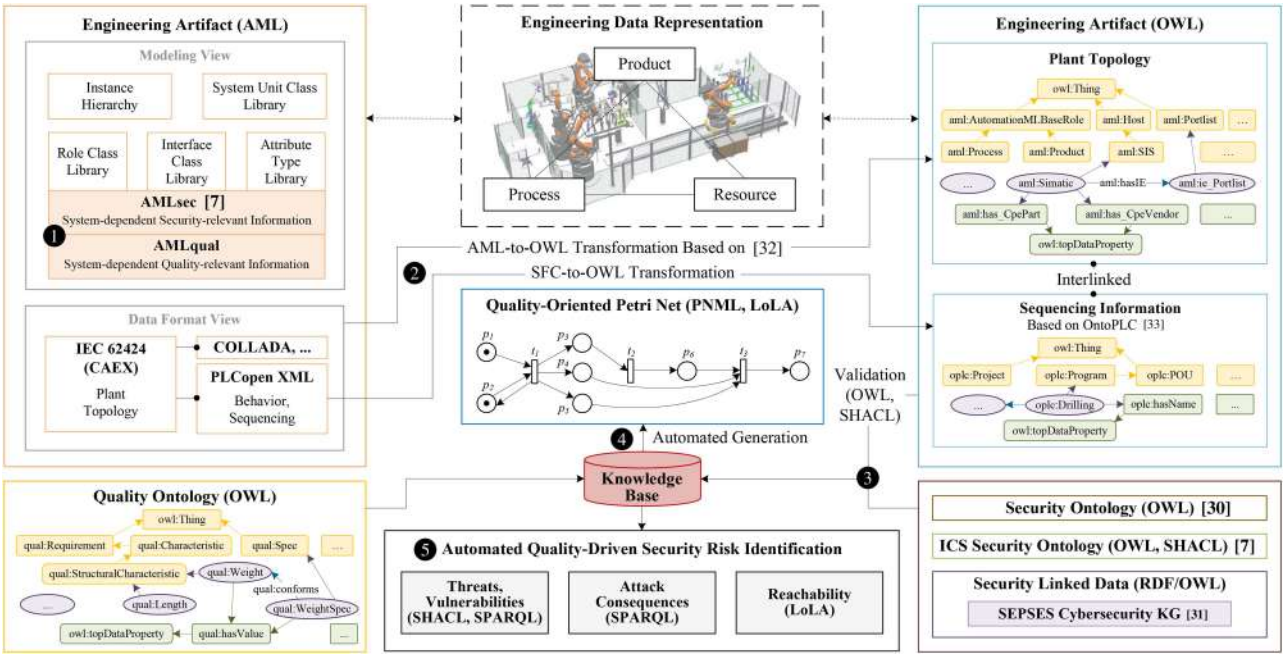


Fig. 2. Overview of QualSec, the quality-driven security risk identification method (based on [7]; robot cell illustration in [34]).

function chart (SFC) language in line with the PLCopen XML specification. To construct the QOPN, we only consider the structure of the SFC network, which can be represented graphically. Other elements of the SFC language, as standardized in the IEC 61131-3, are not relevant to QualSec.

3) *State of a System is Binary:* If an attack against a production system succeeds, it is assumed that the adversary gains full control and can manipulate all quality characteristics that the compromised system can influence during the respective manufacturing step. Similar reasoning applies to QC systems and the outcome of quality checks.
4) *Quality Measurements are Performed In-Line:* Since QualSec incorporates the description of the manufacturing process, we only consider QC efforts that are undertaken along the production line and are modeled as such. Offline quality checks could be accommodated by manually extending the semantic representation of the manufacturing process.

A. Engineering Data Representation

To lift the engineering models contained in AML artifacts to ontologies, we rely on the semantics expressed via AML's libraries of role classes (`RoleClassLib`), interface classes (`InterfaceClassLib`), and attribute types (`AttributeTypeLib`). More precisely, we link the semantics of components modeled in AML to an equivalent representation maintained in our ontologies. The normative libraries specified as part of AML are primarily used for this purpose, thereby reducing the additional modeling effort required to use QualSec. However, certain security-relevant modeling constructs that would significantly enhance QualSec's analysis capabilities are missing in those standard libraries. To overcome this limitation, we reuse AMLsec [7], which comprises libraries that engineers can apply to model security-relevant information (e.g., zones, network protocols, security devices). We carry the idea of realizing semantic matching one step further and introduce a set of libraries named AMLqual that engineers can use to augment their model with quality-relevant information. For example, `AMLqualRoleClassLib` includes, *inter alia*, role classes for QC methods (e.g., ultrasonic testing), to enrich the semantics of `InternalElements` that model the QC system.

Another vital aspect of QualSec is the interlinking of engineering information according to the PPR concept, which can be fully accommodated within the AML format [6]. According to the AML standard, links between modeled products, processes, and resources are established by using an `ExternalInterface` named `PPRConnector`, which is part of the `AutomationMLInterfaceClassLib`. Furthermore, objects within the logic model (i.e., the SFC program), which contains the sequencing information of the manufacturing process, are referenced from CAEX in the usual AML-way by using `LogicElementInterfaces`.

B. Ontological Modeling

As shown in Fig. 2, the KB is composed of the semantically lifted engineering model (i.e., plant topology and sequencing information), the (ICS) security ontology, the quality ontology, and the security-related linked data.

The CAEX-based plant topology within the AML artifact is transformed to OWL using the translation procedure of Hua

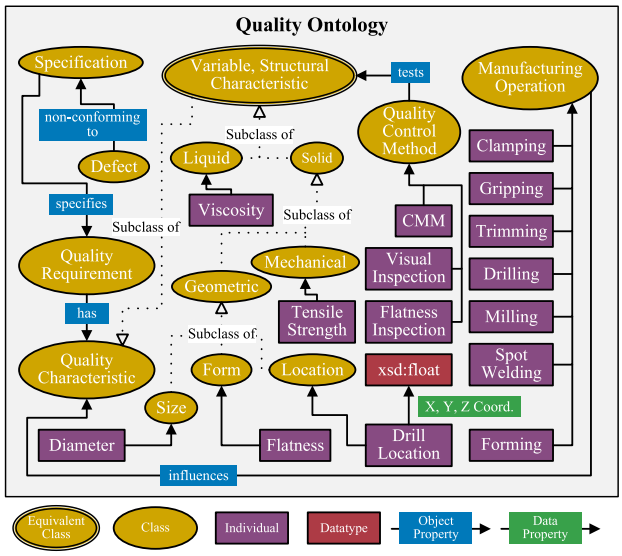


Fig. 3. Visualization of the quality ontology (excerpt).

and Hein [32]. To incorporate the PLCopen XML data into our KB, we have implemented an SFC-to-OWL transformation that instantiates an ontological model from OntoPLC [33]. After lifting the AML artifact to a semantic representation, we perform validation checks using SHACL and then automatically augment the engineering knowledge with security- and quality-specific know-how.

The structure of the security knowledge follows a layered approach, where the middle ontology layer is realized by the security ontology [30] that models rather abstract concepts within the information security domain. The ICS security ontology expands this basic knowledge with information obtained from system-independent (e.g., security standards and guidelines) and system-dependent (e.g., technical requirements of CPPSs) sources. Furthermore, the semantic data model within the KB is interlinked with the *SEPSES Cybersecurity KG* [31] in order to include the latest information on publicly disclosed security issues.

Another vital component of QualSec is the quality ontology. We have designed a comprehensive ontology for the QC domain to capture the knowledge of quality characteristics, methods to check them, and manufacturing processes that influence them (cf. Fig. 3). The rationale behind the quality ontology is to create semantic relations between the PPR information from the engineering model and QC domain knowledge. In this way, we can derive the information that is required to construct the QOPNs that enable quality-driven security risk identification.

To answer Q1, we apply a set of SHACL rules and SPARQL queries that are executed against the KB, yielding risk sources (i.e., threats and vulnerabilities) and attack consequences (i.e., violation of security or safety goals).[2] The employed vulnerability detection rules can be categorized into two classes: First, node and property shapes are used to implement a validation procedure that checks for security weaknesses in the modeled elements of the plant topology (e.g., insecure network protocols and cryptographic algorithms, configuration vulnerabilities). Second, SPARQL-based constraints are employed to detect violations of zone and conduit requirements (ZCR-3.2–3.6) as per the IEC 62443-3-2 [4]. Additionally, we perform a CVE check check by using the SEPSES Cybersecurity KG [31] to determine if the systems intended to be integrated into the plant are affected by known (public) vulnerabilities.

[2]For a more detailed description of this approach, we refer readers to [7].

C. Quality-Oriented PNs

The identification of risks to product quality and consequential events is based on the results of constructing and analyzing PNs that model manufacturing processes. The PN [35] is a well-established formalism with decades of research behind it and represents a convenient tool to model discrete event systems (DESs). In the following, we introduce the notion of QOPNs, specify a generation method for QOPNs, and explain how QOPNs can be analyzed to support the identification of security risks.

1) Preliminaries: Following the definitions given in [36], a marked PN is defined as a 5-tuple (P, T, A, w, x), where (P, T, A, w) is a weighted bipartite graph comprising a finite set of places P, a finite set of transitions T, a set of arcs $A \subseteq (P \times T) \cup (T \times P)$, and a weight function on the arcs $w : A \to \{1, 2, 3, \dots\}$. Further, x is a marking of the set of places that is associated with a row vector $\mathbf{x} = [x(p_1), x(p_2), \dots, x(p_n)] \in \mathbb{N}^n$. The marking row vector $\mathbf{x}$ defines the state of the PN and a transition $t_j \in T$ is enabled, if and only if, $x(p_i) \geq w(p_i, t_j)\ \forall p_i \in I(t_j)$, where $I(t_j) = \{p_i \in P : (p_i, t_j) \in A\}$.

Recall that QualSec incorporates a formal representation of the manufacturing process that is first translated from SFC to OWL and then processed further to construct a QOPN. The beauty of QOPNs is that they capture the dependencies among process steps, quality characteristics, and attacks against them, leading to an enhanced understanding of propagation effects.

In general, a manufacturing process consists of n production steps $o_1, \dots, o_n$ that are executed by m production systems to fulfill l jobs. Each production step o influences h characteristics of the machined part or product, which are then checked by k QC steps to determine whether they meet their stipulated quality specifications. Since the quality-driven security risk identification is performed from a process-centric point of view, the QOPN is based on the process-oriented Petri net (POPN) [37]. In a POPN, a place represents the status of a resource or job order, or an operation, while a transition denotes either the start or end of an operation [37]. The QOPN is a classical PN (P, T, A, w, x), as defined above, that extends the notion of the POPN. In Table I, we assign meaning to P and T to ensure proper interpretation of QOPNs.

It is worth reiterating that we do *not* aim to fully translate SFC programs in their complete form to PNs or one of the PN dialects. Instead, we utilize the sequencing information expressed via the SFC structure, which encodes the description of the manufacturing process, to construct QOPNs that aid security risk identification.

TABLE I
NOTATION AND SEMANTICS OF QOPNS

Places

P = $\bigcup_{i=1}^{13} S_i$, $S_i \cap S_j = \varnothing$ for all $i, j \in \{1, \ldots, 13\}, i \neq j$, where
S_1 = $\{o_1, \ldots, o_n\}$ is a set of places denoting production steps,
S_2 = $\{r_1, \ldots, r_v\}$ is a set of places denoting the status of resources (i.e., production system or QC system ready), $v = m + k$,
S_3 = $\{u_1, \ldots, u_v\}$ is a set of places denoting that resources are vulnerable,
S_4 = $\{\bar{u}_1, \ldots, \bar{u}_v\}$ is a set of places used as a complement to S_3 (i.e., resources are *not* vulnerable),
S_5 = $\{y_1, \ldots, y_m\}$ is a set of places denoting that manipulating one or multiple quality characteristics through a compromised production system has been completed,
Q_o = $\{q_1, \ldots, q_h\} \in S_6$ is a set of places denoting quality characteristics influenced by production step o,
$\bar{Q}_o$ = $\{\bar{q}_1, \ldots, \bar{q}_h\} \in S_7$ is a set of places denoting that quality characteristics, which are influenced by production step o, have been compromised,
S_8 = $\{c_1, \ldots, c_k\}$ is a set of places denoting QC steps,
S_9 = $\{a_1, \ldots, a_{k*2}\}$ is a set of places denoting whether a defect has been detected by a QC system,
S_{10} = $\{z_1, \ldots, z_k\}$ is a set of places whose user-defined markings predefine that the corresponding (benign) QC system would detect any maliciously introduced defects,
S_{11} = $\{\bar{z}_1, \ldots, \bar{z}_k\}$ is a set of places used as a complement to S_{10},
S_{12} is a set of auxiliary places to model various structures (e.g., XOR-joins), and
S_{13} = $\{s, f, d\}$, where s is a place denoting the job order status, f is a place denoting the finished product, and d is a place denoting the defects.

Transitions

T = $\bigcup_{i=1}^{7} G_i$, $G_i \cap G_j = \varnothing$ for all $i, j \in \{1, \ldots, 7\}, i \neq j$, where
G_1 = $\{\alpha_1, \ldots, \alpha_{n*2}\}$ is a set of transitions denoting the start or end of a production step,
G_2 = $\{\beta_1, \ldots, \beta_{k*3}\}$ is a set of transitions denoting the start or end of a QC step (includes two variants of the end step to cover defect and no defect conditions),
G_3 = $\{\gamma_1, \ldots, \gamma_m\}$ is a set of transitions denoting attacks against production systems,
G_4 = $\{\delta_1, \ldots, \delta_{k*2}\}$ is a set of transitions denoting whether a QC system successfully detected a defect ($\delta^{\dagger}$) or failed to detect it ($\delta^{\ddagger}$) assuming that neither the QC system nor any quality characteristic under test was compromised beforehand,
G_5 = $\{\epsilon_1, \ldots, \epsilon_{k*2}\}$ is a set of transitions denoting whether a QC system detected a defect ($\epsilon^{\dagger}$) or did not detect it ($\epsilon^{\ddagger}$) after a quality characteristic was compromised (yet, the QC system itself remained intact),
G_6 = $\{\zeta_1, \ldots, \zeta_{k*2}\}$ is a set of transitions denoting whether a compromised QC system was manipulated in a way to suppress the detection of a maliciously introduced defect ($\zeta^{\dagger}$) or to detect a non-existent defect with the objective to waste material ($\zeta^{\ddagger}$), and
G_7 is a set of auxiliary transitions (similarly to S_{12}).

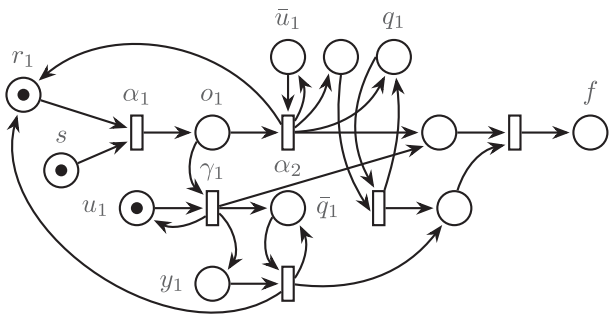


Fig. 4. Minimal QOPN (unlabeled nodes $\in S_{12} \cup G_7$).

2) Modeling and Construction: A QOPN is composed of one or multiple QOPN templates that are assembled according to the formal process description at hand. To achieve a valid QOPN, the SFC model to be transformed must at least contain the sequence *Initial Step* $\rightarrow$ *Production Step* $\rightarrow$ *Terminal Step*, which leads to the template shown in Fig. 4.

The minimal QOPN depicted in Fig. 4 contains only one quality characteristic, q_1, and is shown in its initial state, where

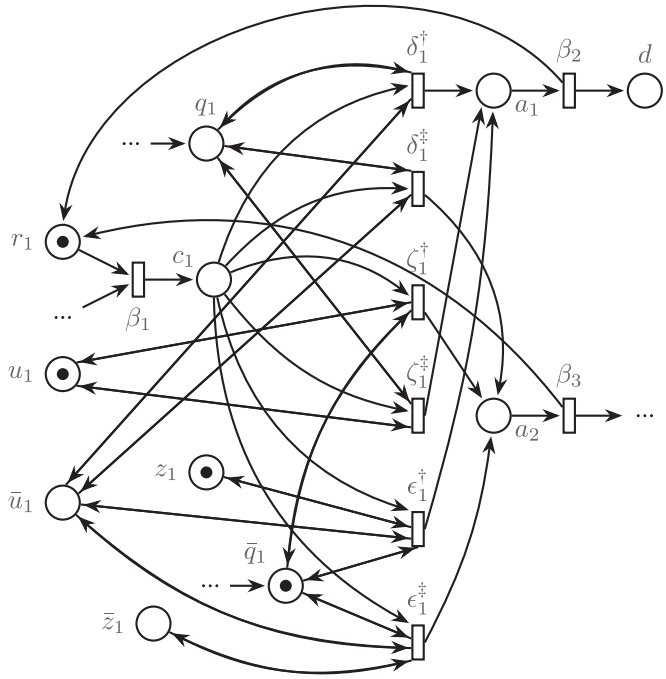


Fig. 5. QOPN template for a QC step with a single quality characteristic under test.

$x(u_1) = 1$ and $x(\bar{u}_1) = 0$ (the complement of $x(u_1)$) were specified arbitrarily for demonstration purposes. Note that the initial state of the generated QOPN depends on prior results of the vulnerability analysis (in particular, to denote vulnerable resources) and optionally on user input (e.g., to predefine the outcome of a QC step). Furthermore, a sequence of manufacturing operations may be followed by one or multiple QC steps to check whether the involved quality characteristics meet the specified requirements. This case is covered by the QOPN template shown in Fig. 5. Owing to the sets G_4, G_5, G_6, and the PN structure given in Fig. 5, various attack scenarios involving QC systems can be modeled. Again, the QOPN depicted in Fig. 5 includes only one quality characteristic, and $x(u_1) = x(z_1) = x(\bar{q}_1) = 1$, as well as $x(\bar{u}_1) = x(q_1) = x(\bar{z}_1) = 0$, were specified arbitrarily for the purpose of illustrating the PN structure.

The templates were designed to ensure boundedness of the constructed QOPN, that is, $\forall \mathbf{x} \in \mathsf{Reach}(QOPN), \forall p \in P : x(p) \leq \kappa$, where $\mathsf{Reach}(QOPN)$ is the reachable state set of the QOPN and κ is a positive number. This property is an essential requirement for applying reachability-based analysis techniques due to the fact that the PN's reachability graph must be finite.

3) Analysis: To answer Q2 and Q3, we reformulate these questions as reachability queries on QOPNs in Computation Tree Logic (CTL). The formulae for checking the desired reachability properties are expressed as EFϕ, where the state predicate ϕ takes the following forms:

Q2 $((\exists s \in \mathcal{S} : \lambda(s) > 0) \wedge (f > 0))$, where $\mathcal{S} = \{u \in S_3 \mid x(u) = 1\}$ and λ is a relation from $\mathcal{S}$ to S_7. Informally, we describe this reachability problem as follows: Is it possible that the manufacturing process

finishes without detected defects, even though some quality characteristics were compromised by exploiting vulnerable assets? After checking reachability, we analyze and filter the witness states to obtain a subset of S_7 that provides an answer to this question.

Q3 $((\exists s \in \mathcal{S} : \lambda(s) > 0) \wedge (f > 0))$, where $\mathcal{S} = \{\bar{u} \in S_4 \mid x(\bar{u}) = 1\}$ and λ is a relation from $\mathcal{S}$ to S_7. This reachability problem can be understood as checking if the manufacturing process may finish without detected defects, while some quality characteristics were indirectly compromised by exploiting vulnerable assets in preceding manufacturing steps. Similarly to Q2, we process the witness states after reachability checking to answer this question.

Q4 cannot be answered with a single reachability query and requires an iterative procedure, as shown in Algorithm 1. This algorithm takes a generated QOPN as input and produces a set U', which is a proper subset of S_3 containing places that correspond to resources of QC systems that need to be vulnerable and successfully compromised to evade quality checks. After initializing the result set U' and the set $\mathcal{T}$ that will contain transitions demonstrating the execution path starting from the initial marking, the state predicate ϕ is defined. Since we want to check if there is an execution path, where a product defect is found during a QC inspection, we define the state predicate such that the number of tokens on the place d denoting the detected defects is greater than zero. Based on this, the reachability query is expressed in CTL as the following formula: $\mathrm{EF}(d > 0)$. As long as there is a reachable state satisfying ϕ, the body of the loop is executed. In line 5, $\mathcal{T}$ is filled with the witness path, which is then processed in reverse: In each iteration, it is checked if the current element in the loop is a member of ${G^\dagger}_5$ (i.e., the transition denotes the detection of a defect). In the body of the `if`-statement, we retrieve the place denoting that the resource of the QC system that detected the defect is vulnerable, add it to the result set, retrieve the complementary place (i.e., resource not vulnerable), and adapt the marking such that the QC system is now indicated as vulnerable. Note that the procedure outlined in Algorithm 1 presupposes that at least one quality characteristic can be compromised through the exploitation of a vulnerable asset employed for a production step, since an answer to Q4 should reveal which QC system(s) an adversary would need to manipulate in order to conceal introduced product defects.

Algorithm 1: Reachability Analysis for Q4.

Input: A QOPN $N \leftarrow (P, T, A, w, x)$
Result: A subset of places of N corresponding to resources that need to be compromised in order to disguise an attack on product quality $U' \subset S_3$

```
1  U' ← ∅                                   // result set
2  T ← ∅                                    // witness path set
3  φ ← (d > 0)                              // state predicate
4  while N satisfies EFφ do
5      T ← GetWitnessPath()
6      for i ← |T| to 1 do
7          if T(i) ∈ G†5 then
8              u_i ← GetResourcePlace(T(i))
                                            // vulnerable resource
9              U' ← U' ∪ {u_i}
10             ū_i ← GetComplementaryPlace(u_i)
11             x(u_i) ← 1; x(ū_i) ← 0
12             break
```

D. Implementation

We created the AMLqual libraries with the AutomationML Editor.[3] The quality ontology was modeled with Protégé[4] [38]. Since we build upon the results of Eckhart et al. [7], we have extended their prototype to incorporate our quality-driven risk identification method. In particular, we have implemented the SFC-to-OWL translation, the QOPN construction, and the export to Petri Net Markup Language (PNML) and LoLA file formats in Scala. To conduct reachability analyses, which is an integral part of QualSec, we utilize LoLA 2 [39], [40].

AMLqual, the source code of the implemented prototype, and the AML files used for the case study are publicly available on GitHub.[5]

V. Case Study

This section presents the results of a case study that was conducted to showcase QualSec. The engineering data used in the case at hand is based on the official AML example of a robot cell [34], which aims to demonstrate how AML can be used to model the topology, behavior, and geometry of a robotic spot welding cell. To obtain a more comprehensive model, we extended these artifacts in the following ways:

(i) A description of a stamping process was integrated into the existing SFC (which only models the sequence of joining activities).
(ii) The plant topology was supplemented with PPR relations and communication-related information.
(iii) IT/OT assets were populated with system-dependent, security- and quality-relevant information using AMLsec and AMLqual.

The process considered in the case study comprises activities of vehicle manufacturing. More precisely, we focus on the stamping and joining processes for the inner front door panel, which represent a crucial part of the body in white (BiW) production line. It is evident that the structural characteristics of closures strongly influence the quality of the complete BiW; hence, conducting a quality-driven security analysis already during the engineering of the involved CPPS is prudent.

Fig. 6 illustrates the manufacturing steps from a PPR-centric perspective, where the process view is modeled in the SFC language. Due to space limitations, we cannot present an illustration of the plant topology considered in the case study. We, therefore, refer readers to the web version of the figure.[6]

[3][Online]. Available: https://www.automationml.org/download-archive
[4][Online]. Available: https://protege.stanford.edu
[5][Online]. Available: https://github.com/sbaresearch/amlsec
[6][Online]. Available: https://github.com/sbaresearch/amlsec/blob/master/appendix/qualsec/plant-topology.pdf

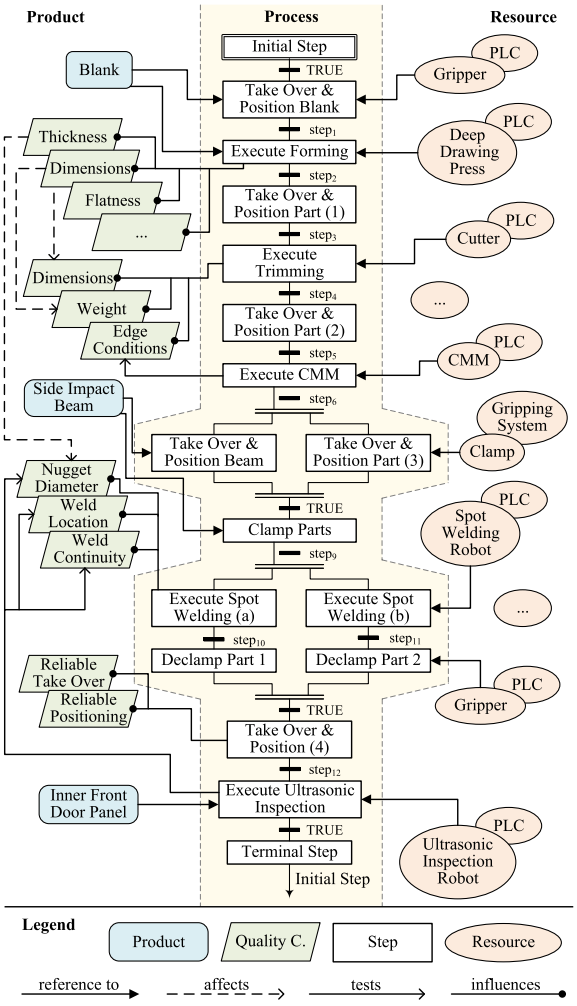


Fig. 6. PPR-centric view of the manufacturing process considered in the case study (gripping and positioning operations are combined into a single step for the sake of brevity).

TABLE II
EXCERPT OF THE QUALSEC ANALYSIS RESULTS

Legend:
vulnerable asset
■ pertains operation or QC
▲ compromised
⊛ affected by compromised qual.
covered by QC
QC evasion

Step	Asset		Quality Characteristics: Edge Conditions	Dimensions	Weight	Thickness	Nugget Diameter	Weld Continuity
Forming	S71516F_7	●	▲	▲	▲	▲		
Trimming	S71518_1	○	■	⊛	⊛			
CMM	S71518_2	○	■					
Spot Welding (a)	KRC4_1	○				⊛	■	
Spot Welding (b)	KRC4_2	○					■	
Ultrasonic Insp.	S71516F_11	●					■	

A. Results

In the following, we describe the most important results that we obtained by executing the QualSec prototype with the described input of the case study:

Q1 The results of the threat, vulnerability, and consequence identification indicate that 47 of the 370 assets of the plant topology (67 of which have the class OTComponent) have 193 vulnerabilities that may be exploited by 9 distinct threats, possibly leading to 80 consequences.

Q2 The CVE check revealed that the PLC `S71516F_7`, which controls the deep drawing press, has a known vulnerability. If this vulnerable asset is compromised, the sheet metal forming step could be influenced to manipulate several quality characteristics of the stamped pieces, including their thickness, dimensions, and edge conditions. Possible defects resulting from this attack would remain undetected because the employed coordinate-measuring machine (CMM) only tests the edge conditions of the trimmed pieces.

Q3 Since the subsequent manufacturing operation relies on the correct dimensions of the formed blanks, an attack launched against the deep drawing press could also affect the dimensions and weight of the trimmed parts. Furthermore, an incorrect blank thickness would require a different size of the weld nugget formed as part of the spot welding step. Although the nugget diameter is checked through ultrasonic testing, the PLC `S71516F_11` controlling the spot welding quality inspection robot is vulnerable and can therefore be circumvented if successfully attacked.

An excerpt of these results is displayed in Table II. In a second iteration, the plant topology has been adapted based on the answers to Q1–Q3 given above to make the CPPS more resilient. More specifically, the vulnerability in `S71516F_11` has been mitigated and the CMM now also tests the dimensions of the stamped and trimmed parts.

Q4 To validate if the performed adaptations yield a security improvement, we execute the reachability analysis outlined in Algorithm 1, intending to identify those QC assets that potentially detect malicious product changes. The results showed that an attack against the deep drawing press could only be disguised by compromising the PLCs `S71518_2` and `S71516F_11`, which control the CMM and ultrasonic testing robot, respectively. Ideally, these devices are hardened to detect attacks that target the product quality.

B. Discussion

In the following, we reflect on the results of the case study and critically evaluate the usefulness of QualSec. To this end, we briefly reiterate the gaps in the literature and analyze how well QualSec achieves its goals to address them.

1) *Efficient Security Risk Identification:* Systems integrators are in need of a method that assists engineers in addressing security issues during the integration phase [2]. Our work is based on [7], which represents a first step toward a fully automated identification of security risks using engineering data. We improved the method proposed in [7] by incorporating the model of the manufacturing process

(i.e., sequencing information) into our KB to enrich its results. In this way, the security vulnerabilities identified for answering Q1 can be associated via PPR links to individual steps of the manufacturing process, which may support risk analysis and risk evaluation. However, note that the vulnerability analysis operates at the plant topology level. This limits the scope of analysis to the plant model and public sources (e.g., industrial security standards, advisories, CVEs). Furthermore, we only consider the structure of SFC programs to construct PNs (more specifically, QOPNs), whereas other transformation techniques (e.g., [41]) provide more comprehensive coverage.

2) *QC and Security:* One of the first serious discussions of the relationship between QC and CPS security appeared in 2018 when Elhabashy et al. [11] proposed a cyber-physical attack taxonomy featuring a QC perspective. In a later work [12], they identified weaknesses in QC systems that adversaries might exploit to conceal the physical effects of attacks. Both works [11], [12] emphasize the necessity of taking QC aspects into account when designing CPPSs in order to make them more resilient to such attacks. QualSec aims to address this need by providing a risk-based approach that helps engineers better understand the impact of potential cyber-physical attacks in terms of product quality. The answers to Q2 and Q3 obtained through QualSec allow users to pinpoint compromised quality characteristics of workpieces in attack scenarios and analyze the dependencies among them. The method's results also indicate under which conditions the QC systems included in the plant topology could potentially detect malicious product changes. Since QualSec is intended to be used as a risk identification tool by systems integrators, its assessment scope is limited to the hierarchical structure of the plant, and it assumes the reasonable worst case. In other words, the presented method was not specifically designed to identify security issues in fine-grained system models (e.g., described in SysML) that would allow for a meaningful representation of vulnerability preconditions and postconditions. Thus, QualSec neglects the product supplier perspective entirely.
3) *What-If Scenarios:* Engineers can use QualSec as a planning tool to perform what-if analyses that allow a safe simulation of attack scenarios involving malicious quality loss. QualSec's results for Q4 help defenders to determine potential chokepoints in the designed QC program that would allow adversaries to bypass QC systems if they are not adequately secured.

VI. Performance Evaluation

The performance and scalability of the prototypical implementation were measured through multiple tests that were carried out using different-sized engineering models (cf. Table III). The smallest dataset (A) corresponds to the engineering model that was used for the case study, which contains the plant topology for one site[7] and the corresponding logic model depicted in Fig. 6. For datasets B–F, we expanded the base model by increasing the number of sites (Vienna `InternalElement`) and the process description (SFC) in steps of two.

TABLE III
Overview of the Datasets Used for the Evaluation

	A	B	C	D	E	F
Engineering Data						
`InternalElements` (in K)	0.87	1.74	3.49	5.23	6.97	8.71
AML Size (in MB)	1.00	2.00	4.00	6.00	8.10	10.10
Steps in SFC	23	44	86	128	170	212
After AML & SFC Trans.						
Triples (in K)	18.96	34.01	64.09	94.17	124.26	154.34
Knowledge Base Size (in MB)	2.20	4.00	7.60	11.20	14.80	18.40
After Method Execution						
Triples (in MM)	0.06	0.12	0.32	0.60	0.97	1.42
Knowledge Base Size (in MB)	5.50	11.90	30.70	57.40	92.00	134.60
QOPN Places (in K)	0.23	0.46	0.91	1.36	1.81	2.26
QOPN Transitions (in K)	0.12	0.24	0.47	0.71	0.95	1.18
QOPN Arcs (in K)	0.75	1.50	3.00	4.50	6.00	7.50
Assets (in K)	0.37	0.74	1.48	2.22	2.95	3.69

We measured the execution time of 60 experiments that were conducted by performing five runs per dataset with two cluster configurations. The first cluster consisted of the following three nodes: Node 1 hosted the triple store (Apache Jena Fuseki), a database for storing events (Apache Cassandra), and actors to provide a front-end and manage work items. Nodes 2 and 3 were used to run the work executor actors that perform the actual QualSec method. The second cluster consisted of two additional work executor nodes (i.e., five nodes in total). All nodes of both cluster configurations were cloud-hosted virtual machines running Fedora 35 x64 with 16 vCPUs and 32 GB RAM.

Fig. 7 summarizes the performance evaluation. In Fig. 7(a), we show the average execution time of the main steps of the setup phase (viz., AML-to-OWL transformation, SFC-to-OWL transformation, and model augmentation), the generation of the QOPN, and the reachability analyses for answering Q2–Q4. Note that these reported measurements were made with both cluster configurations (i.e., 10 runs per dataset) since the respective tasks were not processed in parallel by multiple work executor actors. The average execution time for the risk identification logic and the QualSec method in total are plotted per cluster setup in Fig. 7(b) and (c), respectively.

Building upon earlier work [7], we answer Q1 by executing a set of SPARQL queries and SHACL rules. Consequently, the performance of the threat, vulnerability, and attack consequence identification depends on the following factors:

1) The implementation of the SPARQL, SHACL, and inference engines.
2) The executed queries and rules.
3) The size and structure of the semantic data.

As can be seen from Fig. 7(b), scaling out the QualSec application with additional work executor nodes in a cluster

[7]See footnote 6.

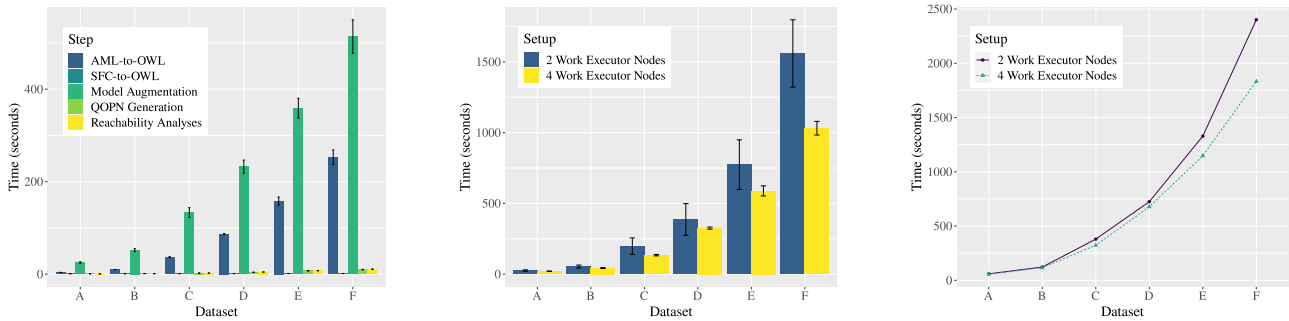


Fig. 7. Performance assessment results of our implemented prototype (error bars indicate standard deviations). (a) Setup, QOPN generation, case study. (b) Validation and risk identification. (c) Total.

can yield considerable performance improvements, especially for larger datasets.

Due to the fact that answering Q2–Q4 necessitates the construction of reachability graphs, the presented method suffers from the well-known *state explosion problem* [42]. Thus, albeit the reachability graphs are finite given the boundedness of QOPNs, the size of the state space can be unmanageable. Increasing the practicality of reachability analysis of PNs is a long line of research that has spawned various techniques to reduce the state space (e.g., stubborn sets [43]). LoLA [40] implements, inter alia, partial order reduction (the stubborn set method) and symmetry reduction, which can also be applied in combination [44]. We observe that the state space reduction techniques implemented in LoLA [40] alleviate state explosion, at least to the extent that Q2–Q4 can be answered within reasonable time (avg. 0.51 ± 0.03 s for dataset A). In fact, as can be seen from Fig. 7(a) and (b), the execution time of the QOPN generation mechanism and reachability analysis to answer Q2–Q4 is negligible compared to the security risk identification phase that answers Q1.

VII. Conclusion

In this article, we have presented a method named QualSec that automates the identification of security risks pertaining to CPPSs based on engineering data. The novelty of QualSec was that it stimulates a quality-driven perspective on security that places special emphasis on the quality characteristics of the manufactured products. Our proposed method can reveal security issues in the plant topology and expose weaknesses in QC that adversaries may exploit to introduce defects during manufacturing deliberately. QualSec utilizes PPR knowledge modeled in CAEX and SFC as part of AML to create a semantic KB. Threats, vulnerabilities, and attack consequences are then automatically identified by executing several SHACL rules and SPARQL queries against the KB. Furthermore, the structure of the modeled manufacturing process was used to construct a QOPN automatically. This QOPN serves as a basis for reachability analysis to answer risk-related questions. Systems integrators can apply QualSec to initiate proper mitigation of security risks during the engineering phase. The resulting CPPSs may be more secure by design and thereby inhibit attackers from compromising the quality of manufactured goods, possibly contributing to a decline in the number of faulty products entering the market.

Further research should be undertaken to improve QualSec in the following ways: The current version of our method is intended to be used during the engineering of CPPSs and, therefore, heavily relies on the engineering data exchange format AML. However, since a QOPN is constructed based on a semantic representation of the production process, the input format does not necessarily have to be PLCopen XML. Incorporating additional sources into QualSec would extend the method's scope to cover the operation phase.

Another possible improvement of QualSec would be to increase the degree of detail of the systems' state. In this article, we make the (relatively strong) assumption that the successful exploitation of a vulnerability results in full control of the system and allows an adversary to manipulate all quality characteristics that the compromised system can influence. The rationale behind this assumption is twofold:

(i) The abstraction level of the plant model available at the engineering phase may hinder the definition of postconditions of exploiting vulnerabilities.

(ii) Users might be primarily interested in worst-case scenarios.

Nevertheless, enriching the KB may enable a finer-grained analysis of how quality characteristics can be influenced based on the privileges gained by an adversary.

There is also room for improvement with respect to the engineering data sources used for risk identification. In its current version, QualSec processes the plant topology in CAEX and the sequencing information in PLCopen XML, which are both part of AML. Utilizing COLLADA interfaces to incorporate geometry and kinematics information into QualSec appears to be an appealing extension of our work. In this way, the attack consequence identification component could be enhanced to address safety aspects more thoroughly.

Finally, we want to suggest some ideas to advance the PN-based analysis further. Probabilistic PNs may be applied to better reflect various quality inspection strategies (e.g., random sampling). Additionally, attaining a more rigorous translation

from SFC to PN, also including timing information (time PN), would be worthwhile.

Acknowledgment

The authors would like to thank Walid Fdhila for informative discussions on the submitted manuscript and Yameng An for providing the initial version of OntoPLC [33].

References

[1] M. Eckhart, K. Meixner, D. Winkler, and A. Ekelhart, "Securing the testing process for industrial automation software," *Comput. Secur.*, vol. 85, pp. 156–180, 2019.

[2] P. Kieseberg and E. Weippl, "Security challenges in cyber-physical production systems," in *Proc. Softw. Qual., Methods Tools Better Softw. Syst.*, 2018, pp. 3–16.

[3] M. Eckhart, A. Ekelhart, A. Lüder, S. Biffl, and E. Weippl, "Security development lifecycle for cyber-physical production systems," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc.*, 2019, pp. 3004–3011.

[4] *Security for Industrial Automation and Control Systems – Part 3-2: Security Risk Assessment and System Design*, Int. Electrotech. Commission, Geneva, Switzerland, Standard IEC 62443-3-2:2020, 2020.

[5] *IT-Security for Industrial Automation - General Model*, Verlag des Vereins Deutscher Ingenieure, Düsseldorf, Germany, Standard VDI/VDE 2182-1, 2011.

[6] M. Schleipen and R. Drath, "Three-view-concept for modeling process or manufacturing plants with AutomationML," in *Proc. IEEE Conf. Emerg. Technol. Factory Autom.*, 2009, pp. 1–4.

[7] M. Eckhart, A. Ekelhart, and E. Weippl, "Automated security risk identification using AutomationML-based engineering data," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 3, pp. 1655–1672, May./Jun. 2022.

[8] R. Drath, A. Lüder, J. Peschke, and L. Hundt, "AutomationML - the glue for seamless automation engineering," in *Proc. IEEE Conf. Emerg. Technol. Factory Autom.*, 2008, pp. 616–623.

[9] N. Schmidt and A. Lüder, "AutomationML in a nutshell," AutomationML e.V., Tech. Rep., Nov. 2015.

[10] S. Faltinski, O. Niggemann, N. Moriz, and A. Mankowski, "AutomationML: From data exchange to system planning and simulation," in *Proc. IEEE Int. Conf. Ind. Technol.*, 2012, pp. 378–383.

[11] A. E. Elhabashy, L. J. Wells, J. A. Camelio, and W. H. Woodall, "A cyber-physical attack taxonomy for production systems: A quality control perspective," *J. Intell. Manuf.*, vol. 30, no. 6, pp. 2489–2504, 2018.

[12] A. E. Elhabashy, L. J. Wells, and J. A. Camelio, "Cyber-physical attack vulnerabilities in manufacturing quality control tools," *Qual. Eng.*, vol. 32, no. 4, pp. 676–692, 2020.

[13] L. J. Wells, J. A. Camelio, C. B. Williams, and J. White, "Cyber-physical security challenges in manufacturing systems," *Manuf. Lett.*, vol. 2, no. 2, pp. 74–77, 2014.

[14] L. D. Sturm, C. B. Williams, J. A. Camelio, J. White, and R. Parker, "Cyber-physical vunerabilities in additive manufacturing systems," *Int. Solid Freeform Fabr. Symp.*, vol. 7, pp. 951–963, 2014.

[15] S. Belikovetsky, M. Yampolskiy, J. Toh, J. Gatlin, and Y. Elovici, "dr0wned cyber-physical attack with additive manufacturing," in *Proc. 11th USENIX Workshop Offensive Technol.*, 2017, pp. 1–16,.

[16] L. Apvrille and Y. Roudier, "SysML-Sec: A SysML environment for the design and development of secure embedded systems," in *Proc. Int. Conf. Asia-Pacific Council Syst. Eng.*, 2013, pp. 1–16.

[17] Y. Roudier and L. Apvrille, "SysML-Sec: A model driven approach for designing safe and secure systems," in *Proc. 3rd Int. Conf. Model-Driven Eng. Softw. Dev.*, 2015, pp. 655–664.

[18] R. Oates, F. Thom, and G. Herries, "Security-aware, model-based systems engineering with SysML," in *Proc. 1st Int. Symp. ICS SCADA Cyber Secur. Res.*, 2013, pp. 78–87.

[19] L. Lemaire, J. Lapon, B. De Decker, and V. Naessens, "A SysML extension for security analysis of industrial control systems," in *Proc. 2nd Int. Symp. ICS SCADA Cyber Secur. Res.*, 2014, pp. 1–9.

[20] L. Lemaire, J. Vossaert, J. Jansen, and V. Naessens, "Extracting vulnerabilities in industrial control systems using a knowledge-based system," in *Proc. 3rd Int. Symp. ICS SCADA Cyber Secur. Res.*, 2015, pp. 1–10.

[21] M. Glawe, C. Tebbe, A. Fay, and K.-H. Niemann, "Knowledge-based engineering of automation systems using ontologies and engineering data," in *Proc. Int. Joint Conf. Knowl. Discov., Knowl. Eng. Knowl. Manage.*, 2015, pp. 291–300.

[22] C. Tebbe, M. Glawe, A. Scholz, K.-H. Niemann, A. Fay, and J. Dittgen, "Wissensbasierte Sicherheitsanalyse in der Automation," *atp magazin*, vol. 57, no. 04, pp. 56–66, 2015.

[23] M. Glawe and A. Fay, "Wissensbasiertes Engineering automatisierter Anlagen unter Verwendung von AutomationML und OWL," *at-Automatisierungstechnik*, vol. 64, no. 3, pp. 186–198, 2016.

[24] C. Tebbe, M. Glawe, K.-H. Niemann, and A. Fay, "Informationsbedarf für automatische IT-Sicherheitsanalysen automatisierungstechnischer Anlagen," *at-Automatisierungstechnik,* vol. 65, no. 1, pp. 87–97, 2017.

[25] Y. Cherdantseva et al., "A review of cyber security risk assessment methods for SCADA systems," *Comput. Secur.*, vol. 56, pp. 1–27, 2016.

[26] W. Knowles, D. Prince, D. Hutchison, J. F. P. Disso, and K. Jones, "A survey of cyber security management in industrial control systems," *Int. J. Crit. Infrastruct. Prot.*, vol. 9, pp. 52–80, 2015.

[27] M. H. Henry, R. M. Layer, K. Z. Snow, and D. R. Zaret, "Evaluating the risk of cyber attacks on SCADA systems via petri net analysis with application to hazardous liquid loading operations," in *Proc. IEEE Conf. Technol. Homeland Secur.*, 2009, pp. 607–614.

[28] M. H. Henry, R. M. Layer, and D. R. Zaret, "Coupled petri nets for computer network risk analysis," *Int. J. Crit. Infrastruct. Prot.*, vol. 3, no. 2, pp. 67–75, 2010.

[29] C.-W. Ten, C.-C. Liu, and G. Manimaran, "Vulnerability assessment of cybersecurity for SCADA systems," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1836–1846, Nov. 2008.

[30] S. Fenz and A. Ekelhart, "Formalizing information security knowledge," in *Proc. 4th Int. Symp. Inf., Comput., Commun. Secur.*, 2009, pp. 183–194.

[31] E. Kiesling, A. Ekelhart, K. Kurniawan, and F. Ekaputra, "The SEPSES knowledge graph: An integrated resource for cybersecurity," in *Proc. Int. Conf. Semantic Web*, 2019, pp. 198–214.

[32] Y. Hua and B. Hein, "Interpreting OWL complex classes in AutomationML based on bidirectional translation," in *Proc. 24th IEEE Int. Conf. Emerg. Technol. Factory Autom.*, 2019, pp. 79–86.

[33] Y. An, F. Qin, B. Chen, R. Simon, and H. Wu, "OntoPLC: Semantic model of PLC programs for code exchange and software reuse," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1702–1711, Mar. 2021.

[34] AutomationML, "AutomationML example: Robot cell," AutomationML, Tech. Rep., Mar. 2017. [Online]. Available: https://www.automationml.org/wp-content/uploads/2021/06/AML_RobotCell_en_public.zip

[35] C. A. Petri, "Kommunikation mit Automaten," Ph.D. dissertation, Universität Hamburg, 1962.

[36] C. G. Cassandras and S. Lafortune, "Petri nets," in *Introduction to Discrete Event Systems*, 2nd ed. New York, NY, USA: Springer, 2008, pp. 223–267.

[37] M. Zhou and N. Wu, "Process-oriented Petri net modeling," in *System Modeling and Control with Resource-Oriented Petri Nets*, 1st ed. Boca Raton, FL, USA: CRC Press, 2010, pp. 43–55.

[38] N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Fergerson, and M. A. Musen, "Creating semantic web contents with Protégé-2000," *IEEE Intell. Syst.*, vol. 16, no. 2, pp. 60–71, Mar./Apr. 2001.

[39] K. Schmidt, "LoLA: A low level analyser," in *Proc. 21st Int. Conf. Appl. Theory Petri Nets*, 2000, pp. 465–474.

[40] K. Wolf, "Petri net model checking with LoLA 2," in *Proc. 39th Int. Conf. Appl. Theory Petri Nets Concurr.*, 2018, pp. 351–362.

[41] N. Wightkin, U. Buy, and H. Darabi, "Formal modeling of sequential function charts with time Petri nets," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 2, pp. 455–464, Mar. 2011.

[42] A. Valmari, "The state explosion problem," in *Proc. Lectures Petri Nets I: Basic Models: Adv. Petri Nets*, 1998, pp. 429–528.

[43] A. Valmari, "Stubborn sets for reduced state space generation," in *Proc. 10th Int. Conf. Appl. Theory Petri Nets*, 1991, pp. 491–515.

[44] K. Wolf, "Generating Petri net state spaces," in *Proc. 28th Int. Conf. Appl. Theory Petri Nets Models Concurr.*, 2007, pp. 29–42.

A Survey on Cyber Security Threats in IoT-Enabled Maritime Industry

Imran Ashraf, Yongwan Park, Soojung Hur, Sung Won Kim, Roobaea Alroobaea, Yousaf Bin Zikria, *Senior Member, IEEE*, and Summera Nosheen

***Abstract*— Impressive technological advancements over the past decades commenced significant advantages in the maritime industry sector and elevated commercial, operational, and financial benefits. However, technological development introduces several novel risks that pose serious and potential threats to the maritime industry and considerably impact the maritime industry. Keeping in view the importance of maritime cyber security, this study presents the cyber security threats to understand their impact and loss scale. It serves as a guideline for the stakeholders to implement effective preventive and corrective strategies. Cyber security risks are discussed concerning maritime security, confidentiality, integrity, and availability, and their impact is analyzed. The proneness of the digital transformation is analyzed regarding the use of internet of things (IoT) devices, modern security frameworks for ships, and sensors and devices used in modern ships. In addition, risk assessment methods are discussed to determine the potential threat and severity along with the cyber risk mitigation schemes and frameworks. Possible recommendations and countermeasures are elaborated to alleviate the impact of cyber security breaches. Finally, recommendations about the future prospects to safeguard the maritime industry from cyber-attacks are discussed, and the necessity of efficient security policies is highlighted.**

***Index Terms*— Maritime security, IoT, cyber security threats, vulnerability, malware.**

I. Introduction

TECHNOLOGICAL developments have shown unprecedented speed over the past decade and revolutionized many fields by incorporating novel technologies, policies, and operational procedures. Analogous to several other domains, advanced digitization, information, and operation technology have also made their way in the maritime industry.

The maritime freight-forwarding industry serves as the foundation for international trade carrying around 80% of goods globally and contributing 70% of trade value [1], [2]. Consequently, large investments from multinational companies like Maersk, IBM, and Google, etc., accelerated the revolutionization of the maritime industry. Not only that, Maersk and IBM are working on projects to commercialize the blockchain technology for digital global trade platforms [3]. Shipping automation and incorporation of intelligent systems in maritime is to be deployed by Google, and Rolls Royce [4]. Similarly, projects on digitizing the platforms are carried out under Det Norske Veritas, and Germanischer Lloyd [5]. With the digitalization of the maritime operational platforms, safe navigation, low manning requirements, and security are visioned. With a large increase in the operations of the maritime freight industry over the past decade, further, expansion is expected in the near future. Figure 1 shows the statistics of container throughput for worldwide ports for this decade, indicating a substantial increase in the throughput from 622 million twenty-foot equivalent units (TEUs) in 2012 to an expected 945 million TEUs in 2024 [6].

The maritime industry has evolved from traditional mechanical systems to electromechanical and digital systems involving changes in industrial control systems over the past decade. Consequently, the modern maritime industry operates on semi-automatic/automatic controlled systems, automated harbors, satellite communication, and navigation systems. Such systems combine sophisticated hardware and software systems operated through mobile networks involving the maritime industry stakeholders. Marine communication is carried out using board systems involving shore stations and satellites. Digital selective calling (DSC) is used for distress alerts, safety calls, and routine priority messages, digital selective calling (DSC) is used, which can be integrated with very high frequency (VHF) radios used for ship-to-ship communication. Similarly, satellite communication is used for areas where the shore stations have no coverage [7]. Maritime communication systems contain equipment and devices, a large number of which are connected to the internet or telecommunication systems [8] and can be attacked remotely using both simple as well as sophisticated cyber attacks. Predominantly, the maritime organizations are

Manuscript received 14 December 2021; revised 1 March 2022; accepted 30 March 2022. Date of publication 15 April 2022; date of current version 8 February 2023. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) by the Ministry of Education under Grant NRF-2021R1A6A1A03039493; in part by NRF Grant by the Korean Government through the Ministry of Science and ICT (MSIT) under Grant NRF-2022R1A2C1004401; and in part by the Taif University Researchers Supporting Project, Taif University, Taif, Saudi Arabia, under Grant TURSP-2020/36. The Associate Editor for this article was A. K. Bashir. *(Imran Ashraf and Yongwan Park are co-first authors.) (Corresponding authors: Yousaf Bin Zikria; Summera Nosheen.)*

Imran Ashraf, Yongwan Park, Soojung Hur, Sung Won Kim, and Yousaf Bin Zikria are with the Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea (e-mail: imranashraf@ynu.ac.kr; ywpark@yu.ac.kr; sjheo@ynu.ac.kr; swon@yu.ac.kr; yousafbinzikria@ynu.ac.kr).

Roobaea Alroobaea is with the Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia (e-mail: r.robai@tu.edu.sa).

Summera Nosheen is with the Faculty of Engineering, School of Computer Science, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: summera.nosheen@sydney.edu.au).

Digital Object Identifier 10.1109/TITS.2022.3164678

TABLE I
A SUMMARY OF CYBER SECURITY THREATS, THREAT ACTORS AND OBJECTIVES

Threat	Level	Threat actors	Objectives
Cyber vandalism	1	Hackers, vandalist, angered employees, activist	Data stealing, destroying, or public posting for media coverage.
Cyber Theft	2	Individual, small groups (political, ideological), spammers	Information, disruption of destruction of business operations, profit or ideological gains
Cyber Incursion	3	Organized enterprise, government entity, terrorist groups	Information of weaknesses, backdoor planting, access, alter or destroy information.
Cyber Sabotage	4	Organized professional organizations, military secret operatives	High-level information regarding secret R&D critical for organization/government, cracking security procedures, infiltration
Cyber Conflict	5	Government operatives, highly skilled terrorist groups, sophisticated hacker group	Infrastructure destruction, high importance mission-critical information
Cyber Internal	6	Employees, workers, third party service providers	Non-intentional mistakes, carelessness, lack of skill to open opportunities for 1 to 5 discussed threats.

not well prepared to handle cyber attacks, as pointed out in [9]. For different kinds of breaches, the preparedness varies with respect to the size and scope of the organization. For example, large companies are well prepared for data breaches which are primarily attributed to the higher ratio of data breaches that occurred in large companies. The capability of handling cyber attacks is increased for those organizations who report such attacks and devise countermeasures to prevent similar future attacks. In this regard, this study makes the following contributions

- This study conducts an extensive review of the security threats for the IoT-enabled maritime industry.
- Comprehensive background on maritime security threats space is provided where different threats, threat actors, and objectives for threats are discussed.
- Cybersecurity threats related to the maritime industry are analyzed regarding different elements of maritime infrastructure like vessels, offshore units, etc., and the onboard devices like navigation systems, data recorders, logistics, etc.
- For assessing the potential threat and risk of cyberattacks, various risk analysis methods are elaborated with their advantages and disadvantages. In addition, different threat mitigation methods are discussed.
- A brief and compact prospective discussion is provided for the shortcomings of existing defense strategies for handling the maritime risks, and future directions are outlined.

A taxonomy of the research papers covered in this study is provided in Figure 2. The rest of the study is organized in the following fashion. Section II provides the background of the cyber security threats and their various types. Maritime cyber security threats are discussed in Section III. Risk impact analysis of cyber security threats is performed in Section IV while risk mitigation schemes are given in Section V. Overview of probable threats with respect to Industry 4.0 is described in Section VI. Future research directions are provided in Section VII while the conclusion is given in Section VIII.

II. BACKGROUND ON CYBER SECURITY THREATS

Keeping in view the proneness of the electromechanical and digital systems, involving connected hardware and software components, the associated risks and threats are large and complex, necessitating the extensive evaluation of systems' vulnerabilities. The security threats become worse when geopolitical disputes and piracy attacks are considered. Maritime security threats can be broadly categorized under two groups: intentional threats and unintentional threats.

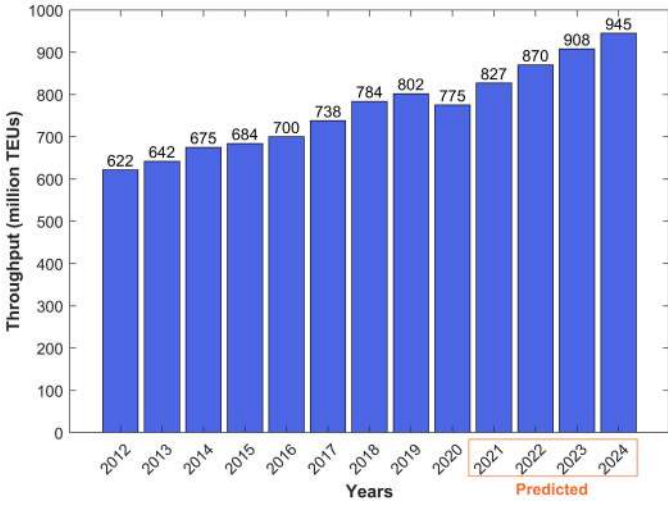


Fig. 1. Container throughput in million TEUs for maritime world wide [6].

A. Intentional Threats

Intentional direct threats are cyber security threats caused by a large number of adversaries and involve different methods and techniques.

1) Cyber Vandalism: Representing an ideological motivation, such individuals/groups steal sensitive information to exploit their target. Often inspired by different individuals, cyber vandalists, also called hacktivists, misuse the stolen data for malicious purposes, such as blackmail, extortion, and ransom, etc. [10].

2) Cyber Sabotage: Cyber sabotage, also called espionage, threats come from industry rivals and market competitors, often targeting the intellectual properties of a target company [11]. It is the planned and organized intrusion to steal confidential information, alter if it provides an institutional benefit, or destroy data/products to outwit the competitor. Espionage aims at obtaining a competitive edge by empowering own skills by stealing intellectual property or disrupting the competitors' business operations [12].

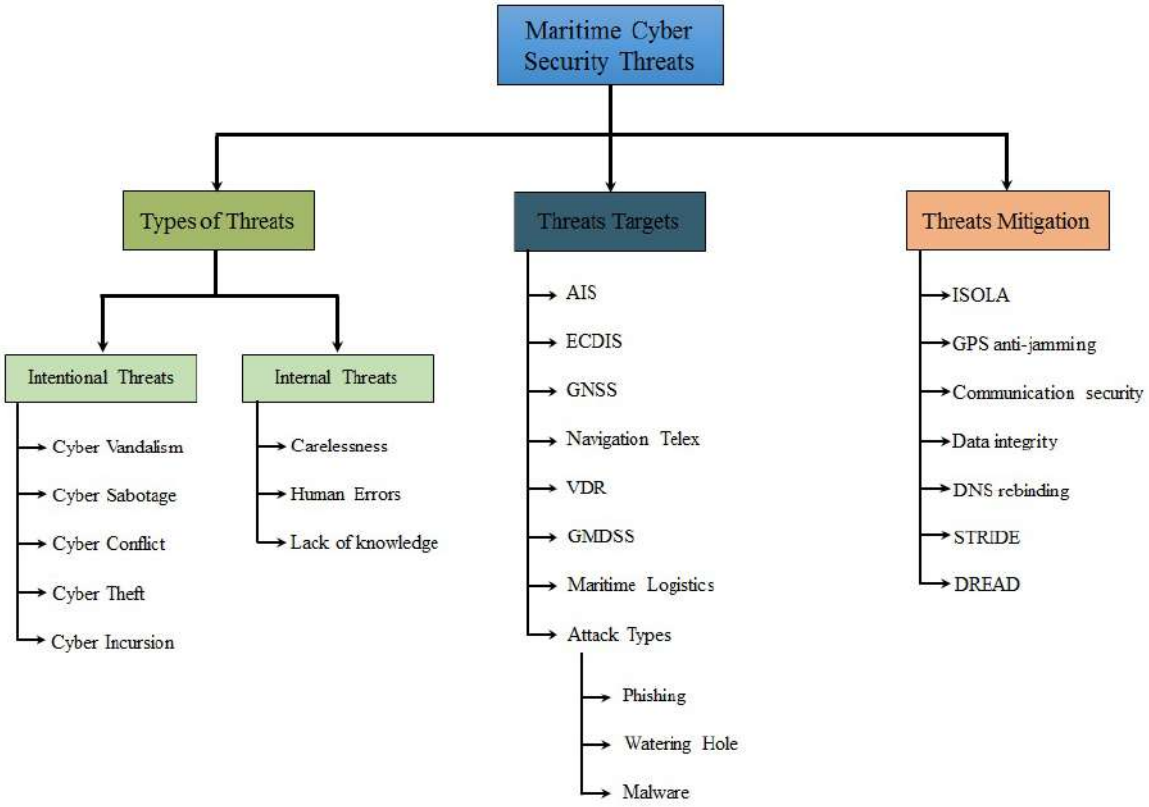


Fig. 2. Taxonomy of the papers discussed in this study.

3) Cyber Conflict: The scale and scope of the intentional attacks become wide when it is state-sponsored or government-driven. Countries may launch cyber attacks on the maritime industry of an opponent or competing country [12]. Primarily such attacks are made for obtaining state secrets and similarly other important information that may provide leverage. Similarly, secret business agreements and similar other commercial information of high importance can be targeted [13]. State-sponsored attacks are launched for economic dominance, information control, or national destabilization [14].

4) Cyber Theft: Cyber thieves, also called Terrorist groups, are often formed by certain religious, political, and social doctrines and take actions to target the opposing groups, nations, and countries. The maritime sector can also target such groups where the attacks are carried out using electronic and computerized media for obtaining unauthorized access to confidential information. Attacks are aimed at both destroying these resources, as well as using them for ransom and gaining the upper hand [12].

5) Cyber Incursion: Individuals or criminal organizations may also launch Cyber-attacks for criminal activities. Such attacks are launched for extortion, fraudulent activities, and illegal access to the intellectual property of an organization [12]. By gaining access to different controlling systems, weapons, drugs, and contraband operations are performed for economic benefits and stealing secret information for blackmail, ransom, and information selling to other groups [15].

B. Internal Cyber Threats

Besides the intentional cyber security threats for the maritime industry, the harm can be done unintentionally due to the negligence of employees or third-party service providers. Threats from internal employees can occur due to carelessness, human errors, or lack of knowledge about particular tools or procedures [16]. The intensity of internal threats varies with respect to the importance of the system being exposed to the security threat. Adversaries can misuse exposed systems to control and exploit them for secret information. Internal threats are often the outcome of improper training, lack of skills to handle a system, human judgmental error, and ignorance [12]. Third-party software and hardware systems can also jeopardize maritime security if software containing back doors, poorly tested software and error-containing systems are installed.

A schematic diagram of cyber security risks in the maritime industry and the associated risk level is portrayed in Figure 3. The number represents the risk level, with a higher number indicating the higher risk. Numbers from 1 to 6 are attributed to ‘low’, ‘moderate’, ‘high’, ‘very high’, ‘severe’, and ‘extreme’ risk for these threats. Cyber internal threats indicate the highest threat level and expose the companies to the maximum risk.

Table I provides the overview of the types of cyber security threats for the maritime industry, along with the possible threat actors and their objectives. The cyber internal threat category is ranked with the highest risk level as employees’ carelessness, lack of proper training, and knowledge may expose an organization’s infrastructure to all the threats described here.

Fig. 3. Cyber security threats and associated risk level.

III. Analyzing Cyber Security Threats Related to Maritime

Basic components of the maritime infrastructure are depicted in Figure 4 indicating three important components: vessels, ground infrastructure, and communication network. Vessels contain on-board systems such as global maritime distress and safety systems (GMDSS), maritime administrative systems, communication systems, etc., prone to different kinds of cyberattacks. Similarly, off-shore systems comprise public infrastructure, including automatic identification systems for vessels and crew managers, private service providers, off-shore security systems, etc. Different adversaries can attack to obtain unauthorized access. A schematic diagram of on-board and off-board systems is given in Figure 5.

A. Automatic Identification System Related Attacks

An automatic identification system (AIS) provides safe navigation in the sea and collision avoidance by providing navigation-related information of other ships such as ship type, course, speed, ship status (anchor, or underway), etc. AIS aims at reducing the risks of possible collisions with other ships by communicating with them. However, communication makes AIS the most vulnerable system of the ship [17], [18].

With technological advancement, the AIS data can be reproduced, and a virtual ship can be placed with false speed, heading, course, and other information to deceive other ships. Weather information can be generated and sent to other ships to change their route. AIS attacks occur due to a lack of appropriate procedures to ensure integrity and encryption protocols which makes it easy for the attackers to intercept AIS transmission [19]. For example, an Iranian oil ship used falsified AIS data and pretended to be Tanzanian to navigate to Syria [20]. Using a very high frequency (VHF), an attacker can intercept AIS transmission, tamper the AIS data to steal identity information, communicate with a ship by impersonating port authorities, block the communication with other ships, and direct the vessel to the desired location by impersonating as competent maritime authority [21], [22]. AIS can also be the target of a denial-of-service attack, fake close point to alert collision alert, and data flood by transmitting at higher frequency [22], [23].

B. Electronic Chart Display and Information System

ECIDS has been a mandatory part of the ships since January 1, 2011, and contains several important functions in hardware and software for safe navigation. ECDIS is used for displaying ships course for the crew using the bridge-placed operating system. ECDIS contains position, compass, speed, etc., and is connected to ship systems and sensors and is updated via USB or the internet. Despite being an essential part of ships, it is found to be the easy target of adversaries [24]. The primary source of malicious code execution on ECDIS is the obsolete baseline operating systems or operating systems that do not allow upgrades [25].

C. Global Navigation Satellite System

Similar to navigation at land, GNSS provides important information for safe sailing at sea through guided navigation by GPS. After the AIS, GNSS has been regarded as the most vulnerable asset in the maritime sector [25].

Spoofing and jamming are the two most prominent threats to GPS technology. Spoofing involves using the port, access control address, and internet protocol (IP) to conceal the original identity for performing malicious activities. During the jamming, the GPS signals are disrupted or disturbed by intercepting the boat's frequency. Unlike spoofing, which is an impersonating act, jamming involves electronic or mechanical intervention to disrupt radar or radio communications [26]. Jamming attacks are usually carried out by commercial devices that are low cost and easy to buy online [27]. Spoofing attacks are complex as compared to jamming, as they require simulating the satellite signals that require high power and complicated apparatus [28].

Research shows that the navigational systems are the primary target of maritime cyber attacks due to their vulnerability, followed by ECDIS and engine control [29].

D. Navigation Telex

Navigation Telex (NAVTEX) provides urgent navigation and meteorological information for safe navigation by the port authorities. The information is disseminated by telex in the ship that operates at specific frequencies, and information is available via the website as well [30]. NAVTEX is connected to the internet, storage devices, and other systems prone to attacks. Attacks may result in incorrect messages to misguide the ship and blocking the service to send the messages from the attacker to guide the ship to the location of the attacker's choice [31].

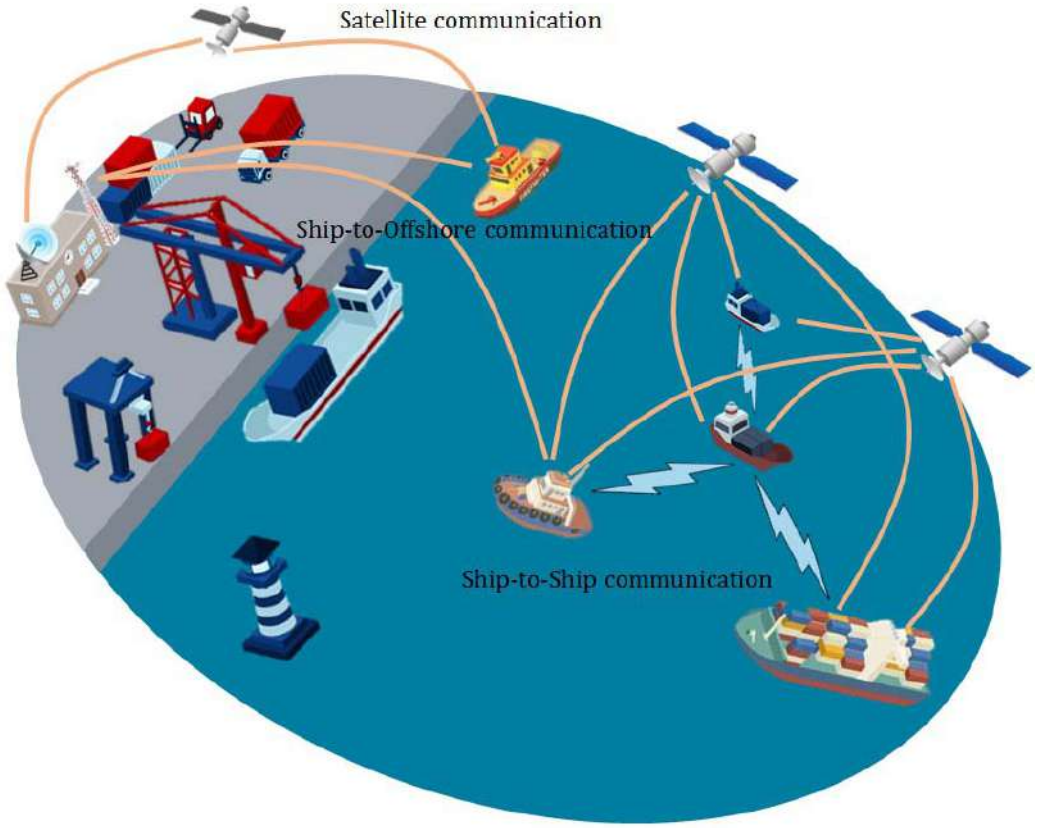


Fig. 4. Basic components of maritime infrastructure.

E. Voyage Data Recorders

A voyage data recorder (VDR) is used to store the voyage details of the ship and can be a potential tool for investigating ship accidents. It serves a similar purpose, as BlackBox does for the airplane, with superior functions. It records the speed, direction, position, conversations, etc. of the last 12 hours that can be used to analyze ship performance, accident analysis, and damage analysis. VDR is prone to intruder attacks, and the attacker needs to be inside the ship as it is connected to a local area network (LAN). Attacks happen due to inappropriate authentication mechanisms, weak encryption protocols, and obsolete firmware [32], [33]. VDR can be attacked for denial of service for obfuscation through the USB, CD, and DVD, etc. [34].

F. Global Maritime Distress and Safety System

GMDSS is the fundamental system for distress management and involves sending distress messages to shores and requesting search and rescue support. It also broadcasts the maritime safety information (MSI) for other ships in the vicinity that could help the distressed ship to a safe route []. Malware infections are targeted on GMDSS, resulting in partial damage or complete destruction. The control can also be taken to guide the ship to a designated location by the attacker. The identity of another ship can be spoofed using the GMDSS to initiate communication with other ships for influencing cargo safety. GMDSS interactions with SCC (shore control center) can be compromised to steal sensitive information of ship operations. Owing to the importance of GMDSS during emergency and rescue operations, any disruption can risk the rescue operations [35]. Similarly, jamming attacks can cause damage and denial-of-service for GMDSS [36]. To mitigate the impact of cyber attacks on GMDSS, counterpart systems are a potential solution [37].

G. Threats to Maritime Logistics Environment

With the advancements in technology, traditional supply chain and logistics systems have been transformed into supervisory control and data acquisition (SCADA) systems where the flow of goods can be remotely controlled. This infrastructure involves internet of things (IoT) platforms, satellites, and ICT procedures to control and monitor the maritime logistics and supply chain (MLSC). Consequently, SCADA infrastructure and cyber-physical systems (CPS) are prone to cyber-attacks from adversaries. MLSC systems comprise several CPS that has been the target of adversaries during the recent events [38]–[40].

SCADA systems in the current maritime sector involve interoperable components integrated with ICT systems and involve communication. Sensors and devices used for position tracking and monitoring, such as IoT sensors and cameras, satellite communication, etc., are susceptible to different cyberattacks [41]. SCADA systems can be the victim of five different kinds of cyberattacks.

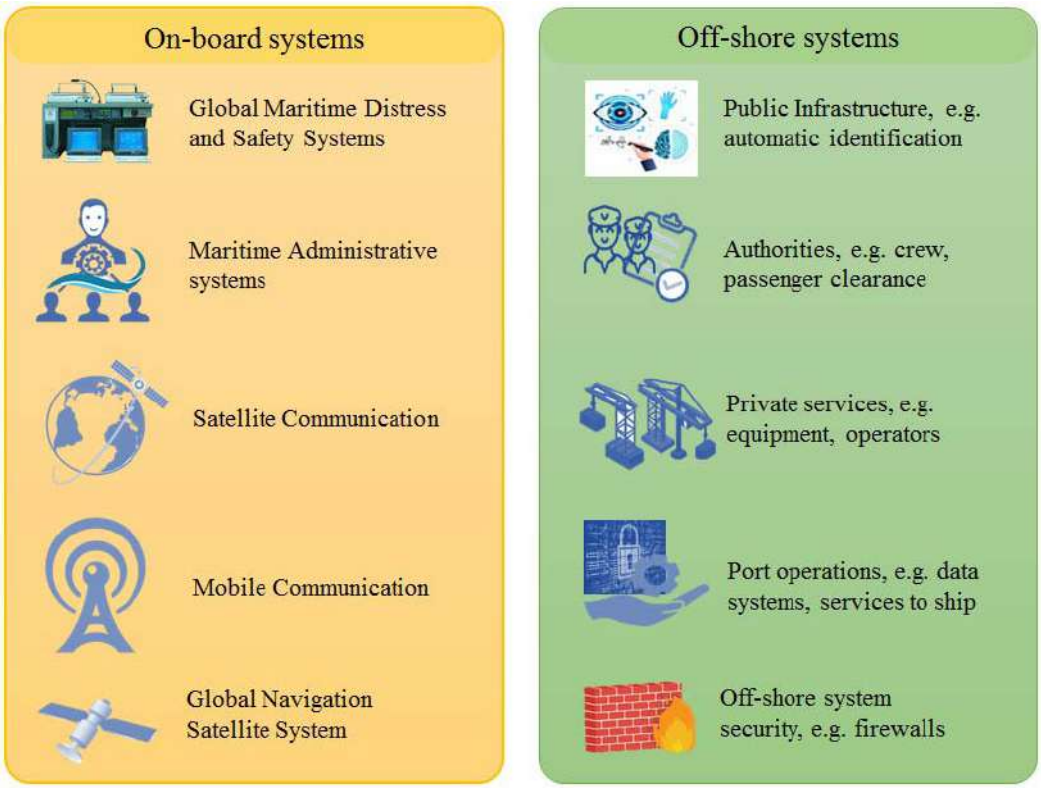


Fig. 5. Maritime systems for on-board and off-shore platforms.

1) Attacks can be directed to a communication stack such as a network layer.
2) Transport layer can be attacked using SYN flood attack types which involves sending transmission control protocol (TCP) connection requests faster, making it impossible for the machine to handle it. It leads to a denial-of-service (DoS) outcome.
3) Attacks like packet replay on the application layer. Such attacks normally happen due to weak security controls.
4) Adversaries attack hardware to obtain unauthorized access and remotely control the devices. Hardware attacks traditionally occur where the authentication controls or appropriate or missing.
5) Software cyber attacks include attacking the software working as an application layer between the sensors and application packages. For example, structure query language (SQL) can be the victim of SQL injection attacks.

In addition to the above-discussed SCADA attacks, the use of social media platforms for accessing the alerts, news regarding hazards, and similar other events can affect the operational capability of such system in emergency response scenarios [42].

H. Cyberattacks in Maritime

1) Phishing Attacks: The phishing attack is the most commonly used cyberattack, including social engineering and malware attacks. The former utilizes email services and fake websites to inflict damage or steal information, while the latter uses different malware installed on a personal computer. Phishing attacks aim at getting the users' personal information such as username, and password, etc., by tricking the user into visiting a fake website [43]. Phishing also includes a sub-category, spear phishing targets the company's employees through emails very similar to the company's legitimate emails. The email contains an attachment that can steal sensitive information stored on the computer once it is clicked to view.

2) Watering Hole Attack: Watering hole attacks target a specific group for a security exploit by using the group's specific websites known to be visited. Such attacks are specifically targeted on the employees of an organization/crew members to gain access to their personal computer by infecting the legitimate websites [44]. Malicious codes are placed on famous websites by exploiting their vulnerabilities and weaknesses and redirecting the users to attackers' websites [45]. Although uncommon, watering hole attacks are harder to detect as they come from legitimate and famous websites. Systems that analyze the compromised websites have been proposed to alleviate the cyber risks of watering hole attacks [46]–[48].

3) Malware: Malware is a group of computer code programs intended to steal or destroy the data on a computer using viruses, spyware, and ransomware. Malware is used for recording a user's activity and stealing confidential information for blackmail and publishing online [49]. With the increase in the number of IoT devices for the modern maritime industry, experts have regarded malware as an attractive choice to

penetrate and breach cyber security [50]. Malware is also used for identity fraud and to commit crimes and terrorist activities in the maritime sector. Similarly, ransomware is also malware containing the zip or other files where opening these files can block access to resources. The attacker requires a ransom amount to allow access. The malware aims at creating man-in-the-middle attacks by exploiting (SSL) or (TSL) weaknesses to download important data from the user's computer [51], [52].

IV. Cyber Risk Analysis Methods

International maritime organization (IMO) is the central agency from the United Nations (UN) to devise policies and procedures for the maritime industry's safety and security, including the risks to the maritime sector and maritime induced risks for the environment. For safeguarding the ships from cyber attacks, it has defined protocols and procedures for a preventive and corrective course of actions, including the elements of cyber risk management [53], as shown in Figure 6. IMO defines five elements for cyber risk management, including identification, protection, detection, responding to risks, and recovering. In addition, the national institute of standards and technology of the United States (US) further elaborates this framework and provides detailed discussions on how to use it [54]. Similarly, the institute of engineering and technology (IET) [55] provides the code of practicing cyber security for ships, and the Baltic and international maritime council (BIMCO) drafts the guidelines for onboard ships [21]. Several models have been contrived to analyze risk impact analysis for maritime cyber risks based on these elements. On the other hand, several individual works outline the guidelines for cyber security for commercial maritime and policies for managing cyber risk [56].

Several models have been designed to analyze the cyber risks with the maritime industry. Maritime risk assessment utilizes qualitative and quantitative methods where the former prioritizes the risks based on their probability. At the same time, the latter performs numerical analysis by awarding risk values to each risk. Predominantly, maritime physical risks analysis relies on probability analysis based on empirical statistics [57], [58]. A qualitative risk analysis is performed for inertial navigation system-related cyber risks in [59]. In addition to the crew interviews, testing is also performed to analyze different vulnerabilities. Results show that remote desktop, terminal service, and remote protocols are vulnerable to arbitrary remote code and man-in-the-middle attacks, respectively. A more critical risk is the server message block service which can be exploited to arbitrary code execution and disclosure of sensitive information. An interview and survey-based method is adopted by [60] for ECDIS cyber vulnerabilities. Unsupported windows, server message block (SMB) vulnerability, improper handling of remote procedure call (RPC), SMB remote execution, and SMB security update are critical risks for ECDIS in maritime ships.

The authors perform cyber risk analysis with a framework based on IMO and IET guidelines in [61] following an on-board survey and cyber security testing for analyzing ECDIS-related cyber risks. Cyber security testing involves vulnerability scanning and penetration testing techniques. The study finds out that the Apache webserver poses a high level of risk as it is obsoleted. As a result, the functionality of the ECDIS can be fully destroyed. Similarly, an experimental ship assessment is carried out in [62] involving the cyber security survey and cyber vulnerability computational scanning to analyze the ECDIS vulnerabilities. Results suggest that obsolete operating systems, server service vulnerability, SMB vulnerability, and SBM security updates are the cyber threats that can be exploited to run arbitrary code from a remote location. Along the same direction, study [63] performs cyber security testing for ECDIS vulnerabilities. Web servers are outdated, printer sharing and operating systems are vulnerable to unauthorized access, leading to a denial of service, crashing ECDIS, stealing sensitive information, man-in-the-middle attacks, etc. In addition, the study analyzes the cyber security risks associated with the third-party service provided and finds out that third-party abandoned and out-of-date components and components involving insecure setup are the major threats.

A survey is conducted in [64] for cyber security vulnerabilities in maritime involving mariners, port officers, IT system experts, and third-party service providers. Survey results highlight the crew-training standards inappropriate (74%), followed by the cyber-attacks with 55%. A total of 60% are found to be explaining the lack of cyber security training. Additionally, 50% of the participants blamed IT as the vulnerable technology for cyber-attacks, while 41% regarded IT and OT as equally responsible. Regarding the cyber crimes, malware, phishing scams, and web-based attacks have been placed at the top three with 31%, 13%, and 13%, respectively of all the cyber crimes in the maritime. The authors of [65] investigate the factors responsible for cyber threats in maritime through a survey. An 80% of the participants considered the crew training insufficient, while 56% ranked cyberattacks as the leading problem for the maritime sector. The majority of the participants (57%) did not receive training regarding cyber security, and 80% suggested the importance of maritime cyber training over general cyber security training. Malware, phishing, and web attacks have been regarded as the leading cyber attacks with 26%, 16%, and 16%, respectively, of all the cyber attacks in the maritime sector.

In the same fashion, the role of human behavior on the cyber security of maritime systems is studied in [66], where the crew members are divided into different groups such as introvert, extrovert, and intuitive, etc. Interviews with the crew members indicate that majority of the people attached with the IT have a medium or low level of knowledge. Despite the installed security systems on the ships, the crew members are not well trained to operate the sophisticated programs. Often, cyber incidents happen due to operators' mistakes due to lack of proper training, carelessness, or poor skill set. A future prospect of the maritime cyber risk is presented in [67] by conducting a survey where 93% of the respondents suggest that the frequency and intensity of the cyber attacks will increase. In addition, the perceptions and potential of social media as a tool for cyber attacks are evaluated, indicating that 74% of the participants believe social media is a potential source of cyber attacks. An 87% believe that the cyberattacks

can be handled more prudently if properly reported and investigated to mitigate future attacks. Study [68] discusses the cyber threats to critical maritime infrastructure, including on-board systems and port operations. Analyzed incidents include high-value cargo theft by infecting authentication data, software malware to shut down port operations, and software infection to interrupt port operations. The study discusses several challenges associated with maritime cyber attacks handling.

The maritime cyber risk analysis (MaCRA) model is one of the risk assessment models in the maritime sector [69] that combines cyber and maritime factors for risk analysis. By considering ship functions, configurations, users, and environmental factors, the framework provides the maritime cyber risks associated with a particular ship type and assists in devising appropriate security procedures. The maCRA model is extended for risk analysis in the autonomous ships by [34] to provide anticipated risks for the futuristic ships. In this regard, the risks are discussed with respect to navigation systems and cargo systems, considering the reward, ease of exploit, and system vulnerability.

GPS jamming has significant repercussions for navigation. [70] shows that the positioning error during GPS jamming is too high to produce catastrophic outcomes if the sailing is continued. Similarly, GPS jamming makes AIS useless as AIS uses GPS signals for slot timing sources which are required for VHS communication based on self-organized time division multiple access (SOTDMA). Jamming GPS also has a strong impact on radar communications, and radar-based detection has erroneous estimations [70]. In addition, if the GPS data is used for slot timing in digital communication such as cellular telephone and satellite communication, GPS jamming would affect these systems.

Risk assessment methods for SCADA can be qualitative, quantitative, and hybrid, combining the first both. Fault tree events analysis [71], object-based event scenario tree [72] and probabilistic risk analysis tools [73] follow a semi-qualitative approach while [74], [75] present quantitative models for risk assessment. For SCADA-related risks assessment, several important research works can be found that extensively studied different approaches for the past two decades [60], [76]–[78]. These research works cover risk assessment methods for static and real-time systems, including monitoring, detection, impact analysis, and countermeasures.

The study [79] presents an automated threat modeling approach regarding cyber security threats to the maritime industry. It comprises three modules each for feature extraction, cyber threat intelligence (CTI)-based detection, and CTI-based attack categorization. The proposed approach performs automated CTI contrary to traditional systems where threat-related features are manually extracted [80]. The model provides increased accuracy as compared to the state-of-the-art approaches.

V. Threat Mitigation Methods

In general, the maritime sector lacks a timely response to introduce the appropriate countermeasures for resolving

Fig. 6. Elements of cyber risk management, adopted from [21].

technical vulnerabilities, which increase the susceptibility of the onboard systems [81]. Due to the better maintenance off-shore systems in the maritime sector, they experience a low number of cyberattacks compared to their counterparts. Secondly, onboard systems rely on obsolete underlying operating systems or those operating systems that do not allow upgrades. The upgrade failures may occur due to conflicting IT and OT technologies standards where the upgrade of one may not support the other. Out of date systems put the entire ship at the hand of the adversaries. Maritime needs to prioritize the critical systems and ensure their safety first, such as navigation, ECDIS, and VDR [82].

Several frameworks and techniques have been presented to mitigate the probability of cyber risks by building detection and correction procedures for cyber attacks. For example, A novel framework, innovative and integrated security system onboard covering the life cycle of a passenger ships voyage (ISOLA), is presented in [83] that performs risk analysis for cruising ships at sea. The analysis covers both vulnerabilities and threats for onboard and off-shore cyber attacks and recommends several data fusion solutions to mitigate the risk impact. The authors present an integrated framework in [84] to monitor the air-sea-ground space for oil ships. Comprising of sensing, network, and application layers, the sensing layer is used to collect the data from air, sea, and ground transmitted via the network layer. The spaceborne synthetic aperture radar (SAR) is used for data collection. The collected observations combined with the forecasting model can provide reliable and accurate trajectory predictions in case of distress situations.

With increasing threats to GPS spoofing and jamming, an authentication scheme is presented by [85] for 6G-IoT-enable maritime transportation. The proposed approach is

TABLE II
A BRIEF SUMMARY OF MODELS USED FOR CYBER RISK ANALYSIS

Ref.	Model	Application	Objectives
[34]	maCRA for autonomous ships	Autonomous ships	Anticipation of probable risks for futuristic autonomous ships in maritime sector.
[59]	Qualitative	INS risk analysis	Analyzing risks associated with navigation tools, charts and interfaces.
[60]	Qualitative	ECDIS risk analysis	Risk analysis for ECDIS components such as SMB, RPC, etc.
[61]	maCRA	Ship with different functionalities, users and configurations	Risk analysis for different kinds of ships, by considering ease of vulnerabilities and exploit reward.
[62]	Mixed	ECDIS risk analysis	Risk analysis using vulnerability scanning and penetration testing techniques.
[63]	Mixed	ECDIS vulnerability	Onboard ship security survey and computational scanning for cyber vulnerability.
[64]	Computational penetration testing	ECDIS third-party service vulnerability	Computational penetration for cyber security threats associated with ECDIS third party services.
[65]	Survey	Maritime cyber risk analysis	Highlighting the most vulnerable component of the maritime including the crew, IT, and operation technology.
[66]	Interviewing	Human factors in cyber risk	Evaluating and highlighting the human factors leading to cyber attacks.
[67]	Survey	Factors for cyber risk	Analyze factors responsible for cyber risks in maritime such as training, IT procedures vulnerability, etc.
[68]	Survey	Cyber risk analysis	Finding factors related to maritime cyber attack to mitigate risk impact, such as social media, human factors, etc.

based on a lightweight message exchange protocol with increased security following initialization, vessel registration, and mutual authentication. The protocol is validated by using the Real-Or-Random model. Results indicate the superior performance of the proposed approach with respect to security and security-to-efficiency trade-off. Along the same lines, an attribute-based data aggregation scheme is proposed in [86] that focuses on the security of isolated IoT-enabled maritime ships. In the proposed scheme, onboard sensors are incorporated for the aggregation of the maritime terminal. The zero-knowledge proof ensures that only legitimate participants can participate in the communication. Results prove the security reliability of the scheme and the reduced computation cost.

The study [87] proposes a framework to detect and defend against the domain name system (DNS) rebinding attacks. A Markov chain model is used to model the DNS rebinding attack. The important attributes are extracted and used with a novel detection model. Experimental results show that the model is suitable for onboard local IoT devices and provides a defense mechanism against DNS rebinding attacks. Similarly, a security and privacy-preserving protocol is proposed in [88] to secure the communication between the maritime electric vehicles and charged grids. The proposed solution is based on blockchain technology and utilizes encryption and consensus algorithms to ensure secure communication [89]. Another endeavor to secure the data sharing between maritime ships and offshore servers is [90] that designed an identity-based information-sharing scheme. The scheme utilizes the blockchain in the fog environment, and smart contracts are used to control secure access to the data. With the proposed scheme, increased security is obtained with reduced computational complexity. Similarly, the authors propose a data integrity framework for maritime transportation systems in [91]. The data blocks are encoded using the erasure coding that provides security against malicious attacks. The data is stored on the cloud and can be recovered in case of data loss. The proposed approach proves to have a low computational head.

Several threat modeling approaches have been devised and adopted for maritime cyber risk analysis and mitigation. STRIDE covers six security threats: spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privileges and performs qualitative analysis of cyber risks [92]. Threat analysis is carried out by developing attack scenarios regarding security objectives such as integrity, authorization, etc. STRIDE is especially useful for discovering vulnerabilities in the systems under design, thus enabling the authorities to eliminate such vulnerabilities in the design process [93], [94]. DREAD is another model for risk mitigation that weights the risks by considering five aspects, including damage potential, reproducibility, exploitability, affected users, and discoverability [35]. Damage refers to the content inflicted to the system regarding the affected things (both users and systems). Reproducibility is the attackers' ability to reproduce it, and exploitability is the extent to which the systems are vulnerable. In contrast, the ability of the attacker to find the system's vulnerability is discoverability. Unlike STRIDE, which focuses on a qualitative analysis, DREAD quantifies the risks by performing a quantitative risk analysis. The values of DREAD elements are determined into high, medium, and low that are used to assign a cyber attack weight for each of the CPS [95].

A hybrid framework based on STRIDE and DREAD is presented in [96] for minimizing the threat of cyber attacks in the maritime sector. By analyzing various CPS's qualitative and quantitative risk factors, the study suggests appropriate controls to alleviate the risk of maritime cyber-attacks. The authors present MITIGATE, a threat mitigation scheme for maritime supply chain [97]. It can be used for MLSC infrastructure and the SCADA system to analyze the risk of cyber risk in a dynamic environment.

VI. INDUSTRY 4.0

Industry 3.0, which focused on automation, computers, and electronics, has been shifting towards Industry 4.0. It includes cyber-physical systems, the internet of things, networks, and

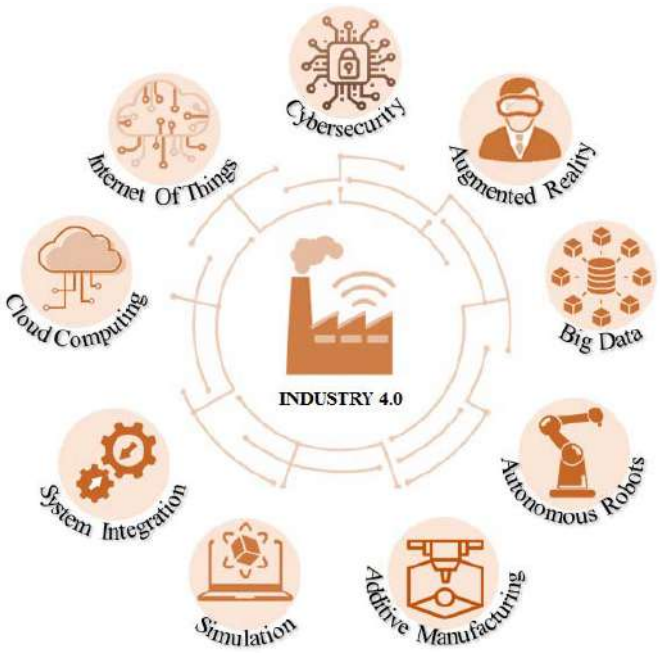


Fig. 7. Industry 4.0 envisions digital transformation, adopted from [98].

many more, as shown in Figure 7 is now performing digital transformation of maritime and its related industries. Using information and communication technology, Industry 4.0 aims at integrating machines and processes to make intelligent networks.

Technology and systems are becoming complicated and connected with every passing day. The concept of digital twins [99] and virtual reality based on the simulation present significant opportunities for the maritime sector to offer training and knowledge for crew members, third-party staff, and other people related to maritime [100]. Although digital twins are not very useful for analyzing cyber risks, virtual reality can play a significant role. It can be used to study maritime vulnerabilities arising in the foreseen Industry 4.0, where everything is connected.

Industry 4.0 is heavily reliant on the concept of IoT, where different small devices communicate via the internet, and the IoT networks can be very complex, and massive [101]. With the availability of cheaper computing power and the proliferation of mobile devices, a massive number of devices will be connected and communicating regarding ships/ports. This ubiquity will also increase the vulnerability of the communication network as more and more devices are connected [102]. So, real-time connected systems are to be modeled to study the probable cyber risks and analyze their impact. This need is further enhanced with the inception of autonomous ships, which are built on the IoT network [103]. With autonomous ships, cyber-physical systems become more prevalent and imminent because a higher number of devices will be used in physical operations.

The major cyber threats are directed remotely via the internet. However, with short-range communication in IoT devices for Industry 4.0, the intrusion threats are expected to be higher than remote threats necessitating tightly secured and well-encrypted protocols. Three important steps for the safekeeping of maritime IT and OT systems are the IT security procedures, cyberattack response and recovery, and preparedness for cyberattacks [104]. The manager should be trained to accept and embrace the IT security mechanisms and protocols to implement IT hygiene. Cyber security training should be considered an integral part of maritime security, and appropriate response and recovery procedures should be in place [105]. Additionally, the procedures should be updated periodically to ensure that they are up-to-date. Last but most importantly, a risk-free cyber environment does not exist. No matter how advanced the technologies become, related vulnerabilities and cyber risks emerge in new forms, which necessitates the importance of being prepared to expect the threats and respond to them accordingly.

VII. Discussions and Future Directions

A. Ship Diversity and Disparate Environment

Many challenges in maritime cyber security bar appropriate cyber security measures and mechanisms. A major challenge is the diversity of the ships and the disparate environments they operate. With ships from different classes, the installed systems, operated environments, requirements for onboard systems, and security procedures vary significantly, making it very difficult to define standard security mechanisms that would fit all. Another problem is the lack of reliable cyber security protocols for ship equipment like GPS and ECDIS [106] due to heterogeneous vendors and manufacturers where the implementation of a security protocol may be very different. The third complexity arises from the third-party service providers that deal with the maritime operational systems. The short visitations during the ashore stay of the ship limit their capability to fix problems appropriately.

B. Improper Cyber Security Risk Assessment

One major shortcoming for the secured maritime industry is the improper risk assessment of cyber security threats. For example, different nations in the European Union (EU) implement disparate security policies and practices, complicating risk assessment comparison. In addition, targeted risk assessment procedures should be developed with respect to the nature of the MLSC infrastructure, where processes are both distributed and interconnected. Research shows that the training and knowledge of the crew member are not up to the mark to deal with the cyber risks. The majority of maritime professionals suggest a lack of knowledge specifically in the field of maritime cyber security [50]. Lack of training and expertise for cyber security led to 88% to 90% of the shipping accidents, as stated in [107], [108]. Similarly, the reliance on obsolete and outdated systems in the maritime is a major problem [109], [110].

C. Lack of Real-World Testing

Poor crew skills, complexity and sophistication of on-board systems, outdated and vulnerable information systems, inappropriate integration of IT and OT procedures, network/system

heterogeneity, and lack of updating the cyber security procedures are the leading challenges for elevated maritime cyber security risks. Lack of real-world testing systems can make it very difficult to analyze the risk impact of cyber attacks fully. Especially, systems for penetration testing in the dynamic environment are needed for futuristic cyber attacks analysis and prevention. Ethical hacking should also be promoted to make beforehand preparations to counter cyber risks [111]. GPS jamming and spoofing is the leading cyberattack that caused potential damage to the maritime industry. Relying on one navigation guide technology seems a bottleneck and inappropriate. With more sensors on-board such as radar and LIDAR (light detection and ranging) in future ships, these sensors can be used for navigation and utilize other resources for navigation guides.

D. Increased Dependence on Cyber Technology

Recent automation and digitization have evolved the maritime sector by combining IT and OT more than ever. With the advanced digital technology, the maritime infrastructure relies on cyber technology increasing its proneness to different kinds of cyberattacks. Maritime-related cyberattacks are challenging due to a lack of information on the cyberattacks, economic and disruptive impact, and insightful investigations. Cyber attacks on the maritime can target navigation, cargo movement, ECDIS elements, off-shore AIS, third-party service providers, and other processes and threaten human lives, ecosystem, and maritime trade. Cyber attackers aim to obtain media attention, ransom, destroy an organization's resources, sell confidential data, and sabotage. In addition, ship transportation to the desired location, intervening in cyber security defense, and gaining critical information regarding national infrastructure are the primary goals of different types of adversaries. However, most of these attacks happen due to obsolete operational systems, especially software, and the carelessness of the maritime staff. The proper training and knowledge of Crews can significantly enhance the defense against such attacks, and so can the up-to-date operational procedures. A recent increase in maritime cyber security threats requires next-generation cyber security dealing procedures in real-time which means that the equipment and protocols to perform real-world experiments using vulnerability testing and penetration testing are need of the hour.

E. Need to Adopt Emerging Solutions

To ensure increased defense against evolving cyber attacks, novel and emerging solutions must be adopted. In this regard, two technologies can play a pivotal role in alleviating the risk of cyber attacks on maritime ships: satellite IoT and high altitude platform (HAP) solutions. With increased GPS jamming and spoofing attacks on maritime ships, satellite IoT can work as a complementary solution with wide coverage and therefore can be advantageous in many ways [112]. Such low orbit satellites can provide communication at lower latency with lower transmission loss and supplement the GNSS [113], [114]. The third generation partnership project (3GPP) incorporates the solutions for new radio (NR) to support non-terrestrial networks (NTN) communications [115]. In the same way, HAP systems can provide broadband connectivity and telecommunication services to remote areas where connectivity to the core network is not possible. In case of distress situations, HAP systems can provide the connectivity for mobile and core network for backhauling [116]. Since HAP systems require minimal ground infrastructure, they can be pivotal for disaster, distress, and emergency response cases.

VIII. Conclusion

With rapid technological advancements, the maritime sector has prospered regarding technology like sensors, communication, and security. Despite the potential benefits of embracing such digital transformation, the proneness of the maritime industry has been substantially increased as well, opening new ways and paradigms for cyber attacks. This study analyzes the cyber security threats for the maritime industry regarding the devices used for sensing, communication, navigation, and emergency response in case of distress. It is observed that the ships lack the technical staff to handle the under attack situation. The ship crew does not possess competence or is not well trained to handle cyberattacks, and the cyber security aspect of ships is overlooked. Despite several systems being in place, the relied-on systems/software are often obsolete, not fully operational, or unsuitable for real-world situations. In addition, security devices and frameworks are heterogeneous and lack standard operating procedures.

References

[1] European Community Shipowners' Associations, *Shipping and Global Trade Towards an EU External Shipping Policy*. Accessed: Nov. 22, 2021. [Online]. Available: https://www.ecsa.eu/sites/default/files/publications/2017-02-27-ECSA-External-Shipping-Agenda-FINAL.pdf

[2] M. Kalouptsidi, *The Role of Shipping in World Trade*. Accessed: Nov. 22, 2021. [Online]. Available: https://econofact.org/the-role-of-shipping-in-world-trade

[3] A. Roger. (2018). *Maersk and IBM to Form Joint Venture Applying Blockchain to Improve Global Trade and Digitise Supply Chains*. [Online]. Available: https://www.forbes.com/sites/rogeraitken/2018/01/16/ibm-forges-global-joint-venture-with-maersk-applying-blockchain-to-digitize-global-trade/?sh=3d6b0a36547e

[4] B. S. Rivkin, "Unmanned ships: Navigation and more," *Gyroscopy Navigat.*, vol. 12, no. 1, pp. 96–108, Jan. 2021.

[5] L. Register, "Cyber-enabled ships shipright procedure assignment for cyber descriptive notes for autonomous & remote access ships," Lloyd's Register, London, U.K., Tech. Rep., 2017.

[6] M. Placek. (2021). *Container Throughput at Ports Worldwide From 2012 to 2020 With a Forecast for 2021 Until 2024*. [Online]. Available: https://www.statista.com/statistics/913398/container-throughput-worldwide/

[7] A. Chakrabarty. *What Marine Communication Systems Are Used in the Maritime Industry*. Accessed: Nov. 27, 2021. [Online]. Available: https://www.marineinsight.com/marine-navigation/marine-communication-systems-used-in-the-maritime-industry/

[8] Y.-C. Lee, S.-K. Park, W.-K. Lee, and J. Kang, "Improving cyber security awareness in maritime transport: A way forward," *J. Korean Soc. Mar. Eng.*, vol. 41, no. 8, pp. 738–745, Oct. 2017.

[9] A. R. Lee and H. P. Wogan, "All at sea: The modern seascape of cybersecurity threats of the maritime industry," in *Proc. OCEANS MTS/IEEE Charleston*, 2018, pp. 1–8.

[10] J. J. George and D. E. Leidner, "From clicktivism to hacktivism: Understanding digital activism," *Inf. Org.*, vol. 29, no. 3, Sep. 2019, Art. no. 100249.

[11] A. Bagchi and J. A. Paul, "Espionage and the optimal standard of the customs-trade partnership against terrorism (C-TPAT) program in maritime security," *Eur. J. Oper. Res.*, vol. 262, no. 1, pp. 89–107, Oct. 2017.

[12] H. Boyes, R. Isbell, and A. Luck, "Code of practice: Cyber security for ports and port systems," *Inst. Eng. Technol.*, vol. 28, p. 2016, Oct. 2016.

[13] A. Oruc and M. S. M. MIMarEST, "Claims of state-sponsored cyberattack in the maritime industry," in *Proc. Int. Naval Eng. Conf. & Exhib.*, 2020, doi: 10.24868/issn.2515-818X.2020.021.

[14] D. Volz, "Chinese hackers target universities in pursuit of maritime military secrets," *Wall Street J.*, Jan. 2019. Accessed: Nov. 29, 2021. [Online]. Available: https://www.wsj.com/articles/chinese-hackers-target-universities-in-pursuit-of-maritime-military-secrets-11551781800

[15] D. J. Bodeau, R. Graubart, and J. Fabius-Greene, "Improving cyber security and mission assurance via cyber preparedness (cyber prep) levels," in *Proc. IEEE 2nd Int. Conf. social Comput.*, Oct. 2010, pp. 1147–1152.

[16] I. Progoulakis, P. Rohmeyer, and N. Nikitakos, "Cyber physical systems security for maritime assets," *J. Mar. Sci. Eng.*, vol. 9, no. 12, p. 1384, Dec. 2021.

[17] J. DiRenzo, D. A. Goward, and F. S. Roberts, "The little-known challenge of maritime cyber security," in *Proc. 6th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, 2015, pp. 1–5.

[18] B. Mednikarov, Y. Tsonev, and A. Lazarov, "Analysis of cybersecurity issues in the maritime industry," *Int. J. Inf. Secur.*, vol. 47, no. 1, pp. 27–43, 2020.

[19] G. C. Kessler, J. P. Craiger, and J. C. Haass, "A taxonomy framework for maritime cybersecurity: A demonstration using the automatic identification system," *Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 12, no. 3, p. 429, 2018.

[20] BIMCO. (2016). *The Guidelines on Cyber Security onboard Ships*. [Online]. Available: https://www.liscr.com/sites/default/files/online_library/Guidelines_on_cyber_security_onboard_ships_version_1-1_Feb2016(3).pdf

[21] *The Guidelines Cyber Secur. Onboard Ships*, BIMCO, Copenhagen, Denmark, 2016.

[22] M. Balduzzi, A. Pasta, and K. Wilhoit, "A security evaluation of AIS automated identification system," in *Proc. 30th Annu. Comput. Secur. Appl. Conf.*, 2014, pp. 436–445.

[23] B. Hyra, "Analyzing the attack surface of ships," M.S. thesis, DTU Comput. Dept. Appl. Math. Comput. Sci., Technical Univ. Denmark, Lyngby, Denmark, 2019. [Online]. Available: https://backend.orbit.dtu.dk/ws/portalfiles/portal/174011206/190401_Analyzing_the_Attack_Surface_of_Ships.pdf

[24] B. Svilicic, D. Bráic, S. Žuékin, and D. Kalebic, "Raising awareness on cyber security of ECDIS," *TransNav, Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 13, no. 1, pp. 231–236, 2019.

[25] T. Pseftelis and G. Chondrokoukis, "A study about the role of the human factor in maritime cybersecurity," *SPOUDAI-J. Econ. Bus.*, vol. 71, nos. 1–2, pp. 55–72, 2021.

[26] J. Bhatti and T. E. Humphreys, "Hostile control of ships via false GPS signals: Demonstration and detection," *NAVIGATION, J. Inst. Navigat.*, vol. 64, no. 1, pp. 51–66, 2017.

[27] M. Filić, "Foundations of GNSS spoofing detection and mitigation with distributed GNSS SDR receiver," *Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 12, no. 4, pp. 1–8, 2018.

[28] M. S. K. Awan and M. A. Al Ghamdi, "Understanding the vulnerabilities in digital components of an integrated bridge system (IBS)," *J. Mar. Sci. Eng.*, vol. 7, no. 10, p. 350, 2019.

[29] A. Androjna and M. Perkoviá, "Impact of spoofing of navigation systems on maritime situational awareness," *Trans. Maritime Sci.*, vol. 10, no. 2, pp. 361–373, Oct. 2021.

[30] K. Korcz, "Maritime radio information systems," *J. KONES*, vol. 24, pp. 1–8, Dec. 2017.

[31] K. Tam, K. Moara-Nkwe, and K. Jones, "The use of cyber ranges in the maritime context: Assessing maritime-cyber risks, raising awareness, and providing training," *Maritime Technol. Res.*, vol. 3, no. 1, pp. 16–30, Jul. 2020.

[32] R. Santamarta, "Maritime security: Hacking into a voyage data recorder (VDR)," IOActive, Seattle, WA, USA, 2016. Accessed: Nov. 29, 2021. [Online]. Available: https://ioactive.com/maritime-security-hacking-into-a-voyage-data-recorder-vdr/

[33] A. Cantelli-Forti, "Forensic analysis of industrial critical systems: The costa concordia's voyage data recorder case," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Jun. 2018, pp. 458–463.

[34] K. Tam and K. Jones, "Cyber-risk assessment for autonomous ships," in *Proc. Int. Conf. Cyber Secur. Protection Digit. Services*, Jun. 2018, pp. 1–8.

[35] G. Kavallieratos, S. Katsikas, and V. Gkioulos, "Cyber-attacks against the autonomous ship," in *Computing Security*. Cham, Switzerland: Springer, 2018, pp. 20–36.

[36] K. Tam and K. D. Jones, "Maritime cybersecurity policy: The scope and impact of evolving technology on international shipping," *J. Cyber Policy*, vol. 3, no. 2, pp. 147–164, May 2018.

[37] F. X. M. de Osés and A. U. Juncadella, 'Global maritime surveillance and oceanic vessel traffic services: Towards the E-navigation," *WMU J. Maritime Affairs*, vol. 20, no. 3, pp. 1–14, 2021.

[38] L. O'Donnell-Welch. (2021). *Cybercriminals Target Transport and Logistics Industry*. [Online]. Available: https://duo.com/decipher/cybercriminals-target-global-logistics-industry

[39] A. Kinsey. (2021). *Cyber Security Threats Challenge International Shipping Industry*. [Online]. Available: https://www.maritimeprofessional.com/news/cyber-security-threats-challenge-international-369770

[40] C. Clmpanu. (2020). *All Four of the World's Largest Shipping Companies Have Now Been Hit by Cyber-Attacks*. [Online]. Available: https://www.zdnet.com/article/all-four-of-the-worlds-largest-shipping-companies-have-now-been-hit-by-cyber-attacks/

[41] E.-M. Kalogeraki, N. Polemi, S. Papastergiou, and T. Panayiotopoulos, "Modeling SCADA attacks," in *Smart Trends System, Security Sustainability*. Cham, Switzerland: Springer, 2018, pp. 47–55.

[42] D. Kravets. (2009). *Feds: Hacker Disabled Offshore Oil Platforms' Leak-Detection System*. [Online]. Available: http://www. wired.com/2009/03/feds-hacker-dis

[43] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: State of the art and future challenges," *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3629–3654, Dec. 2017.

[44] J. Allen *et al.*, "Mnemosyne: An effective and efficient postmortem watering hole attack investigation system," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2020, pp. 787–802.

[45] K. A. Ismail, M. M. Singh, N. Mustaffa, P. Keikhosrokiani, and Z. Zulkefli, "Security strategies for hindering watering hole cyber crime attack," *Proc. Comput. Sci.*, vol. 124, pp. 656–663, Oct. 2017.

[46] K. Borgolte, C. Kruegel, and G. Vigna, "Delta: Automatic identification of unknown web-based infection campaigns," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013, pp. 109–120.

[47] K. Borgolte, C. Kruegel, and G. Vigna, "Meerkat: Detecting website defacements through image-based object recognition," in *Proc. 24th Secur. Symp.*, 2015, pp. 595–610.

[48] Z. Li, S. Alrwais, X. Wang, and E. Alowaisheq, "Hunting the red fox online: Understanding and detection of mass redirect-script injections," in *Proc. IEEE Symp. Secur. Privacy*, May 2014, pp. 3–18.

[49] P. R. Toth *et al.*, "Small business information security: The fundamentals," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Interagency Rep. (NISTIR) 7621 Rev. 1, 2016.

[50] J. I. Alcaide and R. G. Llave, "Critical infrastructures cybersecurity and the maritime sector," *Transp. Res. Proc.*, vol. 45, pp. 547–554, Jan. 2020.

[51] Z. Cekerevac, Z. Dvorak, L. Prigoda, and P. Cekerevac, "Internet of Things and the man-in-the-middle attacks–security and economic risks," *MEST J.*, vol. 5, no. 2, pp. 15–25, 2017.

[52] A. Mallik, "Man-in-the-middle-attack: Understanding in simple words," *J. Pendidikan Teknol. Inf.*, vol. 2, no. 2, pp. 109–134, 2019.

[53] *Guidelines on Maritime Cyber Risk Management*, Int. Maritime Org. London, U.K., 2017.

[54] M. P. Barrett *et al.*, "Framework for improving critical infrastructure cybersecurity version 1.1," Nat. Inst. Standards and Technol., Gaithersburg, Maryland, USA, Tech. Rep., 2018. [Online]. Available: https://www.nist.gov/cyberframework, doi: 10.6028/NIST.CSWP.04162018.

[55] H. Boyes and R. Isbell, *Code Practice: Cyber Security for Ships*. London, U.K.: Institution of Engineering and Technology, 2017.

[56] A. Rana, "Commercial maritime and cyber risk management," *Saf. Defense*, vol. 5, no. 1, pp. 46–48, 2019.

[57] J. Montewka, S. Ehlers, F. Goerlandt, T. Hinz, K. Tabri, and P. Kujala, "A framework for risk assessment for maritime transportation systems—A case study for open sea collisions involving RoPax vessels," *Rel. Eng. Syst. Saf.*, vol. 124, pp. 142–157, Apr. 2014.

[58] J. Nordström *et al.*, "Vessel TRIAGE: A method for assessing and communicating the safety status of vessels in maritime distress situations," *Saf. Sci.*, vol. 85, pp. 117–129, Jun. 2016.

[59] R. Svilicic J. Zec, "A study on cyber security threats in a shipboard integrated navigational system," *J. Mar. Sci. Eng.*, vol. 7, no. 10, p. 364, Oct. 2019.

[60] Y. Cherdantseva *et al.*, "A review of cyber security risk assessment methods for SCADA systems," *Comput. Secur.*, vol. 56, pp. 1–27, Feb. 2015.

[61] B. Svilicic, J. Kamahara, J. Celic, and J. Bolmsten, "Assessing ship cyber risks: A framework and case study of ECDIS security," *WMU J. Maritime Affairs*, vol. 18, no. 3, pp. 509–520, Sep. 2019.
[62] B. Svilicic, J. Kamahara, M. Rooks, and Y. Yano, "Maritime cyber risk management: An experimental ship assessment," *J. Navigat.*, vol. 72, no. 5, pp. 1108–1120, Sep. 2019.
[63] B. Svilicic and I. Rudan, "Shipboard ECDIS cyber security: Third-party component threats," *Pomorstvo*, vol. 33, no. 2, pp. 176–180, Dec. 2019.
[64] K. Tam and K. Jones, "Situational awareness: Examining factors that affect cyber-risks in the maritime sector," *Int. J. Cyber Situational Aware.*, vol. 4, pp. 40–68, 2019.
[65] K. Tam and K. Jones, "Factors affecting cyber risk in maritime," in *Proc. Int. Conf. Cyber Situational Awareness, Data Analytics Assessment*, 2019, pp. 1–8.
[66] R. Hanzu-Pazara, G. Raicu, and R. Zagan, "The impact of human behaviour on cyber security of the maritime systems," *Adv. Eng. Forum*, vol. 34, pp. 267–274, Oct. 2019.
[67] S. Karamperidis, G. Koligiannis, and F. Moustakis, "Building a digital armour for the maritime sector against cyber-attacks," Tech. Rep., 2020.
[68] A. Androjna and E. Twrdy, "Cyber threats to maritime critical infrastructure," *Cyber Terrorism Extremism as Threat to Crit. Infrastructure Protection*. Ljubljana, Slovenia: Ministry Defence Republic, 2020.
[69] K. Tam and K. Jones, "MaCRA: A model-based framework for maritime cyber-risk assessment," *WMU J. Maritime Affairs*, vol. 18, no. 1, pp. 129–163, Mar. 2019.
[70] A. Grant, P. Williams, N. Ward, and S. Basker, "GPS jamming and the impact on maritime navigation," *J. Navigat.*, vol. 62, no. 2, pp. 173–187, Apr. 2009.
[71] P. A. S. Ralston, J. H. Graham, and J. L. Hieb, "Cyber security risk assessment for SCADA and DCS networks," *ISA Trans.*, vol. 46, no. 4, pp. 583–594, 2007.
[72] G. D. Wyss and F. A. Durán, "OBEST: The object-based event scenario tree methodology," Sandia National Labs., Albuquerque, NM, USA, Tech. Rep. SAND2001-0828, 2001.
[73] M. A. McQueen, W. F. Boyer, M. A. Flynn, and G. A. Beitel, "Quantitative cyber risk reduction estimation methodology for a small SCADA control system," in *Proc. 39th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2006, p. 226.
[74] D. I. Gertman, R. Folkers, and J. Roberts, "Scenario-based approach to risk analysis in support of cyber security," in *Proc. Int. Topical Meeting Nucl. Plant Instrum. Controls, Hum. Mach. Interface Technol.*, 2006, pp. 1–5.
[75] C. Beggs and M. Warren, "Safeguarding Australia from cyber-terrorism: A proposed cyber-terrorism SCADA risk framework for industry adoptionn," *J. Inf. Warfare*, vol. 7, no. 1, pp. 24–35, 2008.
[76] J. D. Markovic-Petrovic and M. D. Stojanovic, "An improved risk assessment method for SCADA information security," *Elektron. Elektrotechn.*, vol. 20, no. 7, pp. 69–72, Sep. 2014.
[77] M. Keshk, N. Moustafa, E. Sitnikova, and G. Creech, "Privacy preservation intrusion detection technique for SCADA systems," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, Nov. 2017, pp. 1–6.
[78] T. Marsden, N. Moustafa, E. Sitnikova, and G. Creech, "Probability risk identification based intrusion detection system for scada systems," in *Int. Conf. Mobile Netw. Manage.* Cham, Switzerland, Springer, 2017, pp. 353–363.
[79] P. Kumar, G. P. Gupta, R. Tripathi, S. Garg, and M. M. Hassan, "DLTIF: Deep learning-driven cyber threat intelligence modeling and identification framework in IoT-enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Nov. 16, 2021, doi: 10.1109/TITS.2021.3122368.
[80] R. Arul, S. Basheer, A. Abbas, and A. K. Bashir, "Role of deep learning algorithms in securing Internet of Things applications," in *Deep Learning for Internet Things Infrastructure*. Boca Raton, FL, USA: CRC Press, 2021, pp. 145–164.
[81] A. Androjna, T. Brcko, I. Pavic, and H. Greidanus, "Assessing cyber challenges of maritime navigation," *J. Mar. Sci. Eng.*, vol. 8, no. 10, p. 776, Oct. 2020.
[82] D. Trimble, J. Monken, and A. F. Sand, "A framework for cybersecurity assessments of critical port infrastructure," in *Proc. Int. Conf. Cyber Conflict*, 2017, pp. 1–7.
[83] P. M. Laso *et al.*, "ISOLA: An innovative approach to cyber threat detection in cruise shipping," in *Developments and Advances in Defense and Security*. Singapore, Springer, 2022, pp. 71–81.
[84] Q. Pan *et al.*, "Space-air-sea-ground integrated monitoring network-based maritime transportation emergency forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2843–2852, Mar. 2021.
[85] S. A. Chaudhry *et al.*, "A lightweight authentication scheme for 6G-IoT enabled maritime transport system," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 22, 2021, doi: 10.1109/TITS.2021.3134643.
[86] C. Wang, J. Shen, P. Vijayakumar, and B. B. Gupta, "Attribute-based secure data aggregation for isolated IoT-enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 1, 2021, doi: 10.1109/TITS.2021.3127436.
[87] X. He *et al.*, "DNS rebinding threat modeling and security analysis for local area network of maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 22, 2021, doi: 10.1109/TITS.2021.3135197.
[88] A. Barnawi, S. Aggarwal, N. Kumar, D. M. Alghazzawi, B. Alzahrani, and M. Boulares, "Path planning for energy management of smart maritime electric vehicles: A blockchain-based solution," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 15, 2021, doi: 10.1109/TITS.2021.3131815.
[89] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, "A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid," *IEEE Trans. Comput.*, early access, Nov. 24, 2021, doi: 10.1109/TC.2021.3130402.
[90] B. B. Gupta, A. Gaurav, C.-H. Hsu, and B. Jiao, "Identity-based authentication mechanism for secure information sharing in the maritime transport system," *IEEE Trans. Intell. Transp. Syst.*, early access, Nov. 15, 2021, doi: 10.1109/TITS.2021.3125402.
[91] D. Liu, Y. Zhang, W. Wang, K. Dev, and S. A. Khowaja, "Flexible data integrity checking with original data recovery in iot-enabled maritime transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, Nov. 15, 2021, doi: 10.1109/TITS.2021.3125070.
[92] A. Shostack, *Threat Modeling: Designing for Security*. Hoboken, NJ, USA: Wiley, 2014.
[93] D. Seifert and H. Reza, "A security analysis of cyber-physical systems architecture for healthcare," *Computers*, vol. 5, no. 4, p. 27, Oct. 2016.
[94] G. Kavallieratos and S. Katsikas, "Attack path analysis for cyber physical systems," in *Computing Security*. Cham, Switzerland: Springer, 2020, pp. 19–33.
[95] P. Nespoli, D. Papamartzivanos, F. G. Mármol, and G. Kambourakis, "Optimal countermeasures selection against cyber attacks: A comprehensive survey on reaction frameworks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1361–1396, 2nd Quart., 2018.
[96] G. Kavallieratos and S. Katsikas, "Managing cyber security risks of the cyber-enabled ship," *J. Mar. Sci. Eng.*, vol. 8, no. 10, p. 768, Sep. 2020.
[97] E.-M. Kalogeraki, S. Papastergiou, H. Mouratidis, and N. Polemi, "A novel risk assessment methodology for SCADA maritime logistics environments," *Appl. Sci.*, vol. 8, no. 9, p. 1477, Aug. 2018.
[98] I. Sceep. (2021). *Industry 4.0 and the Fourth Industrial Revolution Explained*. [Online]. Available: https://www.i-scoop.eu/industry-4-0/
[99] F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, S. C.-Y. Lu, and A. Y. C. Nee, "Digital twin-driven product design framework," *Int. J. Prod. Res.*, vol. 57, no. 12, pp. 3935–3953, 2018.
[100] S. Frydenberg, K. Nordby, and J. O. Eikenes, "Exploring designs of augmented reality systems for ship bridges in Arctic waters," *Hum. Factors*, vol. 26, p. 27, Dec. 2018.
[101] S. D. Pizzo, A. De Martino, G. De Viti, R. L. Testa, and G. De Angelis, "IoT for buoy monitoring system," in *Proc. IEEE Int. Workshop Metrol. Sea, Learn. Measure Sea Health Parameters (MetroSea)*, Oct. 2018, pp. 232–236.
[102] A. Zolich *et al.*, "Survey on communication and networks for autonomous marine systems," *J. Intell. Robot. Syst.*, vol. 95, no. 3, pp. 789–813, 2019.
[103] K. Tam, K. Forshaw, and K. Jones, "Cyber-SHIP: Developing next generation maritime cyber research capabilities," in *Proc. Conf. ICMET Oman*, 2019.
[104] H. Oe and H. Nguyen, "Opportunities, challenges, and the future of cruise ship tourism: Beyond covid-19 with ubiquitous information sharing and decision-making," *Int. J. Manage. Decis. Making*, vol. 20, no. 3, pp. 221–240, 2021.
[105] G. Kavallieratos, V. Diamantopoulou, and S. K. Katsikas, "Shipping 4.0: Security requirements for the cyber-enabled ship," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6617–6625, Oct. 2020.
[106] L. Kelion. (2018). *Ship Hack 'Risks Chaos In English Channel*. Accessed: Dec. 2, 2021. [Online]. Available: https://mfame.guru/ship-hack-risks-chaos-in-english-channel/
[107] C. Heij and S. Knapp, "Predictive power of inspection outcomes for future shipping accidents—An empirical appraisal with special attention for human factor aspects," *Maritime Policy Manage.*, vol. 45, no. 5, pp. 604–621, Jul. 2018.

[108] C. Park, W. Shi, W. Zhang, C. Kontovas, and C. Chang, "Cybersecurity in the maritime industry: A literature review," in *Proc. 20th Commemorative Annu. Gen. Assem.*, 2019, pp. 79–86.
[109] R. Sen, "Cyber and infomation threats to seaports and ships," *Maritime Secur.*, vol. 4, pp. 281–302, Dec. 2016.
[110] K. D. Jones, K. Tam, and M. Papadaki, "Threats and impacts in maritime cyber security," *Eng. Technol. Ref.*, vol. 1, 2012.
[111] R. Chia, "The need for ethical hacking in the maritime industry," *Soc. Nav. Architects Mar. Eng.*, vol. 38, pp. 108–121, Sep. 2019.
[112] D. Yang, Y. Zhou, W. Huang, and X. Zhou, "5G mobile communication convergence protocol architecture and key technologies in satellite Internet of Things system," *Alexandria Eng. J.*, vol. 60, no. 1, pp. 465–476, Feb. 2021.
[113] M. Jia and Q. Guo, "Editorial: Intelligent cognitive internet of integrated space and terrestrial things," *Mobile Netw. Appl.*, vol. 24, no. 6, pp. 1924–1925, Dec. 2019.
[114] Y. Qian, L. Ma, and X. Liang, "The performance of chirp signal used in LEO satellite Internet of Things," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1319–1322, Aug. 2019.
[115] *Solutions for NR to Support Non-Terrestrial Networks (NTN)*, document 38.821, 3GPP, 2019.
[116] M. Q. Vu, N. T. Dang, and A. T. Pham, "HAP-aided relaying satellite FSO/QKD systems for secure vehicular networks," in *Proc. IEEE 89th Veh. Technol. Conf.*, Apr. 2019, pp. 1–6.

Imran Ashraf received the M.S. degree (Hons.) in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, Gyeongsan, South Korea, in 2018. He has worked as a Post-Doctoral Fellow at Yeungnam University. He is currently working as an Assistant Professor with the Information and Communication Engineering Department, Yeungnam University. His research areas include positioning using next-generation networks, communication in 5G and beyond, location-based services in wireless communication, smart sensors (LIDAR) for smart cars, and data analytics.

Yongwan Park received the B.E. and M.E. degrees in electrical engineering from Kyungpook University, Daegu, South Korea, in 1982 and 1984, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo, USA, in 1989 and 1992, respectively. He worked at the California Institute of Technology as a Research Fellow from 1992 to 1993. From 1994 to 1996, he served as a Chief Researcher for developing IMT-2000 system at SK Telecom, South Korea. Since 1996, he has been a Professor of information and communication engineering at Yeungnam University, South Korea. From January 2000 to February 2000, he was an Invited Professor at the NTT DoCoMo Wireless Laboratory, Japan. He was also a Visiting Professor at UC Irvine, USA, in 2003. From 2008 to 2009, he served as the Director of the Technology Innovation Center for Wireless Multimedia, Korean Government. From 2009 to March 2017, he also served as the President of the Gyeongbuk Institute of IT Convergence Industry Technology (GITC), South Korea. He is also serving as the Chairman of 5G Forum Convergence Service Committee, South Korea. His current research interests include 5G systems in communication, OFDM, PAPR reduction, indoor location-based services in wireless communication, and smart sensors (LIDAR) for smart car.

Soojung Hur received the B.S. degree from Daegu University, Gyeongbuk, South Korea, in 2001, the M.S. degree in electrical engineering from San Diego State University, San Diego, in 2004, and the M.S. and Ph.D. degrees in information and communication engineering from Yeungnam University, South Korea, in 2007 and 2012, respectively. She is working as a Research Professor with the Mobile Communication Laboratory, Yeungnam University. Her current research interests include the performance of mobile communication, indoor/outdoor location, and unnamed vehicle.

Sung Won Kim received the B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, South Korea, in 1990 and 1992, respectively, and the Ph.D. degree from the School of Electrical Engineering and Computer Sciences, Seoul National University, in August 2002. From January 1992 to August 2001, he was a Researcher at the Research and Development Center of LG Electronics, South Korea. From August 2001 to August 2003, he was a Researcher at the Research and Development Center of AL Tech, South Korea. From August 2003 to February 2005, he was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, USA. In March 2005, he joined the Department of Information and Communication Engineering, Yeungnam University, Gyeongsangbuk-do, South Korea, where he is currently a Professor. His research interests include resource management, wireless networks, mobile computing, performance evaluation, and machine learning.

Roobaea Alroobaea received the bachelor's degree (Hons.) in computer science from King Abdulaziz University (KAU), Saudi Arabia, in 2008, and the master's degree in information system and the Ph.D. degree in computer science from the University of East Anglia, U.K., in 2012 and 2016, respectively. He is currently an Associate Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests include human–computer interaction, software engendering, cloud computing, the Internet of Thing, artificial intelligent, and machine learning.

Yousaf Bin Zikria (Senior Member, IEEE) is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University, South Korea. He authored more than 100 refereed articles, conference papers, book chapters, and patents. His journal article's cumulative impact factor (IF) is more than 320. He published papers at the top venue, including IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, *IEEE Wireless Communications Magazine*, IEEE NETWORK, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Future Generation Computer Systems* (Elsevier), *Sustainable Cities and Society* (Elsevier), and *Journal of Network and Computer Applications* (Elsevier). He has managed numerous FT/SI in SCI/E indexed journals. His research interests include the IoT, 5G, machine learning, wireless communications and networks, WSNs, routing protocols, CRAHN, CRASN, transport protocols, VANETS, embedded systems, and network and information security. He also held the prestigious CISA, JNCIS-SEC, JNCIS-ER, JNCIA-ER, JNCIA-EX, and Advance Routing Switching and WAN Technologies certifications. He is listed in the world's top 2% of researchers published by Elsevier and Stanford University. GoogleScholar: https://scholar.google.com/citations?user = K90qMyMAAAAJhl = en Website: https://sites.google.com/view/ybzikria Researchgate: https://www.researchgate.net/profile/Yousaf-Zikria

Summera Nosheen received the Ph.D. degree from the School of Electrical Engineering and Computing, The University of Newcastle, NSW, Australia, in 2021. She is currently with the Faculty of Engineering, The University of Sydney, NSW, Australia. She received the Commonwealth Department of Education, Science and Training and The University of Newcastle Research Training Program (RTP) tuition fee and stipend scholarships. Her research interests include wireless networks, quality of service, quality of experience, and MAC layer resource allocation.

Cyber LOPA: An Integrated Approach for the Design of Dependable and Secure Cyber-Physical Systems

Ashraf Tantawy, *Member, IEEE*, Sherif Abdelwahed, *Senior Member, IEEE*, and Abdelkarim Erradi, *Member, IEEE*

***Abstract*—Safety risk assessment is an essential process to ensure a dependable cyber-physical system (CPS) design. Traditional risk assessment considers only physical failures. For modern CPSs, failures caused by cyber attacks are on the rise. The focus of latest research effort is on safety–security lifecycle integration and the expansion of modeling formalisms for risk assessment to incorporate security failures. The interaction between safety and security lifecycles and its impact on the overall system design, as well as the reliability loss resulting from ignoring security failures, are some of the overlooked research questions. This article addresses these research questions by presenting a new safety design method named cyber layer of protection analysis (CLOPA) that extends the existing layer of protection analysis (LOPA) framework to include failures caused by cyber attacks. The proposed method provides a rigorous mathematical formulation that expresses quantitatively the tradeoff between designing a highly reliable and a highly secure CPS. We further propose a co-design lifecycle process that integrates the safety and security risk assessment processes. We evaluate the proposed CLOPA approach and the integrated lifecycle on a practical case study of a process reactor controlled by an industrial control testbed and provide a comparison between the proposed CLOPA and current LOPA risk assessment practice.**

***Index Terms*—Cyber-physical system (CPS), hazard and operability (HAZOP), IEC 61511, layer of protection analysis (LOPA), NIST SP 800-30, risk assessment, supervisory control and data acquisition, safety instrumented system (SIS), safety integrity level (SIL), security.**

I. Introduction

A CYBER physical system (CPS) is an integration of a physical process with computation and networking required for physical system monitoring and control. The integration of process dynamics with those of computation and networking brings a plethora of engineering challenges. As the majority of CPSs are deployed in mission-critical applications, the dependability and resilience to failures is a key design property for modern CPS.

To ensure that a given CPS is dependable, a risk assessment is carried out at both design time and operation time. The risk assessment process highlights the system weaknesses and helps define the safety requirements that need to be met to achieve the target reliability measures. The classical approach to perform the risk assessment is to consider physical failures only. As state-of-the-art CPS designs move to open-source hardware and software, cyber attacks have become a source of failure that cannot be ignored.

Realizing the critical nature of CPS cyber attacks and their impact on the safety of people and environment, as well as the potential catastrophic financial losses, the research community developed several approaches to integrate security aspects into the safety risk assessment process. This integration has been done mainly by extending the reliability modeling formalism to incorporate security-related risks. One of the overlooked research questions is how safety and security interact with each other and how this interaction would impact the overall system design. Putting this research question in a different format: *Is there a tradeoff between designing a highly reliable and a highly secure system?* a related research question is: *If we ignore the cyber security attacks in the design process, what is the impact on the overall system reliability? Is the reliability gain worth the complexity introduced by integrating security both at design and runtime?* A follow-up research question is: *Under what conditions can we ignore security failures?*

In order to better understand the interaction between safety and security lifecycles in the system design process, we consider in this article the safety risk assessment process and study the impact of overlooking failures caused by cyber attacks. We refer to such failures as security failures in the rest of this article. By formally introducing the failures caused by attacks into the risk assessment process, we can define the reliability requirements for the cyber components of the system as a function of both the failure rate of physical components and the resilience to cyber attacks. This formal requirement specification enables us to understand the design tradeoff between higher reliability of physical components versus higher resilience of cyber components, and the sensitivity of the overall system performance to both types of failures. In addition, we can gain insight into the interplay between safety and security and how to integrate both lifecycles

Manuscript received February 24, 2021; revised September 22, 2021 and January 16, 2022; accepted March 11, 2022. Date of publication April 22, 2022; date of current version June 2, 2022. This work was supported by the Qatar National Research Fund (a member of the Qatar Foundation) under Grant NPRP 9-005-1-002. The statements made herein are solely the responsibility of the authors. Associate Editor: W. Dong. *(Corresponding author: Ashraf Tantawy.)*

Ashraf Tantawy is with the School of Computer Science and Informatics, De Montfort University, LE1 9BH Leicester, U.K. (e-mail: ashraf.tantavy@dmu.ac.uk).

Sherif Abdelwahed is with the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284 USA (e-mail: sabdelwahed@vcu.edu).

Abdelkarim Erradi is with the Department of Computer Science and Engineering, Qatar University, Doha 2713, Qatar (e-mail: erradi@qu.edu.qa).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TR.2022.3163652.

Digital Object Identifier 10.1109/TR.2022.3163652

during the design process. More specifically, we consider the layer of protection analysis (LOPA), a widely adopted risk assessment method that follows a hazard identification study, such as hazard and operability (HAZOP). LOPA is carried out to identify whether an additional safety instrumented system (SIS) is needed for specific hazardous scenarios to achieve the target risk level. As a modern SIS is typically an embedded device, it has both physical and security failure modes. We mathematically derive the SIS design constraints in terms of both physical and security failure probabilities. Additionally, we propose an integrated safety–security design process that shows the flow of information between both lifecycles.

We can classify the research work on combining safety and security for CPSs into two broad categories that try to answer the following research questions: 1) given the independent safety and security lifecycles, what are the similarities/differences and how could the two lifecycles be aligned or unified? This research direction usually focuses on answering the question "what to do," rather than "how to do it"; and 2) for a given CPS, how can we carry out risk assessment (qualitative/quantitative) that considers both physical failures and cyber attacks? Consequently, how can we unify the process of safety and security requirements definition and verification? This research direction focuses on common modeling techniques that can incorporate both safety and security failures and often extends model-based engineering body of knowledge and tools to incorporate security requirements in the design process. In Section VI, we survey the main results for each research direction. A more thorough survey is presented in [1] and [2].

Our contribution: The work presented in this article addresses both research directions with a new approach. First, we integrate both safety and security lifecycles based on a rigorous mathematical formulation that captures their interaction. The formulation enables the designer to assess how a security design decision would impact system safety. This is in contrast to the existing research work that does not explicitly model the dynamic safety–security lifecycle interaction. Second, we develop an integrated safety–security design lifecycle and show in detail how to apply it to a real-world design, distinguishing the work from abstract research on risk assessment that does not carry over to the design stage. Finally, our approach is founded on LOPA, a practical approach that is extensively used in industry, giving the approach the merit for industrial implementation.

The rest of this article is organized as follows. Section II introduces the background information required for problem setup, including IEC 61511 safety lifecycle and the LOPA method, cyber dependence between control and safety systems, and cyber security risk assessment. Section III proposes a new LOPA mathematical formulation called cyber layer of protection analysis (CLOPA) that incorporates failures due to cyber attacks. Section IV proposes an integrated safety–security lifecycle process. Section V presents a case study for the design of a safety system for a chemical reactor, comparing classical LOPA approach to the proposed CLOPA formulation. Section VI summarizes the related work on safety–security co-design. Finally, Section VII concludes this article.

II. Safety and Security Risk Assessment

There are two main embedded systems that control and safeguard a given physical system: the control system and the safety system. In the process industry, the control system is referred to as the basic process control system (BPCS), and the safety system is referred to as the SIS. In practice, both the systems typically have a programmable controller architecture with one or more back planes, processor cards, and a variety of input–output interface cards [3]. For larger systems, BPCS and SIS architectures comprise multiple distributed nodes connected via a communication backbone. Fig. 2 depicts the two systems and their connectivity over a control network. In the following, we briefly discuss the SIS design lifecycle, BPCS and SIS security lifecycles, and their interaction.

A. IEC 61511 Safety Lifecycle Process

Fig. 1 shows the SIS design lifecycle according to IEC 61511 standard [4]. The design starts with hazard and risk assessment, where systems hazards are identified. HAZOP study, what if analysis, and fault tree analysis are the most common methods at this stage [5]. The risk assessment phase ranks each identified risk according to its likelihood and consequence, either quantitatively or qualitatively, and associates a risk ranking for each hazard. The resulting list of hazards and associated risk ranking is used as an input to the second phase focused on the allocation of safety functions to protection layers. This phase deals only with hazards that exceed a threshold risk rank that an organization is willing to accept. For each hazardous scenario, there is a target mitigated event likelihood (TMEL) measure that is defined based on the risk rank. The purpose of this phase is to check if the TMEL is met with existing protection layers. If not, an additional protection layer is recommended, often in the form of a new safety instrumented function (SIF) with a predefined safety integrity level (SIL) to cover the gap to the TMEL. The SIF comprises one or more sensors, a logic solver, and one or more actuators. The logic solver is commonly referred to as the SIS. An example SIF is illustrated in Fig. 6 for an overflow hazardous scenario of a reactor system, which will be discussed in detail in Section V. Risk matrix, risk graph, and LOPA are the most commonly used methods for the allocation of safety functions to protection layers [3].

The third phase is the development of the safety requirement specification (SRS), which documents all the functional and timing requirements for each SIF. The fourth phase is the detailed design and engineering. Phases 5–8 are concerned with system installation and commissioning, operation, modification, and decommissioning. Phase 2 is where the CPS control and safety systems are considered in the risk assessment process. Therefore, we study this phase in depth in this article. Since LOPA is the predominant approach for this phase, we limit our discussion to LOPA methodology. Other approaches could be adopted in a similar way.

The underlying assumption in LOPA analysis is that all protection layers, including the new SIF, are independent. In other words, if one layer failed, this does not increase or decrease

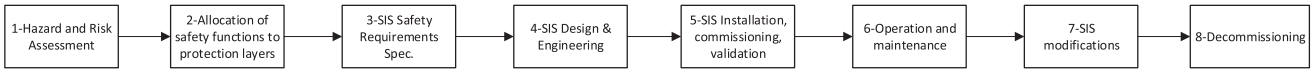


Fig. 1. IEC 61511 SIS design lifecycle (adopted from [4]).

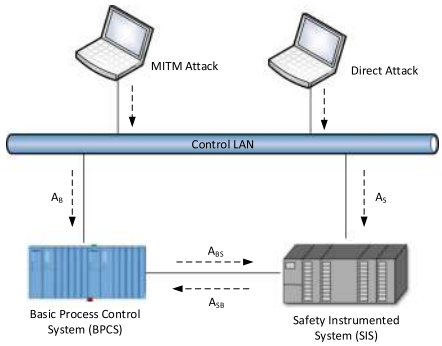


Fig. 2. Snapshot of an industrial control system architecture showing BPCS–SIS connectivity and potential attack vectors.

the likelihood of failure of the other layers. This assumption simplifies the mathematical analysis significantly, as it allows the multiplication of individual probabilities to obtain the required joint probability. The simplicity of LOPA calculations is probably one of the key reasons behind its widespread adoption by industry. Unfortunately, when cyber security is considered as a potential failure in LOPA analysis, the independence assumption between the control and safety systems no longer holds, as explained in the next section.

B. Control and Safety System Cyber Dependence

Each of the control and safety systems has two modes of failure: BPCS physical failures, B_p, BPCS security failure, B_c, SIS physical failure, S_p, and SIS security failure S_c. For physical failures, IEC 61511 standard strongly recommends complete separation between the control and safety systems of any plant. This separation includes sensors, computing devices, and final elements such as valves and motors. Separation also includes any common utility such as power supplies. The industry adopted this separation principle; hence, BPCS and SIS physical failures could be accurately assumed to be independent, i.e., $P[B_p, S_p] = P[B_p]P[S_p]$.

One exception to the separation between BPCS and SIS is the cyber communication link between the control and safety systems. Fig. 2 shows a snapshot of a typical industrial control system architecture showing the communication link between the BPCS and the SIS. BPCS–SIS communication could be over the control LAN or via a dedicated point to point serial link. The communication protocol is typically an open standard such as Modbus or DNP3 [6], [7]. This type of communication exists in many industrial installations to exchange plant data, as the data from field devices connected to the safety system are not accessible from the BPCS and *vice versa*. Given this architecture, we can define two attack vectors for SIS compromise: 1) a direct attack that exploits an existing controller vulnerability could be launched against the SIS node. This could be via any node on the control LAN or using man-in-the-middle (MITM) attack that exploits the BPCS–SIS communication. We designate this attack event by A_S in Fig. 2; and 2) by compromising the BPCS first and then exploiting the BPCS–SIS link to compromise the SIS. We designate this pivot attack by the sequence of events A_B and A_{BS} in Fig. 2. Furthermore, we designate the attack event from the SIS to the BPCS by A_{SB}. The attack sequence $A_B \rightarrow A_{BS}$ may be easier if the SIS is highly secured such that a direct attack may be infeasible. This is particularly true if we consider the fact that the BPCS is a trusted node to the SIS.

The above analysis shows a clear dependence between the control and safety systems that violates the original LOPA independence assumption. The security failures for the BPCS and the SIS are no longer independent because of the data communication coupling. We can formulate the different security failure probabilities as in (1)–(3) using basic probability laws with the aid of Fig. 2, where $P[A_i]$ is interpreted as the probability of success of attack A_i. Furthermore, the considered attack A_i should have the impact of stopping the BPCS or SIS from performing its intended control or safeguard function as related to the hazard under study. This is important because not all attacks that exploit controller vulnerabilities result in a process hazard. Therefore, the attacks considered represent a subset of the complete set of attacks that could exploit the BPCS or SIS vulnerabilities. Accordingly, from hereafter, $P[A_B]$, $P[A_S]$, $P[A_{BS}]$, and $P[A_{SB}]$ refer to the relevant attacks that cause a process hazard. This concept is revisited throughout this article and is made more clear in the case study when sample attacks are presented. For a case study example on the fact that not all cyber failures have a system reliability consequence, we refer the interested reader to [8] for a study on the impact of different software failure modes on system reliability for the electric power grid domain. Finally, it can be easily shown that if the BPCS–SIS communication link does not exist, or fully secured, then $P[A_{SB}] = P[A_{BS}] = 0$, and (3) reduces to the independent case $P[S_c, B_c] = P[S_c]P[B_c]$

$$P[B_c] = P[A_B] + P[A_S]P[A_{SB}] - P[A_B]P[A_S]P[A_{SB}] \tag{1}$$

$$P[S_c] = P[A_S] + P[A_B]P[A_{BS}] - P[A_B]P[A_S]P[A_{BS}] \tag{2}$$

$$P[S_c, B_c] = P[A_B](P[A_S] + P[A_{BS}]) + P[A_S]P[A_{SB}] - P[A_B]P[A_S](P[A_{BS}] + P[A_{SB}]). \tag{3}$$

C. Cyber Security Risk Assessment

The calculation of the probability of cyber attacks A_S, A_B, and A_{BS} could be performed during the cyber security risk assessment process. This requires a detailed specification of

the BPCS and the SIS and their connectivity, including the embedded system hardware, operating system, running software services, and the network connectivity. According to the National Institute of Standards and Technology (NIST) SP 800-30 standard, "Guide for Conducting Risk Assessments," the cyber security lifecycle process stages are: 1) asset identification, where the particular cyber components and their criticality levels are identified; 2) vulnerability identification, along with the associated threats and attack vectors; 3) the development of relevant attack trees for each attack scenario identified; 4) penetration testing to validate the vulnerability findings and attack scenarios and to help estimating the effort and probability for individual attack steps for each scenario; and 5) risk assessment to identify the scenarios with unacceptable risk [9]. Fig. 5 shows the BPCS and SIS cyber security lifecycles. In this article, we follow the same cyber security lifecycle, but with the physical process as the main focus. Therefore, for asset identification, the cyber component criticality is primarily identified by its failure impact on the operation of the connected physical component. Similarly, for vulnerability identification, threats and attack vectors are filtered by their impact on the physical process. Attacks that do not disturb the controlled process are ignored as they have no direct impact on the process safety. In addition, such impacts take place with much higher probability at the corporate network level, so they can be ignored with minimum impact on the risk assessment at the control network level. For more detailed discussion on process-driven attack identification, we refer the reader to [10].

For the presented architecture, the BPCS and the SIS are the critical components in direct contact with the process. The calculation of the required BPCS and SIS security failure probabilities could be typically carried out with the aid of attack trees [11]. The attack tree enumerates all the possible routes to compromise the system, and each edge is assigned a probability representing the likelihood of the associated event. Using basic probability laws, the overall probability of a system compromise could be calculated. Section V presents an example of such calculation. We re-emphasize the fact that the scope of cyber security risk assessment and attack trees in this case will be limited to attacks targeting the physical process to cause a process hazard. Although information security attacks with objectives such as stealing information are possible, most of this information is already available at the corporate network level, and an attacker who penetrates down to the control network level to compromise an BPCS or SIS will conceivably have the goal of physical process attack.

III. CLOPA: LOPA With Security Failures

A. Mathematical Formulation

In risk assessment, an initiating event is an unplanned event that when occurring may lead to a hazard. Examples of initiating events include equipment failure, human error, and cyber attacks. A system hazard will take place if one or more of the initiating events occur, and all the associated protection layers against that hazard fail simultaneously. The main objective of LOPA is to calculate the expected number of hazardous events per time interval and compare it to the TMEL. We designate the random variable representing the number of events per unit time for a specific initiating event by N, the random variable representing the simultaneous failure of protection layers when the initiating event occurs by L, where L is Bernoulli distributed with success probability p, and the random variable representing the number of hazards per time interval by H. We then have

$$H = \sum_{i=1}^{N} \mathbb{I}_E(l_i) \tag{4}$$

where $\mathbb{I}$ is the indicator function, and the set $E = \{l : l_i = 1, i = 1:N\}$. We note that for a given $N = k$, H is a binomially distributed random variable with expected value $E[H|N = k] = kp$. Therefore

$$\begin{aligned} E[H] &= \sum_{k=0}^{\infty} E[H|N = k]P[N = k] \\ &= p\sum_{k=0}^{\infty} kP[N = k] = pE[N] = p\lambda \end{aligned} \tag{5}$$

where λ represents the expected value of the number of initiating events per unit time, N. Although N is typically modeled by a Poisson random variable in reliability engineering, we do not assume any specific distribution in the analysis. This is particularly important because some initiating events considered in the article, such as security failures, are not accurately modeled by a Poisson distribution.

Equation (5) is the underlying mathematical concept behind LOPA analysis. Essentially, for each initiating event, the likelihood λ is estimated from field data, and the probability of simultaneous failure of all protection layers is specified. Finally, the expected number of hazards per unit time, $E[H]$, considering all initiating events, is estimated and compared to the prespecified TMEL. If $E[H] >$ TMEL, then a SIS is required with a probability of failure on demand $P[S_p]$ (or equivalently a risk reduction factor (RRF) $= 1/P[S_p]$) that achieves $E[H] \leq$ TMEL.

In order to express the LOPA formula in (5) in terms of all protection layers, including the BPCS and the SIS, we introduce some mathematical notation. We designate the set of initiating events for a given hazardous scenario by $\mathcal{I} = \{I_1, I_2, \ldots, I_n, B_p, \mathcal{A}_r\}$, where n is the number of possible initiating events excluding BPCS failures, B_p denotes the BPCS physical failure event, and $\mathcal{A}_r$ denotes the set of attacks relevant to the hazard under study. We express the associated set of event likelihoods by $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n, \lambda_p, \lambda_c\}$. Furthermore, we denote the set of all possible protection layers by $\mathcal{L} = \{L_1, L_2, \ldots, L_m\}$, where m is the number of existing protection layers, excluding the BPCS and the SIS. BPCS protection is denoted by B, and SIS protection is denoted by S. For each initiating event i, there is a subset of protection layers $\mathcal{L}_i \subseteq \mathcal{L} \cup \{B, S\}$ that could stop the propagation of a hazard from causing its consequences. Table I shows a sample LOPA table using the introduced terminology.

TABLE I
SAMPLE LOPA TABLE

Initiating Event	Likelihood λ_i (/yr)	L_1 ... L_m	BPCS (B)	TMEL
I_1	λ_1	$\leftarrow P[\mathcal{L}_1] \rightarrow$	$P[B]$	10^{-x}
...	...		...	
I_n	λ_n	$\leftarrow P[\mathcal{L}_n] \rightarrow$	$P[B]$	
B_p	λ_p	$\leftarrow P[\mathcal{L}_B] \rightarrow$	1	
$\mathcal{A}_r$	λ_c	$\leftarrow P[\mathcal{L}_B] \rightarrow$	$P[B]$	

$P[\mathcal{L}]$ refers to the combined probability of failure of protection layers applicable to the initiating event from L_1 to L_m.

B. Semantically Relevant Attack Event Formulation

We designate the set of all possible attacks against the BPCS by $\mathcal{A}$. For a given hazard under consideration, the subset of attacks that would lead to this hazard, i.e., the contextually or semantically relevant attacks, is designated by $\mathcal{A}_r$. To estimate the likelihood of relevant cyber attacks (expected number per unit time), λ_c, we assume an attacker profile with an average rate of launching attacks per unit time λ. For every launched attack, the attacker is presented with the complete set of attacks $\mathcal{A}$ and selects only one attack a with probability $P[A = a] = \alpha_a$ that is dependent on the attacker profile, such that

$$\sum_{a \in \mathcal{A}} P[A = a] = \sum_{a \in \mathcal{A}} \alpha_a = 1. \tag{6}$$

The likelihood of cyber attack $a \in \mathcal{A}_r$ is then $\lambda_a = \lambda \alpha_a$. This cyber attack has the potential to cause a system hazard if both the BPCS and the SIS fail jointly to stop the attack (either physical or cyber failure). We designate this probability by $P_a[S, B]$. The exact approach to include every attack $a \in \mathcal{A}_r$ in the LOPA table is to treat each attack as an individual entry, akin to the last row in Table I, with initiating event likelihood λ_a. However, this approach has two main drawbacks: First, the number of attacks could be large, and this would grow the LOPA sheet significantly. Second, the treatment of each attack individually would not allow us to utilize attack modeling techniques such as attack trees that model collectively all the possible attack paths for one attack objective. Therefore, we adopt an alternative approach, where all relevant attacks $a \in \mathcal{A}_r$ could be represented by one entry in the LOPA table. The following lemma summarizes the approximate solution. The proof is included in the Appendix.

Lemma 3.1: Assume a given hazard scenario H, a control system BPCS, an average rate of launching attacks against BPCS λ, Hazard H semantically-relevant attack set $\mathcal{A}_r$, and probability α_a of selecting attack $a \in \mathcal{A}_r$. Then, the impact of all initiating events $a \in \mathcal{A}_r$ on the LOPA calculation could be approximated by a single initiating event with likelihood $\lambda_c = \lambda \sum_{a \in \mathcal{A}_r} \alpha_a$ and a BPCS failure probability with respect to the combined set of attacks $a \in \mathcal{A}_r$, where each attack probability is weighted by the factor $\gamma_a = \alpha_a / \sum_{a \in \mathcal{A}_r} \alpha_a$.

The lemma enables us to use attack trees with leaf nodes weighted by γ_a to calculate the BPCS security failure probability in response to the combined set of attacks $\mathcal{A}_r$ with likelihood $\lambda \sum_{a \in \mathcal{A}_r} \alpha_a$. For the special case where the cyber attacker profile results in random selection of the attack $a \in \mathcal{A}$, e.g., an attacker with no knowledge about the system, the likelihood reduces to $\lambda |\mathcal{A}_r| / |\mathcal{A}|$ and the leaf node weights reduce to $\gamma_a = 1/|\mathcal{A}_r| \, \forall a$. We use this special case in the case study in Section V.

C. Cyber LOPA Formulation

With the introduced notation, the expected number of hazards in (5), which should be less than the TMEL, could be expanded as

$$E[H] = P[S, B] \left(\sum_{i=1}^{n} (\lambda_i P[\mathcal{L}_i]) + \lambda_c P[\mathcal{L}_B] \right) + \lambda_p P[S] P[\mathcal{L}_B] \leq \text{TMEL} \tag{7}$$

where $\mathcal{L}_B$ is the set of protection layers for BPCS physical or security failure event, and we assume that all protection layers are independent of the BPCS and the SIS, while keeping the dependence between the BPCS and the SIS. In addition, higher order probability terms resulting from multiple initiating events are ignored due to their insignificance.

To calculate the joint failure probability $P[S, B]$, we use basic probability laws and the fact that the BPCS and the SIS have both the physical and cyber modes of failure, as explained in Section II-B, to obtain

$$\begin{aligned} P[S, B] = {} & P[S_p] \left(P[B_p](1 - P[B_c] - P[S_c]) + P[B_c] \right) \\ & + P[S_c, B_c] \left(1 - P[S_p] - P[B_p] + P[S_p] P[B_p] \right) \\ & + P[S_c] P[B_p]. \end{aligned} \tag{8}$$

Substituting (8) into (7), we obtain the general LOPA equation in (9). We call this expanded version of LOPA hereafter CLOPA

$$P[S_p] \leq \frac{\beta - (\alpha_1 P[S_c] + \alpha_2 P[S_c, B_c])}{\alpha_1 - \alpha_1 P[S_c] + \alpha_2 P[B_c] - \alpha_2 P[S_c, B_c]} \tag{9}$$

where

$$\alpha_1 = P[B_p] \left(\sum_{i=1}^{n} (\lambda_i P[\mathcal{L}_i]) + \lambda_c P[\mathcal{L}_B] \right) + \lambda_p P[\mathcal{L}_B] \tag{10}$$

$$\alpha_2 = (1 - P[B_p]) \left(\sum_{i=1}^{n} (\lambda_i P[\mathcal{L}_i]) + \lambda_c P[\mathcal{L}_B] \right) \tag{11}$$

$$\beta = \text{TMEL}. \tag{12}$$

In order to define the CLOPA formula in terms of the actual design variables $P[A_S]$ and $P[A_{BS}]$ that represent the probability of security failures of actual CPS components, we substitute (1)–(3) into (9) to obtain

$$P[S_p] \leq \frac{\beta - \gamma_1 P[A_S] - \gamma_2 P[A_{BS}](1 - P[A_S])}{\gamma_3 - \gamma_3 P[A_S] - \gamma_2 P[A_{BS}](1 - P[A_S])} \tag{13}$$

where:

$$\gamma_1 = \alpha_1 + \alpha_2 [P[A_B] + P[A_{SB}](1 - P[A_B])] \tag{14}$$

$$\gamma_2 = (\alpha_1 + \alpha_2) P[A_B] \tag{15}$$

$$\gamma_3 = \alpha_1 + \alpha_2 P[A_B]. \tag{16}$$

Equation (13), along with (10)–(12) and (14)–(16), represents the general CLOPA formulation to design the SIS. It represents

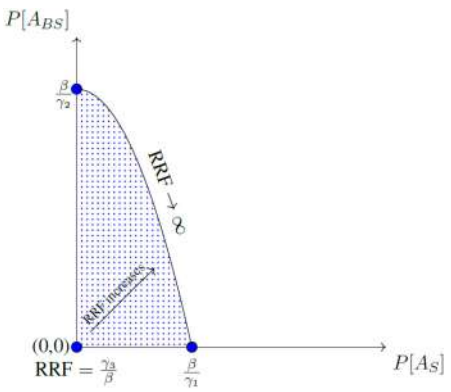


Fig. 3. CLOPA design region (shaded). Any point in the shaded region results in a feasible SIS. Points near the boundary require an SIS with a very high RRF value and, hence, difficult to obtain in practice.

an upper bound on the probability of physical failure for the safety system in terms of the security failure probabilities, showing clearly the coupling between the safety system and security system design. The design of the SIS should satisfy (13), where the design variables are $P[S_p]$, $P[A_S]$, and $P[A_{BS}]$. The rest are model parameters that are predetermined, including the BPCS failure marginal probabilities. This is because the BPCS design is independent of the SIS design and usually takes place earlier in the engineering design cycle. Note that we assume here that $P[A_{SB}]$ is a known parameter. This is because by completely defining the BPCS and its hardware and software specifications, the probability of a cyber attack compromising process safety could be estimated, even though the SIS is not yet completely defined. Table 1 in the Appendix summarizes the model variables, parameters, and how the model parameters are calculated.

It should be noted that with existing LOPA methodology, security failures are ignored, i.e., $P[A_B] = P[A_S] = P[A_{BS}] = P[A_{SB}] = 0$. Substituting these zero values into (13), we obtain the classical LOPA formulation

$$P[S_p] \leq \frac{\beta}{\alpha_1} = \frac{\text{TMEL}}{P[B_p]\sum_{i=1}^{n}(\lambda_i P[\mathcal{L}_i]) + P[\mathcal{L}_B](\lambda_p + \lambda_c)}. \tag{17}$$

D. Design Space

Using the fact that $P[S_p] \geq 0$ for a realizable safety system in (13), we obtain

$$\gamma_1 P[A_S] + \gamma_2 P[A_{BS}] - \gamma_2 P[A_S]P[A_{BS}] \leq \beta. \tag{18}$$

Fig. 3 shows the shaded region defined by the inequality in (18). The boundary curve is defined by (18) when the following equality holds:

$$P[A_{BS}] = \frac{\beta}{\gamma_2}\left(\frac{1 - \left(\frac{\gamma_1}{\beta}\right)P[A_S]}{1 - P[A_S]}\right). \tag{19}$$

The first-order derivative of the boundary curve is negative for $\gamma_1/\beta > 1$ and positive otherwise. Since $\gamma_1 > \beta$ to require a safety system [proof is straightforward by inspecting (10), (11), (14), and (17)], the boundary curve is concave as in Fig. 3. We note that any point in the shaded region results in a feasible SIS. Points on the boundary curve result in $P[S_p] = 0$ or equivalently RRF $\to \infty$. Points closer to the boundary would have high values for the RRF, requiring a very highly reliable SIS that may not be achievable in practice. Points closer to the origin result in lower RRF. It can be easily shown that the contour lines for (13), where $P[S_p] = C$, could be expressed as

$$P[A_{BS}] = \frac{C\gamma_3 - \beta}{\gamma_2(C-1)}\left(\frac{1 - \left(\frac{C\gamma_3 - \gamma_1}{C\gamma_3 - \beta}\right)P[A_S]}{1 - P[A_S]}\right). \tag{20}$$

The contour line that represents the design boundary in Fig. 3 can be derived from (20) by setting $C = 0$.

We can extract the following information from this graph.

1) The maximum probability of security failure for the safety system by directed attacks is β/γ_1. This probability results in an unrealizable safety system, as the required RRF $\to \infty$.
2) The maximum probability of security failure for the safety system by pivot attack via the BPCS is β/γ_2. Likewise, this probability does not result in a realizable safety system.
3) Finally, the minimum value of the RRF is achieved at the origin for a perfectly secured safety system, where $P[A_S] = P[A_{BS}] = 0$, with the RRF given by

$$P[S_p]_{\max} = \frac{\beta}{\gamma_3}, \quad \text{RRF}_{\min} = \frac{\gamma_3}{\beta}. \tag{21}$$

Clearly, points outside the shaded region result in nonrealizable SIS. This result re-emphasizes the interplay between the safety and security systems of a CPS.

The design space highlights the major difference between LOPA and CLOPA. In LOPA, the SIS requirement is related to reliability in the form of the required SIL. In CLOPA, an additional requirement for the SIS is its security resilience, in the form of an upper bound on the probability of a security failure (cyber attack success), either directly or indirectly via the BPCS.

E. Classical LOPA Error

To obtain the error resulting from using classical LOPA, we subtract (17) from (13) to obtain

$$e_{\text{RRF}} = \frac{\zeta_1 + \zeta_2 P[A_S] + \zeta_3 P[A_{BS}](1 - P[A_S])}{\beta\left[\beta - \gamma_1 P[A_S] - \gamma_2 P[A_{BS}](1 - P[A_S])\right]} \tag{22}$$

where

$$\zeta_1 = \beta(\gamma_3 - \alpha_1) = \beta\alpha_2 P[A_B] \tag{23}$$

$$\zeta_2 = \alpha_1\gamma_1 - \beta\gamma_3 \tag{24}$$

$$\zeta_3 = \gamma_2(\alpha_1 - \beta). \tag{25}$$

The minimum error occurs for a perfectly secured safety system, i.e., $P[A_S] = P[A_{BS}] = 0$:

$$\min e_{\text{RRF}} = P[A_B]\left(\frac{\alpha_2}{\beta}\right). \tag{26}$$

The error will be zero, i.e., classical LOPA result matches CLOPA, if the probability of BPCS security failure via a direct attack is zero.

IV. Safety–Security Co-design

A. Design Process

The current industrial practice is to perform safety and security risk assessments independently, treating the physical and cyber components of a CPS as two separate entities. As illustrated in Section III, accurate safety risk assessment requires knowledge about the cyber components and their security failure probabilities. Formally, the objective is to design an SIS architecture $\mathcal{A}$ that satisfies (13) in terms of both physical and security failure probabilities. Suppose that the architecture $\mathcal{A}$ could be represented by a set of design variables represented by the vector $\mathbf{x}$. If we can relate the physical and security failure probabilities to the vector $\mathbf{x}$ by $P[A_S] = f(x), P[A_{BS}] = g(x), P[S_p] = h(x)$, then we can use these functions to substitute the relevant probabilities into (13), and our design problem will be to find a set of values for the vector $\mathbf{x}$ that satisfies the CLOPA constraint (13). Unfortunately, this design approach is not followed by industry for several reasons. First, abstracting a given architectural design $\mathcal{A}$ into a set of design variables is a very difficult task, not to mention that these design variables have to be linked to both physical and security failures. Second, finding an exact or approximate representation of the functions $f(.), g(.)$, and $h(.)$ that relate the failure probabilities to the design variables may not be possible, as it is not always clear how a design decision would result in a higher or lower probability of failure. Finally, even if we were able to make a perfect modeling, the resulting problem to solve may turn into a discrete optimization problem that is not possible to solve in polynomial time.

Owing to these modeling limitations, the current industrial practice to design SISs (excluding cyber attacks) is to follow an iterative process and rely on engineering judgment during the design process. More precisely, the required risk reduction factor RRF_d is initially calculated; then, the engineering design proceeds to achieve RRF_d using both experience and industrial standard guidelines [4]. After the design is completed, design verification is conducted to calculate the RRF of the proposed design RRF_v. If the resulting $\text{RRF}_v \geq \text{RRF}_d$, then the design stops. Otherwise, the design is refined until the condition $\text{RRF}_v \geq \text{RRF}_d$ is satisfied. In the following, we will adopt the same iterative design approach for CLOPA.

Fig. 4 illustrates the iterative design process. We start with initial values ($P[A_S], P[A_{BS}], \text{RRF}_d$) that satisfy the CLOPA constraint in (13). We then proceed with the SIS design to produce an architecture $\mathcal{A}$. The architecture is then verified to estimate its probability of failure on demand or equivalently its RRF_v. The architecture is also used to carry out a security risk assessment to estimate the probability of security failures $P[A'_S]$ and $P[A'_{BS}]$. If the new set of obtained values ($P[A'_S], P[A'_{BS}], \text{RRF}_v$) still satisfy the CLOPA equation, the design stops. Otherwise, a new iteration will start to adjust the design in order to achieve the CLOPA constraint. This adjustment could be either by adding more security controls or by increasing the reliability of the system using fault-tolerant techniques. Algorithm 1 summarizes the iterative design process.

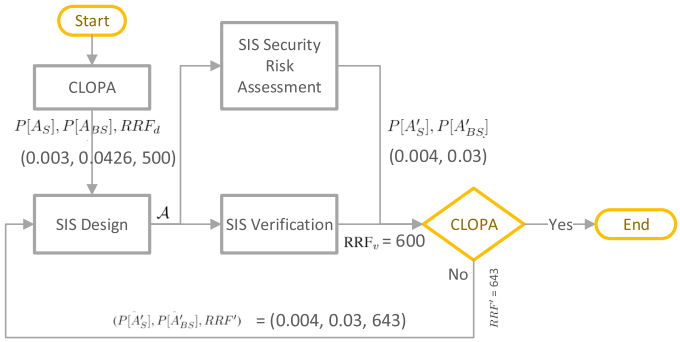


Fig. 4. CLOPA iterative design process. CSTR case study design values are shown.

One question is how can we choose the initial values for RRF , $P[A_S], P[A_{BS}]$? This initial design point could be selected with the aid of the design contour plot as in Fig. 3, where the design point strikes a balance between security and reliability. What is *reasonable* regarding the RRF is well documented in the standards using SILs, and the extra cost to move from one SIL to a higher SIL is well quantified in industry. What is not very clear, though, is what is an achievable value for the probability of security failures. This is still not a well-developed field, and the argument of how to assess such probabilities is still going on in the research community.

Another important question is whether there is any formal guarantees that Algorithm 1 will terminate. To answer this question, we need to know, or at least approximate, how the architectural design $\mathcal{A}$ impacts the RRF, $P[A_S]$, and $P[A_{BS}]$. As pointed out earlier, this is very hard in practice. Without such relationship, the question of convergence to a solution for algorithm termination cannot be precisely answered. However, in practice, modifying the SIS design to increase the RRF is usually done by changing sensor and actuator configuration or reliability figures, as they are often the weakest links in the reliability chain, while the logic solver is minimally changed [4]. Accordingly, for all practical purposes, we can assume that the design process will converge after few runs.

B. Integrated Safety–Security Lifecycle

As the analysis in this article shows a clear coupling between safety and security design requirements, we propose the integrated lifecycle in Fig. 5. In the following, we present a brief description of the lifecycle steps in the order of their execution, according to the numbering labels in Fig. 5.

① *SIS safety lifecycle—HAZOP:* The first step is to carry out the hazard analysis for the physical system, often using HAZOP. This process identifies important assets that may be subject to, or contribute to, risk scenarios. Then, the process identifies all feasible hazards and associated risk ranking, as well as the associated cyber components for each identified hazard. This constitutes an input to the BPCS security lifecycle. If we designate the set of

Algorithm 1: Integrated Safety-Security Lifecycle Design Algorithm.

input : BPCS
output: $\mathcal{A}$, $\boldsymbol{\theta}_S$
$([P[A_B], P[A_{SB}]) \leftarrow$ `BPCS-SecCycle`(BPCS) ;
$\boldsymbol{\theta}_B \leftarrow (P[A_B], P[A_{SB}])$;
$(P[A_S], P[A_{BS}], \text{RRF}_d) \leftarrow$ `DesignContour`$(\boldsymbol{\theta}_B)$;
$\boldsymbol{\theta}_S \leftarrow (P[A_S], P[A_{BS}], \text{RRF}_d)$;
do
  $\mathcal{A} \leftarrow$ `SIS-SafeCycle`$(\boldsymbol{\theta}_S)$;
  $\text{RRF}_v \leftarrow$ `SIS-Verify`$(\mathcal{A})$;
  $(P[A'_S], P[A'_{BS}]) \leftarrow$ `SIS-SecCycle`$(\mathcal{A})$;
  $\text{RRF}' \leftarrow$ `CLOPA`$(P[A'_S], P[A'_{BS}], \boldsymbol{\theta}_B)$;
  $\boldsymbol{\theta}_S \leftarrow \left(P[A'_S], P[A'_{BS}], \text{RRF}'\right)$;
while $RRF_v < RRF'$;
return $\mathcal{A}$, $\boldsymbol{\theta}_S$;

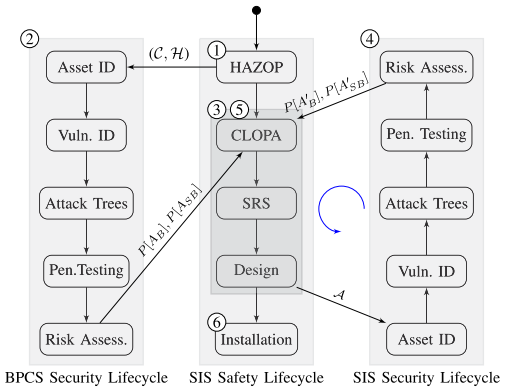


Fig. 5. Integrated safety and security lifecycles. The process starts at ① HAZOP, followed by ② BPCS complete security lifecycle, then ③ SIS safety lifecycle up to the end of the design stage, followed by ④ complete SIS security lifecycle, ⑤, CLOPA check, and possibly several iterations of steps ③, ④, and ⑤ and then terminates at the SIS installation stage.

hazards by $\mathcal{H}$, and the set of cyber components by $\mathcal{C}$, then the output from this process is the function $f : \mathcal{H} \mapsto \mathbb{R}$ representing the risk ranking and the relation $R \subseteq \mathcal{H} \times \mathcal{C}$ representing the cyber components for each hazard.

② *BPCS cyber security lifecycle:* The BPCS cyber security lifecycle, including vulnerability analysis, attack tree generation, penetration testing, and risk assessment, is performed on the BPCS. Ideally, the security risk assessment should be carried out for each process hazard scenario identified during HAZOP to identify the relevant vulnerabilities that may cause a process disruption. However, for the given centralized architecture, the BPCS is typically controlling a large number of control loops; hence, it may not be necessary to repeat the security risk assessment process for each control loop, as vulnerabilities may be applicable to several hazardous scenarios. The output of this process is the BPCS security failure probabilities $P[A_B]$ and $P[A_{SB}]$.

③ *SIS safety lifecycle—CLOPA and SIS design:* The first iteration of CLOPA and SIS design will proceed according to Algorithm 1 and Fig. 4. The CLOPA calculates the design requirement for the SIS in terms of its reliability as defined by the RRF, and its cyber security resilience as defined by $P[A_S]$ and $P[A_{BS}]$. The SIS design then proceeds according to IEC 61511 standard [4] to produce an architecture $\mathcal{A}$. The design includes the hardware architecture, redundancy scheme, and software architecture. The specific design architecture can vary across industries and organizations, but the design has to achieve the required RRF, $P[A_S]$ and $P[A_{BS}]$, as calculated by CLOPA. After the design is completed, SIS verification is carried out to calculate the RRF_v. It should be highlighted that the SIS is one component only of the SIF. The SIF includes the sensor, SIS, and the actuator. Therefore, the verification is carried out on the whole SIF. For a detailed discussion on SIS design and verification, the reader is referred to [3].

④ *SIS cyber security lifecycle:* Using the resulting SIS design hardware and software architecture $\mathcal{A}$, the SIS security lifecycle is carried out. Since the SIS is not yet implemented at this stage, SIS penetration testing is not possible and, hence, omitted from the security lifecycle. The output from this process is the SIS security failure probabilities $P[A'_S], P[A'_{BS}]$, derived from SIS vulnerabilities that may lead to a process hazard. It is noted that the SIS security lifecycle at the right of Fig. 5 proceeds from bottom to top for a better presentation.

⑤ *Safety lifecycle—CLOPA:* The CLOPA calculation is carried out using the values obtained from the safety verification and SIS security lifecycle, $(P[A'_S], P[A'_{BS}], \text{RRF}_v)$, to verify that the architecture $\mathcal{A}$ satisfies the CLOPA constraint. The process SIS safety lifecycle $\rightarrow$ SIS Cyber security lifecycle $\rightarrow$ CLOPA (designated by the blue arrowed arc in Fig. 5) repeats until the CLOPA constraint is satisfied.

⑥ *Installation:* The finalized design then moves to the installation phase.

V. Integrated Design Example

In this section, we present an integrated design example for a process control system to illustrate the proposed CLOPA and integrated lifecycle. The system described in this section is a real testbed located in Qatar University and comprises the process simulator and the full plant control system. As the integrated design lifecycle is substantial, with some steps outside the scope of this article (e.g., SIS architectural design and security risk assessment), it is not possible to present the design process in full details. However, we try to focus on the big picture as related to the proposed CLOPA, while discussing briefly each design step. Wherever needed, we refer the reader to relevant references for further details.

A. CPS Description

We consider the continuous stirred tank reactor (CSTR) process illustrated in Fig. 6. The reactor vessel has an inlet stream carrying the reactant A, an outlet stream carrying the product

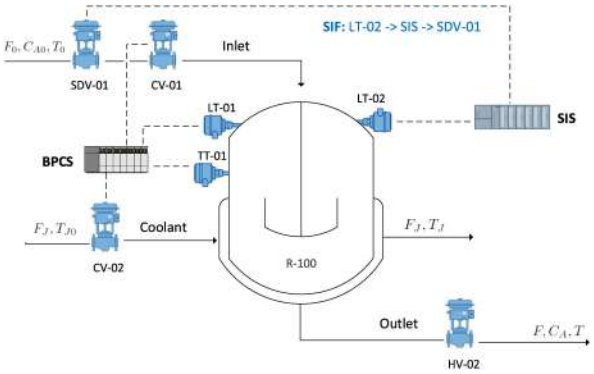


Fig. 6. Reactor piping and instrumentation diagram. ISA standard symbols are not strictly followed for illustration purposes.

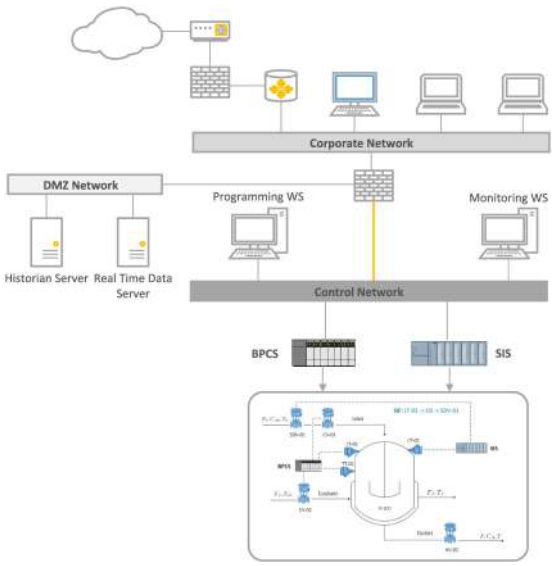


Fig. 7. CPS architecture for an industrial control system testbed, following NIST 800-82 guidelines. The firewall blocks any direct communication between the corporate network and the control network. Plant floor information is accessible only via the DMZ.

B, and a cooling stream carrying the cooling fluid into the surrounding jacket to absorb the heat of the exothermic reaction. A first-order reaction takes place where a mole fraction of reactant A is consumed to produce product B. The process has a level control loop (LT-01 $\rightarrow$ BPCS $\rightarrow$ CV-01) to maintain the liquid level in the reactor, and a temperature control loop (TT-01 $\rightarrow$ BPCS $\rightarrow$ CV-02) to control the reaction rate. The SIF (LT-02 $\rightarrow$ SIS $\rightarrow$ SDV-01) protects the reactor from the overflow hazard and will be explained later in this section. For more detailed explanation about the process including the state space model, the reader is referred to [12].

The CSTR process is controlled by the industrial control system shown in Fig. 7, which follows NIST 800-82 standard with one firewall and a DeMilitariZed (DMZ) zone [13]. The corporate network cannot communicate directly with the control network. The only allowable information flow paths via the firewall are from the control network to the data logging servers in the DMZ zone and from the corporate network to the DMZ for information retrieval. The BPCS and the SIS have Modbus/TCP communication over the control network [6].

B. Integrated Lifecycle

In the following discussion, we follow the integrated lifecycle in Fig. 5, and as per the itemized steps in Section IV-B, we have the following.

① *SIS safety lifecycle—HAZOP:* Table II shows the HAZOP sheet for the CSTR process. Each row contains: 1) the possible hazard; 2) all possible initiating events for each hazard whether mechanical or electronic failures; 3) consequences if the hazard occurred, including safety, financial, and environmental losses; 4) existing safeguards that could prevent the hazard from propagating and causing the consequences; and 5) the risk rank, which is typically a function of the consequences. There are two identified hazards for the reactor process: high level causing an overflow hazard, and high temperature that may lead to reactor runaway and possible meltdown. Both the hazards have high and very high risk rankings; therefore, the two risk scenarios qualify for further LOPA assessment. In the following, we limit our discussion to the high-level hazard scenario only. High-temperature hazard could be treated similarly.

② *BPCS cyber security lifecycle:* We need to calculate $P[A_B]$ and $P[A_{SB}]$ for the BPCS, the probability that the BPCS fails due to a direct attack and a SIS-pivot attack, respectively, in a way that generates the high-level process hazard. We conducted vulnerability identification on the CPS network in Fig. 7, constructed the attack trees, and carried out penetration testing to verify the vulnerability findings. We assumed an attacker profile where attacks are selected randomly. The total number of semantically relevant attacks is found to be $|\mathcal{A}_r| = 42$. We assume that relevant attacks represent 10^{-4} of all possible attacks, i.e., $|\mathcal{A}_r|/|\mathcal{A}| = 10^{-4}$. Therefore, according to Lemma 3.1, we obtain an initiating event likelihood $10^{-4}\lambda$ and a weight factor $\gamma_a = 0.024$ for each attack a at the leaf nodes of the attack tree. As the full details of vulnerability analysis, attack design, and penetration testing are beyond the scope of this article, we refer the interested reader to [10] and [14].

To compromise the BPCS, we assume the more realistic situation with no insider threat and no direct communication from the corporate network to the control network. In this scenario, the attacker has to detour to compromise the real-time (RT) server in the DMZ and use it as a pivot to attack the BPCS, either directly or via the monitoring workstation (designated HMI hereafter) that has legitimate communication with the BPCS. We start with the assumption that one of the corporate network PCs that has legitimate access to the RT server is compromised. There are several well-known attack vectors in the IT security domain to achieve such compromise, such as a spam email, a web service vulnerability, or an external malware USB, just to name a few. Fig. 8 shows an abstract attack tree that summarizes the BPCS compromise paths, where the database server compromise is a prerequisite attack step. In the following, we expand each of the leaf nodes in this abstract attack tree into the corresponding detailed attack trees. More detailed treatment of each attack tree as well as penetration testing could be found in [10].

TABLE II
PARTIAL HAZOP SHEET FOR THE REACTOR PROCESS

Hazard	**Initiating Event** (Cause)	**Consequences**	**Safeguards** (IPL)	Risk Rank
High Level (Reactor overflow)	BPCS failure OR Human error (misaligned valves)	2 or more fatalities (safety), Product loss (financial), Environmental contamination (environment)	Reactor dike (Mitigation)	High
High Temperature (Reactor Meltdown/explosion)	BPCS failure OR Coolant inlet control valve fully (partially) closed OR Inlet valve stuck fully open	10 or more fatalities (safety), Product loss (financial), Environmental contamination (environment)	None	V. High

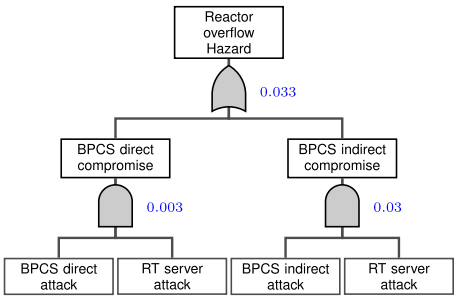


Fig. 8. Abstract attack tree to compromise the BPCS to generate overflow process hazard for the CSTR reactor. Leaf nodes are further expanded in Figs. 9–11.

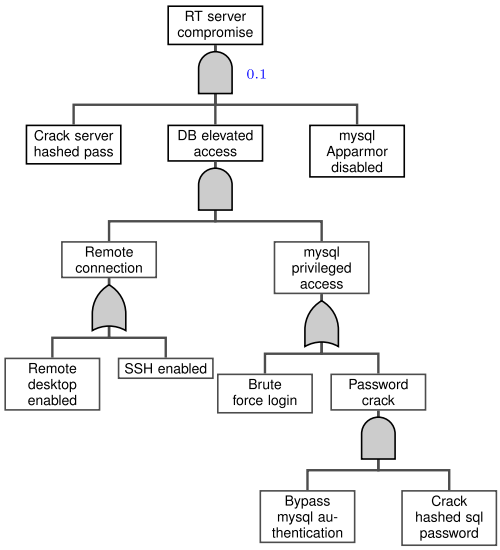


Fig. 9. RT server attack tree. Database vulnerabilities are exploited to gain root access and use the RT server as a pivot to attack the control network. DB elevated access is a prerequisite to crack the server hashed password. Leaf nodes that represent a distinct attack have a weight factor $\gamma_a = 0.024$ that is combined with their success probability. Therefore, Bypass mysql and Crack hashed sql password, Brute force login, SSH enabled, remote desktop enabled, and each has a weight factor combined with their success probability.

Fig. 9 shows the RT server attack tree. The basic idea is to exploit mysql database vulnerabilities via SSH to obtain the Linux server password Hash Dump and possibly crack the password to achieve privilege escalation and gain full control over the RT server. The probability of success of such an attack depends on several factors, including mysql configuration settings to allow brute-force login attack, mysql login password strength, configuration of mysql security monitoring app, and whether SSH is enabled. For the purpose of this case study, we choose this probability arbitrarily as 0.1. It should be highlighted that the attack tree does not have the sequence semantics to represent a sequence of attack steps. For example, the remote connection step has to be executed before the mysql privileged access in Fig. 9. We represent this sequence by the AND gate aggregator, noting that in some other cases the AND gate may represent simultaneous attack steps. For more information on attack trees and their semantics, the reader is referred to [15].

Fig. 10 shows the BPCS attack tree, which is divided into two main parts: DoS attack and integrity attack. The DoS attack may not lead to a reactor overflow unless there is a concurrent process disturbance that could not be controlled with the DoS-induced delayed BPCS control response. The probability of such disturbance could be estimated from plant information. The integrity attack injects a low-level measurement value for LT-01 to drive the BPCS controller to increase valve CV-01 opening or directly forces control valve CV-01 to open 100%. This will cause a reactor overflow if the SIS is not activated. The injection of the malicious value in the control loop could be accomplished by either gaining access to the controller and overwriting the control program or more simply sending Modbus packets to the controller with the malicious values. Modbus attack is much easier to launch but requires configuration data to identify the Modbus register address for either LT-01 or CV-01. The probability of BPCS indirect attack is chosen arbitrarily as 0.03.

Finally, Fig. 11 shows the attack tree for BPCS attack via the HMI. The attack is launched by remote desktop connection to the HMI and legitimately controlling CV-01 via the GUI. This indirect attack is easier than targeting the BPCS directly as it does not require knowledge about the controller configuration or Modbus register addresses associated with the sensor and valve of the targeted control loop. This is because all the information is already programmed in the GUI software. The probability of BPCS indirect attack is estimated to be 0.3. Using Fig. 8 and the three presented attack trees in Figs. 9–11, the total probability of BPCS attack that leads to an overflow hazard could be estimated by $P[A_B] \approx 0.033$.

It should be highlighted that the assignment of a probability measure to the success of attack actions is subject to debate in the research community, and there is no published agreed-upon data as in the case of reliability failure data. One approach is to use attack databases, such as NIST National Vulnerability Database (NVD) [16], to estimate the probability of a cyber attack success based on attributes such as required knowledge level and attack

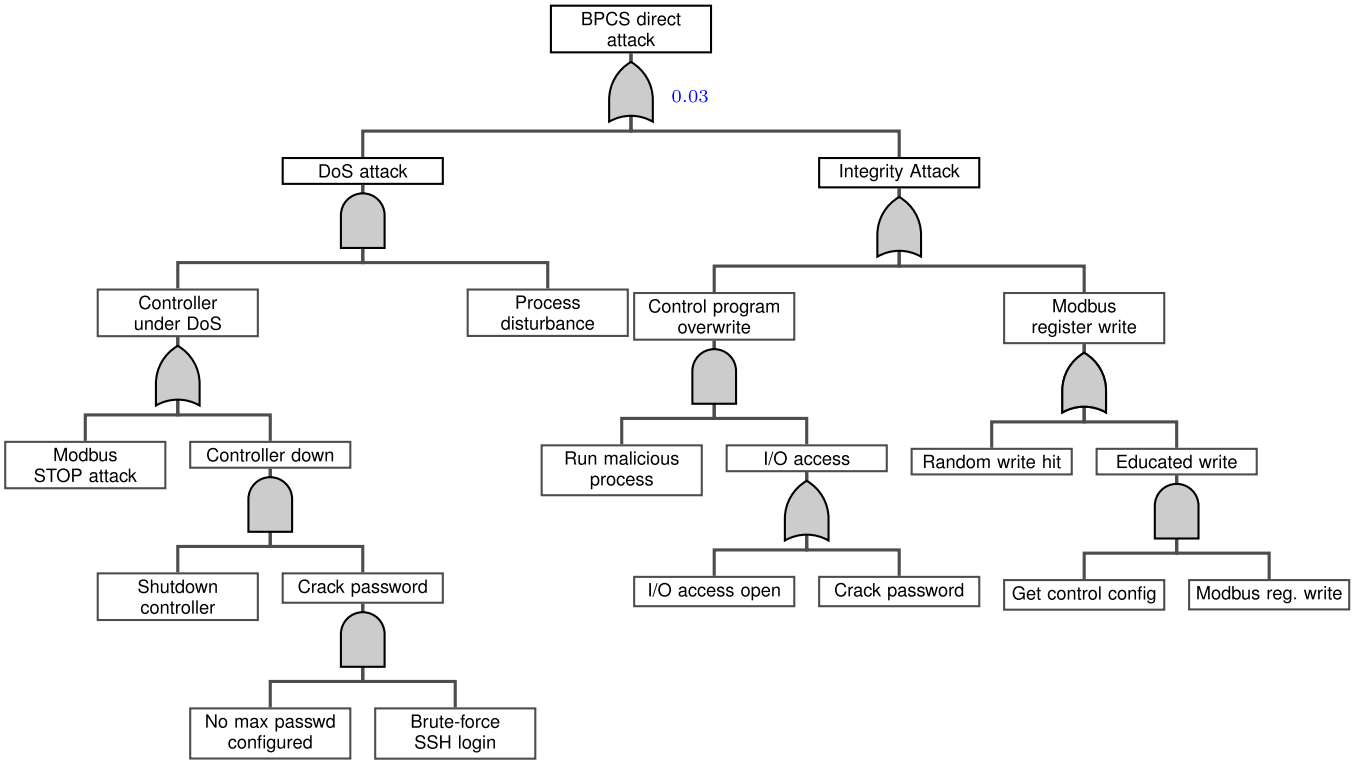


Fig. 10. Attack tree for BPCS compromise to generate a reactor overflow hazard. A DoS attack synchronized with a process disturbance or an especially crafted integrity attack would cause the CSTR to overflow. Leaf nodes that represent a distinct attack have a weight factor $\gamma_a = 0.024$ that is combined with their success probability.

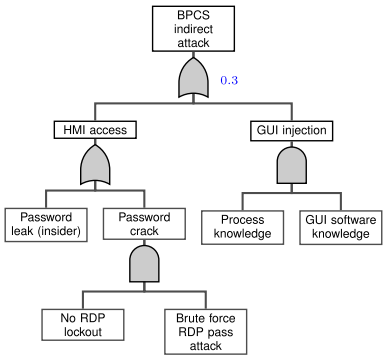


Fig. 11. HMI-BPCS indirect attack tree. The compromised HMI is used to embed the attack against the BPCS using the legitimate traffic between the GUI and the BPCS control program. Leaf nodes that represent a distinct attack have a weight factor $\gamma_a = 0.024$ that is combined with their success probability.

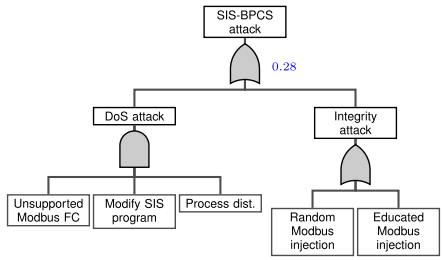


Fig. 12. SIS-BPCS attack tree. The SIS is used as a pivot to launch either a DoS attack or a crafted integrity attack that results in reactor overflow. Leaf nodes that represent a distinct attack have a weight factor $\gamma_a = 0.024$ that is combined with their success probability.

difficulty. However, this approach has the drawback that it does not take into account the specifics of each organization. In this article, we rely on the experience obtained during the penetration testing carried out by the research team in combination with NVD to assign the probability measures. This does not impact the analysis as the presented case study is meant for illustration purposes to explain the design process.

To calculate the probability of BPCS cyber attack leading to a process hazard given an SIS cyber compromise $P[A_{SB}]$, we focus on Modbus attack vectors for both integrity and DoS attacks. Integrity attacks target sensor LT-01 or valve CV-01 as before, either randomly or using leaked Modbus register configuration. DoS attack could be launched by utilizing non-programmed Modbus function code hoping that it would crash the BPCS Modbus master. Fig. 12 summarizes the attack tree, and the probability is chosen arbitrarily as $P[A_{SB}] \approx 0.2813$. To summarize, the desired outcome from the BPCS security lifecycle is $(P[A_B], P[A_{SB}]) = (0.033, 0.2813)$.

It should be noted that complete attack trees for the given BPCS and CPS architecture could span multiple pages. However, full attack trees may obscure the analysis and will serve no additional insight. Therefore, the simplified attack trees presented here act as a better illustration of the design methodology. For more in-depth treatment of the cyber risk assessment for the presented case study, refer to [10].

TABLE III
LOPA SHEET FOR THE CSTR OVERFLOW HAZARDOUS SCENARIO

Initiating Event	Likelihood λ_i (/yr)	Tank Dike	Safety Procedure	Human Intervention	BPCS ($P[B_p]$)	TMEL
Inlet flow surge	10^{-1}	10^{-2}	1	10^{-1}	10^{-1}	10^{-6}
Downstream flow blockage	10^{-1}	10^{-2}	10^{-1}	10^{-1}	10^{-1}	10^{-6}
Manual valves misalignment	10^{-1}	10^{-2}	10^{-1}	10^{-1}	10^{-1}	10^{-6}
BPCS physical Failure	$10^{-1}(\lambda_b)$	10^{-2}	1	10^{-1}	1	10^{-6}
BPCS attack Failure	$10^{-2}(\lambda_c)$	10^{-2}	1	10^{-1}	1	10^{-6}

Numbers in each cell represent the probability of failure of the associated protection layer.

TABLE IV
CSTR CLOPA—CALCULATED PARAMETER VALUES

LOPA Parameter	Value	Source
$\Sigma_{i=1}^{3}\lambda_i$	0.3	LOPA Sheet
$P[\mathcal{L}]$	0.001	LOPA Sheet
λ_b	0.01	LOPA Sheet
λ_c	0.01	LOPA Sheet
$P[B_p]$	0.01	LOPA Sheet
α_1	1.13×10^{-4}	CLOPA Parameter-Calculated Eq. (10)
α_2	1.17×10^{-4}	CLOPA Parameter-Calculated Eq. (11)
β	10^{-6}	CLOPA Parameter-Calculated Eq. (12)
γ_1	1.4868×10^{-4}	CLOPA Parameter-Calculated Eq. (14)
γ_2	7.5785×10^{-6}	CLOPA Parameter-Calculated Eq. (15)
γ_3	1.1686×10^{-4}	CLOPA Parameter-Calculated Eq. (16)

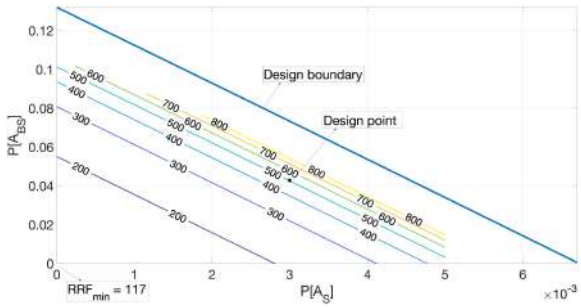


Fig. 13. CSTR case study: CLOPA design region with the contour plot for the RRF.

③ *SIS safety lifecycle—CLOPA:* Table III shows the LOPA sheet for the CSTR overflow hazard identified from the HAZOP, where the initiating event likelihoods and failure probabilities are adopted from [17] and [18]. The BPCS cyber attack likelihood is calculated as $\lambda_c = 10^{-4}\lambda = 0.01$ per year, assuming $\lambda = 100$ per year. Note that human intervention is considered a protection layer assuming that there is sufficient time for the operation team to manually isolate the reactor in the field. Some conservative approaches omit any human intervention or safety procedure from the LOPA.

From the LOPA sheet, we extract the event likelihood values to calculate the CLOPA model parameters using (10)–(12) and (14)–(16), along with $(P[A_B], P[A_{SB}]) = (0.033, 0.2813)$ from the BPCS security lifecycle. Table IV summarizes the parameter values. Substituting into the CLOPA constraint (13), we obtain

$$P[S_p] \leq \frac{1 - 148.68P[A_S] - 7.6P[A_{BS}](1 - P[A_S])}{117(1 - P[A_S]) - 7.6P[A_{BS}](1 - P[A_S])}. \quad (27)$$

Our objective now is to design an SIF with architecture $\mathcal{A}$ that satisfies (27) in order to achieve the required process safety objective as defined by the TMEL in the LOPA analysis. Our initial design for the SIF will comprise a level sensor (LT-02), a logic solver (SIS), and a shutdown valve (SDV-01), as illustrated in Fig. 6. The SIF will take an independent action upon reactor overflow and will close the inlet shutdown valve. The architecture of the SIF could vary through design iterations to achieve the required safety. As an example, sensors may be duplicated or sometimes triplicated to achieve higher reliability, and the SIS architecture may include redundant CPU modules. We note that for a perfectly secured SIS ($P[A_S] = P[A_{BS}] = 0$), $P[S_p] \leq 1/117$, or equivalently RRF ≥ 117. This is the minimum achievable RRF. Since for practical systems there is no zero probability of cyber security attack failures, our SIS design is expected to have an RRF > 117.

Using the calculated LOPA parameter values, the design region (18) and boundary (19) are defined by

$$P[A_{BS}] \leq 0.132\left(\frac{1 - 148.68P[A_S]}{1 - P[A_S]}\right) \quad (28)$$

where the design boundary is defined when the equality holds. The contour lines for the RRF in (20) are defined by

$$P[A_{BS}] = \left(\frac{15.42}{C-1}\right)\left(\frac{(C - 0.008) - (C - 1.27)P[A_S]}{1 - P[A_S]}\right) \quad (29)$$

for different values C of the RRF. The design region and the contour lines are plotted in Fig. 13. We note that as we approach the design boundary, either by increasing $P[A_S]$ or $P[A_{BS}]$, the RRF rapidly increases such that it is not possible to plot the contour lines in this region in a visible way. The design in this region is very sensitive to input variations (i.e., a very small variation in probabilities will result in a very large change in RRF). Therefore, the design point should be selected as far as possible from the design boundary. To further illustrate the increase in RRF, Fig. 14 shows a 3-D plot for the RRF as it varies with both $P[A_S]$ and $P[A_{BS}]$. It should be evident from the 3-D plot that for small values of $P[A_S]$, the function gradient is smaller, resulting in a less-sensitive design to probability variations. At larger values of $P[A_S]$ near the design boundary, the RRF increases exponentially with $P[A_{BS}]$. These results could be verified by calculating the gradient of (13).

To proceed with the design process, we pick the point $P[A_S] = 0.003$ as a reasonable probability value for SIS direct attack failure that is away from the steepest ascent region in Fig. 14. We now need to choose a practical value of $P[A_{BS}]$

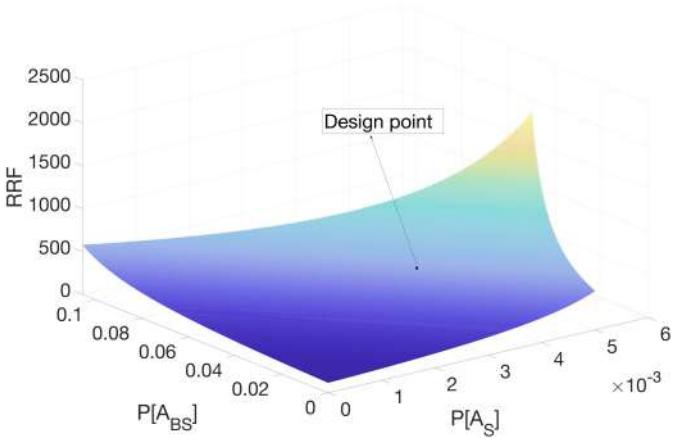


Fig. 14. CLOPA RRF as it varies with SIS security failure probabilities. Steepest ascent region to the right should be avoided when selecting the operating point.

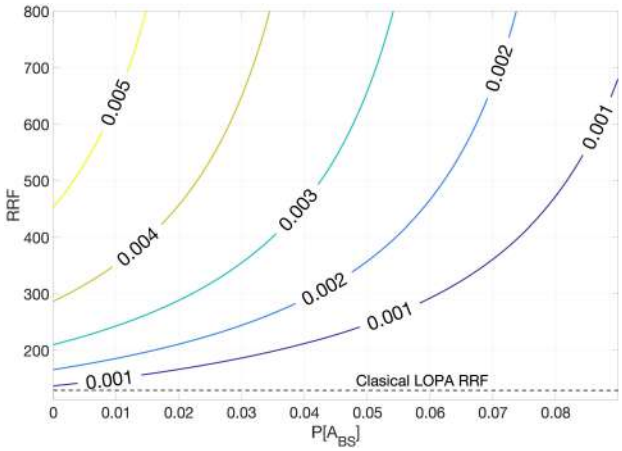


Fig. 15. Increase of the RRF with $P[A_{BS}]$. Each curve corresponds to the fixed value indicated for $P[A_S]$.

that results in an achievable target RRF. With the help of Fig. 13 and contour lines, $P[A_S] = 0.003$ intersects the contour line for RRF $= 500$ at $P[A_{BS}] = 0.0426$. Alternatively, the value of $P[A_{BS}]$ could be obtained from (29) by setting $C = 500$ and $P[A_S] = 0.003$. The design point (0.003, 0.0426, 500) is indicated in Figs. 13 and 14. The design and verification of the SIF then resume according to IEC 61511 to develop an architecture $\mathcal{A}$ that satisfies the combined CLOPA requirement: RRF ≥ 500, $P[A_S] \leq 0.003$, and $P[A_{BS}] \leq 0.0426$. The detailed design and verification of the SIF are outside the scope of this article (refer to [4] for more details). To complete the case study, we will assume that the design engineer came up with an architecture $\mathcal{A}$ that was verified using vendor data, resulting in reliability $\text{RRF}_v = 600$, with a design margin from the required RRF $= 500$.

④ *SIS cyber security lifecycle:* The resulting SIF architecture $\mathcal{A}$ is used to carry out the SIS security lifecycle, similar to the BPCS security risk assessment in step 2 of the design process. As the SIS detailed design and verification are not in the scope of this article, we will assume for the sake of illustration that the architecture $\mathcal{A}$ results in a cyber system configuration that has a higher probability of SIS security attack failure $P'[A_S] = 0.004$ while reducing the BPCS pivot attack failure probability to $P'[A_{BS}] = 0.03$ via securing the BPCS–SIS link.

⑤ *Safety lifecycle—CLOPA:* The architecture $\mathcal{A}$ results in $P'[A_S] = 0.004$, $P'[A_{BS}] = 0.03$, and $\text{RRF}_v = 600$. We need to verify if these values satisfy the CLOPA constraint (27). Plugging the probability values results in $P[S_p] \leq 1.54 \times 10^{-3}$, or equivalently RRF ≥ 643. As $\text{RRF}_v = 600 < 643$, the architecture has to be modified either by reducing further the cyber attack failure probabilities or by increasing the system reliability via fault tolerance techniques. It may take the design engineer multiple iterations until the design achieves the CLOPA constraint. In practice, the iterations do not involve a complete architectural redesign, but rather changing the redundancy scheme or security hardening in order to achieve the design objective. To conclude the case study example, we will assume that the design engineer came up with an architecture that preserves the aforementioned probability values while increasing the RRF to 650. This concludes the design process and the system moves to the implementation phase. The case study design values are superimposed on the iterative design process in Fig. 4 as an illustration. ■

C. Classical LOPA Error

Classical LOPA ignores cyber attack probabilities altogether. For the given problem, it results in RRF $= 113$ as per (17). The minimum CLOPA RRF occurs for a perfectly secured safety system where $P[A_S] = P[A_{BS}] = 0$, achieving RRF $= 117$. Therefore, the minimum error between LOPA and CLOPA RRF estimation is 4. The error gets worse as security failure probabilities increase. For the given design point $P[A_S], P[A_{BS}] = (0.003, 0.0426)$, the classical LOPA error is $e_{\text{RRF}} = 378$. This is a significant amount of error that results in the design of a less reliable system that will not achieve the target risk level. Fig. 15 better illustrates the error increase with increasing the security failure probability $P[A_{BS}]$ for different values of $P[A_S]$. For small values of $P[A_S]$, the curves show slow increase in the RRF with $P[A_{BS}]$. As $P[A_S]$ increases, the RRF increase becomes exponential. A similar contour figure for fixed $P[A_{BS}]$ values could be generated. The design point for the case study $P[A_S] = 0.003$ was chosen as a tradeoff between an achievable cyber attack probability value and a moderate rate of increase for the RRF. The 3-D plot for the error in RRF versus $P[A_S], P[A_{BS}]$ is identical to Fig. 14, except by shifting down the 3-D curve by 113, the LOPA RRF value; therefore, it is omitted to avoid repetition.

D. Sensitivity Analysis

Calculating the probability of a security failure is a debatable subject in the research community, especially with the lack of statistical data that are available for physical failures. One question that comes to mind is the robustness of the developed CLOPA model to probability variations. We conducted a numerical analysis to calculate the partial derivatives of the RRF with respect to $P[A_S], P[A_{BS}]$. The two partial derivative plots are very similar to Fig. 14 and omitted for space limitation. For

small probability values, the change in the RRF is in the range of 15% for 10^{-3} change in $P[A_S]$. As probabilities increase and we approach the decision boundary, the change in the RRF jumps to around 80% for 10^{-3} change and increases exponentially as we get closer to the decision boundary. A similar behavior is exhibited with $P[A_{BS}]$ change (figure omitted for brevity). However, the change in the RRF has much lower percentage, ranging from 7% for small probability values, and increasing to around 37% as we approach the decision boundary. We highlight the following three key observations.

1) For small cyber failure probability values, the model sensitivity is acceptable since the SIL levels have an order of magnitude ratio, so a small percentage change would likely keep the system requirement in the same SIL category. However, this requires that the probability error is in the range of 10^{-3}.
2) The model is more sensitive to direct attack failure probabilities than BPCS pivot attacks.
3) We should always try to design our system as far as possible from the decision boundary. The model sensitivity with respect to probability changes increases as we approach the decision boundary.

VI. Related Work

HAZOP has been the dominant risk assessment method for the process industry for over 30 years [5], [19], [20]. LOPA has been used in conjunction with HAZOP to design SISs and specify the SIL for each SIF [21]. Because of the wide adoption of LOPA by industry due to its systematic approach and quantitative risk assessment capability, LOPA has been included as one of the methods in IEC 61511-3 standard with several illustrating examples [4]. The LOPA approach has been applied to physical security risk analysis in [22]. However, to the best of authors' knowledge, there is no research work on integrating security attacks in the LOPA framework for SIS design.

There are emergent standardization initiatives to address safety and security coordination in CPSs. IEC 62443-4-1 (Security for Industrial Automation and Control Systems—Part 4-1: Secure Product Development Lifecycle Requirements) is a standard developed by the ISA-99 committee with the purpose to extend the existing safety lifecycle at different phases to include security aspects to ensure safe CPS design [23]. IEC TC65 AHG1 is a recently formed group linked to the same technical committee developing IEC 61508 and IEC 62443 to consider how to bridge functional safety and cyber security for industrial automation systems [24]. IEC 62859 (Nuclear Power Plants—Instrumentation and Control Systems—Requirements for Coordinating Safety and Cyber Security) is a standard derived from IEC 62645 for the nuclear power industry to coordinate the design and operation efforts with respect to safety and cyber security [25]. DO-326 (Airworthiness Security Process Specification) is a standard for the avionics industry that augments existing guidelines for aircraft certification to include the threat of intentional unauthorized electronic interaction to aircraft safety [26]. A taxonomy of dependable and secure computing is introduced in [27] in order to facilitate the communication among different research communities. The concepts and taxonomy presented are a result of a joint committee on "Fundamental Concepts and Terminology" that was formed by the Technical Committee on Fault-Tolerant Computing of the IEEE Computer Society and the IFIP WG 10.4 "Dependable Computing and Fault Tolerance." A preliminary work on the research in this article that combines the two research directions stated below is presented in [28].

A. Lifecycle Integration

Kornecki and Liu [29] use fault tree analysis to combine both safety and security failures in one unified risk assessment framework for the aviation industry. The outcome of the risk assessment is used to define both safety and security requirements. A road map for cyber safety engineering to increase air traffic management system resilience against cyber attacks is proposed in [30]. The V-shaped model to develop embedded software for CPS is augmented with security actions in [31]. The integration of IEC 61508 safety standard and IEC 15408 for IT security is described in [32]–[34] for building automation systems. Sørby [35] describes in more detail the integration of IEC 61508 safety lifecycle and the CORAS approach to identify security risks [36]. An approach to align safety and security during the different stages of system development lifecycle is proposed in [37]. The approach, called Lifecycle Attribute Alignment, ensures compatibility between safety and security controls developed and maintained during the system development lifecycle. HAZOP, a predominantly used method for safety risk assessment in the process industry, is modified in [38] to include security failures. The authors introduce new guide words, attributes, and modifiers for security components akin to traditional HAZOP limited to safety failures. Failure Mode and Effect Analysis is extended in [39] to include security vulnerabilities, suggesting the name Failure Mode Vulnerability and Effect Analysis. For a survey on the integration of safety and security in the CPS, refer to [2].

B. Model-Based Risk Assessment

Several graphical methods have been used to combine safety and security analysis. Goal structuring notation (GSN) is a graphical notation used to model requirements, goals, claims, and evidence of safety arguments [40]. The SafSec research project for the avionics industry elaborates on the use of GSN to integrate both safety and security arguments in one representation [41]. A similar approach is used in [42], where the authors apply the nonfunctional requirement (NFR) approach to quantitatively assess the safety and security properties of an oil pipeline CPS. NFR is a technique that allows simultaneous safety and security graphical representation and evaluation at the architectural level.

The simplicity and wide adoption of fault and attack trees promoted the research work to merge both modeling tools. The integration of fault trees and attack trees is considered in [43] in order to extend traditional risk analysis to include cyber attack risks. A quantitative analysis is proposed by assigning probabilities to tree events. Similarly, fault tree analysis is used in [29] to analyze safety/security risks in aviation software. Steiner and Liggesmeyer [44] extend component fault trees to

contain both safety and security events. Both the qualitative and quantitative analyses are performed to assess the overall risk. The quantitative analysis is enabled by assigning probabilities to safety events and categorical rating (low, medium, and high) for security events. Kumar and Stoelinga [45] translate the combined fault-attack tree into stochastic time automata to enable quantitative risk analysis. The use of bow-tie diagrams and analysis in place of fault trees is reported in [46], where it is integrated with attack trees for combined safety–security risk assessment.

Given the limited semantics of fault trees, Boolean-logic-driven Markov process (BDMP) graphical formalism introduced in [47] has been used to integrate safety and security events. The approach integrates fault trees with the Markov process at the leaf node level and associates a mean time to success for security events and a mean time to failure for safety events. This allows both a qualitative and a quantitative risk assessment for the given system. The formalism also enables the modeling of detection and response mechanisms without a need for model change. The work in [48] applies BDMP formalism to a pipeline case study, illustrating different types of safety–security interdependencies. In [49], Stuxnet attack is modeled using the BDMP and a quantitative risk analysis is carried out on the industrial control system.

Petri nets have also been proposed to overcome the limitations of fault trees. A formalism for safety analysis named state/event fault trees is reported in [50]. In this formalism, both deterministic state machines and Markov chains are combined, while keeping the visualization of causal chains known from fault trees. This formalism is extended in [51] to include an attacker model to deal with both safety and security. Similarly, stochastic Petri nets have been used in [52] to model the impact of intrusion detection and response on CPS reliability and in [53] to assess the vulnerabilities in supervisory control and data acquisition systems. Bayesian belief networks are also considered as one of the model-based approaches. In [54], a Bayesian belief network is used to assess the combined safety and security risk for an oil pipeline example.

The Unified Modeling Language (UML) commonly used in software engineering has also been used for safety and security risk assessment. Misuse cases for UML diagrams have been used to define safety requirements in [55] and security requirements in [56], independently. A combined process for Harm Assessment of Safety and Security has been proposed in [57] based on both UML and HAZOP studies. UMLsafe [58] and UMLsec [59] are two UML extensions that enable modeling of safety and security requirements, respectively. The combined UMLsafe/UMLsec is proposed in [60] for safety–security codevelopment. SysML-sec, a SysML-based model driven engineering environment, is used in [61] for the formal verification of safety and security properties.

System-theoretic process analysis (STPA) was developed as a new hazard analysis technique to evaluate the safety of a system [62]. Friedberg *et al.* [63] extend the STPA to include system security aspects in the analysis. The expanded approach is named STPA-SafeSec and demonstrated on a use case in the power grid domain. The system-theoretical accident model and process (STAMP) is applied to the Stuxnet attack in [64], showing that the attack could have been avoided if the STAMP was applied during design time.

VII. Conclusion

Classical safety assessment methods do not take into account failures due to cyber attacks. In this article, we showed quantitatively that overlooking security failures could bias the risk assessment, resulting in underdesigned protective systems. In addition, the design of safety and security subsystems for complex engineering systems cannot be carried out independently, given their strong coupling as demonstrated in this article. Although the design becomes more complicated when considering cyber attacks, the development of new software tools or the modification of existing industrial tools could automate the process.

In this article, we considered the control system (BPCS) design as given, following common industrial practice. The joint optimization of both BPCS and SIS designs, from both safety and security perspectives, is a potential extension for the presented work. In addition, the presented integrated lifecycle relies in part on designer's experience to make design decisions to achieve the system requirements. Optimal system design that captures possible safety and security design choices with associated financial cost could provide a better quantitative approach to find the optimal system operating point rather than relying on design heuristics. Furthermore, the integration of both the safety and security lifecycles into model-based design toolchains is crucial for adoption by industry.

Finally, the work presented in this article discusses the impact of cyber security failure on system safety. A closely related problem is how safety failures could impact cyber security. There is not much work in this direction, perhaps because the focus in CPSs is always on safety, considering the security of the cyber system as a secondary issue. Nevertheless, this is an important problem. On the one hand, a simple safety failure may be injected to cause a security compromise that may be exploited to produce a higher security compromise that could lead to a greater safety hazard. On the other hand, both directions, i.e., Safety $\rightarrow$ Security and Security $\rightarrow$ Safety, are closely related and interacting, and therefore, optimizing a CPS performance with respect to safety/security or both cannot be fully achieved without understanding the two types of interactions.

VIII. Source Code

The source code for the CLOPA in the form of MATLAB m files to regenerate the research results including the case study is located at https://github.com/Ashraf-Tantawy/CLOPA.git.

Appendix A

Proof of Lemma 3.1

Proof: The aggregate likelihood of all attacks to cause a hazard taking into account BPCS and SIS protection could be approximated by (neglecting higher order probability terms)

$$\Lambda = \lambda \sum_{a \in \mathcal{A}_r} \alpha_a P_a[S, B]. \tag{30}$$

TABLE A1
CLOPA MODEL PARAMETERS

Symbol	Description	Type	Calculation Method/ Data Source
λ_i	Initiating event i likelihood (/yr)	Parameter	Reliability data
λ_p	BPCS physical failure event likelihood (/yr)	Parameter	Reliability data
λ_a	BPCS cyber attack a likelihood (/yr)	Parameter	Refer to Section III-B
λ_c	BPCS semantically-related attacks likelihood (/yr)	Parameter	Refer to Section III-B, Lemma III.1
λ	BPCS cyber attack likelihood for all attacks	Parameter	Statistical attack data
α_a	Probability of selecting attack a by the attacker	Parameter	Attacker profile model
γ_a	Weight factor for the attacks for the equivalent BPCS	Parameter	Refer to Lemma III.1
$P[\mathcal{L}_i]$	Probability of failure of all protection layers for initiating event i	Parameter	Reliability data
TMEL	Target Mitigated Event Likelihood	Parameter	Determined by the corporate policy
$P[B_c]$	Probability of BPCS security failure	Intermediate design variable	BPCS security risk assessment
$P[B_p]$	Probability of BPCS physical failure	Parameter	Reliability data
$P[A_B]$	Probability of BPCS direct security failure	Parameter	BPCS security risk assessment
$P[A_{SB}]$	Probability of BPCS SIS-pivot security failure	Parameter	BPCS security risk assessment
$P[S_c]$	Probability of SIS security failure	Intermediate design variable	SIS security risk assessment
$P[S_p]$	Probability of SIS physical failure	Design variable	SIS security risk assessment
$P[A_S]$	Probability of SIS direct security failure	Design variable	SIS security risk assessment
$P[A_{BS}]$	Probability of SIS BPCS-pivot security failure	Design variable	SIS security risk assessment
$P[S_c, B_c]$	Probability of simultaneous SIS and BPCS security failure	Intermediate design variable	BPCS & SIS security risk assessment
$\alpha_1 - \alpha_2$	-	Auxiliary parameters	Eq. (10), (11)
$\gamma_1 - \gamma_3$	-	Auxiliary parameters	Eq. (14) to (16)
$\zeta_1 - \zeta_3$	-	Auxiliary parameters	Eq. (23) to (25)
β	-	Auxiliary parameters	Eq. (12)

Variables designated as "Design variable" are with respect to CLOPA, but could be a design variable of another assessment, such as $P[A_B]$, derived from BPCS security risk assessment. Variables designated as "intermediate design variables" could be expressed in terms of design variables.

Using (8) to expand the joint probability, we obtain

$$\Lambda = \lambda \sum_{a \in \mathcal{A}_r} \alpha_a \left(\eta_1 + \eta_2 P[B_c^a] + \eta_3 P[S_c] P[B_c^a | S_c]\right) \quad (31)$$

where $P[B_c^a]$ represents the probability of BPCS security failure with respect to attack a, and η_1, η_2, and η_3 are probability terms not dependent on the attack a. Expanding, we obtain

$$\Lambda = \lambda \left(\sum_{a \in \mathcal{A}_r} \alpha_a\right) \times \quad (32)$$

$$\left(\eta_1 + \eta_2 \sum_{a \in \mathcal{A}_r} \gamma_a P[B_c^a] + \eta_3 \sum_{a \in \mathcal{A}_r} \gamma_a P[S_c] P[B_c^a | S_c]\right) \quad (33)$$

where $\gamma_a = \alpha_a / \sum_{a \in \mathcal{A}_r} \alpha_a$. Ignoring higher order probabilities, we obtain

$$\Lambda \approx \lambda \left(\sum_{a \in \mathcal{A}_r} \alpha_a\right) \times \quad (34)$$

$$\left(\eta_1 + \eta_2 P\left[\sum_{a \in \mathcal{A}_r} \gamma_a B_c^a\right] + \eta_3 P\left[S_c, \sum_{a \in \mathcal{A}_r} \gamma_a B_c^a\right]\right). \quad (35)$$

Comparing (31) and (35), the second and third terms in (35) represent an equivalent BPCS with a combined attack vector $\mathcal{A}_r$, where each attack a is weighted by γ_a. In addition, the likelihood of this combined attack vector is $\lambda(\sum_{a \in \mathcal{A}_r} \alpha_a)$.

REFERENCES

[1] S. Kriaa, L. Pietre-Cambacedes, M. Bouissou, and Y. Halgand, "A survey of approaches combining safety and security for industrial control systems," *Rel. Eng. Syst. Saf.*, vol. 139, pp. 156–178, 2015.

[2] X. Lyu, Y. Ding, and S. H. Yang, "Safety and security risk assessment in cyber-physical systems," *IET Cyber- Phys. Syst.: Theory Appl.*, vol. 4, no. 3, pp. 221–232, 2019.

[3] P. P. Gruhn, *Safety Instrumented Systems: Design, Analysis, and Justification*. Research Triangle Park, NC, USA: Instrum. Soc. Amer., 2006.

[4] *Functional Safety—Safety Instrumented Systems for the Process Industry Sector—Part 1: Framework, Definitions, System, Hardware and Application Programming Requirements*, IEC Standard IEC 61511-1:2016, 2016.

[5] J. Dunjó, V. Fthenakis, J. A. Vílchez, and J. Arnaldos, "Hazard and operability (HAZOP) analysis. A literature review," *J. Hazardous Mater.*, vol. 173, no. 1–3, pp. 19–32, 2010.

[6] A. Swales, *Open Modbus/TCP Specification*, vol. 29. Rueil-Malmaison, France: Schneider Electric, 1999.

[7] I. N. Fovino, A. Carcano, T. De Lacheze Murel, A. Trombetta, and M. Masera, "Modbus/DNP3 state-based intrusion detection system," in *Proc. 24th IEEE Int. Conf. Adv. Inf. Netw. Appl.*, 2010, pp. 729–736.

[8] A. Z. Faza, S. Sedigh, and B. M. McMillin, "Reliability analysis for the advanced electric power grid: From cyber control and communication to physical manifestations of failure," in *Proc. Int. Conf. Comput. Saf., Rel., Secur.*, 2009, pp. 257–269.

[9] G. Stoneburner, A. Goguen, and A. Feringa, "Risk management guide for information technology systems," NIST Special Publication 800-30, 2002.

[10] A. Tantawy, S. Abdelwahed, A. Erradi, and K. Shaban, "Model-based risk assessment for cyber physical systems security," *Comput. Secur.*, vol. 962020, Art. no. 101864.

[11] A. P. Moore, R. J. Ellison, and R. C. Linger, "Attack modeling for information security and survivability," Softw. Eng. Inst., Pittsburgh, PA, USA, Tech. Rep. CMU/SEI-2001-TN-001, 2001.

[12] A. Tantawy, S. Abdelwahed, and Q. Chen, "Continuous stirred tank reactors: Modeling and simulation for CPS security assessment," in *Proc. 11th Int. Conf. Comput. Intell. Commun. Netw.*, 2019, pp. 117–123.

[13] K. Stouffer, J. Falco, and K. Scarfone, "Guide to industrial control systems (ICS) security," NIST Special Publication 800-82, 2011, p. 164.

[14] A. Tantawy, "Automated malware design for cyber physical systems," in *Proc. 9th Int. Symp. Digit. Forensics Secur.*, 2021, pp. 1–6.

[15] S. Mauw and M. Oostdijk, "Foundations of Attack Trees (ser. Lecture Notes in Computer Science), vol. 3935. Berlin, Germany: Springer, 2006, pp. 186–198.

[16] *NVD—Home*, NIST, Gaithersburg, MD, USA, 2016.

[17] INTEF and NTNU, *OREDA Offshore and Onshore Reliability Data Volume 1—Topside Equipment*. Bærum, Norway: DNV, 2015.

[18] *Guidelines for Initiating Events and Independent Protection Layers in Layer of Protection Analysis*. Center Chem. Process Saf., New York, NY, USA, 2015.

[19] F. Crawley and B. Tyler, *HAZOP: Guide to Best Practice*. Amsterdam, The Netherlands: Elsevier, 2015.

[20] B. Skelton, *Hazop and Hazan: Identifying and Assessing Process Industry Hazards*. Rugby, U.K.: IChemE, 1999.

[21] A. M. Dowell, "Layer of protection analysis and inherently safer processes," *Process Saf. Prog.*, vol. 18, no. 4, pp. 214–220, 1999.

[22] F. Garzia, M. Lombardi, M. Fargnoli, and S. Ramalingam, "PSA-LOPA—A novel method for physical security risk analysis based on layers of protection analysis," in *Proc. Int. Carnahan Conf. Secur. Technol.* 2018, pp. 1–5.

[23] *Security for Industrial Automation and Control Systems – Part 4-1: Secure Product Development Lifecycle Requirements*, IEC Standard IEC 62443-4-1, 2018.

[24] H. Kanamaru, "Bridging functional safety and cyber security of SIS/SCS," in *Proc. 56th Annu. Conf. Soc. Instrum. Control Eng. Jpn.*, 2017, pp. 279–284.

[25] *Nuclear Power Plants—Instrumentation and Control Systems—Requirements for Coordinating Safety and Cybersecurity*, IEC Standard IEC 62859:2016, 2016.

[26] C. Torens, "Safety versus security in aviation, comparing DO-178C with security standards," in *Proc. AIAA Scitech 2020 Forum*, 2020, Art. no. AIAA 2020-0242.

[27] A. Avižienis, J. C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Dependable Secure Comput.*, vol. 1, no. 1, pp. 11–33, Jan.–Mar. 2004.

[28] A. Tantawy, A. Erradi, and S. Abdelwahed, "A modified layer of protection analysis for cyber-physical systems security," in *Proc. 4th Int. Conf. Syst. Rel. Saf.*, Rome, Italy, 2019, pp. 94–101.

[29] A. J. Kornecki and M. Liu, "Fault tree analysis for safety/security verification in aviation software," *Electronics*, vol. 2, pp. 41–56, 2013.

[30] C. W. Johnson, "CyberSafety: On the interactions between cybersecurity and the software engineering of safety-critical systems," *Lab. Med.*, vol. 21, no. 7, pp. 411–413, 2012.

[31] A. J. Kornecki and J. Zalewski, "Safety and security in industrial control," in *Proc. Annu. Workshop Cyber Secur. Inf. Intell. Res.*, 2010, Art. no. 77.

[32] T. Novak, A. Treytl, and P. Palensky, "Common approach to functional safety and system security in building automation and control systems," in *Proc. IEEE Int. Conf. Emerg. Technol. Factory Autom.*, 2007, pp. 1141–1148.

[33] T. Novak and A. Treytl, "Functional safety and system security in automation systems—A life cycle model," in *Proc. IEEE Int. Conf. Emerg. Technol. Factory Autom.*, 2008, pp. 311–318.

[34] T. Novak and A. Gerstinger, "Safety- and security-critical services in building automation and control systems," *IEEE Trans. Ind. Electron.*, vol. 57, no. 11, pp. 3614–3621, Nov. 2010.

[35] K. Sørby, "Relationship between security and safety in a security-safety critical system: Safety consequences of security threats," Ph.D. dissertation, Dept. Mech. Ind. Eng., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2003.

[36] K. Stølen *et al.*, "Model-based risk assessment in a component-based software engineering process," in *Business Component-Based Software Engineering*. Boston, MA, USA: Springer, 2003, pp. 189–207.

[37] B. Hunter, "Integrating safety and security into the system lifecycle," in *Proc. Improving Syst. Softw. Eng. Conf.*, 2009, p. 147.

[38] R. Winther, O. A. Johnsen, and B. A. Gran, *Security Assessments of Safety Critical Systems Using HAZOPs* (ser. Lecture Notes in Computer Science), vol. 2187. Berlin, Germany: Springer, 2001, pp. 14–24.

[39] C. Schmittner, T. Gruber, P. Puschner, and E. Schoitsch, *Security application of Failure Mode and Effect Analysis (FMEA)* (ser. Lecture Notes in Computer Science), vol. 8666. Cham, Switzerland: Springer, 2014, pp. 310–325.

[40] *GSN Community Standard Version 1*, Origin Consulting, York, U.K., 2011.

[41] S. Lautieri, D. Cooper, and D. Jackson, "SafSec: Commonalities Between Safety and Security Assurance," in *Constituents of Modern System-Safety Thinking*. London, U.K.: Springer, 2007, pp. 65–75.

[42] N. Subramanian and J. Zalewski, "Quantitative assessment of safety and security of system architectures for cyberphysical systems using the NFR approach," *IEEE Syst. J.*, vol. 10, no. 2, pp. 397–409, Jun. 2016.

[43] I. Nai Fovino, M. Masera, and A. De Cian, "Integrating cyber attacks within fault trees," *Rel. Eng. Syst. Saf.*, vol. 94, pp. 1394–1402, 2009.

[44] M. Steiner and P. Liggesmeyer, "Combination of safety and security analysis—Finding security problems that threaten the safety of a system," in *Proc. 32nd Int. Conf. Comput. Saf., Rel. Secur.*, 2013, pp. 1–8.

[45] R. Kumar and M. Stoelinga, "Quantitative security and safety analysis with attack-fault trees," in *Proc. IEEE Int. Symp. High Assurance Syst. Eng.*, 2017, pp. 25–32.

[46] H. Abdo, M. Kaouk, J. M. Flaus, and F. Masse, "A safety/security risk analysis approach of industrial control systems: A cyber bowtie—Combining new version of attack tree with bowtie analysis," *Comput. Secur.*, vol. 72, pp. 175–195, 2018.

[47] M. Bouissou and J. L. Bon, "A new formalism that combines advantages of fault-trees and Markov models: Boolean logic driven Markov processes," *Rel. Eng. Syst. Saf.*, vol. 82, pp. 149–163, 2003.

[48] S. Kriaa, M. Bouissou, F. Colin, Y. Halgand, and L. Pietre-Cambacedes, *Safety and Security Interactions Modeling Using the BDMP formalism: Case Study of a Pipeline* (ser. Lecture Notes in Computer Science), vol. 8666. Cham, Switzerland: Springer, 2014, pp. 326–341.

[49] S. Kriaa, M. Bouissou, and L. Pièrre-Cambacédès, "Modeling the Stuxnet attack with BDMP: Towards more formal risk assessments," in *Proc. 7th Int. Conf. Risks Secur. Internet Syst.*, 2012, pp. 1–8.

[50] B. Kaiser, C. Gramlich, and M. Förster, "State/event fault trees-A safety analysis model for software-controlled systems," *Rel. Eng. Syst. Saf.*, vol. 92, pp. 1521–1537, 2007.

[51] M. Roth and P. Liggesmeyer, "Modeling and analysis of safety-critical cyber physical systems using state/event fault trees," in *Proc. 32nd Int. Conf. Comput. Saf., Rel. Secur.*, 2013.

[52] R. Mitchell and I. R. Chen, "Effect of intrusion detection and response on reliability of cyber physical systems," *IEEE Trans. Rel.*, vol. 62, no. 1, pp. 199–210, Mar. 2013.

[53] C. W. Ten, C. C. Liu, and G. Manimaran, "Vulnerability assessment of cybersecurity for SCADA systems," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1836–1846, Nov. 2008.

[54] A. J. Kornecki, N. Subramanian, and J. Zalewski, "Studying interrelationships of safety and security for software assurance in cyber-physical systems: Approach based on Bayesian belief networks," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, 2013, pp. 1393–1399.

[55] G. Sindre, "A look at misuse cases for safety concerns," in *Proc. Working Conf. Method Eng.*, 2007, pp. 252–266.

[56] G. Sindre and A. L. Opdahl, "Eliciting security requirements with misuse cases," *Requirements Eng.*, vol. 10, pp. 34–44, 2005.

[57] C. Raspotnig, P. Karpati, and V. Katta, *A Combined Process for Elicitation and Analysis of Safety and Security Requirements* (ser. Lecture Notes in Business Information Processing), vol. 113. Berlin, Germany: Springer, 2012, pp. 347–361.

[58] J. Jürjens, *Developing Saf.-Crit. Syst. With UML* (ser. Lecture Notes in Computer Science), vol. 2863. Berlin, Germany: Springer, 2003, pp. 360–372.

[59] J. Jürjens, *UMLsec: Extending UML for Secure Systems Development* (ser. Lecture Notes in Computer Science), vol. 2460. Berlin, Germany: Springer, 2002, pp. 412–425.

[60] J. Jürjens, "Developing safety-and security-critical systems with UML," in *Proc. DARP Workshop, Loughborough, U.K.*, 2003.

[61] G. Pedroza, L. Apvrille, and D. Knorreck, "AVATAR: A SysML environment for the formal verification of safety and security properties," in *Proc. 11th Annu. Int. Conf. New Technol. Distrib. Syst.*, 2011, pp. 1–10.

[62] J. Thomas, "Extending and automating STPA for requirements generation and analysis," Ph.D. dissertation, Eng. Syst. Division, Massachusetts Inst. Technol., Cambridge, MA, USA, 2013.

[63] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, and S. Sezer, "STPA-SafeSec: Safety and security analysis for cyber-physical systems," *J. Inf. Secur. Appl.*, vol. 34, pp. 183–196, 2017.

[64] A. Nourian and S. Madnick, "A systems theoretic approach to the security threats in cyber physical systems applied to Stuxnet," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 1, pp. 2–13, Jan./Feb. 2018.

Detecting Cyber-Attacks Against Cyber–Physical Manufacturing System: A Machining Process Invariant Approach

Zedong Li, Xin Chen, *Member, IEEE*, Yuqi Chen, Shijie Li, Hangyu Wang, Shichao Lv, and Limin Sun

***Abstract*—The era of the Industrial Internet of Things has led to an escalating menace of cyber–physical manufacturing systems (CPMSs) to cyber-attacks. Presently, the field of intrusion detection for CPMS has significant advancements. However, current methodologies require significant costs for collecting historical data to train detection models, which are tailored to specific machining scenarios. Evolving machining scenarios in the real world challenge the adaptability of these methods. In this article, We found that the machining code of the CPMS contains a complete machining process, which is an excellent detection basis. Therefore, we propose MPI-CNC, an intrusion detection approach based on Machining Process Invariant in the machining code. Specifically, MPI-CNC automates the analysis of the machining codes to extract machining process rules and key parameter rules, which serve as essential detection rules. Then, MPI-CNC actively acquires runtime status from the CPMS and matches the detection rules to identify cyber-attacks behavior. MPI-CNC was evaluated using two FANUC computer numerical control (CNC) machine tools across ten real machining scenarios. The experiment demonstrated the exceptional adaptability capability of MPI-CNC. Furthermore, MPI-CNC showed superior accuracy in detecting cyber-attacks against CPMS compared to existing state-of-the-art detection methods while ensuring normal machining operations.**

***Index Terms*—Computer numerical control (CNC), cyber attack, cyber–physical manufacturing systems (CPMSs), Industrial Internet of Things, intrusion detection.**

I. Introduction

MANUFACTURING industry is an important cornerstone of modern industrial development. With the advent of the Industrial Internet of Things and intelligent manufacturing, the global manufacturing industry is rapidly moving toward networked and intelligent development [1]. Computer numerical control (CNC) system is the core of cyber–physical manufacturing systems (CPMSs) that control the machining process of manufacturing equipment. CNC systems are widely used in important industries, such as the aviation industry, automobile manufacturing, and military industry. Tesla's Giga factory connects CNC systems to the industrial Internet to automatically control production processes, greatly improving production efficiency.

As an increasing number of factories integrate their CNC systems into the Industrial Internet, the security of CPMS has become a paramount requirement and faces formidable challenges. Intrusion detection approaches for CPMS have emerged as a prominent and burgeoning topic. Currently in the field of CPMS security, most researchers focus on training machine learning classification models to detect anomalies by analyzing side channel data, such as current [2], video [3], or audio [4], [5], [6], generated during machining. Some researchers have built digital twin models for CNC systems with a data-driven approach to do consistency checks on the runtime state of CNC systems to detect cyber-attacks [7]. There is also an offline approach to detect whether machining codes have been tampered with, which extracts digital features of machining codes and trains machine learning anomaly classification models [8]. These solutions can effectively detect anomalous processing behaviors for specific machining scenarios.

Regrettably, the absence of adequate security considerations for CNC system manufacturers has resulted in attackers being able to easily launch cyber-attacks by exploiting CNC system vulnerabilities, such as the lack of authentication mechanisms, plain-text transmission, and the existence of unfixed vulnerabilities in the system. Primarily, attackers target the machining code to introduce defects in the product processing. For instance, they may implant a Trojan into the firmware of the CNC system, surreptitiously tamper with the machining code passed into the system, and execute a malicious hole attack [9]. Additionally, attackers have demonstrated the use of steganography to tamper with machining code files in network traffic, diminishing the mechanical strength of the resulting product [10]. In a further form of attack, assailants manipulate key parameters in the memory of the CNC system. For instance, they substitute the processing material by tampering with parameters, integrating smart materials into gas masks to plant physical logic bombs. This causes gas masks to crack and leak during use [11]. Furthermore, attackers have utilized existing open-source tools like C3PO [12] and Industrial

Manuscript received 23 October 2023; revised 11 January 2024; accepted 22 January 2024. Date of publication 25 January 2024; date of current version 9 May 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 110400ZG21. *(Corresponding author: Limin Sun.)*

Zedong Li, Xin Chen, Shijie Li, Hangyu Wang, Shichao Lv, and Limin Sun are with the Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China, and also with the School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: lizedong@iie.ac.cn; chenxin1990@iie.ac.cn; lishijie@iie.ac.cn; wanghangyu@iie.ac.cn; lvshichao@iie.ac.cn; sunlimin@iie.ac.cn).

Yuqi Chen is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: chenyq@shanghaitech.edu.cn).

Digital Object Identifier 10.1109/JIOT.2024.3358798

Security Exploitation Framework (ISF) [13] to send malicious instructions to CNC systems disrupting the processing processes. These instances underscore the pressing need for intrusion detection systems for CNC systems to mitigate such threats effectively.

Motivation: In practical production processes, the CNC system employs various machining codes to handle different products, leading to diverse machining scenarios. These different scenarios require distinct tool paths, raw materials, and machining tools, resulting in different side-channel features with audio, image, current, and voltage. To develop intrusion detection models using side-channel data across different machining scenarios, researchers typically need to gather side-channel data for each new scenario and repeat the training process, incurring significant time and labor costs. However, once attackers successfully deploy an attack script in a CPMS system, they can easily disrupt the different machining scenarios. Consequently, there is an urgent need for an adaptable intrusion detection approach within the CPMS system that can be readily deployed across a variety of machining scenarios to effectively counter existing attack methods.

Insight: Invariant rule-based detection is currently a popular method in the field of industrial control security to effectively detect anomalies due to cyber-attacks. Usually, industrial control devices execute control logic codes to control the normal operation of industrial systems based on the invariant control logic in the control logic codes. Researchers have utilized data-driven [14] or code-driven [15] approaches to extract control logical invariant rules in industrial control systems as the basis for intrusion detection in industrial control systems. They have achieved excellent detection results. Inspired by the invariant rule-based detection, we found that in the field of CPMS, the machining process of the CNC system is invariant, and the machining code contains comprehensive machining process invariant information. Therefore, the complete machining process invariant rules can be extracted by analyzing the machining code. The machining process invariant rules include key elements, such as machining trajectory and machining speed, which can comprehensively describe the machining process of the CNC system and is a reliable basis for intrusion detection.

Method: This article addresses the issue of the limited adaptive ability of the CPMS intrusion detection system by proposing MPI-CNC, an intrusion detection method based on Machining Process Invariant. MPI-CNC automatically and rapidly extracts detection rules from the machining code. The method first parses the machining code to extract tool paths, machining sequences, spindle speeds, and other key machining-related parameters as rules for detecting attacks. MPI-CNC then actively collects runtime machining status, and key parameters from the CNC system during machining. Finally, MPI-CNC verifies the consistency of the runtime machining data based on the detection rules to identify cyber-attacks.

Result: To verify the feasibility of the approach in this article, a prototype was developed based on the FANUC CNC system. We conducted experiments using real CNC machines, analyzed 10 real machining scenarios and 3 attack methods, and evaluated the deployment time cost, detection performance, and interference to the CNC system. Experiments demonstrated that MPI-CNC can be quickly applied to new machining scenarios without preprocessing and detect cyber-attacks accurately in runtime without affecting the normal operation of the CNC. MPI-CNC has better detection performance compared to the other state-of-the-art detection methods. The detection accuracy of machining code injection attack and parameter injection attack reaches 98.81% and 100%, respectively, while the best detection results of other methods are 98.38% [8] and 93.25% [7].

This article contributes as follows.

1) We propose a novel approach for the automatic extraction of detection rules by analyzing machining codes. It can rapidly generate detection rules for different machining scenarios, thereby improving the adaptability capability of CPMS intrusion detection.
2) We conducted a reverse analysis of the FOCAS protocol used in FANUC CNC systems and developed low-interference acquisition request packets that conform to the protocol format. This approach improves the efficiency of data acquisition while reducing interference to the machining process.
3) A prototype CPMS IDS was developed based on the FANUC CNC system. Although this prototype was developed for a specific CNC system, based on this idea, it can be modified to expand and adapt to other CNC systems and protocols.
4) We evaluated the adaptability capability and detection performance of our proposed approach in 10 real machining scenarios and 3 attack scenarios. Experiments show that this approach can be quickly applied to new machining scenarios without preprocessing and detect cyber-attacks accurately in runtime without affecting the normal operation of the CNC.

Roadmap: The remainder of this article is structured as follows. Section II briefly introduces the technical background related to CPMS, and Section III provides an overview of the MPI-CNC. Section IV details the MPI-CNC and the specific implementation. The experimental evaluation is detailed in Section V. Section VI introduces the related work of current CPMS intrusion detection methods. Section VII discusses the limitations of the MPI-CNC. Section VIII is the conclusion.

II. Background

This article is primarily dedicated to proposing an intrusion detection method for CPMS. This chapter serves to provide an overview of the research background, focusing on two essential aspects: 1) the composition and 2) machining process of CPMS, as well as the various forms of attacks encountered by CPMS.

A. Cyber–Physical Manufacturing Systems

The CPMS generally consists of an engineering station, a distributed numeric control (DNC) server, a machine data collection (MDC) server, and manufacturing equipment connected through an industrial switch [see Fig. 1(a)]. The

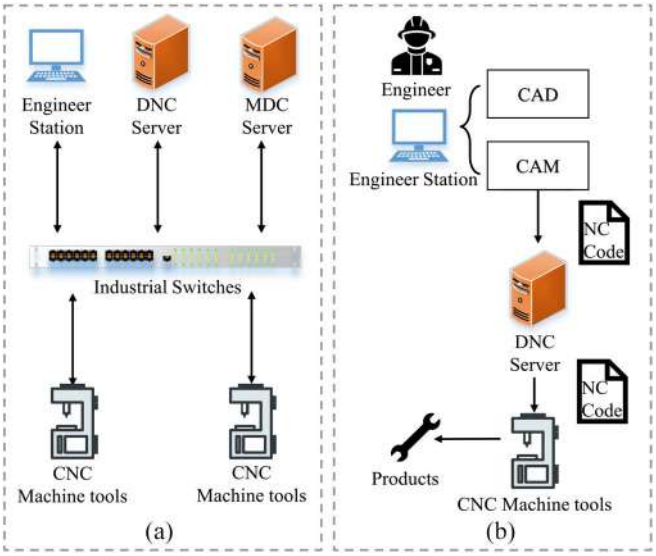


Fig. 1. Network topology and processing process of a CPMS. (a) Network topology of CPMS. (b) Processing process of CPMS.

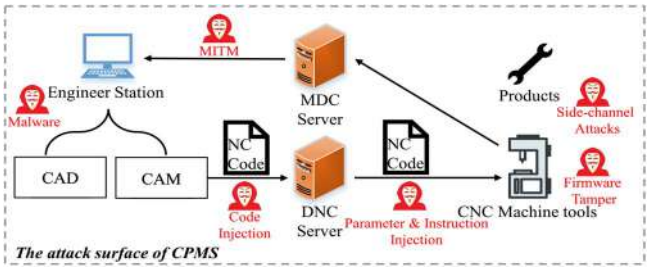


Fig. 2. Attack surface of CPMS.

engineering station is usually an office computer equipped with computer aided design (CAD), and computer aided machining (CAM) programs. The Processing process is shown in Fig. 1(b). Engineers utilize engineering design software to generate machining code (alternative name NC code), which is then uploaded to the DNC server. The DNC server distributes the NC code to the appropriate manufacturing equipment. The CNC system automatically controls the machining process by parsing the NC code. The MDC server interacts with the manufacturing equipment to collect various states of the equipment, including position, speed, temperature, and other information. This data is returned to the monitoring program of the engineering station, allowing engineers to monitor the machining process. Additionally, engineers can send control commands to perform runtime operations during the machining process.

B. Attack Model

In recent times, scholars have analyzed and categorized cyber-attacks targeting CPMS [16], [17]. This article investigates recent cyber-attacks against CPMS, analyzing the attack surface from the perspective of the production process (see Fig. 2). In the production process, the engineer station is connected to a local area network or even the Internet. Attackers can maliciously target the engineer station using spear-phishing [18], BadUSB [19], and other vectors carrying malicious code to exploit system and software vulnerabilities, stealing and tampering with CAD models and NC codes. Devices, such as engineer stations and DNC servers, typically communicate with the CNC system via Ethernet, using communication protocols that often lack authentication, encryption, and other security mechanisms. For example, DNC servers may use the FTP protocol to transmit NC code in clear text. Attackers can perform man-in-the-middle attacks [20], tampering with and stealing NC code from network traffic, and replaying network packets to inject malicious commands and parameters. The attacker can also tamper with the CNC system firmware [9], [21] to interfere with normal processing. However, such attacks are more difficult and require a deep understanding of the underlying code structure of the CNC system. As CPMS is a typical cyber–physical system [22], attackers can use side-channel attack methods, such as electrical measurement interference and acoustic resonance, to interfere with normal processing [23], or infer the production state of the machine tool and workpiece geometry information from leaked physical information [24], achieving a steganography attack.

The attacks were classified into three categories: 1) machining code injection; 2) parameter injection; and 3) instruction injection.

Machining Code Injection: Machining Code injection attack [9], [10], [25] refers to tampering with or replacing the machining code, the NC code, of the CNC system. By modifying key code segments, such as the machining path, spindle speed, or auxiliary control code, attackers can interfere with and disrupt the CNC machining process.

Parameter Injection: Parameter Injection [11], [26] refers to tampering with the parameters of the CNC system. There are many important parameters in the CNC system that affect the machining process, such as the spindle speed ratio value, the rapid feed rate value, and the alarm shielding. Therefore, if attackers can tamper with these key parameters, it will cause serious damage to the CNC system, affecting machining accuracy and potentially damaging the CNC machine.

Instruction Injection: Instruction Injection refers to sending malicious control commands to the CNC system, which disrupts the normal machining process. McCormack et al. [12] introduced an open-source tool called C3PO, which analyzes potential vulnerabilities in network services of 3-D printers and uses network vulnerabilities to send malicious commands to attack remote-controlled CNC systems. Attackers can also use the ISF [13] to inject malicious commands by sending attack scripts to disrupt the production process.

Moreover, the CNC systems also face security threats, such as physical cross-domain attacks and side-channel information leakage [27]. Nevertheless, it is pertinent to note that these threats lie outside the purview of this article. Their inclusion is excluded due to factors, such as their diminished feasibility, limited potential for harm, or their susceptibility, to detection by existing IDSs.

III. Overview

In this research, we propose an innovative intrusion detection method for CPMS, denoted as MPI-CNC. As illustrated in

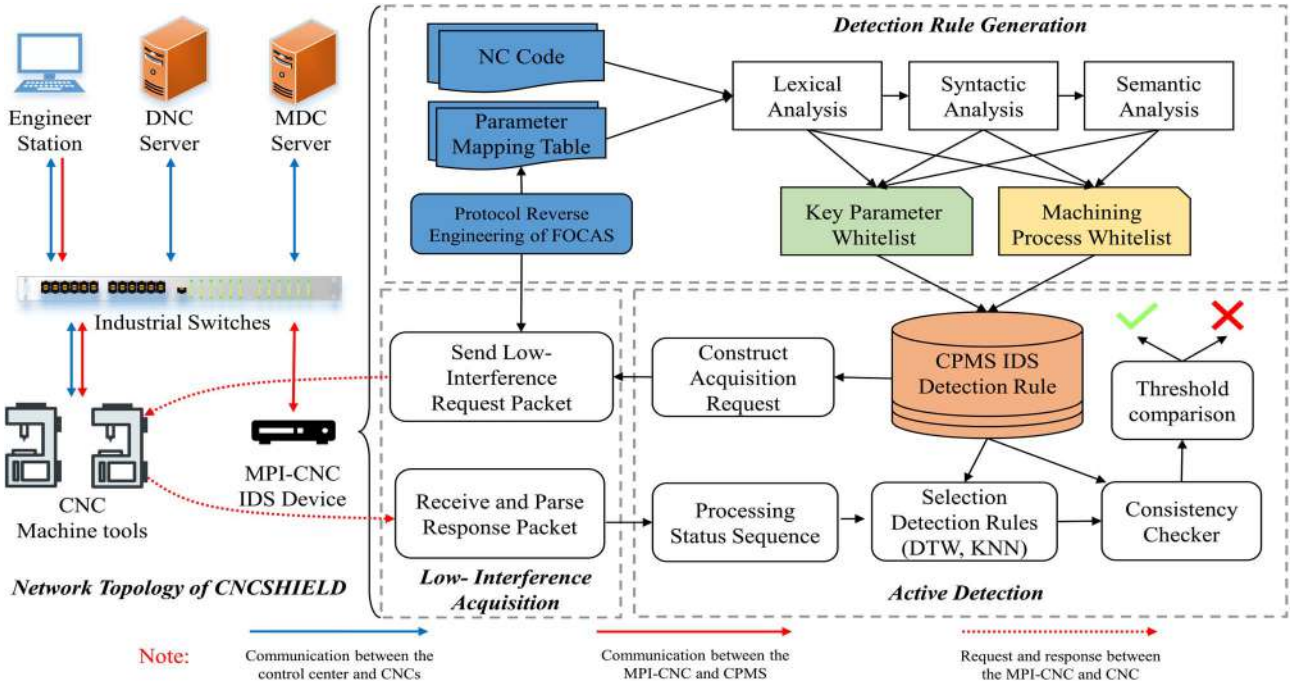


Fig. 3. Systematic approach to build MPI-CNC.

Fig. 3, this method comprises three distinct stages: 1) detection rule generation; 2) low-interference state acquisition; and 3) active detection. In this section, we present a concise overview of the fundamental framework for analyzing intrusion detection methods.

Detection Rule Generation: The detection rule generation module employs static analysis to parse the NC code and extract detection rules. The NC code encapsulates complete machining processes, including key parameter rules and machining process rules. The key parameter rules are utilized to monitor and verify vital parameters within the CNC system, ensuring their accuracy, stability, and safeguarding against malevolent tampering that could lead to diminished machining precision or machine malfunctions. On the other hand, the machining process rules establish reference guidelines by analyzing the invariant characteristics of machining processes in the NC code, enabling the detection of malicious attacks, such as tampering with machining trajectories or program substitution. The extracted detection rules from the NC code furnish a comprehensive depiction of the machining process and enable proactive detection of unexpected anomalies and network attacks.

Low-Interference Acquisition: The low-interference state acquisition module is responsible for the runtime collection of machining states within the CNC system. Typically, CNC system manufacturers provide monitoring software or development kits for monitoring the system's operational status. For instance, the FANUC Focas 1/2 development component facilitates secondary development. It enables runtime remote monitoring through active communication with the CNC system. The development component provides essential information, such as NC programs, tool positions, and spindle speeds. However, direct use of the original development kit for high-frequency data collection can increase the network load of the CNC system, negatively impacting normal machining operations and real-time performance. To circumvent this issue, our study employs reverse engineering to analyze proprietary protocols. We also customize data collection requests and eliminate redundant ones. As a result, we achieve low-interference high-frequency acquisition of runtime machining states within the CNC system.

Active Detection: The active detection module primarily identifies anomalies in the machining process. It processes the runtime machining state data collected by the low-interference state acquisition module and generates alerts. It verifies whether the machining state of the CNC system adheres to the key parameter rules and the machining process rules. This approach prevents code tampering, manipulation of key parameters, and malicious instruction attacks. The active detection stage necessitates determining two critical monitoring parameters: 1) error threshold and 2) monitoring window size. Initially, we experiment with multiple monitoring window sizes based on the state acquisition frequency to determine the optimal size. Subsequently, under specific window sizes, we calculate the cumulative normal error for each monitoring window and set the error threshold using the maximum observed error.

IV. Approach

A. Problem Statement

The CNC manufacturing process is a complex industrial control process in which the CNC system performs closed-loop control of the relative motion of the tool and the workpiece based on multiple sensor data. In this article, we use $u(t)$ in (1) to describe the machining state of the CNC at time

t, where $P(x, y, z)$ indicates the coordinates of the tool in the xyz three axes, $S(t)$ indicates the spindle speed, $F(t)$ indicates the feed rate, and $T(t)$ indicates the current tool number

$$u(t) = (P(x, y, z), S(t), F(t), T(t)). \tag{1}$$

We define (2) with $r(n)$ to describe the machining process indicated by the machining code, which represents the invariant characteristics of the CNC machining process. Specifically, we use $F(x, y, z)$ to represent the curve equation of the machining path, which is commonly straight lines and circular arcs. Start(x, y, z) and End(x, y, z) represent the start and end points of the machining path. The combination of machining path, spindle speed, feed rate, and tool number provides a complete description of the machining process

$$r(n) = (F(x, y, z), \text{Start}(x, y, z), \text{End}(x, y, z), S(n), F(n), T(n)). \tag{2}$$

We use $\mathcal{M}()$ in (3) to describe the CNC machining model, where $\varepsilon(t)$ represents the internal losses of the CNC and reasonable errors due to natural factors

$$u(t + 1) = \mathcal{M}(u(t), r(t), \varepsilon(t)). \tag{3}$$

In the normal machining process, the CNC machining state has reasonable errors $\varepsilon(t)$ due to machine wear and tear, current and voltage jitter, and other factors. However, when the CNC is under a cyber-attack, it can deviate significantly from the CNC machining model $\mathcal{M}()$ and violate the current machining process $r(n)$. Therefore, in this article, we designed an intrusion detection method based on the invariant characteristics of the CNC machining process. Our approach is divided into a detection rule generation module, a low-interference acquisition module, and an active detection engine [see Fig. 3].

B. Intrusion Detection Rule

The machining process is a crucial basis for manufacturing and processing workpieces. The NC code contains the most complete and comprehensive machining process information, such as spindle speed, feed rate, and machining path. The CNC system interprets the NC code into executable instructions to control the various components of the CNC machine tool to complete the machining operations. Through the analysis of the NC code, the following intrusion detection rules can be generated: key parameter whitelist rules and machining process whitelist rules. The approach of generating detection rules based on code analysis demonstrates great applicability in the industrial control field [15], [28].

1) Key Parameter Whitelist Rule: In CNC systems, there are numerous important parameters that can affect the actual production process. The process of sending commands from the CNC system to control the hardware needs to be adjusted to the specific parameters. For instance, the CNC system adjusts the tool's landing position and movement trajectory during the actual machining process based on parameters, such as tool radius compensation and length compensation, or the CNC system controls the feed acceleration based on parameters related to acceleration and deceleration. These parameters directly impact the machining accuracy and stability of the machine tool. If the key parameters in the CNC system are maliciously tampered with by attackers, it can result in decreased machining accuracy or even machine tool failure. Typically, key parameters in CNC systems have specific values or value ranges. For instance, specific tool radius compensation and length compensation parameters have fixed values, and the control parameters for rapid feed acceleration and deceleration generally fall within the range of 140–160 ms. Therefore, we analyze the parameters of FANUC CNC in Table I, and establish whitelist rules for key parameters and their value ranges to monitor the correctness of the key parameters in the CNC system.

TABLE I
FANUC PARAMETERS MAPPING TABLE

Parameter Type	FOCAS Address Mapping	Data Type	Number of parameters
Tool Compensation Parameters	0x000800000001 -0x000800000190	Real	400
Macro Variables	0x001500000001 -0x0015000003e7	Real	633
CNC Parameter	0x008d00000001 -0x008d00006bd9	Bit(axis), Byte(axis), Word(axis), Real(axis)	27609
PMC Parameter	0x800100000000 00000000 -0x800100000bb7 00000009	Bit(axis), Byte(axis), Word(axis), Real(axis)	6572
All Parameters			35214

2) Machining Process Whitelist Rule: The International Organization for Standardization (ISO) has established ISO-6983-1 [29] as the international standard for CNC programming languages. This standard delineates the lexical and syntactic rules governing CNC codes, thereby forming a programming language comprising G-codes and M-codes. Numerous CNC system manufacturers have introduced CNC control products adhering to the ISO-6983-1 standard. For instance, SIEMENS CNC systems, such as SINUMERIK 802D and SINUMERIK 840D, as well as FANUC CNC systems like 0i-md and 0i-mf, support NC programming in compliance with this standard. While CNCs from various manufacturers or models may exhibit diverse representations of NC code programming, they share commonalities, and the discrepancies in syntax and programming concepts are essentially minimal. Consequently, leveraging our proposed detection scheme and considering these shared characteristics, we can customize the intrusion detection system to be applicable to different models of CNC systems.

The CNC system controls the machining process through the NC code. The spindle and multiple servo axes in the CNC system work in coordination to control the movement and rotation of the tool and workpiece, completing automated production machining. Therefore, we parse the NC code and extract the invariant relationships of tool motion trajectories, feed rates, spindle speeds, and other parameters for each step of the machining process flow to generate machining process rules. These rules serve as benchmarks for detecting malicious attacks, such as NC code tampering or parameter injection.

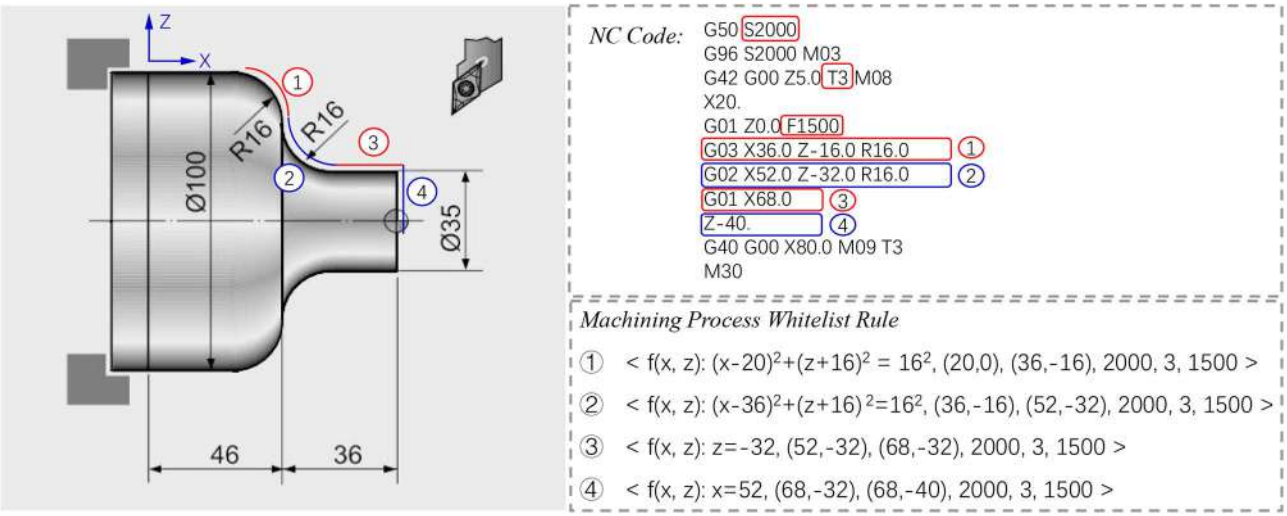


Fig. 4. Case of machining process whitelist rule.

Taking turning machining as an example [see Fig. 4], we demonstrate how to generate detection rules based on the invariance of the machining process using NC codes. The CNC system executes the NC codes to automatically control the machining process during turning. The NC codes specify the spindle speed, feed rate, and tool number for the machining process, and then use G codes to specify the tool's movement trajectory, such as common linear machining($G01$) and circular machining($G02, G03$). Therefore, we perform lexical, syntactic, and semantic analysis on the NC codes to generate the machining process rules, such as rule ①, which indicates that under the condition of spindle speed $S = 2000$ and feed rate $F = 1500$, tool number 3 moves along the curve $(x-20)^2 + (z+16)^2 = 16^2$, with a starting point of $(20, 0)$ and an ending point of $(36, -16)$.

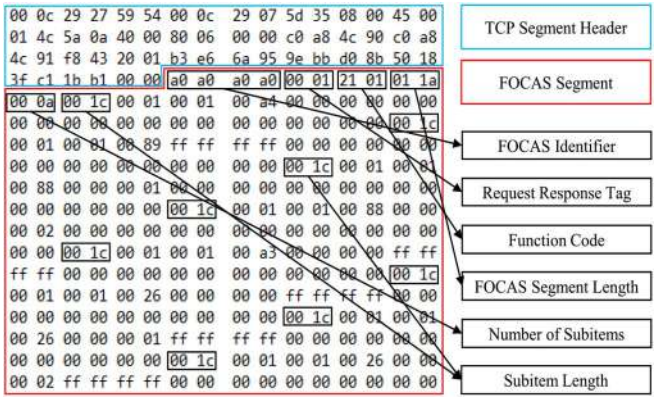


Fig. 5. FOCAS protocol reverse analysis results. This figure shows a binary request packet for collecting the current machining coordinates of a CNC using Focas, which consists of a Focas protocol header and ten subitems.

C. Low-Interference Acquisition

In order to collect the runtime status of the CNC, the conventional method is to use the communication interface provided by the CNC manufacturer. However, we found that the acquisition frequency of Focas, the communication interface provided by FANUC, is too low, which leads to an increase in the alarm delay for intrusion detection and affects the accuracy of the detection rule selection. For this reason, we manually reverse analyzed the protocol format of Focas and designed low-interference acquisition packets.

1) FOCAS Protocol Reverse Engineering: We capture mirror traffic on an industrial switch and conduct reverse protocol analysis on the proprietary protocol Focas for FANUC CNC systems. Focas protocol is an application-layer protocol based on TCP/IP. During the process of establishing a connection, the Focas protocol requires two rounds of TCP handshake to establish the connection. First, the client uses port A (any available port) to initiate a connection establishment request to port 8139 of the CNC system. Then, the client establishes a second connection to port 8193 of the CNC system using port A + 1 or A + 2. Subsequent request and response operations are performed on port A + 1 or A + 2. Reverse protocol analysis reveals the frame format of Focas protocol [see Fig. 5]. The first 4 bytes of the payload section are always a0a0a0a0, serving as the identification for Focas protocol. The 5th and 6th bytes represent the request/response flag. The Focas function code is located in bytes 7 and 8. The 9th and 10th bytes represent the length of the payload data. The 11th and 12th bytes indicate the number of subitems. The first 12 bytes form the header of the Focas protocol. The subitems in the Focas protocol include subitem length, fixed padding, and subfunction code. The payload information of the subitem includes the request parameter address and data format, which are not explicitly described in this article to prevent potential misuse by malicious individuals.

2) Constructing Low-Interference Acquisition Packages: Based on the results of the reverse protocol analysis mentioned above, we found that when using the API interface functions provided by Focas to collect position coordinates, spindle speed, and feed speed of CNC systems, multiple Focas request packets need to be sent to collect the machining status of the CNC system at the same moment. Moreover, these request packets usually contain irrelevant subitems that are unrelated

to attack detection. Under the high-frequency collection, these irrelevant subitems consume a significant amount of network and CNC system computing resources, which affects the CNC system and reduces detection efficiency. To solve this problem, we extracted the detection-related subitems from multiple request packets and combined them into a single request packet, which is then sent to the CNC system to collect multiple machining statuses at the same moment. Upon receiving the response packet, the status data of the CNC system is extracted based on the Focas protocol frame format.

The low-interference acquisition packages based on reverse engineering of the proprietary protocol greatly reduce the network overhead of status collection and minimize the interference on the CNC system. Furthermore, the analysis shows that the S7comm-nck protocol used by SINUMERIK 828, and 848 CNCs can also be used to construct low-interference request packets using the methods in this article. Detailed experimental data can be found in Section V-D.

D. Active Intrusion Detection Method

In this section, we outline the specific methods used for detecting attacks on manufacturing processes based on detection rules [see Fig. 3]. Our approach employs a low-interference, runtime active intrusion detection technique that does not disrupt the normal CNC machining process. During the implementation of this module, we have effectively addressed two key challenges.

1) Common phenomena, such as circuit instability, mechanical jitter, and equipment aging, can occur during the machining process. These issues can lead to inconsistencies in the CNC's execution time for each instruction. As a result, it becomes challenging to accurately and promptly match the collected runtime machining state to the detection rules.
2) The introduction of jitter and other interference due to regular errors, which can lead to an increased false alarm rate in the detection program, necessitating the need to distinguish between regular errors and cyber attacks.

1) Selection Detection Rules: To address challenge 1, We employed the dynamic time warping (DTW) algorithm and the k-nearest neighbors (KNNs) algorithm.

DTW is a dynamic programming algorithm that measures the similarity between time series [30], particularly those of varying lengths. It is commonly used in the fields of speech recognition, gesture recognition, and information retrieval due to its applicability to temporal data. In our study, we utilized the DTW algorithm to align a reference state sequence with a rule label to a captured runtime processing state sequence, with timestamps arranged in chronological order.

KNN is a nonparametric method used in supervised learning [31]. KNN is based on a simple and intuitive concept: if the majority of the k-most similar samples in the feature space of a given sample belong to a certain category, then the sample is also classified as belonging to that category. The algorithm makes its decision by considering only the category of the nearest one or more samples. In our study, we employed the KNN algorithm to classify runtime processing state points. This allowed us to select the appropriate detection rules based on reference state sequences that were labeled with the rules.

2) Consistency Checker-Based Detection Windows and Thresholds: To address challenge 2, we implemented a detection window and alarm threshold in our approach. During the detection process, we collect the runtime state of continuous machining from the CNC machine tool for the duration of the window time and then accumulate the error between each runtime machining state and the machining process rules. An alarm is triggered when the accumulated error exceeds the threshold. If the window expires and the cumulative error does not exceed the threshold, the cumulative error is reset to 0 and a new inspection window is initiated. In this article, we employed (4) to conduct a consistency check, which involves accumulating the Euclidean distance between the runtime machining state and the machining process rule within the inspection window and comparing it with the inspection threshold. Specifically, as in (5), the actual error value is obtained by calculating the distance D between the actual position and the machining trajectory of the machining process rule, and the deviation of the actual feed rate F and spindle speed S from the machining process rule. Additionally, we performed dissimilarity verification between the runtime key parameter matrix $\mathbb{C}_{3\times n}$ and the key parameter rule $\mathbb{K}_{3\times n}$ to ensure key parameter consistency. Equation (3) serves as the theoretical foundation for our machining process consistency verification, enabling us to identify attacks, such as machining code injection, parameter injection, and instruction injection, on the CNC system during the machining process

$$\left\{\sum^{W\ \text{size}} \|y(t)-r(t)\| \le \delta(t)\right\} \wedge \{\mathbb{C}_{3\times n} \oplus \mathbb{K}_{3\times n} = [0]_{3\times n}\} \tag{4}$$

$$\|y(t)-r(t)\| = \sqrt{D^2 + (F-F_r)^2 + (S-S_r)^2}. \tag{5}$$

The active detection engine detects whether the CNC is under attack in an active and low-interference way, and its core part is shown in Algorithm 1. It first establishes a communication connection with the CNC (line 1); then parses the detection rules to simulate the machining path and constructs the active acquisition packet (lines 2–4) and then the attack is detected (lines 5–16); and, finally, when the detection is complete, the connection is disconnected (line 17). In the attack detection phase, the first step is to initialize the detection window and detection threshold (line 6). Next, a low-interference request packet is sent to the CNC, followed by receiving the response data and parsing the protocol to extract the processing state values (lines 7–10). Then, the DTW algorithm is used to match the detection rules for the data in the current detection window. Finally, a consistency check is done on the machining state and key parameters to see if the detection rules are satisfied (lines 12–14). Line 15 indicates the setting of the required frequency for the CNC to achieve low interference.

V. Evaluation

In this section, we focus on answering the following research question.

Algorithm 1: Algorithm of Active Detection

Input : Key Parameter Rule
Machining Process Rule
Detection Window Size
Output: Cyber-Attack Alerts

```
Connect(cncIP, cncPORT);
RuleDB ← LoadRules(KeyParameter, MachiningProcess);
PathSim ← PathSimulation(RuleDB);
AcquisitionPKG ← PackageConstructor(RuleDB);
while ProcessFlag do
    InitDetectionWindow();
    for 0 to DetectionWindowSize do
        Send(AcquisitionPKG);
        ProcessingStatus ← Receive();
    end
    FlaggedStatus ← DTW(ProcessingStatus, PathSim);
    if ConsistencyChecker(FlaggedStatus, RuleDB) is
    False then
        Alarm("Illegal Processes: cncIP, RuleNo.");
    end
    Sleep(t);
end
Disconnect();
```

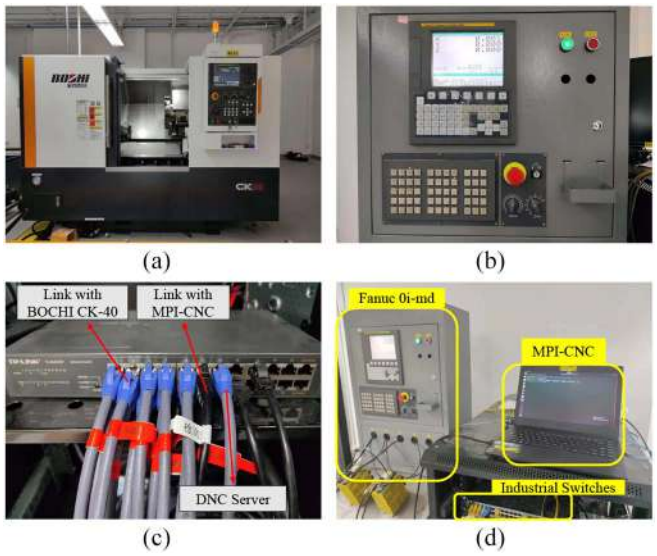


Fig. 6. MPI-CNC experimental environment. (a) BOCHI CK-40 with Fanuc 0i-tf. (b) Fanuc 0i-md. (c) Industrial switches. (d) MPI-CNC deployment environment.

RQ1: What is the time cost of MPI-CNC to generate rules?

RQ2: What is the effectiveness of MPI-CNC detection against cyber attacks on CNC systems?

RQ3: Does MPI-CNC affect CNC machine tool machining efficiency?

We used 3589 lines of C code and 2492 lines of Python code to implement MPI-CNC. MPI-CNC is deployed on a ThinkPad P15Gen2 with an 8 cores Intel Core i9-11950H CPU and 64 GB of RAM. MPI-CNC was evaluated using Fanuc 0i-md CNC and Fanuc 0i-tf CNC.

A. Experimental Environment

Experimental Design: To answer the three research questions above, experiments were conducted using real machining environments with the FANUC CNC system. Due to the lack of real-world cyber-attack data for the FOCAS CNC system, three attack methods discussed in Section II-B were implemented, and the effects of the attacks were demonstrated on real equipment. First, multiple NC programs from realistic machining scenarios that are applicable to the FANUC CNC system were analyzed, and detection rules were generated to verify the accuracy of the automatic parsing of NC code for rule generation. Second, the proposed intrusion detection method was compared with other detection models to evaluate its performance in detecting network attacks in the NC machining process. Finally, in order to demonstrate that our solution has minimal impact on the CNC system's machining process, the variations in machining time were monitored during the detection phase, and the network resource utilization was compared between using low-interference data collection requests and using FOCAS standard interface for collecting CNC system's machining status.

Experimental Environment: In this work, we conducted experiments using a FANUC 0i-tf CNC system [Fig. 6(a)] and FANUC 0i-md CNC system [Fig. 6(b)]. as shown in Fig. 6, We connect the MPI-CNC to the industrial switch connected to the CNC and configure its IP address to be in the same network segment as the other devices so that it can communicate with the CNC normally. Fig. 6(c) partly shows the connection status of BOCHI CK-40, DNC server, and MPI-CNC to each port of the industrial switch. Fig. 6(d) shows the site layout of the Fanuc 0i-md intrusion detection experimental environment, including the CNC, MPI-CNC, and industrial switch.

Cyber-Attack Setting: Specifically, we discuss three attack methods against CNCs in this work, which were implemented on FANUC CNCs due to the lack of available attacks for evaluation purposes. The first attack method is the machining code injection attack, which involves injecting malicious machining instructions and machining paths into the NC code. We introduced 20 tamperings in 10 different machining codes, including creep attacks and trajectory scaling, to evaluate the detection capability of the methods. The second attack method is the remote parameter injection attack, which was implemented by tampering with key CNC parameters through request packets sent to the FANUC CNC based on the Focas protocol inversion results. The third attack method is the malicious command injection attack, which involves tampering with the designated ports of the PMC by sending request packets to the FANUC CNC based on the Focas protocol inversion results and the CNC's interface manual. This allows for malicious commands, such as remote start/stop and on/off coolant, to be injected.

B. RQ1—Time Cost and Accuracy of Generating Detection Rules

We collected 57 NC codes applicable to FANUC CNC systems from real machining scenarios and Internet platforms, such as Github and Traceparts. These codes involve turning and milling processes and consist of 8354 instructions, including instructions for linear machining, circular arc machining,

TABLE II
TIME COST OF GENERATING DETECTION RULES

NC Code	Code Lines	Number of Rules	Times Cost(ms)
O5665-NC	134	90	1.004
O6383	150	93	0.805
NCViewer.nc	5780	5753	0.962
NCtest26.NC	64	61	0.768
7190.3-1A.nc	255	222	0.792
...	...	...	...
Number of NC Code: 57	Total Code Lines: 8354	Total Rules: 7671	Totle Time Cost: 53.579ms

TABLE III
COMPARING THE ACCURACY AND TIME COST OF DTW AND KNN ALGORITHMS IN SELECTING DETECTION RULES

Algorithms	KNN				DTW
	k=3	k=5	k=7	kd-tree	
Accuracy %	96.95	97.82	96.83	97.32	99.38
Times Cost(ms)	273	321	326	137	352

tool changing operations, coolant control, and more. By analyzing the machining instructions in these NC programs, we extracted attack detection rules. Unlike other high-level programming languages, NC programs are not as complex, typically consisting of multiple G codes and M codes. We used the number of G codes that control the machining trajectory to represent the size of the program and generated several key parameter rules based on the relationship between G codes and M codes, as well as one machining process rule per G code.

First, we analyzed the FANUC CNC system user manual and interface manual and combined the results of reverse engineering the FOCAS proprietary protocol to establish a mapping table of G codes, M codes, and system parameter addresses. Then, based on the semantic information of G and M codes in the NC codes, we selected important parameters related to machining and generated key parameter rules. Next, we parsed the G codes that control the machining trajectory, abstracted the curve equation of the machining trajectory based on its semantic information, and combined information, such as the starting point and ending point of the machining, spindle speed, feed rate, and tool number to generate machining process rules.

As shown in Table II, we analyzed all the collected NC codes and recorded all the generated key parameter rules and machining process rules. The 57 NC codes we collected totaled 8354 lines of machining instructions. For these collected 57 NC codes, key parameters related to machining are identified and a total of 6056 machining process rules are generated. In the process of generating inspection rules, we recorded the number of lines of machining code, the number of inspection rules generated, and the time cost of generating the rules for each NC code. The time taken to analyze the generation of inspection rules for a single NC code is 0.94 ms on average, and the time taken for a single inspection rule is 0.007 ms [Table II]. To verify the correctness of the generated rules, we selected 10 representative NC codes and manually verified the accuracy of the automatically generated detection rules using all the key parameter rules and machining process rules. The results showed that the accuracy of the detection rules in a limited number of NC program samples was 100%.

Answer RQ1: The proposed method in this article automatically analyzes NC code and generates comprehensive detection rules without relying on historical manufacturing data. Each NC code takes 0.94 ms to generate accurate inspection rules. The time cost of the method in this article is extremely low and can be quickly used in new machining scenarios.

C. RQ2—CPMS Cyber-Attack Detection Results

In Section II-B, we provide a comprehensive review of the current state-of-the-art research on attacks against manufacturing processes and a summary of three common attack methods. It is worth noting that publicly available CPMS attack methods or attack data sets are typically tailored to specific machining scenarios, devices, and processing processes, and there are currently no generic CPMS attacks. Therefore, in this work, we evaluate the detection effectiveness of the MPI-CNC by implementing three attack methods.

1) Selection Detection Rules Results: We conducted experiments to evaluate the accuracy and time cost of KNN and DTW algorithms in rule selection using 25 283 data points from real machining scenarios. The results are shown in Table III. We tested the performance of the KNN algorithm by setting different values of k in the rule selection experiments and using a kd-tree data structure. The test results showed that when k was set to 5, the rule selection accuracy was 97.82%, which was the optimal parameter for the KNN algorithm in rule selection. It is worth noting that the use of a kd-tree data structure greatly reduced the time cost, with rule selection for 25 283 data points taking only 137 ms. This makes it suitable for complex machining scenarios and real-time detection requirements. When using the DTW algorithm for rule selection, the selection accuracy was as high as 99.38% and the time cost was 352 ms, which falls within an acceptable range and meets real-time alarm requirements. In summary, to improve detection accuracy, MPI-CNC adopts the DTW algorithm as its rule selection algorithm. For complex machining scenarios with high-real-time detection requirements, the KNN algorithm based on a kd-tree data structure should be used.

2) Cyber-Attack Detection Results: We conducted experiments to evaluate the performance of different detection windows in attack detection, and the results are presented in Table IV. The active detection engine successfully detected the machining code injection attack, the key parameter injection attack, and the malicious instruction injection attack. These attacks interfere with the normal machining process and result in changes to the machining trajectory and machining state, which can be directly reflected in the machining process and the key parameters of the CNC system. The active detection engine actively communicates with the FANUC CNC to map its actual machining status and key parameters. This active detection approach makes it difficult for the attacks to be

TABLE IV
COMPARISON OF CYBER-ATTACK DETECTION RESULTS IN DIFFERENT DETECTION WINDOWS AND DETECTION THRESHOLD CASES

Window size /Threshold	Detection accuracy	False alarm rate	Missing alarm rate	Alarm delay(s)
10/0.1	99.15%	2.89%	1.83%	1.45
50/0.5	**98.81%**	**1.09%**	**2.42%**	**2.45**
100/1	98.68%	0.75%	6.17%	3.71
200/2	96.88%	2.61%	7.15%	6.32
500/5	94.87%	4.44%	6.25%	13.75

TABLE V
COMPARISON WITH OTHER CPMS IDS

Attack Type	Machining Code Injection	Parameter Injection	Instruction Injection
Digital Twins[7]	N/A	93.25%	93.25%
KCAD[4]	81.39%	81.39%	N/A
LTDT[3]	95.55%	95.55%	N/A
LSTM-AE[6]	94.79%	94.79%	N/A
Machining Code Analysis[8]	98.38%	N/A	N/A
MPI-CNC	**98.81%**	**100.00%**	**100.00%**

hidden, as it requires the attacker to gain insight into the real machining process and manipulate the CNC firmware, tamper with the network communication module, or employ other sophisticated methods to feedback network data that conforms to the machining process rules and key parameter rules.

As shown in Table IV, we set different detection window sizes of 10, 50, 100, 200, and 500 in the active detection experiments. We set the alarm threshold in millimeters based on the machining accuracy of the CNC machine tool of 0.01 mm multiplied by the window size. The test results show that the detection accuracy of the method proposed in this article is 99.15% with a detection window of 10, and the accuracy decreases with the increase of the detection window, down to 94.87%. The reason for this phenomenon is that as the detection window increases, the detection threshold increases, increasing the missing alarm rate and a decrease in detection accuracy. In Table IV, the missing alarm rate is 1.83% when the detection window is 10. The larger the detection window, the larger the missing alarm rate, and when the detection window is 500, the missing alarm rate is 6.25%. It is worth noting that the false alarm rate becomes larger when the detection window is too large and too small. The alarm delay increases as the detection window increases, mainly because the DTW algorithm takes more time to match more data. Considering the above, we conclude that the optimal detection window size of this method is 50, the detection accuracy is 98.81%, the false alarm rate is 1.09%, the missing alarm rate is 2.42%, and the alarm delay is 2.45 s.

3) Comparison With Other CPMS IDS: In order to demonstrate the effectiveness of our detection scheme, MPI-CNC was compared to its performance with other CPMS ICS models. Specifically, we compared with representative models that are commonly used for detecting manufacturing process attacks, namely, digital twin-based intrusion detection [7], side-channel analysis-based intrusion detection (KCAD [4], LTDT [3], LSTM-AE [6]), and machining code analysis-based intrusion detection [8]. The digital twin-based intrusion detection models require historical processing state data for fitting digital twin models. KCAD and LSTM-AE collect audio data generated by manufacturing equipment during processing to learn anomaly classification models. LTDT analysis of processing video classification anomalies. The machining code analysis-based intrusion detection extracts features of NC codes to train SVM models for offline classification of anomaly codes.

We launched 200 machining code injection attacks against 10 different machining scenarios (machining codes), each introducing 20 tampering points (such as replacement of G02 and G03 with G01; insertion of protrusions or depressions; and modification of endpoints to cause deformation).In addition, 100 parameter injection attacks and command injection attacks were used to test the detection performance of the MPI-CNC. Detection results and attack logs are collected and used to calculate detection accuracy, miss rate and false alarm rate of MPI-CNC, as in

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{6}$$

$$\text{Missing Alarm} = \frac{\text{FN}}{\text{TP}+\text{FN}} \tag{7}$$

$$\text{False Alarm} = \frac{\text{FP}}{\text{TN}+\text{FP}}. \tag{8}$$

The methods have achieved high-detection accuracy using current state-of-the-art techniques, as shown in Table V. The detection accuracy of the digital twin-based detection method against parameter injection attacks and instruction injection attacks is 93.25% [7]. The detection accuracy of the KCAD against processing code tampering and instruction injection attacks is 81.39% [4]. In addition, the LTDT and LSTM-AE models have high-detection accuracies of 95.55% [3] and 94.79% [6]. However, the side channel data features behave differently in different processing scenarios, which makes it difficult to apply to new scenarios quickly. The code analysis-based intrusion detection directly analyzes processing code for offline detection with 98.38% detection accuracy [8], but it lacks runtime detection capability during processing. Compared with the above methods, MPI-CNC can cope with a wider range of attack scenarios. Since we analyze the key parameters of the CNC system and actively detect the key parameter information during the machining process, we can detect parameter injection attacks and instruction injection attacks by 100%. Also, the method in this article analyzes the NC code to generate comprehensive detection rules, so it has a higher detection accuracy similar to the machining code analysis method, with a detection accuracy of 98.81%.

Answer RQ2: Experiments have proved that MPI-CNC can cope with three attack methods, which is more comprehensive than the traditional detection model based on historical data. Moreover, MPI-CNC is significantly better than other detection methods in terms of accuracy of 98.81%.

D. RQ3—Low-Interference Experimental Results

The computational performance of CNC systems is not as high as that of traditional PCs. Therefore, we need to be cautious about whether active detection methods will affect the normal operation of NC systems. In this article, we adopt a low-interference, polling-based approach to collect the

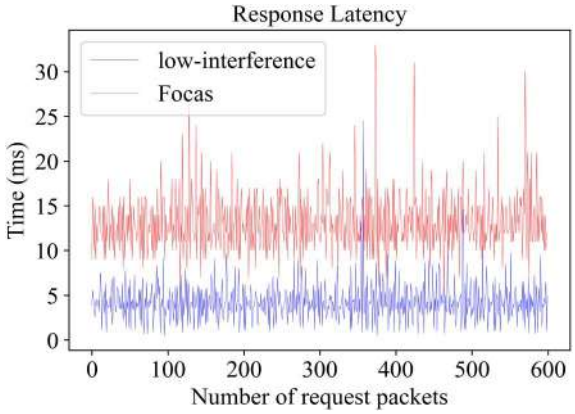


Fig. 7. Comparison of response latency for low interference and Focase.

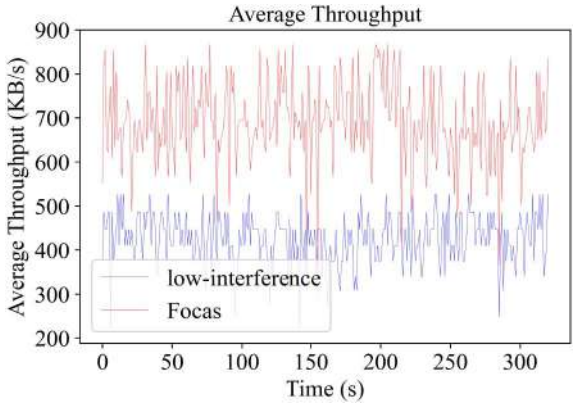


Fig. 8. Comparison of average throughput for low interference and Focase.

machining status of NC systems, and then process the data and detect attack behaviors remotely. In this section, we compare the acquisition delay, network resource utilization, and impact on machine tool processing times between the low-interference collection method proposed in this article and the FOCAS standard interface for collecting the machining status of CNC systems.

1) Acquisition Latency and Network Throughput: As shown in Fig. 7, the average response time for a single collection using the conventional FOCAS collection method is 13.240 ms, with a collection frequency of 75.53 times per second. Using the low-interference collection method proposed in this article, the average response time for a single collection is 4.368 ms, with a collection frequency of 228.94 times per second. The response delay is reduced by 67.00%, and the sampling frequency is increased by 203.12%. Meanwhile, in Fig. 8 the average throughput of the CNC system using the conventional FOCAS collection method is 692.04 KB/s, while the average throughput of the NC system using the low-interference collection method proposed in this article is 426.23 KB/s, resulting in a decrease in network throughput of 38.41%. Experiments have shown that using the self-assembled packet method based on private protocol reverse engineering proposed in this article for collecting the machining status of CNC systems reduces interference to the CNC systems while improving the collection frequency.

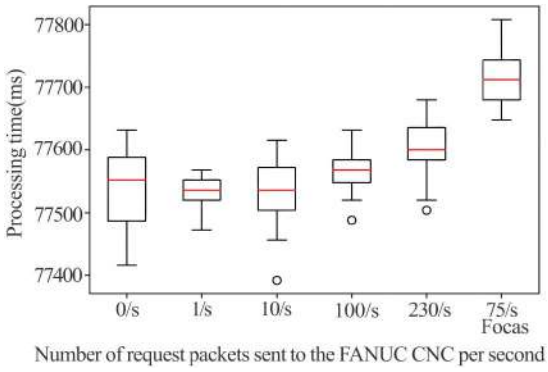


Fig. 9. Influence of network traffic on FANUC CNC.

2) Processing Times Under CPMS IDS: In order to demonstrate that our detection scheme has minimal impact on the machining time of CNC systems, we conducted experiments with different request frequencies during the active detection process to observe changes in machining time. To avoid the forwarding delay of industrial switches, we directly connected the CNC system with Ethernet cables and sent 1, 10, 100, and 230 low-interference request packets per second, as well as 75 Focas standard interface request packets per second. As shown in Fig. 9, under normal conditions without any external interference, the machining time of the CNC system was approximately 1 min and 17.545 s. When the CNC system was subjected to varying degrees of external interference, we found that its machining time was minimally affected. In the case of sending 230 packets per second, the machining time of the CNC system increased by only 0.076% and no packet loss was observed. This indicates that the FANUC CNC system has the capability to process at least 230 packets per second without affecting its normal operation. However, when using Focas standard interface request packets with a maximum rate of 75 requests per second, the machining time of the CNC system increased by 0.217%, which was significantly higher than the low-interference data collection method used in this article. Moreover, the FANUC system can handle a maximum of 75 requests per second. The experiment results show that our detection scheme has minimal impact on the machining time of CNC systems.

Answer RQ3: MPI-CNC significantly improves acquisition frequency and efficiency by reverse engineering dedicated acquisition protocols and customized packet acquisition, reducing CNC network resource usage and increasing machining time by only 0.076%. MPI-CNC does not affect normal machining.

VI. Related Work

Cyber-attacks against CPMS directly affect production efficiency and even threaten the safety of human life. Therefore, IDS research in the field of CPMS has become an academic hotspot. Cyber-attacks on CPMS mainly focus on controlling and disrupting the manufacturing process, which is the key concern of CPMS IDS. By analyzing research results in this

field over the past few years, we have classified the state-of-the-art CPMS IDS into three categories based on detection methods.

CPMS IDS Based on Digital Twins: Digital twin technology utilizes historical data to fit a control model that simulates physical processes [32]. Balta et al. [7] collected and analyzed historical machining data from a 3-D printer to construct controller digital twin models for the 3-D printer's CNC system. By comparing the consistency between the simulated machining state of the controller digital twin model and the actual machining state, they detected cyber-attacks that tampered with the temperature parameters of the 3-D printer's nozzle heaters.

CPMS IDS Based on Side-Channel Analysis: Manufacturing equipment generates a large amount of measurement channel data during the machining process, which can indirectly reflect the machining state. Detection methods based on side-channel analysis are a popular approach in the CPMS IDS field. Chhetri et al. [4] proposed for the first time the use of audio data around manufacturing equipment to train detection models for detecting machining path tampering attacks. Bayens et al. [33] combined analysis of acoustic features of machining equipment, machining location features, and production waste features to verify product consistency. Belikovetsky et al. [5] analyzed audio data from 3-D printer stepper motors and evaluated the similarity between their audio features and audio fingerprints to detect the 3-D printing process. Mamun et al. [3] detecting 3-D printer processing trajectory changes using video stream analysis. Yoginath et al. [2] analyzed the current values of 3-D printer power lines using the Bayesian model to detect creep attacks. Shi et al. [6] extracted features from side-channel data collected by vibration sensors based on the LSTM-autoencoder algorithm and later used the OCSVM classification algorithm for anomaly detection.

CPMS IDS Based on Machining Code Analysis: The ISO has developed the ISO-6983-1 [29] standard as an international standard for NC programming languages. This standard specifies the lexical and syntax rules of NC code, forming a programming language composed of G codes and M codes. NC code contains the most complete and comprehensive control information of the machining process, and the CNC system automatically controls the machining process according to the instructions in the NC code. By analyzing the NC code, anomalies can be effectively detected. Beckwith et al. [8] extracted statistical features from NC code, including the number of G codes and M codes, as well as the frequency of *XYZ* values. They trained a machine learning anomaly classification model and conducted an offline analysis of NC code to identify anomalies. Tsoutsos et al. [34] reverse-engineered NC code to generate 3-D models, and then simulated pressure tests on these models. They discovered vulnerabilities in the NC code during this process.

VII. Discussion

The intrusion detection method proposed in this article has the following limitations.

1) The method may have difficulty in dealing with man-in-the-middle attacks implemented through tampering with the firmware of the CNC system. Such attacks require high-technical skills from the attackers and can effectively bypass the intrusion detection method proposed in this article.
2) The method is effective for application in 2-axis and 3-axis CNC machines, but it may not be able to generate detection rules specifically for 5-axis machine centers.
3) The active detection approach proposed in this article may not be applicable to CNC systems with interface authentication mechanisms. However, it should be noted that currently, most CNC systems do not restrict remote access to machining status information.
4) Low-interference acquisition methods can be applied to CNCs from different vendors, but protocol reversal demands high-technical skill. Automated protocol reversal is a meaningful task that needs to be addressed.

VIII. Conclusion

This article proposes a novel approach for runtime detection of CPMS cyber-attacks, denoted as MPI-CNC. We implement a prototype system on the FANUC CNC machine tools as an example. Specifically, MPI-CNC automatically analyzes NC programs, extracts machining process invariants, and generates attack detection rules, including machining process rules and key parameter rules. Then, using low-interference request packets, MPI-CNC actively communicates with the CNC system to collect process status and key parameters, while setting detection windows and thresholds to detect attack behaviors. In the end, we evaluate MPI-CNC in real machining scenarios using a FANUC CNC machine tool. Experimental results demonstrate that MPI-CNC exhibits excellent adaptability performance, being able to accurately detect various cyber-attacks without affecting the normal operation of the CNC system. Compared with other state-of-the-art detection models, our approach shows superior adaptability performance and detection performance.

References

[1] A. Kusiak, "Smart manufacturing," *Int. J. Prod. Res.*, vol. 56, nos. 1-2, pp. 508–517, 2018.
[2] S. Yoginath et al., "Stealthy Cyber anomaly detection on large noisy multi-material 3-D printer datasets using probabilistic models," in *Proc. ACM CCS Workshop Addit. Manuf. Secur.*, New York, NY, USA, 2022, pp. 25–38.
[3] A. A. Mamun, C. Liu, C. Kan, and W. Tian, "Securing cyber–physical additive manufacturing systems by in-situ process authentication using streamline video analysis," *J. Manuf. Syst.*, vol. 62, pp. 429–440, Jan. 2022.
[4] S. R. Chhetri, A. Canedo, and M. A. Al Faruque, "KCAD: Kinetic Cyber-attack detection method for cyber–physical additive manufacturing systems," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design (ICCAD)*, 2016, pp. 1–8.
[5] S. Belikovetsky, Y. A. Solewicz, M. Yampolskiy, J. Toh, and Y. Elovici, "Digital audio signature for 3-D printing integrity," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 1127–1141, 2019.
[6] Z. Shi, A. A. Mamun, C. Kan, W. Tian, and C. Liu, "An LSTM-autoencoder based online side channel monitoring approach for cyber–physical attack detection in additive manufacturing," *J. Intell. Manuf.*, vol. 34, no. 4, pp. 1815–1831, Apr. 2023.

[7] E. C. Balta, M. Pease, J. Moyne, K. Barton, and D. M. Tilbury, "Digital twin-based cyber-attack detection framework for cyber–physical manufacturing systems," *IEEE Trans. Autom. Sci. Eng.*, early access, May 25, 2023, doi: 10.1109/TASE.2023.3243147.
[8] C. Beckwith et al., "Needle in a haystack: Detecting subtle malicious edits to additive manufacturing G-code files," *IEEE Embed. Syst. Lett.*, vol. 14, no. 3, pp. 111–114, Sep. 2022.
[9] H. Pearce, K. Yanamandra, N. Gupta, and R. Karri, "FLAW3D: A trojan-based cyber attack on the physical outcomes of additive manufacturing," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 5361–5370, Dec. 2022.
[10] M. Yampolskiy, L. Graves, J. Gatlin, J. T. McDonald, and M. Yung, "Crypto-steganographic validity for additive manufacturing (3D printing) design files," in *Proc. Int. Conf. Inf. Secur.*, 2022, pp. 40–52.
[11] T. Le et al., "Physical logic bombs in 3-D printers via emerging 4-D techniques," in *Proc. 37th Annu. Comput. Security Appl. Conf.*, New York, NY, USA, 2021, pp. 732–747.
[12] M. McCormack, S. Chandrasekaran, G. Liu, T. Yu, S. DeVincent Wolf, and V. Sekar, "Security analysis of networked 3-D printers," in *Proc. IEEE Security Privacy Workshops (SPW)*, 2020, pp. 118–125.
[13] B. Shadow. "Industrial security exploitation framework." 2020. [Online]. Available: https://github.com/w3h/isf
[14] R. R. Maiti, C. H. Yoong, V. R. Palleti, A. Silva, and C. M. Poskitt, "Mitigating adversarial attacks on data-driven invariant checkers for cyber–physical systems," *IEEE Trans. Depend. Secure Comput.*, vol. 20, no. 4, pp. 3378–3391, Jul./Aug. 2023.
[15] J. Liu et al., "ShadowPLCs: A novel scheme for remote detection of industrial process control attacks," *IEEE Trans. Depend. Secure Comput.*, vol. 19, no. 3, pp. 2054–2069, Jun. 2022.
[16] P. Mahesh et al., "A survey of cybersecurity of digital manufacturing," *Proc. IEEE*, vol. 109, no. 4, pp. 495–516, Apr. 2021.
[17] Y. Pan et al., "Taxonomies for reasoning about cyber–physical attacks in IoT-based manufacturing systems," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, no. 3, pp. 45–54, Jul. 2017.
[18] T. Lin et al., "Susceptibility to spear-Phishing emails: Effects of Internet user demographics and email content," *ACM Trans. Comput. Human Interact.*, vol. 26, no. 5, pp. 1–28, Jul. 2019.
[19] N. Karsten and L. Jakob, "BadUSB—On accessories that turn evil," presented at Blackhat Conf., Las Vegas, NV, USA, 2014, pp. 1–28.
[20] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2027–2051, Mar. 2016.
[21] S. B. Moore, W. B. Glisson, and M. Yampolskiy, "Implications of malicious 3-D printer firmware," in *Proc. 50th Hawaii Int. Conf. Syst. Sci.*, 2017, pp. 1–10.
[22] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu, "Review on cyber–physical systems," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 1, pp. 27–40, Jan. 2017.
[23] G. Y. Dayanıklı, S. Sinha, D. Muniraj, R. M. Gerdes, M. Farhood, and M. Mina, "Physical-layer attacks against pulse width modulation-controlled actuators," in *Proc. 31st USENIX Secur. Symp.*, Boston, MA, USA, 2022, pp. 953–970.
[24] J. Gatlin et al., "Encryption is futile: Reconstructing 3D-printed models using the power side-channel," in *Proc. 24th Int. Symp. Res. Attacks, Intrusions Defenses*, New York, NY, USA, 2021, pp. 135–147.
[25] L. J. Wells, J. A. Camelio, C. B. Williams, and J. White, "Cyber–physical security challenges in manufacturing systems," *Manuf. Lett.*, vol. 2, no. 2, pp. 74–77, 2014.
[26] T. Zinner, G. Parker, N. Shamsaei, W. King, and M. Yampolskiy, "Spooky manufacturing: Probabilistic sabotage attack in metal AM using shielding gas flow control," in *Proc. ACM CCS Workshop Additive Manuf. (3D Printing) Security*, New York, NY, USA, 2022, pp. 15–24.
[27] S. R. Chhetri, A. Barua, S. Faezi, F. Regazzoni, A. Canedo, and M. A. Al Faruque, "Tool of spies: Leaking your IP by altering the 3-D printer compiler," *IEEE Trans. Depend. Secure Comput.*, vol. 18, no. 2, pp. 667–678, Apr. 2021.
[28] H. Choi et al., "Detecting attacks against robotic vehicles: A control invariant approach," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, 2018, pp. 801–816.
[29] *Automation Systems and Integration—Numerical Control of Machines—Program Format and Definitions of Address Words—Part-1: Data Format for Positioning, Line Motion and Contouring Control Systems, International Organization for Standardization*, ISO Standard 6983-1:2009, 2009.
[30] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978.
[31] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *Int. Statist. Rev./Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.
[32] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models," *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, Oct. 2020.
[33] C. Bayens, T. Le, L. Garcia, R. Beyah, M. Javanmard, and S. Zonouz, "See no evil, hear no evil, feel no evil, print no evil? malicious fill patterns detection in additive manufacturing," in *Proc. 26th USENIX Secur. Symp.*, Vancouver, BC, Canada, 2017, pp. 1181–1198.
[34] N. G. Tsoutsos, H. Gamil, and M. Maniatakos, "Secure 3-D printing: Reconstructing and validating solid geometries using Toolpath reverse engineering," in *Proc. 3rd ACM Workshop Cyber Phys. Syst. Secur.*, New York, NY, USA, 2017, pp. 15–20.

Zedong Li is currently pursuing the Ph.D. degree with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

His research interests include cyber–physical manufacturing systems security and intrusion detection.

Xin Chen (Member, IEEE) is currently pursuing the Ph.D. degree with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

His research interests include cyber–physical manufacturing systems security and intrusion detection.

Yuqi Chen received the B.Sc. degree in computer science from South China University of Technology, Guangzhou, China, in 2015, and the Ph.D. degree from Singapore University of Technology and Design, Singapore, in 2019.

He is an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. Before joining ShanghaiTech, he was a Research Scientist with the System Analysis and Verification Group, Singapore Management University, Singapore. He employs a range of techniques, including testing, reverse engineering, program analysis, and formal methods, to develop practical solutions for securing critical cyber–physical systems. His research interests lie at the intersection of software engineering and security.

Shijie Li is currently pursuing the Ph.D. degree with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

His research interests include industrial control system security and intrusion detection.

Hangyu Wang is currently pursuing the Ph.D. degree with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

His research interests include industrial control system security and access control.

Shichao Lv received the M.S. degree in cryptographyăfrom the University of Electronic Science and Technology of China, Chengdu, China, in 2012, and the Ph.D. degree in information security from the University of Chinese Academy of Sciences, Beijing, China, in 2018.

He is a Ph.D. Professorate Senior Engineer and an M.S. Supervisor from the Institute of Information Engineering, Chinese Academy of Sciences, Beijing. His main research interests include Internet of Things security and industrial control system security.

Limin Sun received the Ph.D. degree from the National University of Defense Technology, Changsha, China.

He is currently a Professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. His main research interests include Internet of Things security and industrial control system security.

Prof. Sun is also the Secretary General of the Select Committee of CWSN and the Director of the Beijing Key Laboratory of IoT Information Security Technology. He is an Editor of the *Journal of Computer Science* and the *Journal of Computer Applications*.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security

Danda B. Rawat, *Senior Member, IEEE*, Ronald Doku and Moses Garuba

***Abstract*—"Knowledge is power" is an old adage that has been found to be true in today's information age. Knowledge is derived from having access to information. The ability to gather information from large volumes of data has become an issue of relative importance. Big Data Analytics (BDA) is the term coined by researchers to describe the art of processing, storing and gathering large amounts of data for future examination. Data is being produced at an alarming rate. The rapid growth of the Internet, Internet of Things (IoT) and other technological advances are the main culprits behind this sustained growth. The data generated is a reflection of the environment it is produced out of, thus we can use the data we get out of systems to figure out the inner workings of that system. This has become an important feature in cybersecurity where the goal is to protect assets. Furthermore, the growing value of data has made big data a high value target. In this paper, we explore recent research works in cybersecurity in relation to big data. We highlight how big data is protected and how big data can also be used as a tool for cybersecurity. We summarize recent works in the form of tables and have presented trends, open research challenges and problems. With this paper, readers can have a more thorough understanding of cybersecurity in the big data era, as well as research trends and open challenges in this active research area.**

***Index Terms*—Big Data Security, Big Data Driven Security, IDS/IPS, Data Analytics.**

I. Introduction and Background

Over the past 15 years, data has increased exponentially in various applications which has led to the big data era (Fig. 1). It is worth noting that big data has some peculiar features which can be leveraged for various purposes (Fig. 2). One of these is the use of big data for detecting risks or attacks. "As our technological powers increase, the side effects and potential hazards also escalate" is a quote by Alvin Toffler which perfectly sums up the world we live in now. Hacking was at first akin to public defacements of things. Hackers hacked for fun and for notoriety. However, these days, attacks are more calculated and motivated. Nations are accusing each other of hacking. There is also a significant rise in industrial espionage which can either be from nation-state or competing entities trying to gather information or to take away a competitor's edge as to increase their own. Additionally, we are seeing this across industries from health care to retail to government to education to the financial sector. Thus, with this much susceptibility and hacking advancements, cybersecurity has become an important field in computer science. Cybersecurity aims at reducing the attack vectors/points to a minimal, because it is impossible secure every attack point. An attacker only has to be successful once which has consequently made the job of securing systems very challenging. The number of attackers out there out-number the people trying to protect it. This is because there is so much information out there that can turn anyone into an attacker. With this in mind, cybersecurity has now gone beyond the traditional way of only focusing on prevention to a more sophisticated PDR paradigm which is: Prevent, Detect and Respond (PDR). Big data is expected to play a major role in this emerging PDR paradigm.

Big data is now a common slogan used to mean the generation of large volumes of data. Enormous amount of data are being generated at an alarming rate. This is due to the growth of the Internet. Laney [1] came up with the term the three V's which he associated with big data. These terms were volume, velocity, and variety. In addition to 3 V's, there is fourth V which is veracity. Volume represents the fact that the data being generated is enormous, velocity represents the fact that data is being generated at an alarming rate, and

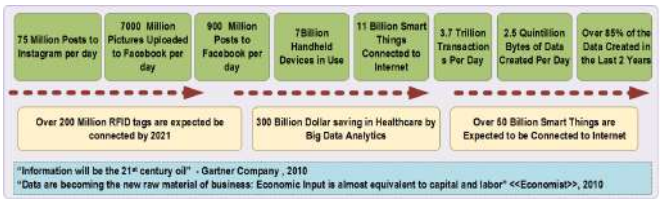


Figure 1. Big data is increasing exponentially making security harder.

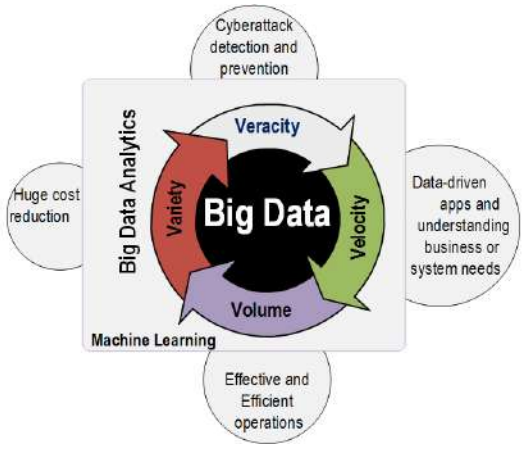


Figure 2. Big data offers typical benefits to business such as informed decisions, competitive advantages and data-driven cybersecurity.

Manuscript received 2018.

Authors are with the Data Science and Cybersecurity Center (DSC^2), Department of Electrical Engineering and Computer Science at Howard University, Washington, DC, USA. Corresponding e-mail: db.rawat@ieee.org

This work is supported in part by the U.S. National Science Foundation (NSF) under grants CNS-1658972, CNS-1650831 and and HRD 1828811. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the NSF.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

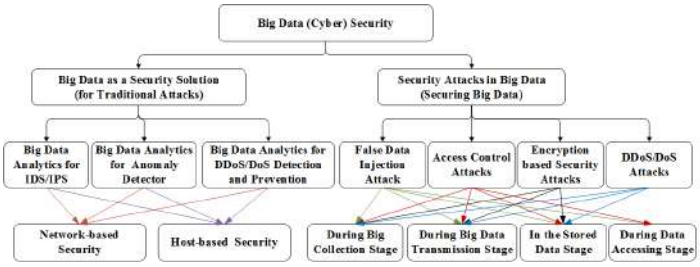


Figure 3. Big data (analytics) as a security solution and security attacks that are unique to big data in a typical big data enabled systems.

variety represents the fact that the data being generated comes in all types of forms. Big Data could be explained simply as data at rest according to Miloslavskaya et al [2]. They also highlighted the difference between big data, data lake, and fast data. Data lake holds a large amount of raw data in its original format. Fast data can be time sensitive data which may either be structured or unstructured, which is usually acted upon right away.

We have more and more data coming, and they are moving from terabytes to petabytes, which are becoming unfamiliar realms [3]. Thus, we need to find new ways of accommodating this data, and there is the need to develop models and algorithms that will enable us to work on these data, to gain insights from it. This is where Big Data Analytics (BDA) comes in. This paper explores research work done on big data enabled security and securing big data (which are categorically presented in Fig. 3).

Although there are related survey papers [4]–[16] on big data security (further details, please refer to Section IV), we present more up to date approaches, insights, perspectives and recent trends on the rapidly advancing research field of big data in the cybersecurity domain. Our approach to this covers the research work done on how big data is used as a security tool and the emergence of big data as high value asset resulting in research work done on how to secure big data. Specifically, the main contributions of this paper include:

- Presenting a comprehensive study on security aspects of big data by categorizing it into two parts: security using big data and big data driven security.
- Presenting a summary of attacks and countermeasures for big data in a tabular form for a side-by-side comparison.
- Presenting a discussion of research challenges, recent trends, insights and open problems for big data in cybersecurity.

The remainder of this paper is organized as follows. We first classify our work into two major sections (Sections II and III). We provide a comprehensive study of security using big data as well as securing big data. For each category, we present the related recent state-of-the-art literature for the different approaches. Section II focuses on the use of big data as a security mechanism. Section III tackles how big data is being protected. Section IV presents relevant survey papers along the line of this paper and the distinction of this paper from the rest of the surveys. Section V presents some research challenges and future directions in this area. Finally, we summarize the paper in section VI.

II. Security using Big Data

Top security companies joined forces to share information with each other in an attempt to gather intelligence from the shared data (SecIntel Exchange). Their goal was to provide reliable security tools for their clients, and to achieve that, they had to learn as much as possible from evolving threats that were developed each day. They understood the power of collaboration for the greater good. This was needed because with the rise of polymorphic malware and other evolving threats, they needed a lot information on these threats in order to fully understand what they were dealing with and how to counteract against it. The traditional approaches of classifying malware were proving to be futile. SecIntel Exchange data provided them with the opportunity to derive actionable insights from voluminous data. Human analysis and traditional methods such as database storage could however not keep up with the pace of the data that was being generated [17]. There was the need to adopt modern approaches. As seen in a case study conducted by Zions Bancorporation [18], it would take their traditional Security Information and Event Management(SIEM) systems between 20 minutes to an hour to query a month's worth of security data. However, when using tools with Hadoop technology, it would only take about one minute to achieve the same results. As such BDA has become an important tool in cybersecurity. Several studies have shown that the traditional approaches and human analysts can not keep up with the big data. BDA is one of the best solutions to combat these issues.

A. Big Data Analytics (BDA) as a Tool to Combat Diverse Attacks

Aypical attacks that can be subdued using big data analytics are depicted in Fig. 4). "If we know the enemy and ourselves, we need not fear the result of a hundred battles" is an excerpt from the Art of War written by the famous Chinese general, Sun Tzu. In other words, it may not be possible to know enough about our enemy, but it is definitely possible to know all that we can about ourselves and the assets we protect. To do that, we have to gather facts about the asset. This is made possible by the data it generates. This data needs to be analyzed and insights need to be drawn. BDA can help prepare, clean, and query heterogeneous data with incomplete and/or noisy records [19], something that would be hard for humans to do. Analyzing data tends to be hard when the data is heterogeneous as [20] discovered. In their work, they presented a platform targeted at achieving real time detection and visualization of cyber threats which they called OwlSight. The platform had several building blocks (data sources, big data analytics, web services and visualization) and had the ability to collect large amounts of information from a variety of sources, analyze the data and output the

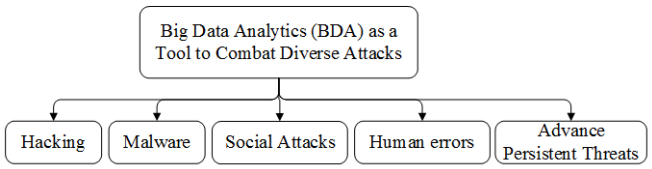


Figure 4. Big data (analytics) can combat diverse attacks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

findings on insightful dashboards. They did face some issues with the heterogeneity of the data. However, for machines to do the work effectively, they need to have some form of human element. Understanding a problem is half the problem solved. The authors in [21] understood this and addressed this issue by coming up with an approach that merged big data analytics with semantic methods with the aim of trying to gain further insights on the heterogeneous data by understanding it semantically. BDA can be used to gather insights making it an essential tool in cybersecurity. However, the features of big data (four V's) also make deriving insights a hard task to accomplish.

In the 2017 Data Breach Investigations Report done by Verizon, it was reported that attacks tend to come from different sources. 62% of the attacks involved hacking, 51% used malware and 43% were social attacks. 14% were a result of human errors. As such, the attacker sometimes relies on human factor in order to execute a successful attack. In such scenarios, people instead of technology become the target of an attack. Email scams and phishing are the most common form of these attacks. In a recent study [22], 52% of successful email attacks get their victims to click within an hour and 30% within 10 minutes. The authors in [23] looked into the role of big data in such attack scenarios. To gain further insights, the authors conducted two studies. The first study involved the Enron email dataset. The second study was carried out on undergraduate students to observe how email phishing broke security systems based on user behaviours. The collected data was then analyzed using Enronic software which was followed by the categorization of email topics. The authors found that, phishers or attackers can understand the behavior of email users using big data analytics, and therefore are able to generate phishing emails that created security threats based on the insights they gathered. The authors planned on proposing a framework for addressing security threat in email communication in the future. In another work, a big data enabled framework was proposed in [24] with the aim of defending against spam and phishing emails by using a global honeynet. Their framework collected data from different sources such as pcap files, logs from a honeynet, black listed sites and social networks for analysis. The framework used Hadoop and Spark for the processing of the collected heterogeneous data which was stored in Hadoop Distributed File System (HDFS). However, this framework does not provide real-time analysis for big data.

Another form of attack is Advanced Persistent Threats (APT) which are sophisticated, well-planned attacks [25]. APTs are very hard to detect, and the challenge of detecting and preventing advanced persistent threats may be answered by using big data analysis. These techniques could play a key role in helping detect threats at an early stage, especially when it uses a sophisticated pattern analysis, that works on different heterogeneous data sources. Given the numerous number of APT attacks that organizations face today, an APT security protective framework has been presented in [26]. The proposed framework integrates deep and 3D defense strategies. To protect against APT attacks, the system classifies data based on the level of confidential data. Botnet attacks is also another area where big data and machine learning techniques are deployed in. The work in [27] studied techniques for mitigating botnet attacks by using big data Analytics. The Advanced Cyber Defense Center (ACDC) orchestrated the sharing of gathered cybersecurity information on botnet attacks with the aim of defending through botnets. The work [28] proposed an architecture to address the current issue of botnet detection. They explored the possibility of employing a Self Organizing Map as an unsupervised learning approach to label unknown traffic. Financial sector is another area where big data analytics is used to prevent malicious actions or cyber attacks. The work in [29] studied using data fusion and visualization techniques in Network Forensic Analysis. Also, Cybersecurity Insurance (CI) is becoming more popular because of the increase of loss mitigation for cyber incidents for financial firms. Big data has now been employed in cybersecurity insurance, and the work in [30] proposed a framework which uses a big data approach in CI to analyze cyber incidents to gain insights in order to make better strategic decisions based on the information gathered. [31] investigated privacy and security issues associated with the sharing of financial data between institutions.

The work in [32] studied a novel Network Functions Virtualization-based (NFV) cybersecurity framework for providing security-as-a-Service in an evolved telco environment. The framework is known as SHIELD. This framework leverages BDA for detecting and mitigating threats in real time. [33] studied the idea of the construction of security monitoring systems for Internet of Things, which is based on parallel processing of data using the Hadoop platform. The proposed systems architecture has different components for the collection of data, storage of data, normalization and analysis, and visualization of data. Storage of data is done on Hadoop to improve the reliability and efficiency of processing of data requests. The work in [34] proposed a Security Information Management (SIM) enhancements using BDA. They devised a blueprint for a big data enhanced SIM, and field tested it using real network security logs. The work in [35] proposed a big data analytics model for protecting virtualized infrastructure in cloud computing. A Hadoop Distributed File system was used for the collection and storage of network logs and application logs from a guest virtual machine. Attack features were then extracted using graph-based event correlation and MapReduce parser identification of the potential paths of attack. A two-step machine learning algorithm using logistic regression and belief propagation were then applied to determine the presence of attacks. SIEM is an important tool in cybersecurity information analytics and a good source of data. The tool developed in [36] analyzes big data (gotten from SIEM) of a Fortune 500 company in order to gain insights about security threats through anomaly detection. They highlight the importance of graph analytics when it comes to intuitively understanding of business needs. Based on this, they apply graph analysis in anomaly detection by adding additional important capabilities of existing tools to their new tool, and then to visualize the network ins and outs. Finally, another use case of big data for security reasons involves a method for analyzing the security of RC4 [37]. Since attacks are diverse and come in multiple

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

forms, BDA has been used as a cybersecurity tool to mitigate those attacks.

An area in cybersecurity where big data is used a lot is in Intrusion detection and prevention systems (IDS/IPS) research. Intrusion attempts are done to usually access information, interfere with the information or to tamper with a system thus making it unreliable and unusable. The IDS concept has been around for two decades but has recently seen a dramatic rise in the popularity and incorporation into the overall information security infrastructure [38]. IDSs are used to determine if there has been a breach or an interference in the network [39]. An IDS is often regarded as a second-line security solution after authentication, firewall, cryptography, and authorization techniques. Similarly, IPS can be classified into two categories: Network-based IPS and Host-based IPS. In network intrusion, prediction and detection is time sensitive, and needs highly efficient big data technologies to deal with problems on the fly [40]. This ensures a proactive rather than a reactive approach to cybersecurity. [41] approached this problem by developing a Proactive Cybersecurity (PCS) system. The PCS is a layered modular platform that makes use of big data collection and processing techniques to a wide variety of unstructured data to identify and thwart cybersecurity attacks. The PCS has a Targeted Vulnerability Predication (TVP) subsystem for detecting threats. Additionally, the model makes use of an Architectural Vulnerability Detection (AVD) subsystem and a risk analysis and recommender (RAR) subsystem for aiding identification and analysis of the identified risks (e.g. [16]). The work in [42] also proposed an architecture that handles IDS/IPS issues in a network. Their architecture stores and manages data from heterogeneous sources and also tries to find insights in the data. DNS data, NetFlow records, HTTP traffic and honeypot data were used in the research. Their approach however only provides offline analysis. Yang [43] presented an alternate approach that detects network anomaly at per-flow level rather than the usual per packet level which tends to bring scalability issues. They build a meta model for a number of machine learning and data mining algorithms. [44] also proposed a network security and anomaly detection framework for the big data systems for Network Traffic Monitoring and Analysis (NTMA) applications. Their framework is known as Big-DAMA. Big-DAMA is a very flexible Big Data Analytics framework (BDAF) that can perform analysis and storage of huge amounts of both heterogeneous structured and unstructured data. Big-DAMA also has batch and stream processing capabilities. Additionally, Big-DAMA utilizes Apache Spark Streaming for stream based analysis and for batch analysis, it uses Spark. For query and storage, it uses Apache Cassandra. Several machine learning algorithms are implemented by Big-DAMA for anomaly detection and network security. Big-DAMA was applied to various network attacks and anomaly detection. It was found to have the ability to speed computations by a factor of 10 in comparison to Apache Spark cluster. Security monitoring using big data has also been extended to other avenues. The work done in [45] also propose a Machine Learning model for Network-based Intrusion Detection Systems in order to detect the network security threats. Different types of ML classifiers are built using data-sets containing the labeled instances of network traffic. The focus of this research was to detect Android threats and give awareness and popularity to the users. This model can be integrated with traditional detection systems to detect advanced threats and reduce false positives. Thus, machine learning models are an essential part of BDA and have especially been used extensively in network anomaly detection.

B. Machine Learning (ML) in Cybersecurity

BDA and machine learning models go hand in hand. To provide security by deriving actionable insights, ML algorithms are needed to learn from the data. ML algorithms fall broadly into three categories: supervised learning, unsupervised learning and semi-supervised learning (which is a combination of supervised and unsupervised learning). The primary differentiator between supervised and unsupervised learning lies in the nature of the data that each uses. Unsupervised learning algorithms are used on data in which the outcome of each training sample is not known. A classic example is in malware detection. To achieve this, we extract the features from malware dataset and find groupings or similarities of the malware. The model uses the features of the data set to find its own groupings. Techniques that are used for unsupervised learning malware analysis are usually clustering algorithms and Principal Components Analysis (PCA). Supervised learning algorithms are trained on data in which the outcome of each training sample is already known. Some techniques used for doing supervised learning are linear and logistic regression, support vector machines, random forests and neural networks which are have commonly been re-branded as deep learning. Deep learning algorithms are very useful for analyzing large amounts of unsupervised data with high variety, which gives it potential in analyzing network data for intrusion detection, especially when it comes to NIDS [46], [47]. [48] tackled this issue when they used a deep learning technique called Self-taught Learning(STL) on the NSL-KDD dataset for intrusion detection on a network.

However, deep learning has some challenges in big data [49]. Its adaptability can be used as a vulnerability when attackers exploit the Machine Learning models. Adversarial examples [50] are machine learning inputs specifically designed to trick the ML model into producing a different output. Various works that have been done on this area try to refine the models [51]. [52] however propose a different approach to detecting adversarial examples. This approach is called feature squeezing which involves the reduction of the search space available to an adversary by merging samples that correspond to many different feature vectors in the original space into a single sample. With the advancement of Generative Adversarial Networks and big data, attackers are using artificial intelligence to circumvent some of the machine-learning automated processes. In lieu of this, a more effective approach is the merging of human and machine elements. Vimod [53] proposed an approach were humans and machine collaborate together. They used high-functioning autistic graduates with specific attributes to monitor networks and network flows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

The other work [54] that incorporated the use of big data to assist humans studied data triage, and how helpful it is in identifying true attack patterns in a noisy data. This approach tries to automatically generate data triage automaton by tracing the actions of security analysts. This approach is different from existing data triage automaton like SIEM, because unlike SIEM, which requires analysts to manually generate event correlation rules, their approach mines data triage rules out of cybersecurity analysts' operation traces. It can be sheen that attackers are using artificial intelligence to trick ML models. Human and machine working together is one of the effective ways to combat these attacks.

III. Securing Big Data

Previous section presented how security can be achieved with big data. This section presents how to secure big data against different attacks. Typical techniques for securing big data are shown in Fig. 5. When data gets really big, securing it becomes really difficult. In [58], authors studied the security issues associated with big data and cloud computing. They identified the fact that most organizations outsource database in the form of big data in to the cloud. Cloud computing however still has many risks associated with it. The goal in [58] was to find security vulnerabilities in the cloud in order to inform vendors about recent vulnerabilities. They noted that confidentiality, integrity and availability in that order as the most important security issues a cloud provider faces. Confidentiality in this scenario would mean the protection of data against unauthorized interference or usage. Integrity would be the prevention of unauthorized and improper data modification. Availability would be akin to data recovery from hardware, software and system errors, and also from data access denials. However, confidentiality is the most important aspect when it comes to big data protection. Several data confidentiality techniques exist with the most notable ones being access control and encryption.

A. Access Control and Encryption Techniques for Big Data

Encryption and access control are similar in the sense that they are both synonymous with privacy and prevention. A notable difference however is that, encryption usually deals with the confidentiality of data. Data can be available to either a trusted or untrusted entity. Encryption ensures that only authorized trusted entities can view the data. Access control however tries to limit the access to data. The data limitations usually happens amongst trusted parties. For this reason, encryption techniques have to be stronger than access control techniques. Encryption imposes very strong limitations over data confidentiality. However, encryption is not an easy task. It tends to be computationally expensive and it has

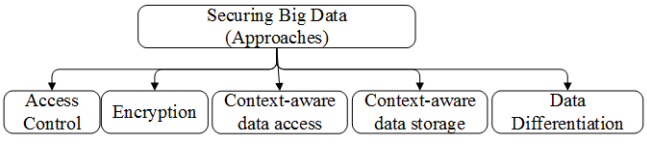


Figure 5. Securing the big data.

scalability issues (many users requiring access to the same data). Access control tends to be more flexible, and is easier to implement. When Big Data is transmitted to the cloud, a security issue emerges. Most organizations would not want their data in the hands of another organization, thus the need for encryption. A common approach is the use of data masking schemes. When the data is transmitted, it is not encrypted because the approaches used to transmit the data requires that the data be decrypted. This exposes the data to attacks. Confidentiality breach is the biggest threat to big data thus the encryption could be used as the primary big data protection technique.

In [59], authors studied the data transmission issues. They proposed computing on masked data to solve this. They proposed an incremental work to improve upon the already existing Fully Homomorphic Encryption (FHE) and other data masking techniques by decreasing the over head associated with other FHE techniques. The work in [60] also tried to improve on a Fully Homomorphic Encryption scheme for big data. It attempted to do this by reducing the public key size with the aim of making their scheme more efficient. The work in [61] proposed a model for the protection of data privacy using a fully homomorphic non-deterministic encryption. The proposed data protection model ensured the prior encryption of data before it was transmitted and therefore avoidance of the loss of data. The proposed system however only accepted numerical data. The output from the system was a result gotten from the computation of the encrypted data which is similar to that of the plaintext. In the future, the authors will look into the improvement of this anonymized data protection approach. The work in [62] explored the use of Format-Preserving Encryption (FPE) data masking scheme for voluminous data. The approach chooses various FPE algorithms depending on the type of data and what needs to be done. Spark framework was used. The authors chose FPE encryption technique because the ciphertext of FPE will still retain the original format of the plaintext instead of unreadable binary string. The ciphertext will now not contain any sensitive information. This approach however has its drawbacks. The encryption speed is slow when compared to other traditional symmetric algorithms. In one FPE method call, the algorithm calls the block cipher many times thereby making it inefficient. Another commonly used encryption is Attribute Based Encryption (ABE). The work in [63] presented a framework for fine-grained data access control to Personal Health Records (PHR) in the cloud that uses Attribute-Based Encryption as an encryption method to ensure that each patient has a unique key based on his/her attributes. The data could be accessed under multi owner settings. It was not only free of errors, but also protected the data from malicious parties aiming at deceiving the data users. In another paper involving ABE, Yang et al [64] addressed some of the shortcomings of ABE encryption in cloud data storage services. They proposed a variant of ABE which is a novel distributed, scalable and fine-grained access control scheme based on the classification attributes of the cloud storage object. Their goal was to improve on the shortcomings of ABE by taking into account the relationships among the attributes. The work in [65] investigated a hybrid approach

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

Table I
SECURING BIG DATA

Method/reference	Goals	Source of Data	Tools/Technologies
Security threats for big data [23]	Mitigating Phishing Attacks	Enron E-mail Dataset	Enronic Software
A big data architecture for security data [24]	Defend Against Spam and Phishing	pcap files, logs from honey net	Hadoop, Spark
Data mining methods for detection of malicious executables [55]	Detect malicious malware	Malicious and benign executable binaries	Machine Learning and Data Mining Algorithms
A practical solution to improve cybersecurity [53]	Security monitoring tool	Network dataset	Data Mining Techniques, High Functioning autistic graduates
Automate Cybersecurity Data Triage [54]	Help security analysts with data triage	The operation traces of security analysts on IDS logs and Firewalls	Data modeling and mining Techniques, Humans
Analyzing and Predicting Security Event Anomalies in BDA Deployment [36]	Improve SIEM by adding important features.	Traditional SIEM systems	Data Mining, Graph Analytics
Network Information Security on Big Data [26]	Advanced Persistent Threat Detection	Network data set	Big Data Analytics, Network event collection techniques, Big Data correlation analysis
Big Data machine learning and graph analytics [56]	Combining batch and stream data processes for efficiency reasons	Hetereogeneous Big Data (any type of data)	Lambda architecture
SIM in light of Big Data [34]	Cyber attack detection	Security logs	Machine learning techniques
Data fusion & visualization [29]	Network forensic investigation	Network logs	Data fusion techniques, Visualization, Self Organizing Map
Owlsight: Platform for real-time detection and visualization of cyber threats [20]	Real time detection and visualization of threats	Heterogeneous network data	Big Data Analytics, Web services, Data visualization
Predicting and fixing vulnerabilities before they occur: a Big Data approach [41]	Proactive Defense (Prevention better than cure approach)	Heterogeneous network data	Big Data Analytics techniques, Machine Learning
Machine learning classification model [45]	Network Intrusion Detection System in Android phones	Android data	Machine Learning Algorithms
A Big Data architecture for large scale monitoring [42]	Intrusion detection and prevention systems	NetFlow records, HTTP traffic and honeypot data	Shark, Spark, Machine Learning algorithms
A Scalable Meta-Model for Big Data Security Analysis [43]	Detect network anomaly at per flow level rather than the usual per packet level which tends to bring scalability issues	Network data	Machine learning and Data Mining Algorithms
Network security and anomaly detection [44]	Intrusion Detection System	Network flow Data	Spark, Cassandra, Machine Learning Algorithms
SHIELD: A novel NFV-based cybersecurity framework [32]	Security as a Service(SecaaS) to protect applications on Software	Heterogeneous cybersecurity data	Big Data Analytics, Machine Learning
Security evaluation of RC4 using Big Data analytics [37]	Analyzing the security of RC4	RC4 Algorithm	MapReduce, Big Data Analytics
Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing [35]	Big Data analytics model for protecting virtualized infrastructure in cloud computing	Network logs and application logs from a guest virtual machine	Machine Learning algorithms
Big Data security analysis approach using computational intelligence techniques [57]	Deduce the security status of the desktop and sources and causes of security breaches	Log file of Windows Firewall	Computational intelligence techniques
Data analytics on network traffic flows for botnet behaviour detection [28]	Issue of botnet detection	Network Traffic Data	Self Organizing Map as an unsupervised learning approach to label unknown traffic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

that combines symmetric cryptography and ABE to secure big data. They wanted to combine the flexibility of attribute-based cryptography and the efficiency of symmetric cryptography. They use Ciphertext-Policy Attribute-Based Encryption (CP-ABE) and AES encryption. In another form of big data encryption scheme, the work in [66] proposed an encrypted MongoDB which utilizes a homomorphic asymmetric cryptosystem which can be used for the encryption of user data and in achieving privacy protection. Thus, the FPE, FHE and ABE are the more popular researched big data encryption techniques.

A model for encrypting both symmetric and asymmetric data was presented in [67] which sought to overcome the limitation of asymmetric encryption techniques such as key exchange problem and the limited size of data and which in turn made it irrelevant for big data applications. Their proposed technique was known as BigCrypt which uses a probabilistic approach to Pretty Good Privacy Technique (PGP). BigCrypt encrypts the message with a symmetric key and encrypts the symmetric key using a public receiver key which is then attached to the message. The message is then sent. At the receiver end, the symmetric key is extracted and then asymmetrically decrypted and used for decrypting the main message. The proposed model was tested on local, web, and cloud server and was found to be efficient. Furthermore, a framework for securing the sharing of sensitive data on a big data platform was proposed in [68]. Sharing sensitive data securely reduces the cost of providing users with personalized services in addition to providing value-added data services. The proposed scheme secures the distributed data, securely delivers it, stores it, and ensures secure usage. Semi-trusted big data is also destroyed. The proposed scheme uses a proxy re-encryption algorithm that is based on heterogeneous ciphertext transformation. The scheme also utilizes a user process protection method based on a virtual machine monitor that supports other system functions. This framework ensures data security while ensuring it is shared safely and securely. Sharma and Sharma in [69] discussed the protection of big data using neural and quantum cryptography. Neural cryptography incorporates the concept of artificial neural networks with classical cryptographic algorithms while quantum cryptography makes use of the phenomenon of quantum physics for securing communications. The authors also provided a comparative analysis between quantum and neural cryptography based on the methodologies that both techniques employ. From the analysis, the authors showed that a quantum computer makes use of quantum mechanisms for computation which are very powerful and can therefore crack complicated problems such as discrete logarithmic problem in a small duration. Neural key exchange protocol is also shown not to depend on any number theory. The analysis also indicates that neural networks probably have higher protection. The work in [70] proposed an efficient group key transfer protocol necessary for ensuring secure group communication on big data. The proposal does not use an online key generation centers (KGC) which is based on 3-LSSS (Linear secret sharing scheme) in that three modular multiplications are needed. Additionally, the protocol uses Diffie-Hellman key agreement. The proposed group key transfer scheme consists of two sections; two party secret establishment section and a section for the group session key transfer. The proposed group key transfer scheme was analyzed to verify its elements of key freshness, key confidentiality, and key authentication. Furthermore, the work in [71] proposed a new encryption scheme that can be used on big data that uses double hashing instead of a single hash. Double hashing they claim eliminates the threat of known cryptanalysis attacks. The work in [72] discussed primarily about the enhancement of CAST block algorithm for the security of big data. Their contribution to the enhancement of the cast block algorithm involved the use of one S-box instead of 6, and an approach to make it more dynamic. The work in [73] presented a framework that is Light-weight Encryption using Scalable Sketching(LESS) for reducing and encrypting the processing of big data on low power platform. This contains two kernels."sketching" and "sketch-reconstruction". Orthogonal Matching Pursuit (OMP) algorithm is implemented on the domain-specific Power Efficient Nano Cluster platform that acts as a hardware accelerator and ARM CPU for big data processing. Finally, the work [74], discuss the security issues of heterogeneous, multimedia big data. They tackle resource constraint issues such as limited computation and energy resources. They proposed data encryption models that deals with this issue by reducing the computation overload on weak nodes and by replacing the current encryption models with an improved version based on SAFE encryption scheme to improve it. The work in [75] mentioned a new approach for the privacy and security protection of clinical data through the use of the art encryption scheme and attribute based authorization framework.

For the access control and privacy of big data, the work in [77] presented a hybrid approach based framework that composes and enforces privacy policies to capture privacy requirements in an access control system. Gao et al [78] presented a cloud security control mechanism based on big data. Cloud computing was observed to have increased the amount of data in the network. Due to this, big data leaks and losses occurred. Therefore, there was the need to provide the necessary level of protection. To that end, they conducted an analysis on big data, analyzed the current big data situation. Gupta et. al. [79] proposed a security compliance model for big data systems. The model provides security and access control to big data systems at the initial stage. The proposed system has four models; the library, low critical log, high critical log, and a self-assurance system. The design of this system ensures real time analysis of big data. The initial level of security provided by the model is facilitated by its web directory and its self-assuring framework that identifies and differentiates genuine users and critical users. The relationship analysis tool of the users blocks users who are deemed not to be genuine. In [76], the authors proposed a framework for privacy policy for big data security. The proposed framework makes use of different techniques including security policy manager, fragmentation approach, encryption approach, and security manager. The characteristics of the privacy policy required flexibility, integration, customizability, and context-awareness. The framework works by receiving data from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

Table II
RESEARCH ON ACCESS CONTROL AND ENCRYPTION TECHNIQUES ON BIG DATA

Method/reference	Problem	Solution
Computing on masked data: a high performance method for improving Big Data veracity [59]	Data not encrypted during Transmission	Improving on FHE by decreasing overhead
A faster fully homomorphic encryption scheme in Big Data [60]	Data not encrypted during Transmission	Improving on FHE by reducing public key size
A Data Masking Scheme for Sensitive Big Data Based on Format-Preserving Encryption [62]	Retain the original format of the plaintext instead of unreadable binary string during transmission	Format-Preserving Encryption (FPE) data masking scheme that chooses various FPE algorithms depending on the type of data and what needs to be done
Big Data Privacy Using Fully Homomorphic Non-Deterministic Encryption [61]	Data Security in the cloud during transmission	Fully Homomorphic non-deterministic encryption
Securing Personal Health Records in Cloud Computing: Patient-Centric and Fine-Grained Data Access Control in Multi-owner Settings [63]	Securing Personal Health Records in the cloud	Attribute-Based Encryption to ensure that each patient has a unique key based on his/her attributes
A Fine-Grained Access Control Scheme for Big Data Based on Classification Attributes [64]	Shortcomings of ABE encryption in cloud data storage services	Improve on the shortcomings of ABE by taking into account the relationships among the attributes
A digital envelope approach using attribute-based encryption for secure data exchange in IoT scenarios [65]	Improving Big Data Security	Better security by combing the flexibility of attribute-based cryptography and the efficiency of symmetric cryptography
CryptMDB: A practical encrypted MongoDB over Big Data [66]	Encryption of user data and in achieving privacy protection	Encrypted MongoDB which utilizes a homomorphic asymmetric cryptosystem
BigCrypt for Big Data encryption [67]	Overcome the limitation of asymmetric encryption techniques	BigCrypt(uses a probabilistic approach to Pretty Good Privacy Technique)
A Multi-level Intelligent Selective Encryption Control Model for Multimedia Big Data Security in Sensing System with Resource Constraint [74]	Security issues of heterogeneous, multimedia Big Data under resource constraints	Proposed a SAFE encryption scheme to replace old encryption models
Secure sensitive data sharing on a Big Data platform [68]	Securing the sharing of sensitive data on Big Data platform	Used proxy re-encryption algorithm based on heterogenous ciphertext transformation
Big Data protection via neural and quantum cryptography [69]	Protecting data	Using neural and quantum cryptography
Novel group key transfer protocol for Big Data security [70]	Secure group communication on Big Data,	Efficient group key transfer protocol using Diffie-Hellman key agreement
Double-Hashing Operation Mode for Encryption [71]	Cryptanalysis attacks	Used double hashing instead of a single hash
Enhancement CAST block algorithm to encrypt Big Data [72]	Enhancement of the cast block algorithm,	Use of one S-box instead of 6 to make it more dynamic
Less: Big Data sketching and encryption on low power platform [73]	Reducing and encrypting the processing of Big Data on low power platform	Light-weight Encryption using Scalable Sketching
Policy enforcement for Big Data security [76]	Privacy policy for Big Data security	Analyzes data, extracts the privacy policies, identifies sensitive data, then fragmentation algorithm executed on sensitive data
Managing the privacy and security of e-health data [75]	Privacy and security protection of clinical data	Art encryption scheme and attribute based authorization framework

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

customer and then analyzing it. It is then followed by the extraction of the privacy policy and finally the identification of sensitive data. Once sensitive data has been identified, a fragmentation algorithm was executed on the sensitive data. The security modules play the role of identifying sensitive data from non-sensitive data and then regulating its access. The work in [80] proposed a privacy protection technology and control mechanism for medical big data. The proposed framework has four main phases; the setup phase, Encrypt and Upload phase, Download phase, and Share File phase. The system first de-identifies the patient personal privacy data, encrypts it using digital signature mechanisms to protect data confidentiality and the authentication of the data. The communication security of the data in the system is protected using the Diffie-Hellman session key while the integrity of the medical records is protected using a digital signature scheme. Access control is not as big as it used to be due to the evolving threats landscape but is still an important research area in big data security today.

B. Alternative Approaches to Securing Big Data

Encryption and access control were the mainstream approaches for big data security. However, researchers have tried other approaches that may or may not involve some form of encryption. The nature of big data makes it difficult to protect everything. Some researchers have tried to determine the important parts of big data to protect those parts only. The work in [81] tried to tackle the issue of securing personal health records by proposing a framework that classifies data based on a person's societal importance and determining the sensitivity levels of the data. Furthermore, [82] tried to secure the attributes of big data that are really important/valuable because protecting everything is a difficult task. They use data masking to protect these high valued attributes. To determine the attributes that are of value, they use a ranking algorithm that prioritizes attributes for big data security. Authors in [83] proposed an attribute selection method for protecting the value of big data by determining attributes that have higher relevance using a ranking algorithm, and providing security measures. In the paper [84], the authors focused on the characteristics of big data and proposed the protection of big data using a security hardening methodology that makes use of attribute relationships. The relationship between the various attributes are expressed using nodes and edges. The proposed model works by limiting the attribute to protect value. The model works by first extracting all the attributes of the targeted big data. The nodes are then arranged circularly followed by the establishment of the relationship between the nodes. The relationship is then set based on either the domain specific criteria or the universal criteria. Finally, the protecting nodes are selected followed by the determination of how to protect the selected nodes. Thus, protecting everything in big data is hard. An easier approach is to find what is important and protect that part only.

Encryption has been used with other techniques as well. The work in [85] proposed a method to secure Multimedia Big Data (MBD) in the healthcare cloud by using a Decoy Multimedia Big Data (DMBD). The DMBD uses fog computing and a pairing based cryptography that will be used to secure the MBD. Fog Computing was utilized for the storage of the decoy files. In their method, the decoy files are retrieved at the onset unlike other methods that usually waits until there is an attack before the decoy files are called. Thus, both attacker and legitimate users both see a decoy file until the legitimacy is confirmed. Aynur in [86] presented a new technique for securing big data in medical applications. The methodology combines three major techniques that include data hiding, image cryptography and steganography. These techniques facilitate safe and de-noised transmission of data. A stream cipher algorithm is used for encrypting the original image. Patient information is then embedded in the encrypted image by means of a lossless data embedding technique together with a key for hiding data to enhance the security of data. Steganography is then applied in embedded image with a private key. When the message gets to the receiver, it is decrypted using inverse methods in reverse order. Efficiently securing big data continues to become a difficult challenge because of big data's variety, volume and veracity issues. The ability to deal with space and time issues by correlating events would play an important role in securing big data. [87] discussed the growth of social media network such as Facebook and cloud computing, and how sharing of multimedia big data has become easier than ever. However, its increased used is faced with issues of piracy problems, illegal copying, and misappropriation. To address these challenges, the authors in this study proposed a system for protecting multimedia big data distribution in social networks. The scheme utilizes a Tree-Structured Harr (TSH) transform. In this scheme, a homomorphic encrypted domain for fingerprinting by means of social media network analysis is applied. The scheme aims at mapping hierarchical social networks into trees structure of the transform of TSH for coding, encrypting, and fingerprinting of JPEG2000. Finally, in [88], authors discussed the use of traditional security framework for protection of the smart grid comes with several disadvantages such as late detection of attacks when damage has already occurred. To address this problem, the authors in this study proposed a security awareness mechanism based on the analysis of big data in the smart grid. The model has three main parts which include the extraction of network security situation factors, network situational assessment and network situational prediction. The method works by integrating fuzzy cluster based analytical tools, reinforcement learning and game theory. The integration of these components facilitates security situational analysis in the smart grid. Simulation tests and experiments showed the proposed system to have high efficiency and low error rate.

Sometimes, we have to protect data from the people and the systems that interact with it. Pissanetzky [89] examined the problem of software vulnerability and the accumulation of unprocessed information in big data. According to the authors, these problems are created by human interventions. To solve these problems, the author proposed the complete elimination of human intervention. In this approach a causal set was taken as the universal language of all information and computations. Additionally, the author also proposed the confinement of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

use of programming languages to the human interface and therefore a creation of an inner layer of mathematical code that is expressed as a causal set. Furthermore, this paper also includes experiments and computational verifications of the theory and proposed applications of this approach to science and technology, computer intelligence, and machine learning. Also, [90] researched on how to protect both the data and the program that processes the data while taking into consideration the big data processing requirements. They propose a model that aims to address the issue by hiding operations performed using steganography and FHE in order to meet the security requirements necessary to protect outsourced data. However, the user's computation cost is somewhat high and the solution does not apply to all applications. The work in [91] addressed the use of cloud computing and how it provides an organization with various services for meeting their various needs. However, data storage in cloud computing could be accessed by cloud operators and therefore compromise information privacy and security. In this respect, this study proposed an approach for splitting and separating the stored data on distributed cloud servers and therefore prevent access by cloud operators. The proposed model was known as Security-Aware Efficient Distributed Storage (SA-EDS) and was based on two algorithms; the Efficient Data Conflation (EDCon) algorithm and Secure Efficient Data Distributions (SED2) algorithm. These algorithms were tested and proved to be efficient. The authors of [92] proposed a Field Programmable Gate Arrays (FPGA) based solution for running BLAST algorithm in a secure manner in MapReduce framework using cloud computing. The proposed system protects data from cloud service provider (CSP) through leveraging on bitstream encryption mechanism and FPGAs tamper resistant property. The authors also put into consideration the risks that arise from keys distribution and propose countermeasures for handling it. The work in [93] studied an approach that assesses the risk behind various applications and provides an explanation of the ability of the application to protect data using a specific security classification level. The proposed method has three main components; Automatic Risk assessment of the Application, Automatic Generation of Criteria for storage of specific data, and Automatic Reporting. The report facilitates the recommendation of the appropriate security level. The work in [94] proposed a hadoop system that would both secure and maintain the privacy of big data. They tried to do this by using four encryption techniques randomly. However, these encryption techniques are time consuming, thus they proposed a buffer system where the buffer stores information whilst the system works on the previous data stored in order to prevent information loss.

Knowing the characteristics of the data is an important aspect of protecting the data. Singh [95] studied the value of real-time BDA and the security challenge that comes with protecting big data. Singh notes that, proper protection of big data should focus on volume, velocity, and variety of big data. Multilevel security for big data should be provided at the application, operating systems, and network levels. However, using the traditional protection mechanism is challenging for large volumes of data that is changing continuously. For this reason, Singh recommends the use of machine learning for protection of big data with focus on supervised and unsupervised learning. Yang [96] examined the visualization of network security under the big data environment. The authors first look at the 5V characteristics of big data including volume, velocity, variety, value, and visualize. These 5V features are then mapped onto network security data followed by a description of the visualization of the data security technology. The network visualization technology proposed include the use of radial traffic analyser and SRNET. They also proposed safety visualization using ClockMap and discussed diversified technologies for visualization of big data. With the increasing volume of big data, security and privacy issues also continue to increase. Peer to peer (P2P) protocols such as BitTorrent are now being used to widen the transfer of big data. However, this increase has also attracted widespread security challenges. Research indicate that P2P are sophisticated in data transfer but experience challenges when distributing big data. Ban et. al. [97] presented a study on the early identification of attacks using the darknet. The system works by first exploring the regularities in communications from the attackers. This is achieved using an itemset mining engine. It then characterizes the activity level of each pattern of attack creating a time series. A clustering algorithm is then applied to extract the most prominent patterns of attack. The attack patterns are clustered into groups having similar activities. Visual hints on the relationship of the various attacks is then provided using a dimension reduction technique. Attacks that feature prominently are then the picked up for further analysis by experts. The authors showed that the proposed system was efficient in early attack detection.

The union of blockchain and big data will make sure that the data that is generated from the blockchain is trustworthy. This is because the provenance of the data is known. Also, the likelihood of the data being interfered with is very low. This is made possible through the blockchain's consensus mechanism and its secure cryptographic hash function which ensures data immutability. Data manipulation would require tremendous amount of hash power in order to be achieved. The centralized way of storing data is prone to data breaches and hacks [109]. This method is susceptible to single point of failure of problems as well. Distributed data storage tries to take data away from the hands of these centralized authorities, thereby taking away various security risks. The work in [106] proposed a model for security sharing based on blockchain technology to address trust issues often associated with circulation of data. The proposed model provides a credible platform for sharing data between data producers and demand parties though building a decentralized security system for the circulation of data. The security system is built using blockchain and smart contract. While blockchain technology ensures the traceability of data, the automated execution of smart contract provides the security for data security sharing. The decentralized architecture ensures the data provider does not suffer from the risks of sharing data from a centralized storage system. On the user's side, transparency in the collection of information is assured by the blockchain operation model and thereby bringing stronger user privacy protection. [110]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

Table III
RESEARCH ON ALTERNATIVE APPROACHES TO SECURING BIG DATA

Method/reference	Problem	Solution
A framework for providing security to Personal Healthcare Records [81]	Securing personal health records	Framework that classifies data based on societal importance and sensitivity levels
A novel data security framework using E-MOD for Big Data [82]	Securing important attributes of Big Data	Ranking algorithm to determine attributes and data masking to protect them
A Security Model for Preserving the Privacy of Medical Big Data in a Healthcare Cloud Using a Fog Computing Facility With Pairing-Based Cryptography [85]	Securing Multimedia Big Data (MBD) in the healthcare cloud	Use fog computing to store Decoy Multimedia Big Data (DMBD)
A space-and-time efficient technique for Big Data security analytics [98]	Space and time issues of Big Data	Bloom filter and its variants
Another Look at Secure Big Data Processing: Formal Framework and a Potential Approach [90]	Protecting both the data and the program that processes the data	Hiding operations performed using steganography and FHE
Attribute relationship evaluation methodology for Big Data security [83]	Attribute selection method for protecting the value of Big Data	Determining attributes that have higher relevance using a ranking algorithm
Real time Big Data analytic: Security concern and challenges with Machine Learning algorithm [95]	real time Big Data analytics and its security challenge,	use of machine learning for protection of Big Data
Research on Network Security Visualization under Big Data Environment [96]	Visualizating network security under the Big Data environment	Use of radial traffic analyser and SRNET
Secure and private management of healthcare databases for data mining [99]	Secure and private management and mining of data in health care	Executing SQL queries on encrypted data and the return differentially-different answers on the outsourced databases
Secure Distribution of Big Data Based on BitTorrent [100]	Security issues accompanying P2P Big Data transmission avenues	Scheme for secure and efficient distribution of Big Data on BitTorrent networks using bittorrent protocols
Secure multimedia Big Data sharing in social networks using fingerprinting and encryption in the JPEG2000 compressed domain [87]	Protecting multimedia Big Data distribution in social networks	Homomorphic encrypted domain for fingerprinting by means of social media network analysis
Security in Big Data of medical records [86]	Securing Big Data in medical applications	Data hiding, image cryptography and stenography
Security-aware efficient mass distributed storage approach for cloud systems in Big Data [91]	Data storage in cloud computing could be accessed by cloud operators and therefore comprise information privacy and security	Splitting and separating the stored data
Security-as-a-service in Big Data of civil aviation [101]	Data protection and privacy preserving services architecture in civil aviation	Civil aviation security data authentication through OpenSSL identity and attribute-based authorization
Towards Early Detection of Novel Attack Patterns through the Lens of a Large-Scale Darknet [97]	Early identification of attacks using the darknet	Itemset mining engine to explore regularities in attack, then machine learning algorithms (clustering) to determine attack patterns and predict attacks
Big Data analysis based security situational awareness for smart grid [88]	Disadavantage of using traditional security framework for protection of the smart grid	Security awareness mechanism based on the analysis of Big Data in the smart grid
Big Data security hardening methodology using attributes relationship [84]	Protection of Big Data using a security hardening methodology	Makes use of attribute relationships to achieve it
On the Future of Information: Reunification, Computability, Adaptation Cybersecurity, Semantics [89]	Problem of software vulnerability and the accumulation of unprocessed information in Big Data	Complete elimination of human intervention

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

Method/reference	Goal	Solution
Privacy preserving large scale DNA read-mapping in MapReduce framework using FPGAs [92]	Running BLAST algorithm in a secure manner in MapReduce framework using cloud computing	a Field programmable gate arrays (FPGA) based solution and a bitstream encryption mechanism
Efficient privacy-preserving dot-product computation for mobile Big Data [102]	Secure privacy-preserving scheme in mobile Big Data	Privacy-preserving dot product
Privacy-Preserved Multi-Party Data Merging with Secure Equality Evaluation [103]	Merging of encrypted data	Data anonymization technique that ensures privacy in the collection and merging of data and secures multiparty sharing of data without the involvement of third parties
Proposition of a method to aid Security Classification in Cybersecurity context [93]	Managing security classification	Assessing the risk behind various applications and providing an explanation of the ability of the application to protect data using a specific security classification level
Toward a cloud-based security intelligence with Big Data processing [104]	Cloud based security intelligence system for Big Data processing	Highly scalable plugin based solution that monitors Big Data systems in real time and therefore reducing the impact of attacks or threats on a distributed infrastructure
Research about New Media Security Technology Base on Big Data Era [105]	Security threats of the new Big Data in digital era	"Blocking as loose" technology for protection, intelligent cleaning of new media Big Data, and mining of Big Data in a safe manner.
Big Data Model of Security Sharing Based on Blockchain [106]	Model for security sharing based on blockchain technology to address trust issues often associated with circulation of data	blockchain and smart contract
Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies [107]	effective vulnerability management for organizations dealing with Big Data	proactive Big Data vulnerability management model based on rigorous statistical models with the capability of simulating anticipated volume and dependence of vulnerability disclosures
New approach for load rebalancer, scheduler in Big Data with security mechanism in cloud environment [108]	Rebalancing and scheduling of loads in Big Data environment,	Proposed scheme uses a load balancing algorithm that merges with MD5 and DES encryption algorithm
Hadoop eco system for Big Data security and privacy [94]	Secure and maintain the privacy of Big Data	Four encryption techniques. Using a buffer system where the buffer stores information whilst the system works on the previous data stored in order to prevent information loss

proposed a system called MeDShare which is a blockchain based and provides data source auditing, and control for shared medical data in cloud repositories. The MeDShare system helps to transfer and share data from one source to another, and are recorded in a tamper-proof manner. The marriage of blockchain and Big Data is imminent as blockchain ensures data integrity.

The work in [102] proposed a secure privacy-preserving scheme using dot product in mobile big data. Privacy-preserving dot product has been used in data mining for a long time as it helps in curbing statistical analysis attacks. It is now being used in big data for its anonymous private profile matching. The paper was just an exploratory research on its use in mobile big data. There is however still room for further improvement. The work in [103] explored the idea of a data anonymization technique to support merging of encrypted data. The technique ensures the protection of privacy in the collection and merging of data and secures multi-party sharing of data without the involvement of third parties. The merging result as proposed in this study does not lead to the violation of the privacy of the individual. Additionally, the proposed mechanism allows for storage of different datasets from different parties in multiple third-party centers without leaking the identity of owners of that data. The anonymized data can be joined securely within a reasonable time. Experiments conducted by the authors indicated that 100,000 entries of data can be merged in about 1.4 seconds using the optimized secure merging procedure. To answer the question of how security classification can be managed on a system. In addition, the work in [104] proposed a cloud based security intelligence system for big data processing. The authors provide a highly scalable plug-in based solution that monitors big data systems in real time and therefore reduced the impact of attacks or threats on a distributed infrastructure. The solution proposed here was named Advanced Persistent Security Insights System (APSIS). APSIS works by taking advantage of a SIEM system including aggregation, correlation, alerting, and forensic analysis. This is exposed to big data but with security intelligence to provide accurate results. APSIS monitors all devices on the network that generate log files and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

therefore assures security. In the future, the authors aim at exploring the proof of concept to evaluate the robustness of the proposed architecture. The work in [105] started by looking at security threats of the new multimedia heterogeneous big data. The first threat was lack of effective mechanisms for the protection of this new media ownership as DRM is facing challenges. Secondly, there is lack of a clean environment for the consumption of new media. To overcome these challenges, Lu proposed the use of "blocking as loose" technology for protection, intelligent cleaning of new media big data, and the mining of big data in a safe manner. [98] summarized how bloom filter and its variants are used to secure big data. After various experiments, they concluded that, bloom filter can be used for efficiency reasons because there are space and time issues when it comes to analyzing and indexing big data which would in turn lead to better security analytics. The research work in [99] proposed a framework for secure and private management and mining of data that addresses both security and privacy issues in health-care data management especially in outsourced databases. The solution works by executing SQL queries on encrypted data and returning deferentially-different answers on the outsourced databases. Laplace mechanism are used to illustrate the computation of private queries. Private decision tree learning is also discussed. An experimental evaluation of the proposed solution shows the system incurs small communication and computation overhead. For this reason, the authors in this study [100] proposed a scheme for secure and efficient distribution of big data on BitTorrent networks. The proposed scheme is built inside the BitTorrent protocol and thus allowing the servers to regulate and trace user's behavior and sensitivity of data.

IV. Existing Surveys on Big Data in Cybersecurity

Bertino [4] presented the security and privacy issues for big data concerning the confidentiality, privacy, and trustworthiness. In data confidentiality, the challenges identified were merging large number of access control policies and enforcing control policies in big data sources. Cybersecurity tasks such as user authentication, access control, and user monitoring are noted to be key in identifying threats and stopping them. The author noted that both security and privacy can be achieved by using advanced technologies such as cryptography. Mishra and Singh [5] examined security and privacy challenges associated with big data analytics for protecting database storage and transaction log files, and secure computations in distributed frameworks. The authors in [6] highlighted the benefits of big data analytics and reviewed security and privacy challenges in big data environments using various BDA tools such as Hadoop, MapReduce, and HDFS. Security and privacy challenges associated with big data environments were also listed as random distribution, security of big data computations, and access control. [7] examined big data emerging issues of security and privacy in relation to the use of big data analytic tools such as Hadoop. The work in [8] presented a review of big data security and privacy challenges while storing, searching and analyzing. In [9], the authors conducted a systematic literature review covering security and privacy for big data by categorizing approaches in terms of confidentiality, data integrity, privacy, data analysis, visualization, data format, and stream processing. Miloslavskaya et al. [10] examined the need for Security Operation Centres (SOCs) for organizations that want to achieve the highest protection for their data. The work in [111] looked at security intelligence centres (SICs) for processing of big data. The work in [112] proposed a framework which combined the techniques of security intelligence and big data analytics to support human analysts for prioritization. The work in [113] studied the security issues identified within the field of multimedia applications. In [11], Arora et. al. performed a survey on big data and its security. The work in [114] highlighted the pros of big data, and then later tackles the challenges faced in China. In [115], Zou analyzed major issues associated with big data and especially the breach of personal information, the potential security risks, and the reduction of control rights of users over their personal information.

Mondek et. al. [116] discussed security analytics in this era of big data and the reality of information security. Mahmood and Afzal [12] presented a review of big data analytics trends, tools, and techniques. The study of security analytics is motivated by the inadequacy of existing cybersecurity solutions to counteract cybersecurity attacks associated with big data. Jayasingh, Patra, and Mahesh [117] discussed security issues and challenges that faces security analysts in big data analytics and visualization. In [118], the authors discussed six changes in the information technology sector that they believe will be the game changers for the next 15 years. The work in [13], [14] presented security solutions for the big data in health-care industry. Health-care generates a lot of data from diverse sources and thus making it difficult to analyze. Similarly, in [119], Patil and Seshadri presented security and privacy issues in big data relating to the health-care security policies. The work in [120] summarized the current health-care security scenarios in big data environments in the USA.

The work in [15] put forward a model of big data security service for data providers, users, and cloud service providers. The work in [121] looked at opportunities, challenges, and security concerns associated with the use of big data in cloud computing. Furthermore, the work in [122] proposed integrated auditing for securing big data in the cloud. The authors presented their study by reviewing the characteristics of big data and security challenges in the cloud. The works in [123]–[125] proposed a security measure for big data, virtualization, and the cloud infrastructure and cloud based big data storage systems. Big data is making its way in the power industry. Smart grid has unique characteristics peculiar to it. The work in [126], [127] highlighted different articles that discuss the peculiarities of smart grid big data and how to properly handle it. Authors in [128] looked at security issues brought by big data applications in the telecommunication industry and especially associated with mobile network operators. In [129], authors surveyed three different techniques, namely homomorphic encryption, verifiable computation and multi-party computation. They discuss relevant security threats in the cloud, and a computation model that captures a large class of big data uses cases. The work in [130] studied the impact

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2019.2907247, IEEE Transactions on Services Computing

of security measures on the velocity of the big data system. This research found out that encryption is not an obstacle to the fast and efficient big data processing like it was before because of the introduction of new technologies. They recommended Encryption zones to be set as default in HDFS.

The work in [131] discussed the issues and challenges brought about by the big data deluge; data that is too big, too fast, and too diverse to the extent that are incompatible with the traditional database system. Paryasto et al. [132] presented the security challenges brought about by big data management through NIST risk management framework. The NIST SP800-30 framework provides a guide for conducting risk management on data. The work in [133] discussed the quality assurance for security applications of big data. The interest in quality assurance arises from the lack of confidence in the outcomes of big data applications. The risks in big data analytics arises out of lack of quality assurance. The work in [134] studied an on-line Cauchy based Clustering for cyber attack monitoring. In [135], authors classified big data during it's analysis phase in order to determine the security level of the data currently being analyzed. The work in [136] presented the various kinds of efforts that had taken towards for introducing a context-based information extraction using National Security Information Sources(NSIS) that enlist various kinds of knowledge inspired by natural activities of living things. The work in [137] showed the analysis of 79,012 articles that are published from the year 1916 to 2016 that relates to security and big data privacy.

In [138], the security of personal information on social media in this era of big data was presented. The study looked into the current situation of social network consumer privacy protection and attributed the security problem to personal information leakage and database defects. In [139], authors surveyed pre-processing techniques for data mining using conventional methods such as filtering, imputation, and embedding. The work in [140] discussed the challenges that exist in the era of wireless big data. Finally, the authors in [141] looked at ICT (considered to be the carrier of big data) supply chain security and big data. We provide a comprehensive study of recent research results by categorizing security with big data and big data security. In our work, we explore the role of big data in cybersecurity (as a tool and as an asset). We present up to date literature in this area and we highlight current and foreseeable challenges and trends in this field. We make it easier on readers by summarizing the problem each paper tried to solve and how they approached it in a tabular form.

V. Research Trends and Open Research Challenges

From the first ever virus known as the "creeper" and the first anti-virus made to neutralize it known as the "Reaper", the cybersecurity landscape has changed. The largest insider attack that ever occurred/happened for over a 30 year duration (1976 - 2006) and involved a former Boeing employee stealing intelligent info and handing them to China. Another well known insider threat was the Edward Snowden saga which involved the leakage of classified information from the NSA which resulted in the people distrusting the government. After, another major cyber security attack was Yahoo failing to report that the accounts of over 3 billion users have been jeopardized. Fast forward to 2017, the attack landscape is starting to shift again from data breaches to data being held for ransom. Ransomwares demanding payment (through cryptocurrency) condemn users to the erasure of their data if the ransom is not paid in beginning to gain traction (WannaCry and NotPretya Ransomware). The threat landscape is changing [142] and research trends need to change in other to combat these cybersecurity attacks. Typical trends, open research challenges and problems are shown in Fig. 6 and described below.

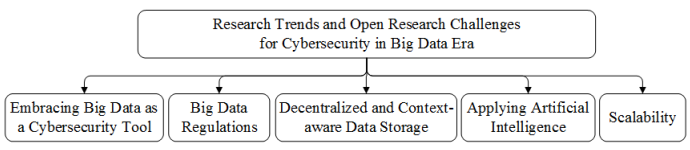


Figure 6. Typical trends, open research challenges and problems.

A. Embracing Big Data as a Cybersecurity Tool

Along with the data generated by IoT devices, the emergence of Bring Your Own Device (BYOD) has made organizations susceptible to various attack vectors. All these devices generate data. Thus organizations are starting to embrace BDA as a tool in their cybersecurity approach. Analyzing the data that passes through the network is essential to protecting the organization. However, some companies still have reservations on employing big data analytics as it tends to be an expensive undertaking. BDA also tends to be a complex field and requires expertise. Furthermore, employees are not comfortable with personal information gathered as this may involve tracking user activity. There are open challenges on how to differentiate the IoT system data, personal data and sensitive data and the protection of each of them using big data analytics.

B. Big Data Regulations

As a result of a myriad data breaches in recent times, new regulations such as Breach of Security Safeguards Regulations in Canada and Europe's General Data Protection Regulation have been implemented. A crucial aspect of the GDPR is the right to be forgotten, which gives an individual the power to enforce the deletion of any information pertaining to him/herself. A research trend we foresee here is self destructing data. Previous work has however been done on this. [143] propose an architecture that aims to solve the issue of personal data privacy. Their research was aimed at protecting the privacy of old data that has been stored on a centralized database which can then be re-used or re-surfaced. Their architecture made sure that copies of such data will become obsolete. This is a research area that might see a lot of growth in years to come, especially due to the emergence of blockchain and decentralized data storage. There are still challenges for big data regulation and policies including the situation where data leaves the organization for cloud storage.

C. Decentralized and Context-aware Data Storage

The most important commodity right now is data. The top companies GAFA (Google, Apple, Facebook, Amazon) have

monopolized data, therefore bringing in the most revenue. New blockchain startups are now basing their business models on how to disrupt these monopolies by highlighting the value of data to the public. The selling point for these startups is that the data stored in a centralized fashion is susceptible to attacks (Facebook, Yahoo, and Equifax hackings) as evident in recent years. A distributed approach to storing data is the safer way to prevent attacks is what is being evangelized. This method is not susceptible to single point of failure problems as well. Companies such as the GAFAs store huge amount of data and they can correctly be termed as data silos. Distributed data storage try to take data away from the hands of these data silos, thereby taking away various security risks. Furthermore, the union of blockchain and big data will make sure that the data that is generated from the blockchain is trustworthy. There are ongoing research and open challenges on decentralized and context-aware data storage for big data.

D. Applying Artificial Intelligence

Artificial Intelligence based polymorphic malware is on the rise. Now, there is an application that can alter malware to trick machine learning antivirus software. In an experiment done by Endgame (a security company), they found out that AI has blind-spots that can be found out by other AI applications. This is evident as seen in Generative Adversarial Networks discovered by google researchers. This shows that organizations should not view machine learning as a fool proof way of defending against malware . More research work is needed in this area because of the rise of GANs. Also, an immediate approach to solve this would be to combine humans and AI in the malware detection approach. AI is not fool proof yet, and we see research trends gearing towards human in the Loop approaches to detect polymorphic malware.

E. Scalability for Cybersecurity Techniques in Big Data era

In big data, protecting everything is hard. The easier approach is to find what is important and protect it. Traditional approaches for securing data might not work in a straightforward way. Thus, finding an optimal approach that is scalable for big data enabled systems is still an active research topic.

VI. Conclusion

In this paper, we have surveyed state of the art literature on big data in cybersecurity. We segmented the work into two parts. The first part was research work involving the use of big data for security purposes. The second part is the research work done on securing big data. We present current trends on the use of BDA as security tool. We also addressed the role of machine learning in this area and some of the challenges machine learning has to overcome before it becomes an important feature in the cybersecurity toolkit. Furthermore, we discussed current literature on techniques used to secure big data. The confidentiality of big data is usually the main focus thus making encryption and access control techniques the main research areas when it comes to big data security. We also discussed the alternative approaches used to secure big data where the proposed approaches rely on other methods than encryption and access control in trying to secure other aspects of the CIA triad. We make it easier on readers by summarizing the problem each paper addresses and their approach to solve it in tabular form. Furthermore, we present future trends in big data security that we foresee, and the challenges associated with it.

References

[1] D. Laney, "3d data management: Controlling data volume, velocity and variety," *META Group Research Note*, vol. 6, no. 70, 2001.

[2] N. Miloslavskaya and A. Tolstoy, "Application of big data, fast data, and data lake concepts to information security issues," in *Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on*, pp. 148–153, 2016.

[3] D. Rawat and K. Z. Ghafoor, *Smart Cities Cybersecurity and Privacy*. Elsevier, December 2018.

[4] E. Bertino, "Big data-security and privacy," in *Big Data (BigData Congress), 2015 IEEE International Congress on*, pp. 757–761, 2015.

[5] A. D. Mishra and Y. B. Singh, "Big data analytics for security and privacy challenges," in *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pp. 50–53, 2016.

[6] Y. Gahi, M. Guennoun, and H. T. Mouftah, "Big data analytics: Security and privacy challenges," in *Computers and Communication (ISCC), 2016 IEEE Symposium on*, pp. 952–957, 2016.

[7] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data emerging issues: Hadoop security and privacy," in *Multimedia Computing and Systems (ICMCS), 2016 5th International Conference on*, pp. 731–736, 2016.

[8] B. Matturdi, Z. Xianwei, L. Shuai, and L. Fuhong, "Big data security and privacy: A review," *China Communications*, vol. 11, no. 14, pp. 135–145, 2014.

[9] B. Nelson and T. Olovsson, "Security and privacy for big data: A systematic literature review," in *Big Data (Big Data), 2016 IEEE International Conference on*, pp. 3693–3702, 2016.

[10] N. Miloslavskaya, A. Tolstoy, and S. Zapechnikov, "Taxonomy for unsecure big data processing in security operations centers," in *Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on*, pp. 154–159, 2016.

[11] S. Arora, M. Kumar, P. Johri, and S. Das, "Big heterogeneous data and its security: A survey," in *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pp. 37–40, 2016.

[12] T. Mahmood and U. Afzal, "Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools," in *Information assurance (ncia), 2013 2nd national conference on*, pp. 129–134, 2013.

[13] S. Rao, S. Suma, and M. Sunitha, "Security solutions for big data analytics in healthcare," in *Advances in Computing and Communication Engineering (ICACCE), 2015 Second International Conference on*, pp. 510–514, 2015.

[14] I. Olaronke and O. Oluwaseun, "Big data in healthcare: Prospects, challenges and resolutions," in *Future Technologies Conference (FTC)*, pp. 1152–1157, 2016.

[15] H.-t. Cui, "Research on the model of big data serve security in cloud environment," in *Computer Communication and the Internet (ICCCI), 2016 IEEE International Conference on*, pp. 514–517, 2016.

[16] E. Damiani, "Toward big data risk analysis," in *2015 IEEE International Conference on Big Data (Big Data)*, pp. 1905–1909, 2015.

[17] C. Sinclair, L. Pierce, and S. Matzner, "An application of machine learning to network intrusion detection," in *Computer Security Applications Conference, 1999.(ACSAC'99) Proceedings. 15th Annual*, pp. 371–377, 1999.

[18] E. Chickowski, "A case study in security big data analysis," *Dark Reading*, vol. 9, 2012.

[19] M. C. Raja and M. A. Rabbani, "Big data analytics security issues in data driven information system," *IJIRCCE*, vol. 2, no. 10, 2014.

[20] V. S. Carvalho, M. J. Polidoro, and J. P. Magalhães, "Owlsight: Platform for real-time detection and visualization of cyber threats," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 61–66, 2016.

[21] Y. Yao, L. Zhang, J. Yi, Y. Peng, W. Hu, and L. Shi, "A framework for big data security analysis and the semantic technology," in *IT Convergence and Security (ICITCS), 2016 6th International Conference on*, pp. 1–4, 2016.

[22] ProofPoint.com, "The human factor report people-centered threats define the landscape," 2018.

[23] T. Zaki, M. S. Uddin, M. M. Hasan, and M. N. Islam, "Security threats for big data: A study on enron e-mail dataset," in *Research and Innovation in Information Systems (ICRIIS), 2017 International Conference on*, pp. 1–6, 2017.

[24] P. H. Las-Casas, V. S. Dias, W. Meira, and D. Guedes, "A big data architecture for security data and its application to phishing characterization," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 36–41, 2016.

[25] A. A. Cárdenas, P. K. Manadhata, and S. Rajan, "Big data analytics for security intelligence," *University of Texas at Dallas@ Cloud Security Alliance*, pp. 1–22, 2013.

[26] W. Jia, "Study on network information security based on big data," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2017 9th International Conference on*, pp. 408–409, 2017.

[27] B. G.-N. Crespo and A. Garwood, "Fighting botnets with cyber-security analytics: Dealing with heterogeneous cyber-security information in new generation siems," in *Availability, Reliability and Security (ARES), 2014 Ninth International Conference on*, pp. 192–198, 2014.

[28] D. C. Le, A. N. Zincir-Heywood, and M. I. Heywood, "Data analytics on network traffic flows for botnet behaviour detection," in *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pp. 1–7, 2016.

[29] H. Fatima, S. Satpathy, S. Mahapatra, G. Dash, and S. K. Pradhan, "Data fusion & visualization application for network forensic investigation-a case study," in *Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on*, pp. 252–256, 2017.

[30] K. Gai, M. Qiu, and S. A. Elnagdy, "A novel secure big data cyber incident analytics framework for cloud-based cybersecurity insurance," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 171–176, 2016.

[31] K. Gai, M. Qiu, and S. A. Elnagdy, "Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 197–202, 2016.

[32] G. Gardikis, K. Tzoulas, K. Tripolitis, A. Bartzas, S. Costicoglou, A. Lioy, B. Gaston, C. Fernandez, C. Davila, A. Litke, *et al.*, "Shield: A novel nfv-based cybersecurity framework," in *Network Softwarization (NetSoft), 2017 IEEE Conference on*, pp. 1–6, 2017.

[33] I. Saenko, I. Kotenko, and A. Kushnerevich, "Parallel processing of big heterogeneous data for security monitoring of iot networks," in *Parallel, Distributed and Network-based Processing (PDP), 2017 25th Euromicro International Conference on*, pp. 329–336, 2017.

[34] F. Gottwalt and A. P. Karduck, "Sim in light of big data," in *Innovations in Information Technology (IIT), 2015 11th International Conference on*, pp. 326–331, 2015.

[35] T. Y. Win, H. Tianfield, and Q. Mair, "Big data based security analytics for protecting virtualized infrastructures in cloud computing," *IEEE Transactions on Big Data*, 2017.

[36] C. Puri and C. Dukatz, "Analyzing and predicting security event anomalies: Lessons learned from a large enterprise big data streaming analytics deployment," in *Database and Expert Systems Applications (DEXA), 2015 26th International Workshop on*, pp. 152–158, 2015.

[37] C. Liu, Y. Cai, and T. Wang, "Security evaluation of rc4 using big data analytics," in *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on*, pp. 316–320, 2016.

[38] K. Scarfone and P. Mell, "Guide to intrusion detection and prevention systems (idps)," *NIST special publication*, vol. 800, no. 2007, p. 94, 2007.

[39] S. Mukkamala, A. Sung, and A. Abraham, "Cyber security challenges: Designing efficient intrusion detection systems and antivirus tools," *Vemuri, V. Rao, Enhancing Computer Security with Smart Technology.(Auerbach, 2006)*, pp. 125–163, 2005.

[40] S. Suthaharan, "Big data classification: Problems and challenges in network intrusion prediction with machine learning," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 4, pp. 70–73, 2014.

[41] H.-M. Chen, R. Kazman, I. Monarch, and P. Wang, "Predicting and fixing vulnerabilities before they occur: a big data approach," in *Proceedings of the 2nd ACM International Workshop on BIG Data Software Engineering*, pp. 72–75, 2016.

[42] S. Marchal, X. Jiang, R. State, and T. Engel, "A big data architecture for large scale security monitoring," in *Big data (BigData Congress), 2014 IEEE international congress on*, pp. 56–63, 2014.

[43] B. Yang and T. Zhang, "A scalable meta-model for big data security analyses," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 55–60, 2016.

[44] P. Casas, F. Soro, J. Vanerio, G. Settanni, and A. D'Alconzo, "Network security and anomaly detection with big-dama, a big data analytics framework," in *Cloud Networking (CloudNet), 2017 IEEE 6th International Conference on*, pp. 1–7, 2017.

[45] S. Kumar, A. Viinikainen, and T. Hamalainen, "Machine learning classification model for network based intrusion detection system," in *Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for*, pp. 242–249, 2016.

[46] U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis, "Network anomaly detection with the restricted boltzmann machine," *Elsevier Neurocomputing*, vol. 122, pp. 13–23, 2013.

[47] M. A. Salama, H. F. Eid, R. A. Ramadan, A. Darwish, and A. E. Hassanien, "Hybrid intelligent intrusion detection scheme," in *Soft computing in industrial applications*, pp. 293–303, Springer, 2011.

[48] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pp. 21–26, ICST (Institute for Computer Sciences, Social-Informatics and ..., 2016.

[49] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, p. 1, 2015.

[50] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[51] N. Papernot, P. McDaniel, A. Sinha, and M. Wellman, "Towards the science of security and privacy in machine learning," *arXiv preprint arXiv:1611.03814*, 2016.

[52] W. Xu, D. Evans, and Y. Qi, "Feature squeezing: Detecting adversarial examples in deep neural networks," *arXiv preprint arXiv:1704.01155*, 2017.

[53] V. Patel, "A practical solution to improve cyber security on a global scale," in *Cybersecurity Summit (WCS), 2012 Third Worldwide*, pp. 1–5, 2012.

[54] C. Zhong, J. Yen, P. Liu, and R. F. Erbacher, "Automate cybersecurity data triage by leveraging human analysts' cognitive process," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 357–363, 2016.

[55] M. G. Schultz, E. Eskin, F. Zadok, and S. J. Stolfo, "Data mining methods for detection of new malicious executables," in *Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on*, pp. 38–49, 2001.

[56] H. H. Huang and H. Liu, "Big data machine learning and graph analytics: Current state and future challenges," in *Big Data (Big Data), 2014 IEEE International Conference on*, pp. 16–17, 2014.

[57] N. Naik, P. Jenkins, N. Savage, and V. Katos, "Big data security analysis approach using computational intelligence techniques in r for desktop users," in *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pp. 1–8, 2016.

[58] K. Kaur, A. Syed, A. Mohammad, and M. N. Halgamuge, "An evaluation of major threats in cloud computing associated with big data," in *Big Data Analysis (ICBDA), 2017 IEEE 2nd International Conference on*, pp. 368–372, 2017.

[59] J. Kepner, V. Gadepally, P. Michaleas, N. Schear, M. Varia, A. Yerukhimovich, and R. K. Cunningham, "Computing on masked data: a high performance method for improving big data veracity," in *High Performance Extreme Computing Conference (HPEC), 2014 IEEE*, pp. 1–6, 2014.

[60] D. Wang, B. Guo, Y. Shen, S.-J. Cheng, and Y.-H. Lin, "A faster fully homomorphic encryption scheme in big data," in *Big Data Analysis (ICBDA), 2017 IEEE 2nd International Conference on*, pp. 345–349, 2017.

[61] T. B. Patil, G. K. Patnaik, and A. T. Bhole, "Big data privacy using fully homomorphic non-deterministic encryption," in *Advance Computing Conference (IACC), 2017 IEEE 7th International*, pp. 138–143, 2017.

[62] B. Cui, B. Zhang, and K. Wang, "A data masking scheme for sensitive big data based on format-preserving encryption," in *Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on*, vol. 1, pp. 518–524, 2017.

[63] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings.," in *SecureComm*, vol. 10, pp. 89–106, Springer, 2010.

[64] T. Yang, P. Shen, X. Tian, and C. Chen, "A fine-grained access control scheme for big data based on classification attributes," in *Distributed Computing Systems Workshops (ICDCSW), 2017 IEEE 37th International Conference on*, pp. 238–245, 2017.

[65] S. Pérez, J. L. Hernández-Ramos, D. Pedone, D. Rotondi, L. Straniero, and A. F. Skarmeta, "A digital envelope approach using attribute-based encryption for secure data exchange in iot scenarios," in *Global Internet of Things Summit (GIoTS), 2017*, pp. 1–6, 2017.

[66] G. Xu, Y. Ren, H. Li, D. Liu, Y. Dai, and K. Yang, "Cryptmdb: A practical encrypted mongodb over big data," in *Communications (ICC), 2017 IEEE International Conference on*, pp. 1–6, 2017.

[67] A. Al Mamun, K. Salah, S. Al-maadeed, and T. R. Sheltami, "Bigcrypt for big data encryption," in *Software Defined Systems (SDS), 2017 Fourth International Conference on*, pp. 93–99, 2017.

[68] X. Dong, R. Li, H. He, W. Zhou, Z. Xue, and H. Wu, "Secure sensitive data sharing on a big data platform," *Tsinghua Science and Technology*, vol. 20, no. 1, pp. 72–80, 2015.

[69] A. Sharma and D. Sharma, "Big data protection via neural and quantum cryptography," in *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, pp. 3701–3704, 2016.

[70] C. Zhao and J. Liu, "Novel group key transfer protocol for big data security," in *Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2015 IEEE*, pp. 161–165, 2015.

[71] S. Almuhammadi and A. Amro, "Double-hashing operation mode for encryption," in *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*, pp. 1–7, 2017.

[72] F. A. Kadhim, G. H. Abdul-Majeed, and R. S. Ali, "Enhancement cast block algorithm to encrypt big data," in *New Trends in Information & Communications Technology Applications (NTICT), 2017 Annual Conference on*, pp. 80–85, 2017.

[73] A. Kulkarni, C. Shea, H. Homayoun, and T. Mohsenin, "Less: Big data sketching and encryption on low power platform," in *Proceedings of the Conference on Design, Automation & Test in Europe*, pp. 1635–1638, European Design and Automation Association, 2017.

[74] C. Xiao, L. Wang, Z. Jie, and T. Chen, "A multi-level intelligent selective encryption control model for multimedia big data security in sensing system with resource constraints," in *Cyber Security and Cloud Computing (CSCloud), 2016 IEEE 3rd International Conference on*, pp. 148–153, 2016.

[75] A. Soceanu, M. Vasylenko, A. Egner, and T. Muntean, "Managing the privacy and security of ehealth data," in *Control Systems and Computer Science (CSCS), 2015 20th International Conference on*, pp. 439–446, 2015.

[76] A. Al-Shomrani, F. Fathy, and K. Jambi, "Policy enforcement for big data security," in *Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on*, pp. 70–74, 2017.

[77] A. Samuel, M. I. Sarfraz, H. Haseeb, S. Basalamah, and A. Ghafoor, "A framework for composition and enforcement of privacy-aware and context-driven authorization mechanism for multimedia big data," *IEEE Transactions on Multimedia*, vol. 17, no. 9, pp. 1484–1494, 2015.

[78] F. Gao, "Research on cloud security control mechanism based on big data," in *Smart Grid and Electrical Automation (ICSGEA), 2017 International Conference on*, pp. 366–370, 2017.

[79] A. Gupta, A. Verma, P. Kalra, and L. Kumar, "Big data: A security compliance model," in *IT in Business, Industry and Government (CSIBIG), 2014 Conference on*, pp. 1–5, 2014.

[80] N.-Y. Lee and B.-H. Wu, "Privacy protection technology and access control mechanism for medical big data," in *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 424–429, 2017.

[81] M. R. Islam, M. Habiba, and M. I. I. Kashem, "A framework for providing security to personal healthcare records," in *Networking, Systems and Security (NSysS), 2017 International Conference on*, pp. 168–173, 2017.

[82] R. Achana, R. S. Hegadi, and T. Manjunath, "A novel data security framework using e-mod for big data," in *Electrical and Computer Engineering (WIECON-ECE), 2015 IEEE International WIE Conference on*, pp. 546–551, 2015.

[83] S.-H. Kim, N.-U. Kim, and T.-M. Chung, "Attribute relationship evaluation methodology for big data security," in *IT Convergence and Security (ICITCS), 2013 International Conference on*, pp. 1–4, 2013.

[84] S.-H. Kim, J.-H. Eom, and T.-M. Chung, "Big data security hardening methodology using attributes relationship," in *Information Science and Applications (ICISA), 2013 International Conference on*, pp. 1–2, 2013.

[85] H. A. Al Hamid, S. M. M. Rahman, M. S. Hossain, A. Almogren, and A. Alamri, "A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography," *IEEE Access*, vol. 5, pp. 22313–22328, 2017.

[86] A. Unal, "Security in big data of medical records," in *IT in Business, Industry and Government (CSIBIG), 2014 Conference on*, pp. 1–2, 2014.

[87] C. Ye, Z. Xiong, Y. Ding, J. Li, G. Wang, X. Zhang, and K. Zhang, "Secure multimedia big data sharing in social networks using fingerprinting and encryption in the jpeg2000 compressed domain," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2014 IEEE 13th International Conference on*, pp. 616–621, 2014.

[88] J. Wu, K. Ota, M. Dong, J. Li, and H. Wang, "Big data analysis based security situational awareness for smart grid," *IEEE Transactions on Big Data*, 2016.

[89] S. Pissanetzky, "On the future of information: Reunification, computability, adaptation, cybersecurity, semantics," *IEEE Access*, vol. 4, pp. 1117–1140, 2016.

[90] L. Xu, P. D. Khoa, S. H. Kim, W. W. Ro, and W. Shi, "Another look at secure big data processing: Formal framework and a potential approach," in *Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on*, pp. 548–555, 2015.

[91] K. Gai, M. Qiu, and H. Zhao, "Security-aware efficient mass distributed storage approach for cloud systems in big data," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*, pp. 140–145, 2016.

[92] L. Xu, H. Kim, X. Wang, W. Shi, and T. Suh, "Privacy preserving large scale dna read-mapping in mapreduce framework using fpgas," in *Field Programmable Logic and Applications (FPL), 2014 24th International Conference on*, pp. 1–4, 2014.

[93] G. Collard, E. Disson, G. Talens, and S. Ducroquet, "Proposition of a method to aid security classification in cybersecurity context," in *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*, pp. 88–95, 2016.

[94] P. Adluru, S. S. Datla, and X. Zhang, "Hadoop eco system for big data security and privacy," in *Systems, Applications and Technology Conference (LISAT), 2015 IEEE Long Island*, pp. 1–6, 2015.

[95] J. Singh, "Real time big data analytic: Security concern and challenges with machine learning algorithm," in *IT in Business, Industry and Government (CSIBIG), 2014 Conference on*, pp. 1–4, 2014.

[96] T. Yang and S. Jia, "Research on network security visualization under big data environment," in *Computer Symposium (ICS), 2016 International*, pp. 660–662, 2016.

[97] T. Ban, S. Pang, M. Eto, D. Inoue, K. Nakao, and R. Huang, "Towards early detection of novel attack patterns through the lens of a large-scale darknet," in *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences*, pp. 341–349, 2016.

[98] S. A. Alsuhibany, "A space-and-time efficient technique for big data security analytics," in *Information Technology (Big Data Analysis)(KACSTIT), Saudi International Conference on*, pp. 1–6, 2016.

[99] N. Mohammed, S. Barouti, D. Alhadidi, and R. Chen, "Secure and private management of healthcare databases for data mining," in *Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on*, pp. 191–196, 2015.

[100] L. Xiao, C. Xu, J. Qin, G. Qin, M. Zhu, L. Ruan, Z. Wang, M. Li, and D. Tan, "Secure distribution of big data based on bittorrent," in *Dependable, Autonomic and Secure Computing (DASC), 2013 IEEE 11th International Conference on*, pp. 82–90, 2013.

[101] W. Zhijun and W. Caiyun, "Security-as-a-service in big data of civil aviation," in *Computer and Communications (ICCC), 2015 IEEE International Conference on*, pp. 240–244, 2015.

[102] C. Hu and Y. Huo, "Efficient privacy-preserving dot-product computation for mobile big data," *IET Communications*, vol. 11, no. 5, pp. 704–712, 2016.

[103] S. Q. Ren, T. H. Meng, N. Yibin, and K. M. M. Aung, "Privacy-preserved multi-party data merging with secure equality evaluation," in *Cloud Computing Research and Innovations (ICCCRI), 2016 International Conference on*, pp. 34–41, 2016.

[104] K. Benzidane, H. El Alloussi, O. El Warrak, L. Fetjah, S. J. Andaloussi, and A. Sekkaki, "Toward a cloud-based security intelligence with big data processing," in *Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP*, pp. 1089–1092, 2016.

[105] Z.-W. Lu, "Research about new media security technology base on big data era," in *Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2016 IEEE 14th Intl C*, pp. 933–936, 2016.

[106] L. Yue, H. Junqin, Q. Shengzhi, and W. Ruijin, "Big data model of security sharing based on blockchain," in *Big Data Computing and Communications (BIGCOM), 2017 3rd International Conference on*, pp. 117–121, 2017.

[107] M. Tang, M. Alazab, and Y. Luo, "Big data for cybersecurity: Vulnerability disclosure trends and dependencies," *IEEE Transactions on Big Data*, 2017.

[108] P. A. Dhande and A. Kadam, "New approach for load rebalancer, scheduler in big data with security mechanism in cloud environment," in *Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference on*, pp. 247–250, 2016.

[109] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security framework," *IEEE Consum. Electron. Mag.*, vol. 7, no. 2, pp. 18–21, 2018.

[110] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "Medshare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

[111] N. Miloslavskaya, "Security intelligence centers for big data processing," in *Future Internet of Things and Cloud Workshops (FiCloudW), 2017 5th International Conference on*, pp. 7–13, 2017.

[112] M. Marchetti, F. Pierazzi, A. Guido, and M. Colajanni, "Countering advanced persistent threats through security intelligence and big data analytics," in *Cyber Conflict (CyCon), 2016 8th International Conference on*, pp. 243–261, 2016.

[113] Q. Jin, Y. Xiang, G. Sun, Y. Liu, and C.-C. Chang, "Cybersecurity for cyber-enabled multimedia applications," *IEEE MultiMedia*, vol. 24, no. 4, pp. 10–13, 2017.

[114] Y. Mengke, Z. Xiaoguang, Z. Jianqiu, and X. Jianjian, "Challenges and solutions of information security issues in the age of big data," *China Communications*, vol. 13, no. 3, pp. 193–202, 2016.

[115] H. Zou, "Protection of personal information security in the age of big data," in *Computational Intelligence and Security (CIS), 2016 12th International Conference on*, pp. 586–589, 2016.

[116] D. Mondek, R. B. Blažek, and T. Zahradnickỳ, "Security analytics in the big data era," in *Software Quality, Reliability and Security Companion (QRS-C), 2017 IEEE International Conference on*, pp. 605–606, 2017.

[117] B. B. Jayasingh, M. Patra, and D. B. Mahesh, "Security issues and challenges of big data analytics and visualization," in *Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on*, pp. 204–208, 2016.

[118] A. Kott, A. Swami, and P. McDaniel, "Security outlook: six cyber game changers for the next 15 years," *Computer*, vol. 47, no. 12, pp. 104–106, 2014.

[119] H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in *Big Data (BigData Congress), 2014 IEEE International Congress on*, pp. 762–765, 2014.

[120] S. Chandra, S. Ray, and R. Goswami, "Big data security in healthcare: Survey on frameworks and algorithms," in *Advance Computing Conference (IACC), 2017 IEEE 7th International*, pp. 89–94, 2017.

[121] S. Anandaraj and M. Kemal, "Research opportunities and challenges of security concerns associated with big data in cloud computing," in *I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on*, pp. 746–751, 2017.

[122] Y. Wang, B. Rawal, and Q. Duan, "Securing big data in the cloud with integrated auditing," in *Smart Cloud (SmartCloud), 2017 IEEE International Conference on*, pp. 126–131, 2017.

[123] S. Bahulikar, "Security measures for the big data, virtualization and the cloud infrastructure," in *Information Processing (IICIP), 2016 1st India International Conference on*, pp. 1–4, 2016.

[124] A. Sharif, S. Cooney, S. Gong, and D. Vitek, "Current security threats and prevention measures relating to cloud services, hadoop concurrent processing, and big data," in *Big Data (Big Data), 2015 IEEE International Conference on*, pp. 1865–1870, IEEE, 2015.

[125] Z. Tan, U. T. Nagar, X. He, P. Nanda, R. P. Liu, S. Wang, and J. Hu, "Enhancing big data security with collaborative intrusion detection," *IEEE cloud computing*, vol. 1, no. 3, pp. 27–33, 2014.

[126] J. Zhao, Y. Wang, and Y. Xia, "Analysis of information security of electric power big data and its countermeasures," in *Computational Intelligence and Security (CIS), 2016 12th International Conference on*, pp. 243–248, 2016.

[127] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423–2436, 2016.

[128] C. Dincer, G. Akpolat, and E. Zeydan, "Security issues of big data applications served by mobile operators," in *Signal Processing and Communications Applications Conference (SIU), 2017 25th*, pp. 1–4, 2017.

[129] S. Yakoubov, V. Gadepally, N. Schear, E. Shen, and A. Yerukhimovich, "A survey of cryptographic approaches to securing big-data analytics in the cloud," in *High Performance Extreme Computing Conference (HPEC), 2014 IEEE*, pp. 1–6, 2014.

[130] L. Dupré and Y. Demchenko, "Impact of information security measures on the velocity of big data infrastructures," in *High Performance Computing & Simulation (HPCS), 2016 International Conference on*, pp. 492–500, 2016.

[131] N. Chaudhari and S. Srivastava, "Big data security issues and challenges," in *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pp. 60–64, 2016.

[132] M. Paryasto, A. Alamsyah, B. Rahardjo, *et al.*, "Big-data security management issues," in *Information and Communication Technology (ICoICT), 2014 2nd International Conference on*, pp. 59–63, 2014.

[133] R. Clarke, "Quality assurance for security applications of big data," in *Intelligence and Security Informatics Conference (EISIC), 2016 European*, pp. 1–8, 2016.

[134] I. Škrjanc, A. S. de Miguel, J. A. Iglesias, A. Ledezma, and D. Dovžan, "Evolving cauchy possibilistic clustering based on cosine similarity for monitoring cyber systems," in *Evolving and Adaptive Intelligent Systems (EAIS), 2017*, pp. 1–5, 2017.

[135] S. Alouneh, I. Hababeh, F. Al-Hawari, and T. Alajrami, "Innovative methodology for elevating big data analysis and security," in *Open Source Software Computing (OSSCOM), 2016 2nd International Conference on*, pp. 1–5, 2016.

[136] K. Dhanasekaran and B. Surendiran, "Nature-inspired classification for mining social space information: National security intelligence and big data perspective," in *Green Engineering and Technologies (IC-GET), 2016 Online International Conference on*, pp. 1–6, 2016.

[137] K. D. Strang and Z. Sun, "Meta-analysis of big data security and privacy: Scholarly literature gaps," in *Big Data (Big Data), 2016 IEEE International Conference on*, pp. 4035–4037, 2016.

[138] L. Yuqing, "Research on personal information security on social network in big data era," in *Smart Grid and Electrical Automation (ICSGEA), 2017 International Conference on*, pp. 676–678, 2017.

[139] J. Hariharakrishnan, S. Mohanavalli, K. S. Kumar, *et al.*, "Survey of pre-processing techniques for mining big data," in *Computer, Communication and Signal Processing (ICCCSP), 2017 International Conference on*, pp. 1–5, 2017.

[140] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE communications magazine*, vol. 53, no. 10, pp. 190–199, 2015.

[141] T. Lu, X. Guo, B. Xu, L. Zhao, Y. Peng, and H. Yang, "Next big thing in big data: the security of the ict supply chain," in *Social Computing (SocialCom), 2013 International Conference on*, pp. 1066–1073, 2013.

[142] E. Damiani, C. Ardagna, F. Zavatarelli, E. Rekleitis, and L. Marinos, "Big Data Threat Landscape," *European Union Agency for Network and Information Security*, Jan 2017. web: https://www.enisa.europa.eu/publications/bigdata-threat-landscape.

[143] R. Geambasu, T. Kohno, A. A. Levy, and H. M. Levy, "Vanish: Increasing data privacy with self-destructing data.," in *USENIX Security Symposium*, vol. 316, 2009.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 19, 2021, accepted May 10, 2021, date of publication June 7, 2021, date of current version July 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087109

Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security

ABEL YEBOAH-OFORI[1], SHAREEFUL ISLAM[2], SIN WEE LEE[2], ZIA USH SHAMSZAMAN[3], (Senior Member, IEEE), KHAN MUHAMMAD[4], (Member, IEEE), METEB ALTAF[5], AND MABROOK S. AL-RAKHAMI[6], (Member, IEEE)

[1]Department of Computer Science and Engineering, University of West London, Ealing London W5 5RF, U.K.
[2]School of Architecture Computing and Engineering (ACE), University of East London, London E16 2RD, U.K.
[3]Department of Computing and Games, Teesside University, Middlesbrough TS1 3BX, U.K.
[4]Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, South Korea
[5]Advanced Manufacturing and Industry 4.0 Center, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia
[6]Research Chair of Pervasive and Mobile Computing, Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Mabrook S. Al-Rakhami (malrakhami@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University through the Vice Deanship of Scientific Research Chairs: Chair of Pervasive and Mobile Computing.

ABSTRACT Cyber Supply Chain (CSC) system is complex which involves different sub-systems performing various tasks. Security in supply chain is challenging due to the inherent vulnerabilities and threats from any part of the system which can be exploited at any point within the supply chain. This can cause a severe disruption on the overall business continuity. Therefore, it is paramount important to understand and predicate the threats so that organization can undertake necessary control measures for the supply chain security. Cyber Threat Intelligence (CTI) provides an intelligence analysis to discover unknown to known threats using various properties including threat actor skill and motivation, Tactics, Techniques, and Procedure (TT and P), and Indicator of Compromise (IoC). This paper aims to analyse and predicate threats to improve cyber supply chain security. We have applied Cyber Threat Intelligence (CTI) with Machine Learning (ML) techniques to analyse and predict the threats based on the CTI properties. That allows to identify the inherent CSC vulnerabilities so that appropriate control actions can be undertaken for the overall cybersecurity improvement. To demonstrate the applicability of our approach, CTI data is gathered and a number of ML algorithms, i.e., Logistic Regression (LG), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT), are used to develop predictive analytics using the Microsoft Malware Prediction dataset. The experiment considers attack and TTP as input parameters and vulnerabilities and Indicators of compromise (IoC) as output parameters. The results relating to the prediction reveal that Spyware/Ransomware and spear phishing are the most predictable threats in CSC. We have also recommended relevant controls to tackle these threats. We advocate using CTI data for the ML predicate model for the overall CSC cyber security improvement.

INDEX TERMS Cyber threat intelligence, machine learning, cyber supply chain, predictive analytic, cyber security, tactic techniques procedures.

I. INTRODUCTION

Cyber Supply Chain (CSC) security is critical for reliable service delivery and ensure overall business continuity of Smart CPS. CSC systems by its inherently is complex and vulnerabilities within CSC system environment can cascade from a source node to a number of target nodes of the overall cyber physical system (CPS). A recent NCSC report highlights a list of CSC attacks by exploiting vulnerabilities that exist within the systems [1]. Organizations outsource part of their business and data to the third-party service providers that could lead any potential threat. There are several examples for successful CSC attacks. For instance, Dragonfly, a Cyber Espionage group, is well known for targeting CSC organization [2], [3]. The Saudi Aramco power station attack halted its operation due to a massive cyberattack [1]. There are

The associate editor coordinating the review of this manuscript and approving it for publication was Po Yang.

existing works that consider CSC threats and risks but a lack of focus on threat intelligence properties for the overall cyber security improvement. Further, it is also essential to predict the cyberattack trends so that the organization can take the timely decision for its countermeasure. Predictive analytics not only provide an understanding of the TTPs, motives and intents of the threat actors but also assist situational awareness of current supply system vulnerabilities.

This paper aims to improve the cybersecurity of CSC by specifically focusing on integrating Cyber Threat Intelligence (CTI) and Machine Learning (ML) techniques to predicate cyberattack patterns on CSC systems and recommend suitable controls to tackle the attacks. The novelty of our work is threefold:

- Firstly, we consider Cyber Threat Intelligence(CTI) for systematic gathering and analysis of information about the threat actor and cyber-attack by using various concepts such as threat actor skill, motivation, IoC, TTP and incidents. The reason for considering CTI is that it provides evidence-based knowledge relating to the known attacks. This information is further used to discover unknown attacks so that threats can be well understood and mitigated. CTI provides intelligence information with the aim of preventing attacks as well as shorten time to discover new attacks.
- Secondly, we applied ML techniques and classification algorithms and mapped with the CTI properties to predict the attacks. We use several classification algorithms such as Logistic Regression (LG), Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT) for this purpose. We follow CTI properties such as Indicator of Compromise (IoC) and Tactics, Techniques and Procedure (TTP) for the attack predication.
- Finally, we consider widely used cyberattack dataset to predict the potential attacks [6]. The predication focuses on determining threats relating to Advance Persistent Threat (APT), command and control and industrial espionage which are relevant for CSC [7]–[9]. The result shows the integration of CTI and ML techniques can effectively be used to predict cyberattacks and identification of CSC systems vulnerabilities. Furthermore, our prediction reveals a total accuracy of 85% for the TPR and FPR. The results also indicate that LG and SVM produced the highest accuracy in terms of threat predication.

The rest of the paper is organised as follows: Section 2 presents an overview of related works including CSC security, cyber threat intelligence and Machine Learning for CSC. Section 3 provides the concepts necessary for the proposed approach and the meta model. Section 4 provides an overview of the proposed approach including the integration of CTI and ML. Section 5 presents the underlying process for the threat analysis and predication. Section 6 implements the process for the threat predication using the widely used Microsoft malware datasets. Section 7 discusses the results and compares the work with the existing works in the literature. Finally, Section 8 provides conclusion and future direction of the work.

II. RELATED WORK

There exists several widely used CTI and ML models in cyber security domain. This section presents the existing works that are relevant with our work.

A. CYBER SUPPLY CHAIN(CSC) SECURITY

The CSC security provides a secure integrated platform for the inbound and outbound supply chains systems with third party service provider including suppliers, and distributors to achieve the organizational goal [10]. Cybersecurity from supply chain context involves various secure outsourcing of products and information between third party vendors, and suppliers [11]. This outsourcing includes the integration of operational technologies (OT) and Information technologies (IT) running on Cyber Physical Systems (CPS) infrastructures. However, there are threats, risks and vulnerabilities that are inherent in such systems that could be exploited by threat actors on the operational technologies and information technologies of the supply inbound and outbound chains systems. The outbound chain attacks include data manipulations, information tampering, redirecting product delivery channels, and data theft. The IT risks include those attacks on the cyber physical and cyber digital system components such as distributed denial of service (DDoS) attacks, IP address spoofing, and Software errors [12]. Regarding CSC security, NIST SP800 [13] proposed a 4 tier framework approach for improving critical infrastructure cybersecurity that incorporates the cyber supply chain risk management framework into it as one of its core components. Tier 1 considers the organizations CSC risk requirement strategy. Tier 2 considers the supply chain associated risk identifications including products and services in the supply inbound and outbound chains. Tier 3 implementation considers the risk assessments, threats analyses, associated impacts and determine the baseline requirements for governance structure. Tier 4 consider real-time or near-time information to understand supply chain risk associated with each product and service. However, the approach and tiers considered risks management but did not emphasize on ML and threat prediction for future trends in the CSC domain. Additionally, [14] proposed a supply chain attack framework and attack patterns that structured and codifies supply chain attacks. The goal of the framework was to provide a comprehensive view of supply chain attacks of malicious insertion across the full acquisition lifecycle to determine the associated threat and vulnerability information.

B. CYBER THREAT INTELLIGENCE (CTI)

Cyber threat intelligence (CTI) gatherings and analysis have become one of the relevant actionable intelligences used to understand both known and unknown threats [4]. The impact of cyberattacks and emerging threats on CSC systems and

its devastating effects on business process, data, Intellectual Property, delivery channel, and cost of recovery has increased the surge for CTI approach. The CTI process includes identification, threat analysis and information disseminating to stakeholders. Considering CTI for cybersecurity, ENISA in [4] explored the opportunities and limitations of current threat intelligence platforms by considering CTI implementation process and threat intelligence programs (TIP) from strategic, tactical and operational goals. The authors proposed a threat intelligence program model that collects, normalize, enrich, correlate, analyse and disseminate threat related information to stakeholders. The strategic CTI goals consider factors that support executive decision makings, tactical goals consider the CTI process and TIP programs that identifying intelligence gap and prioritizing them for risk reduction. The operational goals provide a process that provides an understanding of the threat actors motives, modes of operation, intents, and TTPs and capabilities. However, the processes do not incorporate ML threat predictions. Additionally, [15] proposes a threat intelligence-driven security model that considers six CTI phases and processes lifecycle required to identify intelligence goals. The CTI phases include direction, collection, process, analysis, dissemination, and feedback. The author incorporated internal sources such as network traffic, logs, scans; external sources such as vulnerability database, threat feeds; and human sources such as the dark web and social media into the model for the threat intelligence modelling. The threat intelligence driven security model emphasizes on using network traffics, logs and scans and not ML algorithms for the prediction. Further, [16] develop cyber threat Intelligence metrics that consider assets, requirement business operations, adversary, and consumer intelligence places emphases on value and organizational benefits. The author's approach considers four key stages in the threat intelligence process including intelligence requirements, information collection, analyses, dissemination, and intelligence usage. However, the approach does not consider machine learning for predicting invisible attacks. Furthermore, [17] proposed a CTI model that operationalizes and analyses adversarial activities across the lifecycle of an organization business process to determine actions taken by the attacker. The author's approach was based on the organizational intelligence requirements, information gathering, analyses and disseminate to protect assets for strategic, tactical and operational understanding and situational awareness. However, the works emphasized more on attacker motive and intent and not on ML for the threat predictions. The CTI functional process is to collect metrics and trend analysis for the business risk assessment, prioritization, and decision support with less emphasis on ML for CSC security.

C. MACHINE LEARNING IN CSC SECURITY

There are several works that consider Machine Learning classifiers in various cybersecurity application domains such as spam filters, antivirus and IDS/IPS to predict cyberattack trends [18], [23], [24]. Considering ML for Security [11], proposed ML classification of HTTP attacks using a decision tree algorithm to learn a dataset for performance accuracies and automatically label a request as valid or attack. The authors developed a vector space model used commonly for information retrieval to build a classifier to automatically label the request as malicious in the URL. The approach achieved high precision and recall comparatively. However, the work did not focus on ML and threat prediction in the CSC environment. Further, [20] carried out the feasibility of a study on machine learning models for cloud security to test the models in diverse operation conditions cloud scenarios. The authors compared Logistic Regression, Decision Tree, Naïve Bayes, and SVM classification algorithms techniques to learn a dataset for performance accuracies. The algorithms represent supervised schemes and are used in network security. The result shows an accuracy of 97% in anomalous packet detections. However, the work did consider CSC security from threat prediction in the supply chain environment. Furthermore, [21] surveyed data mining and ML methods for cybersecurity detection methods for cyber analytics in support of intrusion detection in cybersecurity applications. The authors used Artificial Neural Network, Association rules, Fuzzy Association rules and Bayesian Networks classifiers to learn the datasets and provided comparison criteria for the machine learning and data mining models to recognize the types of the attack (misuse) and for detection of an attack (intrusion). However, the techniques and methods used are not ML models and did not focus on ML and threat prediction in the CSC environment. Additionally, [22] review the cybersecurity dataset for ML algorithms used for analysing network traffic and anomaly detection. The author compared the machine learning techniques used for experiments, evaluation methods and baseline classifiers for comparison of the dataset. The results show significant flaws in some dataset during feature selection and are not relevant for modern intrusion detections datasets. However, the review did not stress on the current dataset we used from the Microsoft Malware Threat Prediction website for the prediction. Moreover, [23] explored the classification of logs using ML techniques on a decision tree algorithm to learn a dataset that models the correlation and normalization of security logs. The goal of the ML techniques is to evaluate if the algorithm can predict the performance of classification as an attack or not after a training phase. The dataset used contains anomalous and some identified attacks. The result shows that the DT algorithm was model on internet logs to develop a framework for the normalization and correlation of the classify with an accuracy of 80%. However, the classification model did not compare other classification algorithms such as SVM, LR and RF that are relevant for ML better performance accuracies and threat analysis.

Another initiative [24] explores the viability of using machine learning approaches to predict power systems disturbance and cyberattack discrimination classifiers and focuses specifically on detecting cyberattacks where deception is the core tenet of the event [24]–[30]. The authors in [24]

evaluated the classification performances on, NNge, OneR, SVM, RF, JRpper and Adaboost algorithms to learn the dataset and focused specifically on detecting cyber attacks where deception is the core tenet of the event. For example, in [25], the authors proposed a SCADA power system cyber-attack detection approach by combining a correlation-based feature selection (CFS) method and K-Nearest-Neighbour (KNN) instance-based learning (IBL) algorithm. The combination was useful to reduce the extremely large number of features and to maximize cyberattack detection accuracy with minimum detection time cost. In [26], an ensemble-learning model for detecting the cyberattacks of SCADA-based IIoT platform is proposed. The model was based on the combination of a random subspace (RS) learning method with random tree (RT). The authors in [29] proposed a deep-learning, feature-extraction-based semi-supervised model for cyberattack protection in the trust boundary of IIoT networks. The proposed approach was adaptive to learn unknown attack. However, the works did not consider CSC attacks from supplier inbound and outbound chains.

Regarding ML predictive analytics on various datasets, [28] predicted cybersecurity incidents using ML algorithms to distinguish between the different types of models. The authors used text mining methods such as n-gram, bag-of-words and ML techniques to learn dataset on Naive Bayes and SVM algorithms for classification performance. The experiment was to predict classification accuracies of malware incidents response and actions. The approach did not consider CTI and ML in the CSC system environment. Further, [29] proposed a risk teller system that analyses binary file appearance logs of a machine to predict which machines are at risk of experiencing malware infection in advance. The authors used a random forest algorithm and semi-quantitative methods to build a risk prediction model that creates a profile to capture usage patterns. The results associate each level of risk to a machine infection incident with 95% true positive precision. Besides,[30] characterize the extent to which cybersecurity incidents can be predicted based on externally observable properties of an organization's network. The authors used Verizon's annual data breach investigation report to forecast if an organization may suffer cybersecurity incidents in future. A random forest classifier was used against over 1000 incident reports taken from various datasets. The predictive result achieved an overall accuracy of 90% true positives. However, the work did not provide any inference and map the prediction to existing attacks. All these works above are important and contributed towards the improvement of cyber security by using various ML techniques. However, there is a lack of focus on the overall CSC security context. A limited works emphasize on threat intelligence data for the attack predication. For instance, due to the invisibility nature of cyberattacks, an attack on the CSC system network node has the potential to cascade to other nodes on the supply chain system. Therefore, it is necessary to use ML analytics to predict cyberattacks, threats and the underlying vulnerabilities. Additionally, there is a need to understand an organisational context for the threat analysis. CTI can effectively support to achieve that goal. This work contributes towards this direction. We have integrated CTI for threat gathering and analysis with the ML for the threat prediction so that organizations can determine the suitable control measure for the overall CSC security improvement.

III. FRAMING CONCEPTS

This section presents the conceptual view of the proposed approach by combining concepts from both CTI and CSC.

A. CSC THREAT MODELLING CONCEPTS

This section considers the concepts that are necessary to determine CSC vulnerabilities, goals, requirements, attacks the cyber supply inbound and outbound chains security and the CTI domain [2]. Threat modelling provides a systematic approach to identify and address the possible threats based on a specific context. It provides an understanding of threat actor who can attack the system and possible assets which can be compromised. The proposed approach considers a list of concepts that aid understand the threts and possible mitigation. The concepts provide a view of the relationships between organizational and security goal, requirements, threat actors, attacks, vulnerability, TTPs and indicators of compromise for understanding of the threat. An overview of the concepts is given below:

Goal: A goal represents the strategic aim of an organization. Properties for the goal include the organizational goal, the tangible assets required such as infrastructures to achieve the goal and intangible asset such as credit card information, health record, and other sensitive data for the security goal. The organizational goal is the process, product or service that is carried out. The assets are tangible and intangible assets including the network infrastructures. The security goal is the mechanism, configuration, and control put in place to achieve the goal.

Actor consists of perpetrators, system users, the systems, the third-party vendors, and companies whose services and networks systems are attached to the main organization's supply chain system. The threat actors are those consist of users, agents, cybercriminals, and other systems that aims at compromising the CSC systems and the security goal [8]. The threat actor could be an internal or external attacker. The CSC system includes the various integrations of network nodes that make up the supplier chain system. The third-party vendors include the organization on the supplier inbound and outbound chains that could be attacked, manipulated, or compromised.

Inbound and Outbound Supply Chain: In a CSC environment, the network nodes and communication channels are those that integrate with the inbound and outbound supply chains systems. These are vendors, SMEs, suppliers, and distributors that are on the supply chain. The inbound suppliers are those with external remote access to the CSC system. The outbound chains are those that the organization distributes including individuals, institutions,

and vendors. The organization can experience attacks on the supply inbound and outbound chain that supports the application processes [8]. The threat actor could initial injection attacks or insert a redirect script into the vendor's website and breach the software developed by the manufacturer that is used by the organization's internal employers to distribute services to vendors and individuals. The goal of the attack could be to manipulate, alter or divert products and services after gaining access into the system.

Vulnerabilities: CSC vulnerabilities are the loopholes and configuration flaws that exist on the supply chain system and network nodes that could be exploited by an attack, threat actor or a threat agent. These network vulnerabilities [36] are those that exist on the supply inbound and outbound chains including the network nodes, switches, IP addresses, and firewalls. The vulnerable spots on the CSC system could be identified from various sources including the software, the network, website, the user, processes, the application, and configuration or the third-party vendor. Properties include asset type, source, node, effect and criticality.

Attack: An attack is any deliberate action or assault on the supply chain system with the intent to penetrate a system, to be able to gain access then manipulate and compromise processes, procedures, and delivery channels of electronic products, the information flows, and services [2]. Properties include the type of attack, pattern, prerequisites, and vectors. We consider attack inputs and outputs parameters for our study and the attack concepts for our prediction. Inputs of attack include the tools, capabilities, vectors and knowledge of the vulnerabilities of the domain to exploit. Outputs of the attacks are the patterns, access gained by the threat actor, the methods deployed, TTPs, the loopholes exploited, and the extent of malware propagation and cascading effects. This includes those attacks on cyber physical and cyber digital systems such as hardware, network, IP addresses, and software. The OT and IT delivery mechanisms could be manipulated before the product gets to the consumer [8].

Tactics, Techniques and Procedures (TTPs) consist of the specific adversary behaviour exhibited in an attack [14]. It leverages on resources such as tools, infrastructures, capabilities and personnel. It provides information on the victim's target (who, what or where), that are relevant to exploit targets being targeted, intended effects, kill chain phases, handling guidance and resources of the TTP information [8], [9]. Threats actors' mode of operation is to commit attacks such as Hijacking, social engineering, and footprints, privilege escalation, and reconnaissance penetrate a supply chain.

CSC Requirement: CSC requirements are the constraints and security expectations for the system required to support CSC stakeholders and business needs. The data gathered from stakeholders inform business processes, system infrastructures, internal and external user expectations required for the supply chain system developments and operations [2]. The requirements process and constraints that are generated during the requirements engineering phase forms the basis for the system constraints and statements that support the user and system requirements used to achieve the organizational goal. The requirements consist of attributes such as user categories, stakeholders, description, user ID, acceptance criteria, time constraints, owners and sources. The requirements concepts include properties such as organizational requirements, business requirements, system, user, and operational requirements. The organizational requirements describe the organizational high-level objectives that must be performed to achieve the organizational goal. The business requirements explain the requirement specifications and the properties include customer needs and expectations that must be integrated to meet the system requirements. Systems requirements demand specific properties of the application, architecture and the technical requirements need to be able to describe the features and how the system must function. These system requirements properties include the constraints, assumptions and acceptance criteria and the external entities that will be interacting with the system. They include supply chain systems processes and constraints that are generated during the requirements engineering phase that forms the basis for the system.

Indicators: Indicators are parameters that express an attack of this type, whether it is imminent, in progress or has occurred [32]. Properties required to determine the indicators of compromise includes incident type, source, date & time, impact, motive and intents. The properties are used to determine threat activities, adversary behaviours, TTPs, risky events, or state of the incident to determine what could serve as an indicator of compromise. CSC attack incidents and course of actions provide intelligence about the nature of cyberattack indicators and TTPs that can be deployed on the supply chain especially from the third-party vendor's perspective. Indicators convey specific observable patterns combined with contextual information intended to represent artefacts and or behaviours of interest within a cybersecurity context.

Cyber incident report: Cybersecurity incident is defined as a breach of system security to affect its integrity or availability. It includes unauthorized access or attempted to access a system or causing a disruptive event to essential services. Cybersecurity incident reporting platform provides individuals and organizations with a system to reports cyber incidents they have experienced unexpectedly or any unusual network issues, or suspected fraud or cybercrime activities [31]. Properties for cyber incident reporting include attack type, date and time of the incident, source of the attack, cause of an attack, duration, impact on service, impact on staff and public safety Cyber incident report system is required for cyber threat analysis and to determine the threat level and categorizing. It is used to predict cyberattacks and generate intelligence require to mitigate cyberattacks and for threat information sharing.

Threat information sharing: Threat information sharing is used to provide information necessary to assist an organization in identifying, assessing, monitoring, and responding to cyber threats [32]. Cyber threat information includes

indicators of compromise, tactics, techniques, and procedures used by threat actors, security alerts and threat intelligence reports. It provides findings from the analysis of cyber incidents and suggests actions to take to prevent cyber-attacks, detect, protect, contain, and mitigate cyber incidents. Properties for cyber threat information sharing include information-sharing goals, information sources, scope, sharing community and support. Some rules govern and protect information sharing, such as information sensitivity and privacy, sharing designations, and tracking procedures [32]. It provides a basis for an organization to leverage their combined knowledge, information, experience, and competencies to gain intelligence and understanding of potential threats for remediation and controls.

Controls: Controls are security mechanisms that are put in place to secure organizational business operations and processes. They are security strategies and measures formulated and implemented to ensure that the organizational goal and objectives are achieved [2], [13]. These controls include directive, detective, preventive, corrective and recovery. Directive controls are more strategic and relevant with the specific supplier inbound and outbound chain requirements. These are intended to align organizational and security goals with that of supplier and third-party vendors on the supply chain and provide guidelines for system usage and processes. Preventive controls are policies that are put in place for the technical and physical infrastructures protection. These are derived from standard measures intended to preclude actions violating policy or increasing third party risks to the supply chain system resources. Detective Controls use supply chain attack indicators to identify practices, processes, and tools that identify and possibly react to security violations. These include Firewall, IDS, IPS and the various configurations required for the supply chain systems. Corrective controls involve physical, administrative, and technical measures. Recovery controls includes backup plans, regular updates and contingency planning to ensure integrity or availability of the CSC in the event of an incident. Once an incident occurs on the CSC system that results in the compromise of integrity or availability, the implementation of recovery controls is necessary to restore the system or operation to a normal operating state. These include countermeasures, backups, segmentation, and an incidence response strategy.

The meta-model in Figure 1 explains relationships among the concepts. The organizational goal is determined by the product and services that are produced. The security goal is to ensure that the supply chain systems that support these products and services are secured. CSC organization needs a list of requirements to satisfy for achieve its goals. The TTP as a CTI properties exploits both inbound and outbound vulnerabilities for a successful attack. Cyber incident report provides a detailed about the incident including vulnerability, indicator and incident time frame. This report needs to share among the CSC stakeholders. There are controls which are required to tackle the threats.

IV. THE PROPOSED APPROACH

This section discusses the proposed approach that aims to improve the CSC security. It includes an integration of CTI and ML and a systematic process (presented in the Section 5). Additionally, the underlying concepts of the proposed approach such as actor, goal, TTP, vulnerability, incident, and controls, is also mentioned in Section 3. The approach considers both inbound and outbound chains for the vulnerability so that CSC organisation can focus on the possible system flaws. The approach adopts the CTI process to gather and analyse the threat data and ML techniques to predicate the threat. ML techniques are used on classification algorithms to learn a dataset for performance accuracies and predictive analytics. The rationale for integrating CTI and ML for threat prediction is that the CTI lifecycle process supports input parameters for detecting known attacks whereas ML provides output parameters for predicting known and unknown attacks for future trends.

A. INTEGRATION OF CTI AND ML

The approach combines CTI processes with ML techniques for cyber threat predictive analytics. The goal is to detect vulnerabilities and indicators of compromise on CSC network system nodes using known attacks to predict unknown attacks. We apply the CTI techniques to gather threats (Known attacks) and ML techniques to learn the dataset to predicate cyber threats (unknown attacks) on CSC systems. The inputs are the attacks and TTP that are deployed by threat actors to compromise a system. The attack feature uses properties such as attack type, pattern, attack vectors, and prerequisites to determine the nature of the attack that was deployed. The TTP consists of attack patterns and attack vectors deployed by the threat actor. The TTP parameter includes the capabilities of the threat actor and threat indicators. The threat actor feature uses properties such as user, system and third-party vendors to determine the vulnerable spots and type of tools used for the attack to determine the attack pattern. Tools are the attack weapons or software codes used by the threat actor for reconnaissance and to initiate an attack. For instance, the threat actor could use Nmap tool for scanning a network, Kali Linux tool for penetration and, Metasploit tool for exploiting loopholes in a network. The output parameters are the vulnerabilities and indicators of compromise that are used as threat intelligence. The capability of the threat actor could be determined by the ability to penetrate a system and course Advance Persistent threat (APT) attack and take command and control C&C) the extent of propagation is used to determine the indicators. Finally, we consider various controls such as directive, preventive, detective corrective and recovery required to secure the CSC system.

The rationale for our predictive analytics approach is based on the premise that the cyberattacks phenomenon includes a lot of invincibility, and uncertainties and the makes the threat landscape unpredictable. Similarly, due to the changing organizational requirements, various integrations, varying business processes and the various delivery mechanisms,

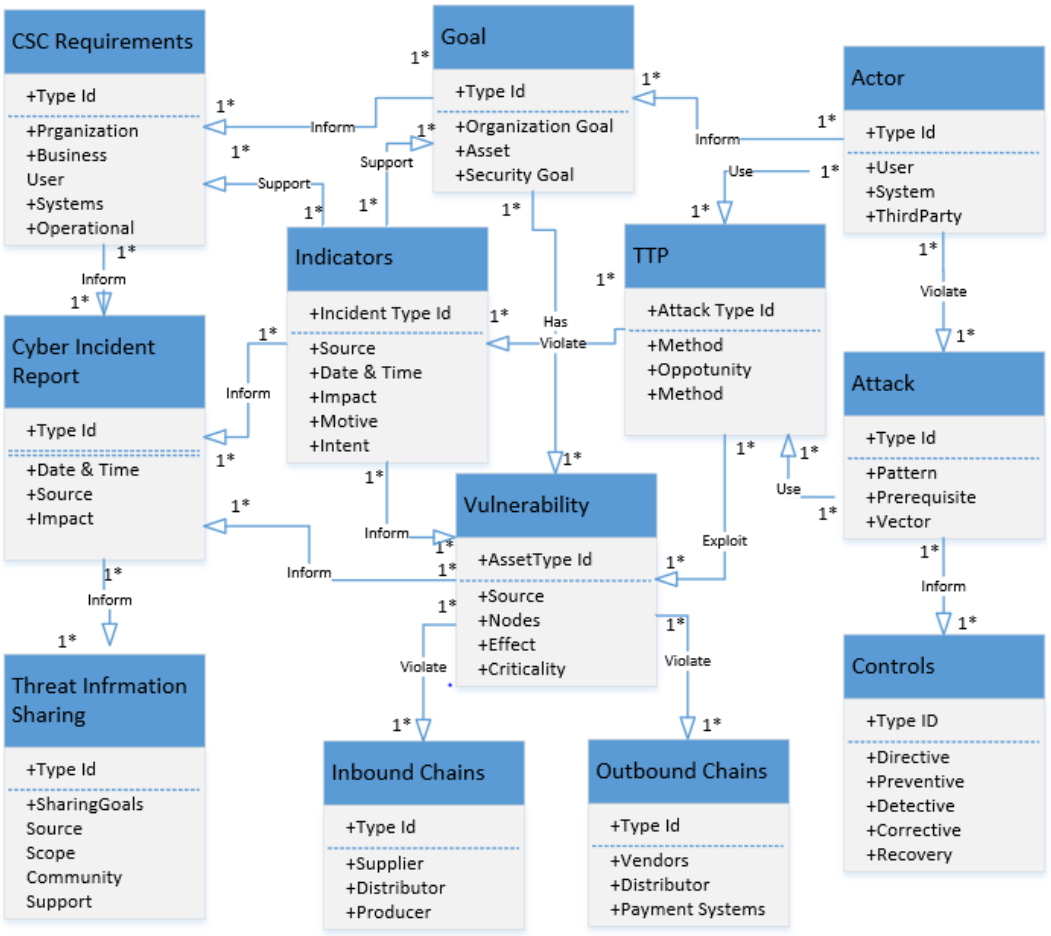


FIGURE 1. Meta-model for the proposed conceptual view of CSC system security.

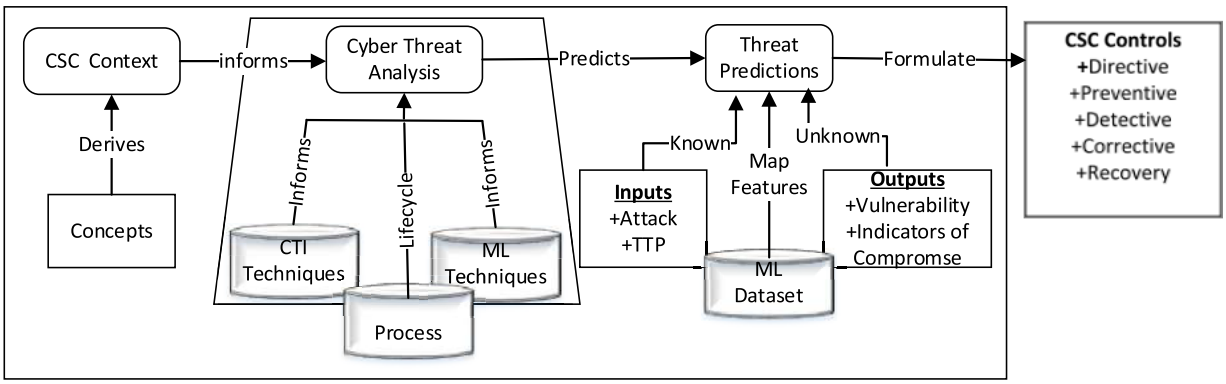


FIGURE 2. Applying CTI and ML for threat intelligence and predictive analytics.

predicting cyberattacks in the CSC organization context has been challenging. To achieve that, first, the proposed approach considers relevant related works and the meta-model concepts to model the CSC attacks and CTI phases. For instance, we identify supply inbound and outbound chain attack indicators and integrate them into CTI phases. Further, the concepts are analysed using the CTI process lifecycle and ML techniques to learn the dataset for our prediction. Furthermore, we use the input and output parameters as indicators for our threat prediction. Finally, the threat prediction results are evaluated to provide informed intelligence regarding the various attacks and future threats that are unknown for appropriate control mechanisms. Figure 2 indicates the proposed approach.

V. THREAT ANALYSIS AND PREDICTION PROCESS

This section discusses the overall process for the CSC threat analysis, prediction, and control in line with the proposed

approach in Section 3. The process includes four sequential phases. It follows a methodical approach and a causal process for each phase to determine strategy, threat analysis, threat prediction, and controls. Each phase includes steps and activities required to achieve the purpose of the phases as shown in Figure 3. The activities include identifying the organization's CSC and security strategy, ML classifications, infrastructures, attack context, input and output parameters for our prediction. The activities for the threat analysis phase include the identification and gathering of threat information, risk assessment and analysis to determine the threat actor, threat profile, TTP and IoC. The activities for the threat prediction phase consider the input parameters for the ML algorithms, predict threats and for performance evaluation by using ML techniques to learn datasets. The control activities include identifying required controls for the CSC systems including internal and external audits to formulate security policies and control mechanisms. We expound on the phases and process further by following the process flow as shown in Figure 3.

A. PHASE 1: DETERMINE STRATEGY

CSC security strategy combines CTI and cybersecurity risk strategy including mechanisms, resources and plans to determine how security goals and controls will be formulated, implemented, and achieved in line with organization goal and objectives. It includes identifying, analysing, reviewing and evaluating organizational assets including infrastructures, resources and implementation procedures. CSC security strategy combines, CTI and cybersecurity risk assessment strategy to gather intelligence and formulate policies. Strategic, tactical and operational management roles and responsibilities are recursive and support each other to ensure security goals are achieved. Strategic management uses intelligence decision to support plans that determine security goals and assign responsibility including executive authorization of blueprints and budget allocation. Tactical management decision regarding the execution of strategic management blueprints including security requirements capturing, third party audit, configuration management plans, uses indicators of compromise to determine controls and validations. The operational level managers ensure the day-to-day implementation of the security goals including monitoring, determining TTPs and escalating threat alerts for remediation and controls. CTI Strategy provides management evidence-based knowledge gathered about threats actors, attacks, patterns, vectors, vulnerabilities, TTPs, motives, intents and capabilities of the adversary. Risk Assessment Strategy considers the organizational goal and assets and develops an overall CSC risk strategy that determines the policies required to guide the organizational business processes. It includes risk assessment, CSC requirements capturing and business function. The risk strategy also considered implementation strategies and procurement policies for OT and IT acquisitions and integrations of assets.

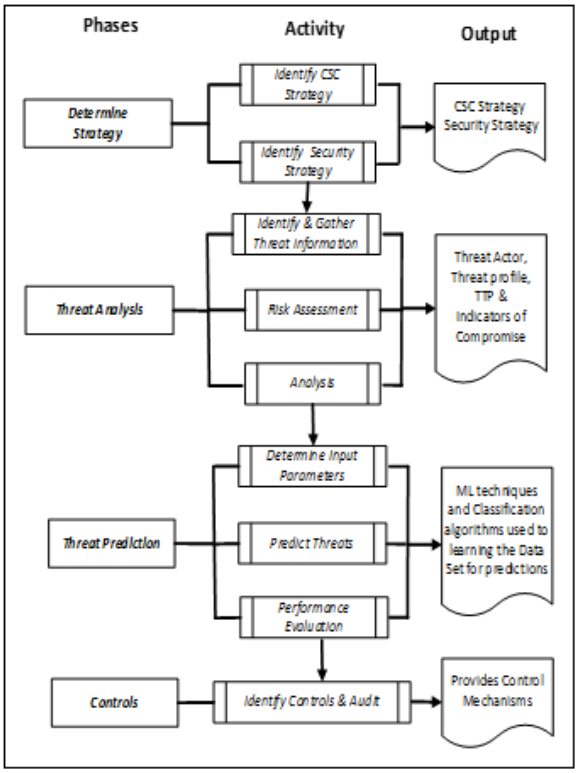


FIGURE 3. Predictive analytics process.

B. PHASE 2: THREAT ANALYSIS

This threat analysis phase follows the CTI techniques to determine and analyse the threats of the CSC context. It requires the CSC strategy information for his purpose and includes three activities.

Activity 1: Identify and Gather Information

This step identifies all vulnerable spots on the supply inbound and outbound chains on the meta-model that is used as indicators for an attack. For instance, in case of a malware attack, this activity looks for the relevant information such as the source of the attack, the tools, patterns and the attack vectors from the analysis of the malware attack that used as our indicator. To determine the indicators of an attack, we use threat activities, adversary behaviours, risky events, or state of the incident to determine what could serve as an indicator. The indicators may be used to identify any inherent vulnerabilities that could be exploited by a threat actor. If necessary, the activity carrying out penetration testing, vulnerability assessment test and threat propagation exercises to determine the supply inbound and outbound chains on the OT and IT by following the below stages [2].

Activity 2: Identify and Gather Information

This step identifies all vulnerable spots on the supply inbound and outbound chains on the meta-model that is used as indicators for an attack. For instance, in case of a malware attack, this activity looks for the relevant information such as the source of the attack, the tools, patterns and the attack vectors from the analysis of the malware attack that used as our indicator. To determine the indicators of an attack,

we use threat activities, adversary behaviours, risky events, or state of the incident to determine what could serve as an indicator. The indicators may be used to identify any inherent vulnerabilities that could be exploited by a threat actor. If necessary, the activity carrying out penetration testing, vulnerability assessment test and threat propagation exercises to determine the supply inbound and outbound chains on the OT and IT by following the below stages [2].

- Stage 1. Reconnaissance: The threat actor uses APT methods to gather intelligence and searches the organization's websites to gather footprints and identify vulnerable spots on the network nodes.
- Stage 2. Experiment: The threat actor uses penetration testing and vulnerability assessment methods various attack patterns, TTP methods, and tools to explore vulnerable spots. The attacks include spear phishing malware or Remote Access Trojan.
- Stage 3. Exploit: the threat actor initiates attack to gain access to the system and other resources of the system. The attack could manipulate, alter and redirect deliveries or initiate and propagate malware.
- Stage 4. Command and Control**:** The threat actor maintains a continuous presence on the system and can change his password to maintain a presence on the CSC using advanced persistent threat attack, remote access command to steal intellectual properties and cause cyber espionage attacks. Most organizations use automated password changing system that prompts users to change their password periodically and that could be exploited by the threat actor. The threat actor can change the password and obfuscate in a Command & Control environment [2].

Activity 3: Risk Assessments

The risk assessment activity includes the process to mitigate CSC risks by determining the probability and impact of CSC attacks and threats as well as the vulnerable spots that could be exploited within the cyber supply inbound and outbound chains and third-party organizations. It identifies all threats that may pose a risk on the system. Risk assesses the CSC security domain and analyse risks access spots that are capture captured. Develop mitigating techniques to control the risks by identifying risks posed by auditing the third-party organizations. Classify them based on their service provisions and levels of integration to the various supply chain network system.

Activity 4: Analysis

This activity focuses on analysis of the threats to determine the actual source of the attack, the type of attack, the attack pattern, the TTP and attack vectors. This will assist to assign the IoC required and what controls are needed. The threat analysis techniques include:

- Stage 1. Threat Activity: Determine the nature of attack, pattern and sources of penetration on the CSC.
- Stage 2: Threat Manipulation: Determines the nature of cybercrimes committed and the extent of the penetration to understand the capabilities, motives and intents of the attacker.
- Stage 3: Threat Impact: Determines the severity of the attack, malware propagation and the cascading effects on the supply chain. These determinants influence the risk factors and the degree of severity of the attacks.

C. PHASE 3: THREAT PREDICATION

The phase considers CSC system nodes that are vulnerable to cyberattacks by integrating CTI and ML to obtain attack predictions of known and unknown attacks using three sequential activities.

Activity 1: Determine Input Parameters

The input parameters mainly consider the attack and TTP to demonstrate how the attackers penetrate a system. In particular, threat actors' properties such as capability and attack vector, tools are used for the input parameters.

- Step 1: Feature Selection: This step includes different ML techniques to select the available features that exist in the data. These feature selection techniques include dimensionality reductions in large datasets for effective and reliable training, testing and prediction. The features we use for our prediction are malware, spyware, spear phishing and Rootkit attacks.
- Step 2: Choosing a Classifier and Performance Metrics: We classify the various algorithms such as LR, DT, SVM and RF in VM to determine (1) the different types of responses based on an attack and (2) different types of response give the TTP deployed. For our study, we use the binary classification as it supports AUC-ROC in distinguishing between the probabilities of the given classes. Further, its precisions can predict correct instances, provides a harmonic mean of precision and recall for the F-score. Determining the right performance metrics to evaluate the algorithms, influences the performance measures and how the algorithm are compared with others. Not using the right metrics could cause overfitting problems and impact on how we evaluate our predictions.

Activity 2: Predict Threats

This activity aims to predicate vulnerabilities and IoC as output feature. The vulnerabilities provide the organization intelligence about areas that are exploitable and the IoC provides the indicators of penetrations, cybercrimes compromises, APTs and C&Cs. Using the cyber threat analysis and the inputs features, we use ML techniques and dataset to predict the output features. The vulnerable spots include network nodes, firewalls, antivirus and anti-malware. The IoC includes the unknown attacks and the extent of cybercrime manipulations, alteration, deletions, exfiltration and redirections that the threat actor could deploy on the system. The stealthy nature of such attacks is so uncertain it cannot be determined on the face value. This includes gathering various attack probabilities and their propagation effects on the CSC using ML techniques to train and test dataset to learn and to gain accurate predictions. The process involves:

- Applying ML techniques to learn the data events from IDS/IPS and firewall logs to collect signatures, threat indicators and, antimalware logs from the various supply chain endpoints. The ML techniques consider LR, SVM, DT, RF and MV algorithms to determine the accuracies of our predictions.
- Determining false positives and false-negative rates.
- Analyse ML results, logs and alerts to understand the attack trends as identified in the initial process to gather intelligence as to what happened, how, why, when, who and where the attack is initiated from.

Activity 3: Performance Evaluation

The performance of the models will be evaluated based on the following values: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Further, the FP and FN will be determined based on the elements of the confusion matrix. We follow the following steps for the performance evaluation.

Step 1: Using Confusion Metrics to Determine TP and FP Outcomes

A confusion matrix is a two-dimensional matrix that evaluates the performance of a classification model with respect to a specific test dataset. It basically compares the actual target values with those predicted by the machine learning model. It provides a better understanding of the values by calculating the data in the matrix and analyse them to determine any positive or negative classifications. Four outcomes are determined when classifying the instances of the dataset. These include True Positive (TP), True Negative (FP), False Positive (FP) and False Negative (FN) rates. For instance, in an event where an instance is positive, and the outcome is classified as positive, its TP else its FP. Where the instance is negative and the outcome is classified as negative, it is counted as TN, else it is FN [15]. We consider the following method to understand the confusion matrix. The accuracy of the confusion metric is the proportion of the total number of predictions that are considered as accurate. We use the following equation below to determine the TPR, TNR, FPR, FNR and the entropy.

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

The recall or true positive rate (TPR) is the proportion of the total number of correct predictions. We consider the equation as:

$$TPR = \frac{TP}{FN + TP} \tag{2}$$

Finally, precision (P) is the proportion of the predicted positive cases that were determined as correct. Hence the formula:

$$P = \frac{TP}{FP + TP} \tag{3}$$

F-measure of F1 – Score (F) is used as the harmonic mean to determine the combinations of precision and recall. We use the formula as:

$$F = \frac{2(Precision\ X\ Recall)}{Precision + Recall} \tag{4}$$

Step 2: Determine Mean Absolute Error (MAE) and Mean Square Error (MSE)

MAE determines the sum of the absolute mean or normal curve of the difference vector between predicted and real values. Whereas MSE determines the mean or normal difference by taking the absolute value of the square root of the mean and convert the units back to the original unit of the output variable and provide a gross idea of the magnitude of the error. For us to predict real numbers or regressions, we used MAE and MSE. The activities include Import AUC-ROC Function, Import Mean Absolute Error, Import Mean Square Error, and Set Entropy Criterion. Entropy is a concept used in information theory to determine the measure of uncertainty about the source of data. It is a unique function that satisfies the four uncertainties axions in a confusion matrix and gives us the degree of disorganization in our data. In an event where a given set of data may contain random collections of unstructured data, and entropy formula is used to separate the positive and negative rates as follows:

$$\text{Entrophy(E)} = -a \log_2 a - b \log_2 b \tag{5}$$

where a = Proportion of positive examples and b = Proportion of negative examples. We use the formula to determine the results in our experiment. We ask the following question to derive the answer from the performance.

- TP = Did the model predicted correctly for the positive class as positive?
- TN = Did the model predicted correctly for the negative class as negative?
- FP = Did the model predicted incorrectly the negative class as positive?
- FN = Did the model predicted incorrectly the positive class as negative?

D. PHASE 4: CONTROL

This final phase aims to identify a list of controls that are to tackle the threat. The controls should ensure that the required security strategic and mechanism are put in place to mitigate the threats. This includes identifying security requirements, internal and external audit as well as threat monitoring and reporting. The process includes identification and review of existing controls, third-party audit and finally information sharing.

VI. IMPLEMENTAION

This section follows the implementation of the proposed approach to determine the applicability of our threat prediction. We only follow threat identification, prediction, and control phases for the implementation.

A. THREAT ANALYSIS

Threat analysis phase uses CTI approach to gather threat. We identify vulnerabilities on the network nodes, IP address,

IEDs and the threats that are linked to the organizational goal that provide us with threat indicators. This includes the TTP used by threat actors and their modes of operations. For our analysis, we adopt the attack concepts and the properties from the meta-model to determine the attack pattern and the TTP deployed on the CSC. The phase involves gathering sources of attacks, vulnerable spots, risks TTPs. Data are gathered from firewalls logs, collecting a signature, threat indicators and events from IDS/IPS, antimalware logs from the various endpoints.

B. THREAT PREDICATION

Further to the discussion in Section 4, threat prediction involves using ML techniques to learn dataset for threat predictions of known and unknown attacks. We follow the ML process for our threat prediction.

1) DESCRIPTION OF DATA

We have considered the widely used dataset from a Microsoft Malware website for the implementation [6]. The dataset is about malware attacks in the Microsoft endpoint system. The data was collected by Microsoft Windows Defender with over 40,000 entries, with 64 columns and each row represents different telemetry data entries. The data represents malware attacks identified on various endpoint nodes from different locations with machine identities, timestamps, organizational identifier and default browser identifiers designed to meet various business requirements. The rationale for using the dataset is that the dataset does not represent Microsoft customer's machine only as it has been sampled to include a much larger proportion of malware infection machines. Therefore, we used this dataset for our predictive analytics as CSC systems integrate various network infrastructures for the business process and interoperability.

The feature description includes MachineIdentifier that considers individual machine ID on the network, GeoNameIdentifier, provides IDs for the geographic region a machine is located in. DefaultBrowsersIdentifier, provides ID for the machine's default browsers. OrganizationIdentifier, provides ID for the organization the machine belongs in. Is protected, provides a calculated field derived from the Spynet Report's AV Products field. Processor considers the process architecture of the installed operating system. HasTpm, indicates true if the machine has TPM (Trusted Platform Module). Over, looks at the version of the current operating system. OsBuild, information indicating the build of the current operating system. Census_DeviceFamily AKA DeviceClass, indicates the type of device that an edition of the OS is intended for desktop and mobile. Firewall, this attribute is true (1) for Windows 8.1 and above if windows firewall is enabled, as reported by the service [6].

2) DATA PREPARATION

The activity involves uploading the data from a website APIs or an HTML file and selecting the data we need then save it as CSV file. We prepare the data by converting the average of the columns of the dataset. Furthermore, we loaded the data from a pre-prepared dataset by calling the categories of the machine learning identifier: The output generated 40,000 training datasets with 62 variables. Handling NaN (Not a Number) in training set by using a command that removes all the NaN in the training set into the dictionary and prints the output. Furthermore, we create a NaN dictionary to handle all the unwanted duplicate data. The output prints $62 - 8 = 54$. (8 columns removed).

3) FEATURE SELECTION

The main features are identified from the primary dataset that are relevant to our work. There were 62 features in the primary data and the focus is on the concepts of attacks, tools and vulnerabilities from our previous work. We characterized threat actor activities, including presumed intent and historically observed behaviour, for the purpose of ascertaining the current threats that could be exploited. Further, we identified eight vulnerable spots and their probability that the cyber attacker could exploit those spots namely the: Firewall, IDS/IPS, Vendors CSC system, Network, IP Addresses, Database, Software, and Websites.

4) BUILDING NEW FEATURES INTO THE DATASET

The features considered as input parameters for the predictions are the attack and TTP as discussed in Section 3.2. To achieve that, we determine the types of attack, tools, vectors, and capabilities for the input. we build the features in line with the existing dataset feature description in [6]. Further, features for predicting the attack inputs and outputs are identified by deriving new features that are in line with the existing datasets and features [6] in Table 2. These features and variables are related to the dataset for our work. Attack patterns are an abstract mechanism for describing how a type of observed attack is executed [32]. The output parameters are determined after our evaluation using the attack pattern, TTPs, vulnerabilities as indicators of compromise. Furthermore, the attack profiles for the ML prediction are built-in dataset. The main goal of our work is to be able to build attack profiles for our ML to predict which node is vulnerable and likely to be attacked. We may not be able to use exact features, but we consider characteristics that are correlated with them and are relevant to represent how the attacks are initiated and the vulnerabilities are exploited for our future prediction. Hence, many features that we analysed were chosen to represent the CTI and security awareness of the stakeholders.

5) CHOOSING AN OPTIMIZATION ALGORITHM FOR THE CLASSIFIERS

For us to choose the classifiers as discussed in Section 4.1.3. activity 1, step 2. we used a pipeline to connect the various classifications. We use the 10-Fold cross-validation to determine the parameter estimation. The 10-Fold cross-validation run and validate the parameter ten times on each algorithm as the values may change and may not generate the accurate

TABLE 1. Matrix to compute the accuracy, precision, recall and the F-score.

Number = 185	Predicted Yes	Predicted No
Actual Yes	TP = 180	FN = 20
Actual No	FP = 40	TN = 120

result when we run it only ones. For the test, we used 10-fold cross validation for more accurate predictive results. The GridsearchCV provides an exhaustive search over specified parameter values for an estimator. We combine all the four algorithms using Majority Voting (MV) algorithm in the classifiers to determine the mean score of the total results. Finally, we use ROC-AUC to distinguish between the accuracies of the binary classification for the predictions [32].

6) EVALUATING THE ACCURACY OF THE THREATS

We consider the following method to understand the confusion matrix as discussed in Section 5. The accuracy of the confusion metrics is the proportion of the total number of predictions that are considered as accurate. Using the equation in Section 5, we evaluate the accuracies (AC) of the metrics to answer the performance of the TP, TN, FP, FN rates in (V) as follows:

$$AC = \frac{180 + 120}{180 + 120 + 40 + 20} = 0.83 \tag{6}$$

Using the Table 3, and the algorithm, we answer the following question to derive the values for the performances. The False positive rate (FPR) determines the rate of negative cases that were incorrectly classified as positive.

• FP = Did the model predicted incorrectly the negative class as positive rates?

$$AC = \frac{40}{120 + 40} = 0.23 \tag{7}$$

The result indicates that FPR of 0.25 negative cases were incorrectly classified as positive. Whereas the true negative rate (TNR) is defined as the number of negative cases that were classified.

• TN = Did the model predicted correctly for the negative class as negative?

$$TNR = \frac{120}{40 + 120} = 0.75 \tag{8}$$

The result indicates TNR of 0.75 were the number of negative cases that were classified as negative.

Further, the false negative rate (FNR) is the proposition of positive cases that were incorrectly classified as negative.

• FN = Did the model predicted incorrectly the positive class as negative?

$$TNR = \frac{20}{180 + 20} = 0.1 \tag{9}$$

The results indicate that the FNR of 0.1 was the proposition of positive cases that were incorrectly classified as negative. The recall or true positive rate (TPR) is the proportion of the total number of correct predictions. We consider the equation as:

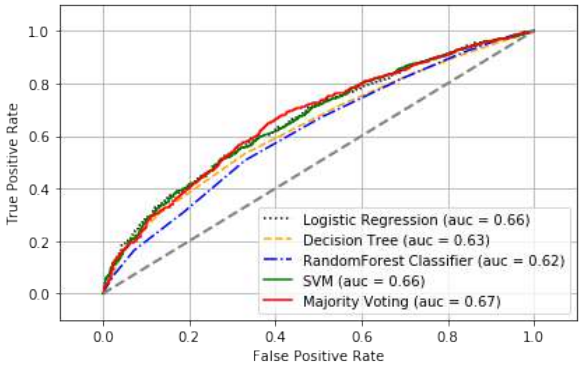


FIGURE 4. Plot the accuracy of all the algorithms in ROC curve for the LG, DT, RF, and SVM in MV.

• TP = Did the model predicted correctly for the positive class as positive?

$$TPR = \frac{180}{180 + 20} = 0.9 \tag{10}$$

The result indicates that the Recall or TPR of 0.9 was the proportion of the total number of instances that were identified correctly from the positive classes. To predict positive cases, we use precision (P) to determine the number of the proportion of instances is considered as correct. Hence the formula:

$$TPR = \frac{180}{180 + 40} = 0.81 \tag{11}$$

The final precision (P) of 0.81 was determined as the proportion of the total number of positive instances that were predicted correctly. The results show that the precision, recall and F-Score used to determine the accuracy and precision of the predictions are considered as accurate between the positive and negative rates. The result indicates that the F-Score of 0.85 was the harmonic mean between precision and recall. The Entropy is 0 if all member of E belongs to the same class, or 1 if they have the same number of samples in each group. The function entropy varies in range from 0 or 1.

7) ACCURACY OF THE ALGORITHMS IN ROC-AUC

Figure 4 depicts the ROC curve that determines the binary classifier system that determines the thresholds of the algorithms. We used AUC_ROC (Area Under Curve – Receiver Operating Characteristics) to model the selection metric for the bi-multiclass classification problem to distinguish between the probabilities of the given classes. AUC_ROC determines the True Positives Rates and False Negatives Rates. We plot the accuracy of all the algorithms in ROC. A 10-fold cross validation was used to determine the accuracy of the LR, DT, SVM and RF algorithms in the ROC. The black, orange, blue and green colours represent the algorithms. The x-axis represented as True Positive Rate and y-axis as False Positive rate. We used a python script to plot the graph as given in Figure 4:

8) 10-FOLD CROSS-VALIDATION

- [*ROCAUC* : 0.66 (+/ − 0.02) *LogisticRegression*]
- *ROCAUC* : 0.63 (+/ − 0.02) [*DecisionTree*]
- *ROCAUC* : 0.62 (+/ − 0.02) [*RandomForest*]
- *ROCAUC* : 0.66 (+/ − 0.02) [*SVM*]
- *ROCAUC* : 0.67 (+/ − 0.02) [*MajorityVoting*]

The results indicate that LG and SVM produced the highest results after we have used the ROC-AUC.

9) DETERMINING THE F-SCORE USING RECALL AND PRECISION RATES

For us to determine the precision, recall, and F-score, we answer the following questions regarding Table 1. Precision: how many positive instances were predicted correctly? Recall: how many instances were identified correctly from the positive classes? F-score: what is the harmonic mean between precision and recall? Using the results from evaluations in (I), we determine the F-Score and used the figures from the recall (0.9) and precision (0.81) to calculate the harmonic mean.

$$F = \frac{2 * 0.81 * 0.9}{0.81 + 0.91} = 0.85 \tag{12}$$

10) INCORPORATING ML AND CASE STUDY FOR EXPERIMENTATION

For us to determine the level of penetration, manipulation and the probability of an attack. We used a case study scenario of the remote CSC attack in [2] as below. The percentages figures were determined using the formula for calculating conditional probabilities in [2] from a low of 1 to a high of 100. The percentage figures in the penetration list are used for the result. The following is the scenario and the table from [2].

11) SCENARIO 1. REMOTE ATTACK ON THE CSC SYSTEM

The organization security team found that an adversary had intruded in the CSC system. The threat actor had compromised the workstation of the CMS that interfaced with suppliers, distributors, and third-party vendors. The organization's electronic products had been altered for some time. The CMS generated inaccurate customer electricity consumptions, which compromised the amount the customers were paying for their utility bills, their online payments, and third-party vendor systems. The organization used two types of payment systems, the prepaid system and post-paid system, that were all integrated into the CMS and HEMS. Using the formula for calculating conditional probabilities [2] and Activity 1 and Table 4, we determined the vulnerable spots, the severities of manipulation in percentages, and threat indicators. The percentages figures were calculated using the formula for calculating conditional probabilities. Further, the figures in penetration list are used to calculate the precision, recall and F-Score in Section 6 for the results.

VII. EXPERIMENTAL RESULTS

This section presents and analyses the results of the threat prediction. We follow a number of assessment parameters such as attack probability, TTP, vulnerable spots, and IoC for this purpose. The attack probability figures are derived from Table 2. The propagation is determined using a probability scale of 0–100%. A percentage score was given after calculating the degree of severity of each manipulation. Form low (≤15%), medium (16% to 59%), or high (above 60%).

- *Prediction of an attack probability.*

Table 3 presents the performance of the classifications of LR, DT, SVM, RF algorithms in identifying the various responses of cyberattacks based on the given malicious attack. From the table, LR achieved an accuracy of 66%, DT, 63% SVM 62% and RF 66%. Comparing the performance of the classifiers, LR and RF both performed better for the Precision, Recall and F-Score, whilst DT and SVM received a low precision, recall and F-score. Comparing that to the attack's categories signifies that Malware, Ransomware and spyware attacks identified different types of responses with 85% accuracy.

- *Prediction of TTP deployed based on the response of the cyberattacks.*

Table 4 presents the performance of the classification algorithms in identifying the various TTPs deployed, and responses based on the given attack vectors. Comparing the TTPs against the attack categories, XSS, session hijacking and RAT attack, DT and SVM achieved a low content for the low precision recall and F-score. However, LR received the highest precision and F-score for malware attack with 83% accuracy for TTPs deployed. Furthermore, ransomware and spyware attacks identified different types of responses for the TTPs with 83% accuracy for the harmonic mean in identifying the attack vectors being rootkit, email attachments and RAT.

- *Prediction of vulnerable spots based on the different types of responses of cyberattacks*

Table 5 presents the performance of the various classifications of the LR, DT, SVM and RF algorithms in identifying the vulnerable spots based on the different types of responses of cyberattacks. The vulnerable spots were identified from the CSC system probable threats table in [2] and used the manipulations figures for precision, recall and F-Score. LR and RF achieved a similar accuracy of 87% for the precision and F-score the successful attacks that signify the probability of exploits on the network nodes. Further, attacks such as malware and ransomware received higher precision based on the exploits and TTPS deployed with 92% accuracy. Whilst spear phishing, session hijacking and DDoS performs lower with the DT and SVM classifiers.

- *Predication of indicators of compromise (IoC).*

Table 6 presents the performance variations of the various classifications algorithms that identify what constitutes as indicators of compromise. With DDoS attack, RF presented the highest precision values of 83% compare to SVM indicating the extent of compromises on the network. LR received

TABLE 2. Probability and threat indicators.

Scenario	Vulnerable Spots	Penetration	Manipulation (%)	Probability	Threat Indicators
1	Firewall	Y	70	High	Wrong Firewall
2	IDS/IPS	Y	60	High	Configuration
3	Vendor	Y	80	High	Audit
4	Network	Y	40	Medium	Sub-netting
5	IP	Y	55	Medium	Segmentation
6	Database	Y	75	High	Sanitizations
7	Software	Y	75	High	Reprogram
8	Website	Y	90	High	SSL/TLS

TABLE 3. Predict the probability of an attack from the various endpoints.

ALGORITHMS	R			DT			SVM			RF		
ACCURACY (%)	**66**			**63**			**62**			**66**		
ATTACKS	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
XSS/Session Hijacking	0.88	0.38	0.65	0.58	0.42	0.68	0.55	0.38	0.63	0.88	0.38	0.65
Spyware/Ransomware	0.90	0.55	0.75	0.85	0.37	0.70	0.65	0.45	0.63	0.90	0.55	0.75
Spear Phishing	0.81	0.17	0.71	0.55	0.28	0.66	0.58	0.36	0.63	0.81	0.17	0.71
Session Hijacking	0.73	0.36	0.62	0.48	0.35	0.61	0.55	0.38	0.63	0.73	0.36	0.62
Rootkit/DDoS	0.56	0.37	0.65	0.57	0.33	0.58	0.53	0.35	0.63	0.56	0.37	0.65
RAT/Island Hopping	0.68	0.30	0.73	0.55	0.22	0.69	0.51	0.25	0.63	0.68	0.30	0.73
Ransomware/Malware	0.88	0.53	0.60	0.59	0.26	0.71	0.54	0.31	0.63	0.88	0.53	0.60
Malware/Spyware	0.81	0.48	0.68	0.58	0.51	0.73	0.55	0.45	0.63	0.81	0.48	0.68
DDoS	0.78	0.36	0.65	0.55	0.33	0.55	0.51	0.32	0.53	0.78	0.36	0.65

TABLE 4. Identify the different TTP deployed based on the response of the cyberattacks.

ALGORITHMS	LR			DT			SVM			RF		
ACCURACY (%)	**66**			**63**			**62**			**66**		
ATTACKS	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
XSS/Session Hijacking	0.82	0.26	0.55	0.55	0.31	0.61	0.55	0.27	0.56	0.82	0.26	0.55
Spyware/Ransomware	0.88	0.51	0.71	0.65	0.33	0.62	0.65	0.31	0.61	0.88	0.51	0.71
Spear Phishing	0.71	0.23	0.61	0.53	0.22	0.56	0.58	0.36	0.59	0.71	0.23	0.61
Session Hijacking	0.63	0.26	0.58	0.52	0.28	0.52	0.56	0.38	0.48	0.63	0.26	0.58
Rootkit/DDoS	0.51	0.27	0.63	0.51	0.31	0.58	0.48	0.35	0.57	0.51	0.27	0.63
RAT/Island Hopping	0.68	0.28	0.68	0.54	0.21	0.61	0.51	0.25	0.58	0.68	0.28	0.68
Ransomware/Malware	0.86	0.44	0.66	0.58	0.22	0.65	0.59	0.31	0.62	0.86	0.44	0.66
Malware/Spyware	0.79	0.41	0.67	0.65	0.51	0.63	0.55	0.45	0.61	0.79	0.41	0.67
DDoS	0.71	0.36	0.61	0.55	0.33	0.55	1.55	0.32	0.53	0.71	0.36	0.61

TABLE 5. Predict vulnerable spots based on the different types of responses of cyberattacks.

ALGORITHMS	LR			DT			SVM			RF		
ACCURACY (%)	**66**			**63**			**62**			**66**		
ATTACKS	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
XSS/Session Hijacking	0.63	0.60	0.61	0.65	0.61	0.62	0.61	0.59	0.60	0.62	0.59	0.61
Spyware/Ransomware	0.85	0.83	0.80	0.86	0.81	0.83	0.82	0.79	0.81	0.83	0.78	0.80
Spear Phishing	0.68	0.62	0.66	0.63	0.59	0.61	0.64	0.60	0.62	0.63	0.61	0.68
Session Hijacking	0.66	0.61	0.64	0.65	0.61	0.64	0.62	0.59	0.60	0.63	0.60	0.62
Rootkit/DDoS	0.64	0.60	0.61	0.63	0.61	0.58	0.61	0.57	0.59	0.64	0.38	0.58
RAT/Island Hopping	0.64	0.61	0.63	0.65	0.62	0.64	0.64	0.61	0.62	0.64	0.33	0.58
Ransomware/Malware	0.84	0.81	0.82	0.85	0.81	0.84	0.61	0.58	0.60	0.75	0.55	0.62
Malware/Spyware	0.82	0.77	0.81	0.86	0.83	0.85	0.85	0.81	0.83	0.66	0.45	0.69
DDoS	0.65	0.61	0.62	0.64	0.60	0.63	0.62	0.59	0.61	0.75	0.33	0.62

the highest precision and F-score for malware and spyware attacks, whereas RF and LR received the similar precision, recall and F-score.

VIII. DISCUSSIONS

The results for the predictive analytics are analysed in AUC_ROC as indicated in Figure 4. A 10-Fold cross-validation was used to run each algorithm to determine the parameter estimation and validated the accuracies. The evaluation of the accuracies of the metrics to answer the performance of the TPR, TNR, FPR, FNR as shown in Table 3. We determine the harmonic mean for the proportion of the total number of accuracies for the precision, recall, and F-score. The proportion for the precision is 220 for the

TABLE 6. Indicators of compromise (IOC). FOR performance variations of the various classifications algorithms.

ALGORITHMS	LR			DT			SVM			RF		
ACCURACY (%)	66			63			62			66		
ATTACKS	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
XSS/Session Hijacking	0.68	0.63	0.66	0.55	0.42	0.61	0.51	0.38	0.63	0.68	0.37	0.71
Spyware/Ransomware	0.80	0.8	0.75	0.85	0.55	0.70	0.65	0.45	0.63	0.78	0.52	0.76
Spear Phishing	0.81	0.17	0.71	0.55	0.65	0.70	0.55	0.45	0.63	0.77	0.17	0.68
Session Hijacking	0.73	0.66	0.62	0.55	0.65	0.70	0.55	0.45	0.63	0.73	0.65	0.62
Rootkit/DDoS	0.56	0.37	0.60	0.55	0.65	0.70	0.55	0.45	0.63	0.56	0.37	0.59
RAT/Island Hopping	0.68	0.30	0.33	0.55	0.65	0.70	0.55	0.45	0.63	0.68	0.30	0.63
Ransomware/Malware	0.70	0.33	0.62	0.55	0.65	0.70	0.55	0.45	0.63	0.72	0.33	0.60
Malware/Spyware	0.74	0.48	0.65	0.55	0.65	0.70	0.55	0.45	0.63	0.71	0.48	0.65
DDoS	0.68	0.56	0.65	0.55	0.65	0.70	1.55	0.45	0.63	0.68	0.56	0.57

TABLE 7. Mapping the attack category and predictive analytics.

Attack Category	CSC Attack Features	Threat Descriptions for Probable Cause of Attack	Threat Predictions (%)
1	XSS/Session Hijacking	Default Browser vulnerabilities and injecting a code in the URL or website	80
2-5	Spyware/Ransomware	Outdated Antivirus/Patches that are not updated regularly	90
6-7	Spear Phishing	Use Reconnaissance to identify vulnerable spots and attach email with a virus	80
8-9	Session Hijacking	Exploit Unchanged Hard-Coded password in software bought off the shelf	75
10-14	Rootkit/DDoS	Attack on BIOS or attach a virus to a USB key to cascade when booting.	80
15-20	RAT/Island Hopping	Attacks from Vendor systems to gain access to the organizational system	70
21-28	Ransomware/Malware	Exploiting outdated OS versions and encryptions especially TLS/SSL	60
29-35	Malware/Spyware	Packet injection and Resonance attacks	70
36-38	DDoS	Exploit IP Address Systems and Packet injections	55

number of positive instances that were predicted correctly. The proportion of recall (0.9) instances was identified correctly from the positive classes. The F-score of (0.85) was the harmonic mean between precision and recall. Hence, an accuracy of 85% is the total number of predictions that are considered accurate for the TPR and FPR. Further, we have a slight variation in our predictions of the TPF and FPR comparing the LR, DT, SVM, and RF algorithms in the pipeline and using MV for running them. However, the accuracy of the proportion of the total number of predictions remains accurate with an average of 65% and 30% as the combine values for the TPR and FPT respectively. Additionally, the results indicate that LG and SVM produced the highest results after we have used the ROC-AUC. The predictive analysis of our evaluation after we have used the CTI to gather information, gain knowledge and understanding of the organizational context and the situational awareness remains acceptable as compared to other literature that focused on ML only for predictions. The Table 7 shows the list the attack categories and threat predictions.

Table 6 combines the probability of attacks identified from previous work and map them with the feature descriptions of the threats to explains the predictive analytics [2]. The mapping includes attack categories, CSC attack features, and the threat describes for probable cause of attacks from the telemetry data and Microsoft endpoint protection threat report for the predictions. The attack categories were determined from the dataset of various threat descriptions from the telemetry data [23] that contains the properties of the various families of malware generated by the Windows defenders. The CSC attack features were derived from the various families of malware that has the probability of infecting the various CSC endpoint nodes. The threat descriptions were gathered by the threat report collected by the Microsoft Windows Defender [23]. The results specify that spyware/ransomware scored 90%. All the attack categories that score 80% indicated that an XSS or session hijacking could be deployed on the CSC website as uses public facing IPs it connects to various vendors. These could lead to spear phishing, rootkit and DDoS attacks. The rest of the threat prediction scores are explained in Table 7.

The paper reveals several observations made from the CSC attacks to using CTI lifecycle processes for intelligence gatherings, and ML for predictive analysis for the overall Smart CPS security improvement. The study revealed that several challenges are facing the organization in securing their systems as attackers are executing arbitrary commands on the supply chain systems remotely and manipulating systems.

A. MAPPING CYBERATTACKS ON CSC FOR PREDICTIVE ANALYTICS OF INDICATORS OF COMPROMISE

Table 8 provides details of how we mapped the cyberattacks on the CSC system for predictive analytics to determine the indicators of compromise. We used the threat modelling concepts in Section 3, and the properties to identify the

TABLE 8. Output parameters for indicators of compromise.

Cyberattack	Attack Pattern	Vulnerability	TTPs	IoCs
Malware	Insert a program in software	Untested Software	Insert Rootkit in code to hide in the system	Cascade to other networks nodes/ bypass antimalware
RAT	Hide in executable program, Backdoor code in an email attachment. HTTP Request Splitting, downloads	Network, Web and application server, Social Engineering, Phishing	Inject entry point identifier in the Explore Phase	Downloads itself when the user opens an email and provides access to the attacker
XSS	Embed malware in web browser content.	Programs that allow the remote host to execute codes and scripts.	Inject XSS payload and response split syntax in the user control input or URL	Injected scripts cascade to resources accessed by the applications
Ransomware	Social Engineering, Trojan, Botnets and Exploit kits to encrypt system files	Targets outdated antivirus and unpatched MS Windows application system	Map user environment, with documents, pictures and recycle bin and report content to C&C.	Calculate entropy of all file contents on the various systems, encrypt and propagate
Session Hijacking	Uses unauthentic HTTP cookies request from users.	Unencrypted websites, HTTP sessions, and open Wi-Fi connections	Insert network traffic that is not encrypted. Man-in-the-Middle attacks	Gain access and commits, APT, C5C and industrial espionage attacks.

TABLE 9. CSC security controls.

CSC Control	Descriptions	Asset	Approach	Implementation
Directive	Strategic management controls derived from the CTI and ML processes intended for policy formulation.	Identify Critical Assets and Security Framework that meet organizational goal	Map CTI gatherings and ML predictive analytics results to security goal	Assign controls to security teams to oversee the implementation. Adopt a framework or standard to support the development
Preventive	Proactive measures that are required to be implemented. Financial, physical, and technical measures intended to preclude actions violating policy or increasing risk to system resources.	Determine attacks that can exploit assets. Assign risks and threat levels to assets using CSCRM.	Determine Mitigations goals including internal and external audit controls	Create awareness by organize training and workshops to train users
Detective	Develop business impact assessment. Involve the use of practices, processes, and tools that identify and possibly react to security violations.	Implement periodic and ad-hoc security assessment using penetration testing and vulnerability assessment to pre-empt cyber threats	Use impact analysis and cost benefit analysis to determine the cost of alternatives of not investing in detection tools	Configure devices and automate passive tools on CSC systems to flag threats, run and monitor reports of firewalls, IDS/IPS, anti-malware and system updates
Corrective	Involve configurations and countermeasures designed to react to the detection of an incident to reduce or eliminate the zero-day attacks.	Design security policies that inform what must be done in the event of an incident	Develop Asset Inventory of all network nodes connected to the CSC organizational network including DHCP security	Implement Policies and business continuity plan to repair CSC systems, hard drive, patches systems, quarantine CSC systems
Recovery	Recovery strategy, Incident response and back up plans, regular updates, and contingency planning to ensure integrity or availability of the CSC system	Design policies and business impact assessment that can assist to restore the system or operation to a normal operating state upon any compromise as soon as possible.	Develop disaster recovery plan that will restore system to its operational state.	Form a team and Organize training and workshops to train staff to understand and be aware of the DRP implementations.

cyberattack, the attack pattern that were used, the vulnerable spots that were exploited, and the TTPs that are deployed by the threat actor on the CSC systems as the indicators of compromise (IoC). Indicators of compromise are parameters used

to express whether an attack-type is imminent, in progress or has occurred. Refer [2] further reading on threat modelling. Threat actors use sophisticated and stealthy methods to inject a virus, worms, bugs or a Trojan into software or in an HTTP request in an 'Island Hopping' attack. The intent is to penetrate the network or gain access to the webserver when a request is being processed. The motive could be to manipulate the vulnerable spots, alter the software and delivery channels and maintain APT and command & control presence.

Using the C&C methods, the attacker can modify products during manufacturing, manipulate it during distributions and the various domain attacks. These attacks could cascade to other nodes on the supply inbound and outbound chains. The table below provides a matrix that blends the input and output parameters for the prediction. Our observation is that the following vulnerabilities exist in the cyber supply chain system:

- The supply chain variables are accessible to the threat actor due to the business applications used for the supply chain variables and that could be exploited using incorrect user data.
- Information retrieved through inputted data is not configured properly due to poor validation.
- The variables are not well encapsulated to prevent software redirect. For instance, setting an input variable as public in a class when developing the software source codes makes the website open to external attackers.

B. MACHINE LEARNING FOR PREDICTIVE ANALYTICS

Machine learning approach to cybersecurity has been effective in analyzing and predicting future attacks and attack trends. We use ML techniques and classification algorithms including LD, SVM, DT, RF, and MV to develop threat intelligence techniques that can predict which nodes on our CSC system are venerable to attacks. We plot the accuracy of all the algorithms in ROC. AUC_ROC to determine the true positives and true negative rates. The results show that the best parameter result was SVM with an accuracy of 0.66. ML provides us with the ability to combine algorithms to determine which of them produced the highest accuracy and output for the best parameter for our prediction. However, it does not provide us with the ability to understand the threat actor's motives and intents.

C. COMPARING RESULTS WITH EXISTING WORKS

A stated in the related works, there have a lot of attention of using ML classifiers for cyber security. A vector space model is used for information retrieval for HTTP attacks using a decision tree algorithm to automatically label the request as malicious in the URL[11]. A number of classification algorithms LR, DT, NB, and SVM are considered for cloud security and tested the models in diverse operational conditions using cloud security scenarios[20]. Further, [21] used data mining and ML methods on Artificial Neural Network, Association rules, Fuzzy Association rules and Bayesian Networks classifiers for cybersecurity detection and analytics in intrusion detection security applications. Furthermore, [22] compared ML datasets used for analyzing network traffic and anomaly detection relevant for modern intrusion detections datasets. Moreover, [23], explored the classification of logs using a decision tree algorithm that models the correlation and normalization of security logs. Similarly, [24] compared NNge. LF, DT, Naïve Bayes, and SVM classification algorithms performance and ML predictions for power system disturbance and cyberattack discriminations. Then, [25] used an instance-based learning classification algorithm to learn a dataset for feature reduction and detection techniques to detect cyberattacks on smart grid. Additionally, [26] used an ensembled learning model based on the combination of a random subspace with random tree to detect cyberattacks on Industrial IoT networks. Likewise, [28] explored mitigating techniques on IoT cybersecurity threats in a smart city by using ML techniques to learn dataset on LR, SVM, DT, RF, ANN and KNN classifiers for anomaly detections. Further, [29] proposed a novel adaptive trust boundary protection for Industrial IoT network by using deep learning on a semi supervised model for detecting unknown cyberattacks. Furthermore, [30] used deep neural network discriminator on a down sample encoder cooperative data generator train the algorithm to capture actual distribution of attack model on industrial IoT attack surface. Additionally, authors in [31] predicted cybersecurity incidents by using Naive Bayes and SVM algorithms to investigate and analyse various datasets collected from SMEs. Finally, [32] model a risk teller system that used ML to predict which machines are at risk of getting infected or are clean and forecast if an organization may experience cybersecurity incidents in the future. Though all the works are relevant and contribute for the cyber security improvement. However, there is a lack of focus on the overall CSC security and ML classifiers are mainly used datasets for the threat predication. The proposed work presents a conceptual view by integrating relevant concepts from CSC and CTI domain. It provides a systematic threat analysis using the CTI techniques and integrates ML classifiers for the threat predication. Additionally, we considered LG, DT, SVM, RF algorithms in Majority Voting to learn the malware threat prediction dataset.

D. CSC SECURITY CONTROLS

There are various security controls in existence, whose effectiveness are based on existing CSC attacks and risks including CIS Controls 2018 and ISO27002:2011. We recommend the approach to address the CSC security using threat intelligence gathered from known and unknown attacks in line with organizational objectives and provide security recommendations. Some organizations provide a recommendation, however, not all may be relevant to the cyber supply chain organizational objective. Table 9 identifies basic concepts that are required to maintain security controls in the supply chain environment. To incorporate cybersecurity controls into a cyber supply chain system, we use knowledge of actual

CSC attacks that have occurred in the past. A compromised supply chain system provides us with the knowledge of previous attacks to continually learn from and build effective and practical defences mechanisms. To ensure proper CSC security controls, the organization must form a strategic team to identify, investigate, review and evaluate the supply chain system processes and applications.

E. THREAT INFORMATION SHARING

Threat information sharing is essential for any cyber physical system and specifically for the CSC context. It helps supply chain organisations and its stakeholders to aware about the current threat trends so that appropriate control can be identified to tackle the attacks. The CTI information includes threat landscapes, TTPs, tools, and intelligence reports. The threat intelligence is shared amongst the various organizations, institutions, vendors and businesses on the CSC system for strategic management decision making. It designates information and creates situational awareness on the various security alerts, assess and monitor threats, risk and existing controls. Due to the sensitive nature of the intelligence and privacy rules, these organizations are required to sign an agreement to ensure the following:

- Establish Information sharing rules
- Establish security system and audit rules
- Establish rules that govern the sharing of sensitive information
- Establish information classification rules. (Need to Know)

Challenges facing information sharing include the sensitivity nature of cyberattacks and the fact that it could lead to reputational damage, and sometimes legal ramifications. Most organizations are reluctant to share information relevant to CSC security.

IX. CONCLUSION

The integration of complex cyber physical infrastructures and applications in a CSC environment have brought economic, business, and societal impact for both national and global context in the areas of Transport, Energy, Healthcare, Manufacturing, and Communication. However, CPS security remains a challenge as vulnerability from any part of the system can pose risk within the overall supply chain context. This paper aims to improve CSC security by integrating CTI and ML for the threat analysis and predication. We considered the necessary concepts from CSC and CTI and a systematic process to analyse and predicate the threat. The experimental results showed that accuracies of the LG, DT, SVM, and RF algorithms in Majority Voting and identified a list of predicated threats. We also observed that CTI is effective to extract threat information, which can integrate into the ML classifiers for the threat predication. This allows CSC organization to analyse the existing controls and determine additional controls for the improvement of overall cyber security. It is necessary to consider the full automation of the process and industrial case study to generalize our findings. Furthermore, we are also planning to consider evaluating the existing controls and the necessary of future controls based on our prediction results.

REFERENCES

[1] National Cyber Security Centre. (2018). *Example of Supply Chain Attacks*. [Online] Available: https://www.ncsc.gov.uk/collection/supply-chain-security/supply-chain-attack-examples

[2] A. Yeboah-Ofori and S. Islam, "Cyber security threat modelling for supply chain organizational environments," *MDPI. Future Internet*, vol. 11, no. 3, p. 63, Mar. 2019. [Online]. Available: https://www.mdpi.com/1999-5903/11/3/63

[3] B. Woods and A. Bochman, "Supply chain in the software era," in *Scowcroft Center for Strategic and Security*. Washington, DC, USA: Atlantic Council, May 2018.

[4] *Exploring the Opportunities and Limitations of Current Threat Intelligence Platforms, Version 1*, ENISA, Dec. 2017. [Online]. Available: https://www.enisa.europa.eu/publications/exploring-the-opportunities-and-limitations-of-current- threat-intelligence-platforms

[5] C. Doerr, TU Delft CTI Labs. (2018). *Cyber Threat Intelligences Standards—A High Level Overview*. [Online]. Available: https://www.enisa.europa.eu/events/2018-cti-eu-event/cti-eu-2018-presentations/cyber-threat-intelligence-standardization.pdf

[6] Research Prediction. (2019). *Microsoft Malware Prediction*. [Online]. Available: https://www.kaggle.com/c/microsoft-malware-prediction/data

[7] A. Yeboah-Ofori and F. Katsriku, "Cybercrime and risks for cyber physical systems," *Int. J. Cyber-Secur. Digit. Forensics*, vol. 8, no. 1, pp. 43–57, 2019.

[8] CAPEC-437, Supply Chain. (Oct. 2018). *Common Attack Pattern Enumeration and Classification: Domain of Attack*. [Online]. Available: https://capec.mitre.org/data/definitions/437.html

[9] Open Web Application Security Project (OWASP). (2017). *The Ten Most Critical Application Security Risks, Creative Commons Attribution-Share Alike 4.0 International License*. [Online] Available: https://owasp.org/www-pdf-archive/OWASP_Top_10-2017_%28en%29.pdf.pdf

[10] US-Cert. (2020). *Building Security in Software & Supply Chain Assurance*. [Online]. Available: https://www.us-cert.gov/bsi/articles/knowledge/attack-patterns

[11] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Towards the prediction of renewable energy unbalance in smart grids," in *Proc. IEEE 4th Int. Forum Res. Technol. Soc. Ind. (RTSI)*, Palermo, Italy, Sep. 2018, pp. 1–5, doi: 10.1109/RTSI.2018.8548432.

[12] J. Boyens, C. Paulsen, R. Moorthy, and N. Bartol, "Supply chain risk management practices for federal information systems and organizations," *NIST Comput. Sec.*, vol. 800, no. 161, p. 32, 2015, doi: 10.6028/NIST.SP.800-161.

[13] *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*, NIST, Gaithersburg, MD, USA, 2018, doi: 10.6028/NIST.CSWP.04162018.

[14] J. F. Miller, "Supply chain attack framework and attack pattern," MITRE, Tech. Rep. MTR140021, 2013. [Online]. Available: https://www.mitre.org/sites/default/files/publications/supply-chain-attack-framework-14-0228.pdf

[15] C. Ahlberg and C. Pace. *The Threat Intelligence Handbook*. [Online]. Available: https://paper.bobylive.com/Security/threat-intelligence-handbook-second-edition.pdf

[16] J. Freidman and M. Bouchard, "Definition guide to cyber threat intelligence. Using knowledge about adversary to win the war against targeted attacks," iSightPartners, CyberEdge Group LLC, Annapolis, MD, USA, Tech. Rep., 2018. [Online]. Available: https://cryptome.org/2015/09/cti-guide.pdf

[17] EY. (2016). *Cyber Threat Intelligence: Designing, Building and Operating an Effective Program*. [Online]. Available: https://relayto.com/ey-france/cyber-threat-intelligence-report-js5wmwy7/pdf

[18] A. Yeboah-Ofori and C. Boachie, "Malware attack predictive analytics in a cyber supply chain context using machine learning," in *Proc. ICSIoT*, 2019, pp. 66–73, doi: 10.1109/ICSIoT47925.2019.00019.

[19] B. Gallagher and T. Eliassi-Rad, "Classification of HTTP attacks: A study on the ECML/PKDD 2007 discovery challenge," Lawrence Liverpool Nat. Lab., Livermore, CA, USA, Tech. Rep., 2009, doi: 10.2172/1113394.

[20] D. Bhamare, T. Salman, M. Samaka, A. Erbad, and R. Jain, "Feasibility of supervised machine learning for cloud security," in *Proc. Int. Conf. Inf. Sci. Secur. (ICISS)*, Dec. 2016, pp. 1–5, doi: 10.1109/ICISSEC.2016.7885853.

[21] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016, doi: 10.1109/COMST.2015.2494502.
[22] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 2186–2193, doi: 10.1109/BigData.2017.8258167.
[23] E. G. V. Villano, "Classification of logs using machine learning," M.S. thesis, Dept. Inf. Secur. Commun. Technol., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2018.
[24] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *Proc. 7th Int. Symp. Resilient Control Syst. (ISRCS)*, Denver, CO, USA, Aug. 2014, pp. 1–8, doi: 10.1109/ISRCS.2014.6900095.
[25] A. Gumaei, M. M. Hassan, S. Huda, M. R. Hassan, D. Camacho, J. D. Ser, and G. Fortino, "A robust cyberattack detection approach using optimal features of SCADA power systems in smart grids," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106658, doi: 10.1016/j.asoc.2020.106658.
[26] M. M. Hassan, A. Gumaei, S. Huda, and A. Almogren, "Increasing the trustworthiness in the industrial IoT networks through a reliable cyber-attack detection model," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6154–6162, Sep. 2020, doi: 10.1109/TII.2020.2970074.
[27] J. Abawajy, S. Huda, S. Sharmeen, M. M. Hassan, and A. Almogren, "Identifying cyber threats to mobile-IoT applications in edge computing paradigm," *Elsevier Sci. Direct Future Gener. Comput. Syst.*, vol. 89, pp. 525–538, Dec. 2018, doi: 10.1016/j.future.2018.06.053.
[28] M. M. Rashid, J. Kamruzzaman, M. M. Hassan, T. Imam, and S. Gordon, "Cyberattacks detection in IoT-based smart city applications using machine learning techniques," *Int. J. Environ. Res. Public Health*, vol. 17, no. 24, p. 9347, Dec. 2020, doi: 10.3390/ijerph17249347.
[29] M. M. Hassan, S. Huda, S. Sharmeen, J. Abawajy, and G. Fortino, "An adaptive trust boundary protection for IIoT networks using deep-learning feature-extraction-based semisupervised model," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2860–2870, Apr. 2021, doi: 10.1109/TII.2020.3015026.
[30] M. M. Hassan, M. R. Hassan, S. Huda, and V. H. C. de Albuquerque, "A robust deep-learning-enabled trust-boundary protection for adversarial industrial IoT environment," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9611–9621, Jun. 2021, doi: 10.1109/JIOT.2020.3019225.
[31] A. Mohasseb, B. Aziz, J. Jung, and J. Lee, "Predicting cybersecurity incidents using machine learning algorithms: A case study of Korean SMEs," in *Proc. INSTICC*, 2019, pp. 230–237, doi: 10.5220/0007309302300237.
[32] L. Bilge, Y. Han, and M. D. Amoco, "Risk teller: Predicting the risk of cyber incidents," in *Proc. CCS*, 2017, pp. 1299–1311, doi: 10.1145/3133956.3134022.
[33] Y. Liu, A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, and M. Liu, "Cloud with a chance of breach: Forecasting cyber security incidents," in *Proc. 24th USENIX Secur. Symp.*, Washington, DC, USA, 2015, pp. 1009–1024.
[34] *Guide to Cyber Threat Information Sharing*, document NIST 800-150, 2018, doi: 10.6028/NIST.SP.800-150.
[35] S. Barnum, "Standardizing cyber threat intelligence information with the structured threat information expression," V1.1. Revision, STIX, USA, Tech. Rep., 2014, vol. 1. [Online]. Available: https://www.mitre.org/publications/technical-papers/standardizing-cyber-threat-intelligence-information-with-the
[36] A. Yeboah-Ofori, S. Islam, and E. Yeboah-Boateng, "Cyber threat intelligence for improving cyber supply chain security," in *Proc. Int. Conf. Cyber Secur. Internet Things (ICSIoT)*, May 2019, pp. 28–33, doi: 10.1109/ICSIoT47925.2019.00012.
[37] A. Boschetti and L. Massaron, *Python Data Science Essentials*, 2nd ed. Dordrecht, The Netherlands: Springer, 2016. [Online]. Available: https://link.springer.com/content/pdf/10.1007/BF00994018.pdf
[38] A. Yeboah-Ofori, "Classification of malware attacks using machine learning in decision tree," *IJS*, vol. 11, no. 2, pp. 10–25, 2020. [Online]. Available: https://www.cscjournals.org/manuscript/Journals/IJS/Volume11/Issue2/IJS-155.pdf
[39] W. Wang and Z. Lu, "Cyber security in smart grid: Survey and challenges," *Elsevier Comput. Netw.*, vol. 57, no. 5, pp. 1344–1371, Apr. 2013.
[40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995, doi: 10.1023/A:1022627411411.

ABEL YEBOAH-OFORI received the B.Sc. degree in computing and information systems from UEL, the M.Sc. degree in information security and computer forensics, and the Ph.D. degree in cyber security from the School of Architecture, Computing and Engineering (ACE), University of East London, U.K. He is currently a Lecturer with the University of West London. He holds a Postgraduate Certificate in Higher Educations Practices (PgCert) and a Fellow of the British Higher Education Academy (FHEA). He is a Prince 2 Project Management Practitioner, Certified Cyber Security and Digital Forensics Investigations practitioner. He has published journal articles, reviewed a few articles, and provided consultancy services. He was invited in 2018 to participate in Cyber Security Maturity Assessment Program with the Global Cyber Security Capacity Centre, USA, Oxford University, and the World Bank. He was invited to an Advisory and Review Workshop 2017 on National Cyber Security Policy and Strategy by MoC and Council of Europe (CoE) as part of GLACY+ activities. His research interests include cyber security, digital forensics, cyber threat intelligence, cyber-attack modeling, cyber supply chain security and risks, and machine learning.

SHAREEFUL ISLAM was a Visiting Researcher with the National Institute of Informatics (NII), Japan, and SBA Research, Austria. He is currently working as a Senior Lecturer and a Programme Leader with the Cyber Security and Network Program, School of ACE, University of East London, U.K. His research interests include in the area of cyber security, requirement engineering, information systems, and risk management. He has pioneered work in developing risk assessment and treatment method using business and technical goals, modeling language for cyber security risk management. The works are implemented in various application domain including cloud migration, critical infrastructure, and information system. He has published more than 70 articles (H-index 23) and he has led and/or participated in projects funded by the European Union (FP7), Innovate U.K., FwF, and DAAD. He has experience of acting as an Evaluator for national and international funding bodies, including the EPSRC, FwF, and CHIST-ERA. He is a Fellow of the British Higher Education Academy (HEA) and a certified PRINCE 2 and Management of RISK (MoR) practitioner.

SIN WEE LEE received the B.Eng. degree (Hons.) in electronics and computing from Nottingham Trent University, U.K., and the Ph.D. degree in neurocomputing from Leeds Beckett University, U.K. He is currently working with the School of Architecture, Computing and Engineering (ACE), University of East London, U.K. He has published more than 40 refereed articles in high-quality journals and international conferences in neural networks, data analytics, and machine learning. His main research interest and field of expertise are in the neural networks and machine learning for data analytics.

ZIA USH SHAMSZAMAN (Senior Member, IEEE) received the Master of Engineering (M.Eng.) degree from the Department of CICE, Hankuk University of Foreign Studies, South Korea, and the Ph.D. degree from the Insight Centre for Data Analytics, National University of Ireland Galway, Ireland. He is currently working as a Senior Lecturer in computer science with the Department of Computing and Games, Teesside University, U.K. He was involved in several research projects funded by FP7, SFI, Cisco Inc., and ETRI. He worked in the ICT industry over seven years and also achieved few professional certifications, such as CEH, CDCP, CCNA, and JNCIA-ER. His research interests include the IoT, the social IoT, CPS, cybersecurity, artificial intelligence, deep learning, semantic web, and ontologies. He is an Advisory Panel Member in Elsevier.

KHAN MUHAMMAD (Member, IEEE) received the Ph.D. degree in digital contents from Sejong University, Seoul, South Korea, in 2019.

He is currently an Assistant Professor with the Department of Interaction Science and the Director of the Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Sungkyunkwan University, Seoul. His research interests include intelligent video surveillance (fire/smoke scene analysis, transportation systems, and disaster management), medical image analysis, (brain MRI, diagnostic hysteroscopy, and wireless capsule endoscopy), information security (steganography, encryption, watermarking, and image hashing), video summarization, multimedia data analysis, computer vision, the IoT/IoMT, and smart cities. He is serving as a reviewer for over 100 well-reputed journals and conferences, from IEEE, ACM, Springer, Elsevier, Wiley, SAGE, and Hindawi publishers. He is an associate editor of four journals and an editorial board member of five journals.

METEB ALTAF received the Ph.D. degree from Brunel University London, London, U.K., in 2009. Since 2009, he has been with the KACST as an Assistant Research Professor. He was appointed as the Director Assistant for Administrative Affairs and the Director Assistant for Scientific Affairs with the National Center for Robotics and Intelligent Systems. After that, he was appointed as the Director of the National Robotics Technology and Intelligent Systems Center before it become known as the National Center for Robotics Technology and Internet of Things. He has been promoted as a Research Associate Professor. In the meantime, he became the Director of the Innovation Center for Industry 4.0, King Abdulaziz City for Science and Technology. He is currently the Director of the Advanced Manufacturing and Industry 4.0 Center. During his career life, he published number of articles in different well-known ISI journals and in well recognized conferences as well as he is lecturing at the Biomedical Technology Department, King Saud University. He has supervised more than 20 research projects locally and internationally as technology transfer projects.

MABROOK S. AL-RAKHAMI (Member, IEEE) received the master's degree in information systems from King Saud University, Riyadh, Saudi Arabia, where he is currently pursuing the Ph.D. degree with the Information Systems Department, College of Computer and Information Sciences. He has worked as a Lecturer with King Saud University, Muzahimiyah Branch, and taught many courses, such as programming languages in computer and information science. He has authored several articles in peer-reviewed IEEE/ACM/Springer/Wiley journals and conferences. His research interests include edge intelligence, social networks, cloud computing, the Internet of Things, big data, and health informatics.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 June 2024, accepted 22 July 2024, date of publication 25 July 2024, date of current version 6 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3433404

RESEARCH ARTICLE

Research on Quantitative Prioritization Techniques for Selecting Optimal Security Measures

JANG JISOO[1], SUBONG JUNG[2], MYUNGKIL AHN[3], DONGHWA KIM[3], JAEPIL YOUN[4], AND DONGKYOO SHIN[1,5,6]

[1]Department of Computer Engineering, Sejong University, Seoul 05006, South Korea
[2]Defense Future Technology Laboratory, LIG System, Seoul 03130, Republic of Korea
[3]Cyber Technology Center, Agency for Defense Development, Seoul 05771, Republic of Korea
[4]Department of Joint Education, Joint Forces Military University (JFMU), Nonsan-si, Chungcheongnam-do 33021, Republic of Korea
[5]Department of Convergence Engineering for Intelligent Drones, Sejong University, Seoul 05006, South Korea
[6]Cyber Warfare Research Institute, Sejong University, Seoul 05006, South Korea

Corresponding author: Dongkyoo Shin (shindk@sejong.ac.kr)

This work was supported by the Defense Acquisition Program Administration and the Agency for Defense Development Project Name: Cyber Warfare Mission Impact Analysis Tool Development Prototype under Project UC220012XD.

ABSTRACT Many organizations and researchers, such as NIST, FIRST, MITRE, etc. in the United States, are conducting various cybersecurity research to counter the evolving cyber threats. Research on improving the security level of systems and networks by checking the network environment is one of the main areas of continuous research. To choose the right security countermeasures, you need to ensure that the defense techniques they contain are appropriate for your systems and networks. However, how to determine this is a difficult and complex issue, and as cyber threats evolve, how to determine this will need to evolve with them. To address these issues, this study quantitatively designed six metrics for defense technologies based on system and network environments and used them to conduct experiments on the entire network, as well as experiments on security countermeasures after a cyber-threat has caused damage in a virtual network environment. The proposed method was able to cover a large number of vulnerabilities relative to the number of mitigation techniques applied, and the prioritized list of mitigation candidates allowed us to select the appropriate list of defense techniques for the network. This research can be developed into an automated technology that collects vulnerabilities for the entire system of the network environment to be applied in the future, measures the defense level, prioritizes the complementary defense technologies, and lists them as defenses.

INDEX TERMS Cybersecurity, cyberspace, cyber warfare.

I. INTRODUCTION

Traditionally, anti-malware and anti-virus tools have been the primary tools and techniques for preventing cybercrime [1]. However, the complexity and diversity of current cybercrime has surpassed the capabilities of these traditional security tools. As a result, cybersecurity researchers believe that the development of new and effective security systems to counter threats is an urgent task [2]. Furthermore, one of the reasons for the in-crease in cyber threats is that cybersecurity policies need to be understood in the context of the ever-changing cybersecurity landscape. To this end, it is important to understand other countries' tactics, and most countries' cybersecurity policies focus on big picture issues such as national security, healthcare, and defense [3]. While cybersecurity technology is constantly evolving through research, cyber threat technology is also evolving. The U.S. has a number of cybersecurity research efforts to address evolving cyber threat technologies, including the National Institute of Standards and Technology's (NIST) Cybersecurity Framework,

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

FIRST's The Common Vulnerability Scoring System (CVSS) 4.0, MITRE's Adversarial Tactics, Techniques and Common Knowledge (ATT&CK), and D3FEND. In addition, various studies have been conducted to block threats with similar patterns by learning known threats through machine learning, and this is a topic that will continue to be researched in the future [4], [5], [6], [7]. However, these studies are limited to responding to new cyber threat technologies because they only enhance security with threats with similar patterns within a set defense technology. To address these issues, this study investigated how to select appropriate cybersecurity technologies against cyber threats. The metrics were designed based on MITRE's ATT&CK [8], which categorizes information about the latest cyberattack techniques into a knowledge graph, and D3FEND [9], which categorizes cybersecurity technologies. The metric can quantify the latest security technologies as updated by D3FEND and ATT&CK. To validate the designed metrics, a virtual network environment with vulnerabilities was designed. Then, cyber-attack scenarios were designed and tested. As a result, we have selected a list of cybersecurity techniques that are optimized for network environments with limited resources. This means that the proposed method can be adapted to continuously evolving network and system environments, security technologies, and threats to improve the overall security level of enterprises and countries.

This research consists of five chapters. Section II describes related work, including MITRE's attack and defense technologies that serve as the background for this research, the current state of research by various organizations and researchers, and defense policies. Section II describes the structure of the methodology proposed in this study, including the design and methodology of metrics to quantitatively measure defensive behavior against cyberattacks. Section III describes the experiments using the method, and Section IV concludes with conclusions, future research directions, and comparisons with other studies.

II. RELATED WORK

A. MITRE's ATT&CK, D3FEND

ATT&CK [8], developed by MITRE Corporation, is a framework used in the field of cybersecurity. It is designed to effectively organize and share knowledge about cyberattacks and provides a cybersecurity standard terminology and taxonomy to provide information about attacker behavior patterns and attack techniques. This allows organizations to develop defensive strategies against specific threats and attack techniques and improve detection and response to security issues. ATT&CK can be broadly categorized into Techniques, Tactics, and Defenses, with "Techniques" and "Tactics" being more related to attacks, and "Defenses" being more related to defense. Tactics represent the attacker's larger strategic goals to achieve their end goal, while Techniques describe the attack techniques as the specific actions within each Tactic. Finally, Finally, "Mitigation," a subset of "Defense," describes defenses and mitigations against specific attack techniques or tactics, providing specific actions or enhancements to detect or prevent attacks. These ATT&CKs are used by security professionals and solution developers to better understand specific attacks and develop defense strategies.

D3FEND [9] is a knowledge graph of cybersecurity countermeasures researched by MITRE, and it does not score cybersecurity technologies by defining digital artefacts, but rather breaks them down by function to help users make more accurate judgements and build security architectures. The framework is constantly being updated, and while the initial release had 5 Tactics, it now has a total of 6 Tactics and 22 sub-techniques, including Models, with further subdivisions below. The D3FEND framework can look up a relevant defense technology by its Technique ID in ATT&CK, and describes the techniques of that defense technology, as well as providing information about the associated digital artefacts. Figure 1 shows the connection between these ATT&CK and D3FEND.

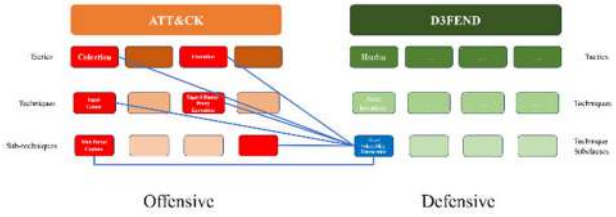


FIGURE 1. Example mapping relationship between ATT&CK and D3FEND.

B. EVALUATE CYBERSECURITY SCORES

Ahmed et al. [10] describe an empirical analysis of a cybersecurity scoring system. Security scores, which are quantitative indicators of an organization's security, generally a higher score indicates that an organization is more secure. However, these scores can vary depending on the organization providing the metric. Additionally, security scores typically use only externally accessible data and are comprised of three sources: external data, publicly available data, and proprietary algorithms. While a high security score indicates that an organization is well secured, even a high-scoring organization may be subject to more attacks than a low-scoring organization if the data it is handling is of a high level of importance compared to other organizations [11]. Therefore, while a security score can be a good indicator of security excellence and a low breach success rate, it is an assessment of an organization's overall security and may be low or too high for the level of criticality of the data [12]. As a result, to ensure fair and accurate assessments, the U.S. Chamber of Commerce has adopted six principles to guide its security ratings. These principles are shown in Figure 2. As a security rating company, BitSight uses data that feeds into a proprietary algorithm based on the six principles to generate a security score ranging from 250 to 900. The metrics consist of a compromised system score comprising five risk vectors, a diligence score focusing on management, such as security

updates to software, and a user behavior score measured by user activity.

Through this analysis, Ahmed et al. [10] point out that no two companies' networks are the same when it comes to measuring security scores, and that the number of users in a network should be considered when measuring scores. They also note that different companies face different types of threats depending on what they need to secure, so security incentives should be based on the criticality of the asset. Finally, it's important to ensure that the network infrastructure is trustworthy.

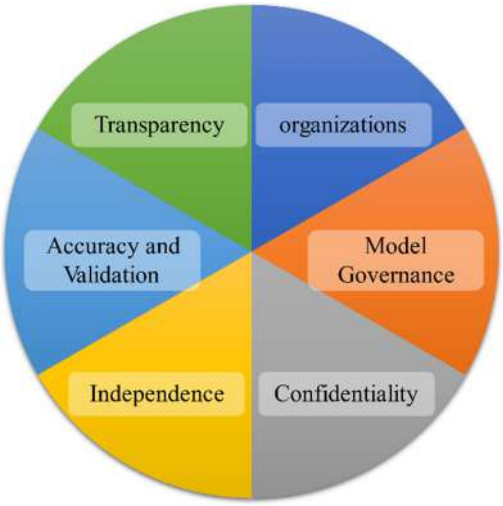


FIGURE 2. Six security rating principles adopted by the U.S. chamber of commerce.

In order to determine and benchmark the cybersecurity risk of an organization, Yampolskiy et al. [13] collected non-intrusive data related to the organization, processed the security information extracted from the collected data and calculated a security score. The calculated security score is assigned based on the correlation between the extracted security information and the overall cybersecurity risk determined by analyzing previously breached companies in the same industry. A patent has been filed to calculate an entity's overall cybersecurity risk score based on the calculated security score and as-signed weights.

C. CYBERSECURITY POLICY-RELATED PROPERTIES

Mishra et al. [14] identified 14 common cybersecurity attributes across seven countries (USA, EU, Australia, Canada, China, India, Malaysia): telecommunications, networks, cloud computing, e-commerce, online banking, smart grid, consumer rights, cybercrime, cryptography, privacy, identity theft, digital signatures, data security, and spam. While these attributes are self-contained, the interdependencies between them can be further specified for specific contexts. To combat cybercrime, the key characteristics of CS need to be identified and well-defined so that a comprehensive policy can be developed. While various stakeholders contribute to the development of CS policy, governments are the primary actors in the creation and revision of policy. Identifying common policies across countries can help academics and policymakers develop cybersecurity policies.

D. CYBERATTACK TARGETS

Cyberattacks are conducted in seven stages: reconnaissance, weaponization, dissemination, exploitation, installation, command and control, and goal achievement [15]. In addition, creating an attack graph for a target network is effective in identifying the attack path from the attack launch point to the target [16]. Identifying vulnerabilities in a network is important to prepare for cyber threats because attackers use vulnerabilities in the target network to identify the optimal attack path.

Common Platform Enumeration (CPE) is a structured naming scheme for software and packages. It consists of 11 attributes, including part, vendor, product, version, update, edition, language, and sw_edition of the software installed on the workstation, expressed as "cpe:2.3:a:microsoft:office:2013:-:-:*:-:-:x64:*". The product, version, update, target_hw, etc. of the CPE name can be used to match the corresponding vulnerability, and multiple CVEs can be matched for a single CPE [17].

Common Vulnerabilities and Exposures (CVEs) are a list of publicly known computer security flaws maintained and overseen by MITRE with financial support from the Cybersecurity and Infrastructure Security Agency (CISA). CVE IDs, the identifiers for CVEs, are assigned by the CVE Numbering Agency (CNA), which includes companies representing major IT vendors. When a security flaw is discovered, it is forwarded to the CNA, which assigns a CVE ID to the information, writes a brief description with references, and distributes it. CVE IDs are issued in the form of a CVE-Year-Serial number [18].

CVSS is an open framework that helps assess security threats by quantifying the nature and severity of software vulnerabilities. It is maintained by FIRST, an international association of incident response and security teams, and currently exists in v3.1 and v4.0 preview. CVSS has three main metrics: foundation, time, and environment, and each metric is composed of subcomponents [18]. The National Vulnerability Database (NVD) allows you to look up the CVSS score by CVE ID and provides a calculator so you can calculate it yourself. It is used by many organizations and vulnerability management programs because it can be used as an indicator of the severity of a vulnerability.

The Common Weakness Enumeration (CWE) is a list of common software and hardware vulnerability types that affect security. A vulnerability is a condition in software, firmware, hardware, and service components that can lead to vulnerability under certain circumstances. The CWE describes and discusses software and hardware weaknesses in a common language and identifies weaknesses in existing software and hardware products. It assesses the coverage of tools targeting these weaknesses and utilizes a common baseline standard for weakness identification, mitigation,

and prevention efforts [19]. These CWEs are related to MITRE's Common Attack Pattern Enumeration and Classification (CAPEC), which focuses on application security and describes common attributes and techniques used by attackers to exploit known weaknesses in cyber-enabled capabilities [20]. In addition, because CAPEC includes the technology numbering of ATT&CK, information about ATT&CK technologies can be obtained through CAPEC and vice versa.

E. CYBER SECURITY STRATEGIES

Varma [21] proposed a methodology to improve cyber resilience by integrating cyber threat detection and mitigation strategies using artificial intelligence (AI). The proposed methodology analyzes various AI-based models and algorithms to evaluate the accuracy and efficiency of cyber threat detection. It analyzes network traffic data using machine learning and deep learning techniques to detect anomalous patterns, and proposes a system that utilizes AI to detect threats in real-time and automatically execute response strategies. Measure detection rate, false positive rate, and response time as performance metrics for threat detection systems. The AI-integrated system is designed to adapt to dynamic cyber threats, and the study demonstrates that AI-based systems are effective in quickly responding to new attack vectors and enhancing an organization's security posture. The proposed system was experimentally validated using various cyber-attack scenarios, and the results showed high detection rates and low false positives compared to traditional security systems. This means that adapting to dynamic cyber threats and choosing a rapid response strategy is crucial to enhance cybersecurity.

Riggs et al. [22] categorize different types of cyber-attacks, including denial of service (DoS), ransomware, man-in-the-middle (MITM) attacks, phishing, and false data injection attacks (FDIA). The researchers also study the specific vulnerabilities associated with these attacks and the mitigation strategies to counter them. For example, DoS attacks can be mitigated through network traffic monitoring and intrusion detection systems (IDS). We proposed a defense-in-depth strategy that incorporates multiple layers of security measures to protect critical infrastructure. This approach involves using intrusion detection systems, encryption, and regular security audits to ensure the resilience of critical systems against cyber threats. They also emphasized the importance of adhering to cybersecurity standards provided by ISO and NIST, which provide frameworks and best practices for developing secure information systems. The authors noted that the rapid increase in cyberattacks on critical infrastructure requires a proactive and adaptive approach to cybersecurity, and by continually updating security measures and leveraging advanced technologies, organizations can better protect their critical assets from evolving cyberthreats.

III. METRICS DESIGN FOR DEFENSIVE SECURITY COUNTERMEASURES

This chapter suggests one metric to quantitatively assess the effectiveness of each defense measure and six metrics to calculate the score.

A. COUNTERMEASURE RECOMMENDATION METHOD PROCESS

An attack vector is created to progress an attack from the network to the cyber attacker's target asset. The assets along the attack path will have multiple vulnerabilities, and there will be multiple defenses that can be applied to the assets. It is possible to select only the vulnerabilities exploited by the attacker and select them as security measures. However, the defensive technologies included in the security countermeasures may not be the optimal security measures for each asset due to cost limitations, lack of equipment, or inability to respond quickly. It is very difficult to select cybersecurity measures while considering these various issues. Therefore, this study proposes a cybersecurity countermeasure recommendation including a three-step algorithm. The algorithm classifies only the defense technologies applicable to the network among the defense technologies identified through the vulnerabilities present in the assets, and finally recommends them through prioritization by measuring the quantitative evaluation score. As preliminary work for the algorithm, we describe the CPE-CVE-CWE-CAPEC-D3FEND mapping methodology.

1) IDENTIFYING CVES VIA CPES

Various vulnerabilities present in an asset can be identified by analyzing the network inside the organization or by knowing the program information used (vendor name, version, product name, etc.), i.e., CPE.

2) CWE MAPPING

For identified CVEs, CWEs are extracted from the 'Observed Examples' column of the CWE dataset or through the CWE-CVE root cause mapping methodology (available on the official page).

3) CWE AND ATT&CK UTILIZING CAPEC

CAPEC has CWE information in the 'Related Weakness' column and ATT&CK attack technique values as 'Entry ID' in the 'Taxonomy Mappings' column.

4) ATT&CK TO D3FEND

D3FEND officially supports mapping with ATT&CK.

By utilizing these mappings, you can effectively find defenses against CVEs identified through network analysis or CPE. The overall structure is shown in Figure 3, and the three-step algorithm is as follows.

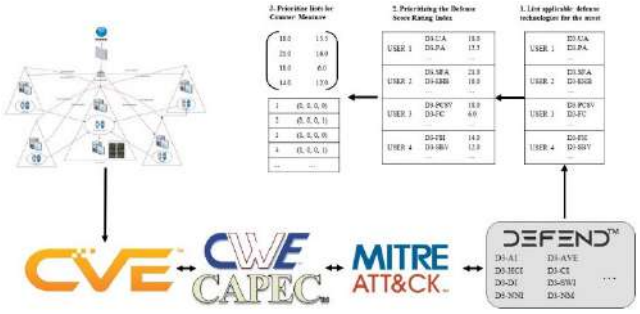


FIGURE 3. Countermeasure recommendation method process.

a: LISTING APPLICABLE DEFENSE TECHNIQUES FOR THE ASSET

Extract a list of applicable D3FEND defense techniques based on the defense techniques based on the vulnerabilities of each asset and the attacker's chosen attack technique (ATT&CK's technique) through the mapping methodology of the previous work.

b: PRIORITIZING THE DEFENSE SCORE RATING INDEX

Sort and prioritize the defense technology rating index calculated for each asset in descending order. The Defense Score Rating Index is described in III-C.

c: PRIORITIZING LISTS FOR CYBERSECURITY COUNTERMEASURES

Provide a prioritized list of defense technologies to recommend various cybersecurity countermeasures to security personnel and administrators. To achieve this, the two levels of asset-specific defense measure lists are determined into a single two-dimensional matrix, which can be prioritized by permutation [23] to provide different combinations of defense countermeasures.

The above procedure allows security personnel to select the appropriate cybersecurity measures for their network environment.

B. DESIGN AND DEFINE METRICS

Two of the six metrics are designed to be related to vulnerabilities. This is a result of accepting the importance of the security update score among the scores mentioned by Ahmed et al. [10]. The rest consisted of factors related to the network environment and position against attack techniques. The six designed metrics are as follows.

1) COST

The cost of applying defenses to your network. This includes both human and physical assets expended to apply the defense behavior. The higher the cost, the better the performance of the mitigation technique, but it is not directly proportional, so it is a good metric for selecting defenses that perform well at a lower cost. The lower the cost, the higher the score.

2) DEFENSE PHASE

Based on the four phases of breach incident response (IR) proposed by the US NIST [24] and the incident response phase consisting of a six-step process proposed by Kral in [25], it is composed of four phases: detection, initial response, recovery response, and investigation and analysis.

3) LEVEL OF DIFFICULTY

The concept of the difficulty of applying defense techniques, which is calculated based on the vulnerability of the asset in the network environment. Vulnerability is calculated based on CVSS and can be measured based on CVSS prediction algorithms [26] for new CVEs due to the constantly evolving cyberspace.

4) ASSET POSITION IN ATTACK PATH

This score is measured by determining the location of network assets targeted by detected threats along the attack path, from the attack launch point to the end goal. If you can proactively stop the threat at an asset close to the origin of the attack, you will score high.

5) EFFECT SCORE

A measure of the effectiveness of a defense technology when applied to a network environment. It is measured by the likelihood that a vulnerability in the network will be eliminated by applying the defense. The effectiveness metric, like the difficulty metric, is based on the CVSS prediction algorithm, which can respond to new vulnerabilities.

6) APPLICABILITY TIME

Ensuring that you can quickly apply defenses and stop threats from the point of attack detection is critical to improving cybersecurity, hence the metric that measures the time it takes to apply defense technique.

7) SINGLE DEFENSE SCORE

The above six metrics are equally weighted, and the higher the score of the remaining metrics relative to the cost metric, the higher the defense evaluation index.

C. CALCULATION METHOD

The six metrics are calculated as follows, and after all calculations, they must be normalized to the same range of values to produce the Defense Assessment Index.

1) COST

It is measured by the network assets, human assets of the network to which the defense technology is applied and is measured by the judgement of the managers and experts of the organization, or by the amount of hiring security experts. However, if the defensive technology is related to security equipment, the cost is calculated by including the cost of such equipment if the organization does not own such equipment, and the human cost is calculated.

When measuring costs, you should consider the following First, the amount of money available depends on the purpose of use (defense, private enterprise, etc.), network environment, security equipment you have, etc. The second is. it is not fixed due to many variables: labor costs, fluctuating market prices of resources, etc. For this reason, it can be measured differently depending on when it is measured and who is measuring it.

The calculated cost is normalized using the min-max normalization algorithm by finding the maximum and minimum cost of all defense technologies. (0 <= Cost <= 1).

2) DEFENSE PHASE

As mentioned in Section III-B, there are four phases and identify the defense phases that can be applied to each defense technique. A defense technique can have multiple defense phases, but for the purposes of this study, it is assumed to have a maximum of two defense phases. Each defense phase is scored from 1 to 5, with detection (4), initial response (5), recovery response (3), and investigation and analysis (1). (1 <= Phase <= 5).

Detection is a defense focused on identifying and alerting to cyber threats and can include network traffic analysis, log monitoring, and anomaly detection. Initial Response is the immediate action taken immediately after a threat is detected. This could be adjusting firewall rules, tightening access controls, or quarantining malicious code. Recovery Response involves steps to repair the damage, such as restoring data backups, reconfiguring systems, and patching vulnerabilities. Investigation and Analysis: Steps to determine the cause of the attack and prevent the same type of attack in the future. Examples include forensic analysis, log analysis, and threat intelligence research.

3) LEVEL OF DIFFICULTY (LVL)

The more vulnerabilities an asset has, the more difficult it is to apply defensive techniques. It is calculated as the sum of the vulnerability scores corresponding to the asset ($AssetCVSS_n$) over the sum of the scores of all vulnerabilities in the attack path ($\sum AssetCVSS$) as shown in Eq. 1. (0 <= Lvl <= 10)

$$Lvl = AssetCVSS_n / \sum AssetCVSS \tag{1}$$

4) ASSET POSITION IN ATTACK PATH (POSITION)

It is measured based on the position of the asset on the cyber attacker's attack path and is calculated as the position of the selected asset relative to the total number of assets on the path. (0 <= Position <= 1).

5) EFFECT SCORE

Based on the vulnerabilities of the asset, it is calculated as the CVSS average of the vulnerabilities after eliminating the vulnerabilities corresponding to the defense technology among the vulnerabilities existing in the asset through the relationship of CPE-CVE-CWE-CAPEC-ATT&CK-D3FEND as shown in Eq 2. (0 <= Effect <= 10).

6) APPLICABLE TIME (TIME)

Calculates the effectiveness of a mitigation technique over the time it takes to apply and complete. If the defensive action can be applied immediately, the effect is good, and the closer the calculated value is to 1, the greater the effect. It is calculated from the time of application, completion, and detection of the at-tack, and the variables as shown in Table 1 are defined based on Minute and calculated as shown in Eq 2. (0 <= Time <= 1)

$$Time = 1 - (TimeDA - TimeAD)/(TimeDC - TimeAD) \tag{2}$$

TABLE 1. The time matric calculations method's parameter and definitions.

Parameter	Definition
$Time_{DA}$	When to apply mitigation techniques
$Time_{DC}$	When defense techniques are applied
$Time_{AD}$	When the attack was detected

7) NORMALIZATION

The above metrics cannot be calculated with the same weight be-cause they all have different ranges of values, so they are normalized to make all the metrics equal, with values between 1-5.

Figure 4 shows how to replace values between 0-1 with values in the range 1-5. Values between 0 and 10 are replaced with values in the range 1-5 by multiplying the value below (the raw value before normalization) by 10.

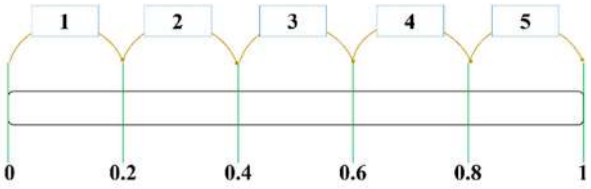


FIGURE 4. Normalization methods.

8) SINGLE DEFENSE SCORE (DS)

The higher the sum of the other metrics relative to the cost, the higher the score. The calculation method is shown in Eq 3.

$$DS = (Step + Lvl + Position + Effect + Time)/Cost \tag{3}$$

The above formula for calculating the defense evaluation index can be further refined by adding weights to the indicators based on the judgement of managers and experts.

Figure 5 is an example of network information and attack scenarios for selecting cybersecurity measures. Figure 6 is an example of a prototype showing the cybersecurity countermeasure priorities calculated based on Figure 5 and the defense techniques included in the countermeasure.

FIGURE 5. Examples of network information and attack vectors.

FIGURE 6. Cybersecurity countermeasures list and examples of defensive technologies included in the cybersecurity countermeasures.

IV. EXPERIMENTS

Due to the unreliability of the prototype, we conducted a logical experiment to verify the proposed method. For this purpose, we constructed a network for experiments. After performing scenarios with cyber-attack vectors on the configured network, we applied the proposed method to verify the results.

A. DESIGNING A NETWORK CONFIGURATION

Design a network for the experiment. The target networks of this study are military networks and corporate networks. Because using a real network environment may leak the organization's vulnerabilities and network information, we constructed a virtual network in this paper. However, in real-world implementations, vulnerability information and attack paths should be measured by security personnel, while other metrics can be helped by external organizations.

Fencl et al. [27] describe an algorithm for network topology design, noting that randomizing the network design can lead to problems with data transmission time delays and topology configuration costs. In addition, the design of network topology is important because if the network is poorly designed, it cannot be guaranteed to be safe from various cyber threats. In addition, conducting experiments to conduct various analyses in a virtual network environment similar to the real network environment is an important step to evaluate the advantages and ultimately deploy the solution before actually using it [28]. Therefore, in this study, we built a virtual network environment rather than a real network environment to conduct our experiments. Although the virtual network environment we designed is based on a private network, the methodology in this study can be applied to different types of networks, including closed networks with limited access and enhanced security measures, corporate networks, and public networks. This adaptability means that the proposed defense metrics can effectively protect against cyber threats regardless of the architecture or accessibility of the network. Conducting experiments in virtual network environments not only mitigates the potential risks associated with real-world testing, but also demonstrates the versatility of the approach to accommodate the unique requirements and challenges presented by different network environments.

The main experiments were conducted in a military network-based virtual environment, a small organization network; however, the military network-based environment is not disclosed in this paper. Therefore, we use a small office/home office network environment designed based on [29], [30], and [31]. Table 2 summarizes the elements required in the designed network, and the designed network is shown in Figure 7.

TABLE 2. Using network topology components.

Component	Example	Required scope
Network devices	Workstation, Printer, Laptop, Switch, etc.	Workstation, Printer, Switch, Server, Router, DB Server
Security devices	Fire Wall, IDC, IPS, etc.	All
Communication information	Packet, Delay time, Connection information, etc.	Connection information
System information	IP, Software information, Data file, User information, etc.	CVE

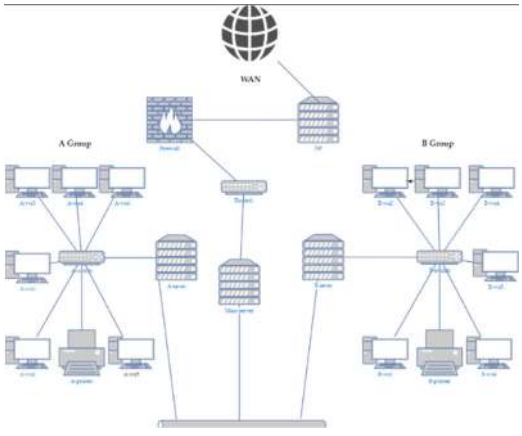


FIGURE 7. Designed network topology.

B. DESIGN ATTACK SCENARIOS

In order to compare the before and after of the proposed method, a cyber-attack must occur. By creating and performing a cyber-attack scenario, it is possible to identify the vulnerability of the network, and by performing the cyber-attack scenario again after applying the proposed

method, it is possible to identify the enhancement of cybersecurity. In addition, in order to demonstrate that the proposed method is a universal method and can be used in various environments, the attack scenarios are subject to the following assumptions and restrictions.

1) Based on ATT&CK's attack techniques, an attacker can use any attack technique that corresponds to the vulnerabilities present in the network. Utilize available attack techniques based on the CVE to D3FEND mapping method mentioned in III-A.
2) Cyber-attack attempts have a 100% success rate and can only be defended by D3FEND's defense technology. This is to evaluate the pure effectiveness of the defense technology by ensuring that it is not affected by rulesets such as security equipment or physical security.
3) To reach the final target network asset, the attack must traverse at least three assets, which means the minimum attack path is three hops, and is designed to allow the attack to progress through a variety of paths.
4) Based on the network topology designed in Section IV-A, the attacker goes through B-ws2, B-ws3, and A-ws5 to reach the final attack target (A-ws4). Figure 8 shows the CPE of the designed network Workstation and some of the CVEs corresponding to the CPE. Furthermore, the defense techniques applied are shown in Table 3 and the attack path is shown in Figure 9. The yellow line shows the direct access path from the network, and the red arrow line shows the flow of the attack path.

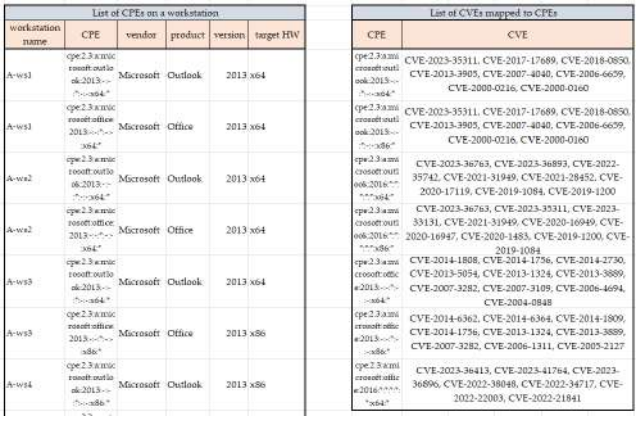


FIGURE 8. The CPE of a designed network workstation (left) and some of the CVEs corresponding to the CPE (right).

These assumptions and limitations allow us to evaluate different cyber-attack scenarios and demonstrate the validity of the proposed methodology. By using different attack types, we can quantitatively evaluate and compare the effectiveness of defense techniques in a network environment.

C. APPLYING THE METHOD

Three of the six indicators in the proposed methodology include the presence of defense equipment and the amount of money spent on defense technologies. Therefore, in order to calculate a single defense score, a preparatory step is required to pre-calculate the three indicators. The preparation phase has the following prerequisites and assumptions.

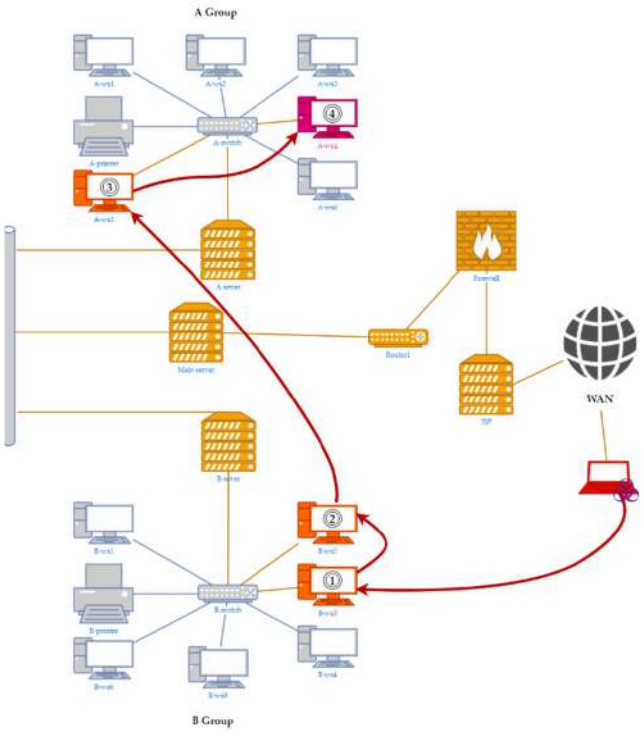


FIGURE 9. Designed attack scenario and route.

TABLE 3. List of defense techniques applied across the network.

Defense Techniques	Description	Device
D3-DNSDL	DNS Access Exclusion Policy	Firewall
D3-NTF	Network traffic filtering	Firewall
D3-BA	Authentication before bootloader programs	Workstations
D3-DE	Disk encryption	Workstations

1) COST

This is fluid as it includes the amount of human resources and equipment, so for the sake of fairness, all costs are calculated at the same amount. However, if defensive equipment is required, the cost of purchasing defensive equipment is taken into account. However, as mentioned in Section III-C, this would result in 0 and 1 for MIN-MAX normalization and 0 and 5 for the defense evaluation index, so we assumed a score of 5 for defense technologies that do not require security equipment and a score of 3 for defense technologies that require security equipment.

2) DEFENSE PHASE

You must set a Defense Phase for each Defense Technique. Set a minimum of one and a maximum of two defense skills.

TABLE 4. Summary of study comparisons.

Proposed Method	Important METRIC	Scope of use	Attack Type Coverage
J. Ahmed et al. [10]	Asset Criticality, Network Reliability, Number of network users	Organization	Botnet Infection, Spam, Malware Server, Unsolicited Communication
Stacy Collett [11]	Data Criticality	Organization	Not specified
Yampolskiy et al. [13]	About Data Security, Weight	Organization	Social Engineering Attacks, Malware, Botnet ifections, Hacker Sites
A. Mishra et al. [14]	Key Common Characteristics of Cybersecurity	Government	DoS, Cybercrime during COVID-19, Cross-border Cyber threats, Attacks on critical infrastructure
V. V. Varma. [21]	Connectivity, Communication Protocols	Organization	DDoS, IP Spoofing
H. Riggs, et al. [22]	Vulnerability Analysis, Anomaly Detection	Organization, Government	Ransomware, APTs
Proposed Method	Future-proofing Technologies, Flexible, Network Environment	Private, Organization, Government	DDoS, IP Spoofing, Brute Force

This is determined based on the description of the defense technology.

FIGURE 10. Three metrics set in the preparation phase: cost, defense level, and time to apply.

3) APPLICABLE TIME

The time from the time of application of the defense technology to the completion of application depends on the ability of the security expert applying the defense technology and the possession of defense equipment. Therefore, this study assumes that all defense equipment is possessed and is calculated based on the Description. It is also assumed that the time from the time of attack detection to the application of the defense technology and the time from the start of application of the defense technology to the completion of application are performed by one security expert.

The Figure 10 shows some of the metric values for the Preparation phase based on the above prerequisites and assumptions.

After all the preparations, we identified the optimal security countermeasures for the attack vectors shown in Figure 9, and the prioritized defense countermeasures for each asset are shown in Figure 11. In B-ws3, A-ws5, and A-ws4, D3-FE, a defense technology related to file encryption, scored the highest, and D3-EDL, which blocks file execution through poli-cy changes, scored the second highest.

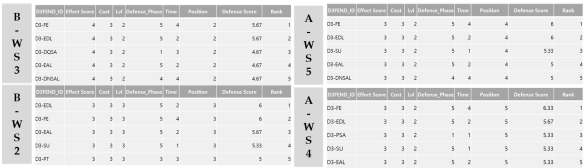


FIGURE 11. Table of defense technology prioritization results for assets.

Based on the workstation's list of defense technologies, the best security countermeasure produced by the permutation was the addition of D3-FE alone (Total 24.0), while the combination of permutations that removed duplicates from all four workstations yielded a score of 22.67: D3-FE (5.67), D3-EAL (5.67), D3-EDL (6.00), and D3-SU (5.33).

V. CONCLUSION

The purpose of this research is to provide effective and efficient security countermeasures for multiple assets with less effort in preparation for cyberattacks or in the event of damage caused by cyber-attacks. Furthermore, this research aims to prepare for the evolving cyber threats in the evolving cyberspace. To validate this, a virtual network environment was built, attack scenarios were written, and experiments were conducted. When cybersecurity countermeasures were selected through the process shown in Figure 4, it was found

that in the case of a small network with fewer paths, only one additional security technology was selected, but it was found to be the most efficient security technology in that network environment. By reversing the mapping relationship of ATT&CK-CAPEC-CWE-CVE-CPE with the defense technologies identified in the experimental results, we found that on average, more than 10 vulnerabilities can be compensated out of the average number of 16.75 vulnerabilities in the assets. We further experimented in a real-world network environment using 10 workspaces and found that they were able to cover an average of 5.4 out of 10 vulnerabilities, which was not significantly different from the results in the virtual environment.

Based on the experimental results, we compared the differences with previous studies. Ahmed et al. [10] emphasized the importance of measuring security scores based on the criticality of assets and network trust. While scores measured by security scoring companies are important, it is important to evaluate the criticality of the data because the criticality of the data determines the likelihood of being targeted by an attacker. Yampolskiy et al. [13] measured the security risk score by collecting data and extracting security information to determine the cybersecurity risk of a company, and Mishra et al. [14] stated that it is important to identify the main characteristics of common CS to develop cyber policies. Varma [21] emphasized the use of AI in cybersecurity to integrate threat detection and mitigation strategies. He designed an AI-integrated system to adapt to dynamic cyber threats and focuses on learning to quickly respond to new attack vectors and strengthen the organization's security posture. Riggs et al. [22] mention the need to incorporate multiple layers of security measures against various cyber threats. In conclusion, most cybersecurity strategy and response techniques studies emphasize collecting network security information, data information, etc. to measure cybersecurity scores in order to prepare for cyber threats. They also emphasize the use of AI techniques to perform automated response strategies to improve cybersecurity.These studies may not be universal and may not be prepared for new threats, and it may be difficult for administrators to justify the response strategies implemented by leveraging AI to perform response strategies or to modify response strategies based on the situation. However, in this study, we used ATT&CK, D3FEND, a knowledge graph-based framework of offensive and defensive techniques that is universal, continuously updated, and adaptable to new threats. We also designed, quantified, and prioritized six metrics for defensive techniques to allow for flexibility in modifying and selecting defensive strategies. This is one of the ways to select the right security measure for the network according to the evolving cybersecurity and attack technologies. In addition, we included CVEs and CVSS in the metrics, which are used globally to measure the security risk of network assets, so it can be used in various environments (individuals, organizations, countries, etc.). Table 4 summarizes a comparison of the results of these studies.

This research aims to help individuals, organizations, countries, etc. select efficient security measures with less effort in the modern world where cybersecurity is becoming increasingly important. To select efficient security measures, we proposed six metrics that can be set by users and automatically calculated according to different environments. We also indexed each column and generated permutations to prioritize them, and applied different defense techniques by removing redundancies, which means that the proposed cybersecurity mitigation procedures can be effectively applied in different network environments. The practical application of the method proposed in this study requires sufficient knowledge of the network environment, and providing this information to an external party may cause greater threats. Therefore, we mainly conducted experiments in a virtual network environment, and confirmed that it can be applied in an environment similar to a real network. For practical application, the administrator should be in charge, and the remaining indicators except vulnerability information and location information along the attack path can be helped by external personnel. In addition, efficiently utilizing frameworks that are continuously updated by leveraging CVSS prediction algorithms [26] or through APIs provided by frameworks (D3FEND, ATT&CK, CVE, etc.) can help adapt to evolving cyber threats. Therefore, the methodology proposed in this study has the following advantages. 1. by using a continuously updated framework and using a commonly used vulnerability management system, it is easy to manage the latest attack, defense, and vulnerability data. 2. By designing and quantitatively evaluating six metrics for defensive technologies, it is possible to understand why defensive technologies are recommended. Personnel can utilize them and use them as a basis for decision making. 3. The proposed methodology can be automated and used after the first network information collection. Finally, it is flexible, as the metrics are measured differently depending on the network information analyzed, and can be used in different network environments. However, this study has some limitations. First, all the frameworks used as mapping relationships may not be well matched due to their continuous updates. Furthermore, they will be unusable if they stop updating. Second, we need to collect system and network information about all the assets that make up the network in order to measure the designed metrics. Finally, while we tried to objectify the network used in our experiments, we conducted our experiments primarily in a virtual environment, which may lead to errors in generalization. In order to weight the designed metrics, it is essential to create various cyber-attack scenarios, collect data through extensive experiments, and then utilize machine learning models to identify and weight metrics that have a real impact on enhancing cybersecurity.

REFERENCES

[1] A. F. Brantly, "The cyber deterrence problem," in *Proc. 10th Int. Conf. Cyber Conflict (CyCon)*, Tallinn, Estonia, May 2018, pp. 31–54.

[2] A. Mishra, Y. I. Alzoubi, A. Q. Gill, and M. J. Anwar, "Cybersecurity enterprises policies: A comparative study," *Sensors*, vol. 22, no. 2, p. 538, Jan. 2022.

[3] T. Rajaretnam, "A review of data governance regulation, practices and cyber security strategies for businesses: An Australian perspective," *Int. J. Technol. Manage. Inf. Syst.*, vol. 2, no. 1, pp. 1–17, 2020.

[4] A. R. Ugale and A. D. Potgantwar, "Anomaly based intrusion detection through efficient machine learning model," *Int. J. Electr. Electron. Res.*, vol. 11, no. 2, pp. 616–622, Jun. 2023, doi: 10.37391/ijeer.110251.

[5] M. S. Akhtar and T. Feng, "Malware analysis and detection using machine learning algorithms," *Symmetry*, vol. 14, no. 11, p. 2304, Nov. 2022, doi: 10.3390/sym14112304.

[6] H. Jiwon, H. Kim, S. Oh, Y. Im, H. Jeong, and H. Kim, "Client-based web attacks detection using artificial intelligence," 2023. Accessed: Jul. 26, 2024, doi: 10.21203/rs.3.rs-2920883/v1. [Online]. Available: https://assets-eu.researchsquare.com/files/rs-2920883/v1_covered_9fd4d387-50c6-49d2-bde6-88f9b563c434.pdf?c=1711504235

[7] T. Z. Difaizi, O. P. L. Camille, T. C. Benhura, and G. Gupta, "URL based malicious activity detection using machine learning," in *Proc. Int. Conf. Disruptive Technol. (ICDT)*, Greater Noida, India, May 2023, pp. 414–418.

[8] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, "Mitre att&ck: Design and philosophy," MITRE Corp., Richmond, VA, USA, Tech. Rep. MP180360R1, 2018. Accessed: Jan. 4, 2024. [Online]. Available: https://www.mitre.org/sites/default/files/2021-11/prs-19-01075-28-mitre-attack-design-and-philosophy.pdf

[9] P. E. Kaloroumakis and M. J. Smith, "Toward a knowledge graph of cybersecurity countermeasures," M.S. thesis, MITRE Corp., 2021. Accessed: Jan. 4, 2024. [Online]. Available: https://d3fend.mitre.org/resources/D3FEND.pdf

[10] J. Ahmed. (2019). *Empirical Analysis of a Cybersecurity Scoring System*. Accessed: Jan. 4, 2024. [Online]. Available: https://digitalcommons.usf.edu/etd/7722

[11] S. Collett. (2016). *Whats in a Security Score?*. Accessed: Jan. 4, 2024. [Online]. Available: https://www.csoonline.com/article/557289/what-s-in-a-security-score.html

[12] J. Vijayan. (2014). *Target Attack Shows Danger of Remotely Accessible HVAC Systems*. Computerworld. Accessed: Jan. 4, 2024. [Online]. Available: https://www.computerworld.com/article/2487452/target-attack-shows-danger-of-remotely-accessible-hvac-systems.html

[13] A. Yampolskiy, R. Blackin, A. Heid, and S. Kassoumeh, "Calculating and benchmarking an entity's cybersecurity risk score," U.S. Patent 10 498 756, Nov. 22, 2016. [Online]. Available: https://patents.google.com/patent/US20160173521A1/en

[14] A. Mishra, Y. I. Alzoubi, M. J. Anwar, and A. Q. Gill, "Attributes impacting cybersecurity policy development: An evidence from seven nations," *Comput. Secur.*, vol. 120, Sep. 2022, Art. no. 102820, doi: 10.1016/j.cose.2022.102820.

[15] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," in *Leading Issues in Information Warfare & Security Research*, vol. 1, Washington, DC, USA: API, 2011, pp. 113–125.

[16] I. Kotenko and A. Chechulin, "A cyber attack modeling and impact assessment framework," in *Proc. 5th Int. Conf. Cyber Conflict (CYCON)*, Tallinn, Estonia, Jun. 2013, pp. 1–24.

[17] B. A. Cheikes, D. Waltermire, and K. Scarfone, "Common platform enumeration: Naming specification version 2.3," U.S. Dept. Commerce, Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. 7695, 2011, doi: 10.6028/NIST.IR.7695. [Online]. Available: https://www.nist.gov/publications/common-platform-enumeration-naming-specification-version-23

[18] M. Adam et al. (2019). *Common Vulnerability Scoring System (CVSS) Version 3.1: Specification Document*. Accessed: Jan. 4, 2024. [Online]. Available: https://www.first.org/cvss/v3.1/specification-document

[19] S. Christey and C. Harris. (Oct. 2009). *Introduction to Vulnerability Theory*. MITRE. [Online]. Available: https://cwe.mitre.org/documents/vulnerability_theory/CWE-Introduction_to_Vulnerability_Theory.pdf

[20] N. Amon and J. Baker. (2021). *Security Control Mappings: A Starting Point for Threat-Informed Defense*. MITRE-Engenuity. Accessed: Jan. 4, 2024. [Online]. Available: https://medium.com/mitre-engenuity/security-control-mappings-a-starting-point-for-threat-informed-defense-a3aab55b1625

[21] V. V. Varma, "Enhancing cyber resilience by integrating AI-driven threat detection and mitigation strategies," *Trans. Latest Trends Artif. Intell.*, vol. 4, no. 4, 2023. Accessed: Jul. 7, 2024. [Online]. Available: https://ijsdcs.com/index.php/TLAI/article/view/396

[22] H. Riggs, S. Tufail, I. Parvez, M. Tariq, M. A. Khan, A. Amir, K. V. Vuda, and A. I. Sarwat, "Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure," *Sensors*, vol. 23, no. 8, p. 4060, Apr. 2023. Accessed: 2024-06-07. [Online]. Available: https://www.mdpi.com/1424-8220/23/8/4060

[23] R. Arboretti, S. Bonnini, L. Corain, and L. Salmaso, "A permutation approach for ranking of multivariate populations," *J. Multivariate Anal.*, vol. 132, pp. 39–57, Nov. 2014, doi: 10.1016/j.jmva.2014.07.009.

[24] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer security incident handling guide," NIST, Special Publication 800.61, Tech. Rep. SP 800-61 Rev. 2, 2012, doi: 10.6028/NIST.SP.800-61r2.

[25] P. Kral. (2011). *The Incident Handlers Handbook*. Sans Institute. Accessed: Jan. 4, 2014. [Online]. Available: https://sansorg.egnyte.com/dl/6Btqoa63at

[26] M. R. Shahid and H. Debar, "CVSS-BERT: Explainable natural language processing to determine the severity of a computer security vulnerability from its description," in *Proc. 20th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Pasadena, CA, USA, Dec. 2021, pp. 1600–1607, doi: 10.1109/ICMLA52953.2021.00256.

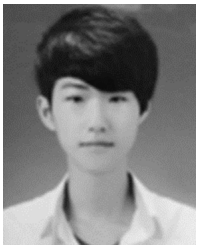
[27] Fencl et al., "Network topology design," *Control Eng. Pract.*, vol. 19, no. 11, pp. 1287–1296, 2011.

[28] A. Bavier, N. Feamster, M. Huang, L. Peterson, and J. Rexford, "In VINI veritas: Realistic and controlled network experimentation," in *Proc. Conf. Appl., Technol., Architectures, Protocols Comput. Commun.*, Italy, Aug. 2006, pp. 3–14, doi: 10.1145/1159913.1159916.

[29] L. Yang and Y. Ding, "The design of network topology big data platform in cloud computing," in *Proc. 2nd Int. Conf. Adv. Technol. Intell. Control, Environ., Comput. Commun. Eng. (ICATIECE)*, Bangalore, India, Dec. 2022, pp. 1–5, doi: 10.1109/ICATIECE56365.2022.10047353.

[30] J. L. Harrington, "Part two: Design and connectivity," in *Ethernet Networking for the Small Office and Professional Home Office*. Amsterdam, The Netherlands: Elsevier, 2010.

[31] W. Odom, *CCNA 200-301 Official Cert Guide*, vol. 2. Indianapolis, IN, USA: Cisco Press, 2019, Accessed: Jan. 4, 2024. [Online]. Available: https://www.cefracor.org/sites/www.cefracor.org/files/webform/documents/offre-complete/fichier/pdf-ccna-200-301-official-cert-guide-library-wendell-odom-pdf-download-free-book-eee99cc.pdf



JANG JISOO received the B.S. degree in computer science from Seoul Hoseo Occupational Training College, Seoul, South Korea, in 2021, and the M.S. degree in computer science from Sejong University, Seoul, in 2023, where he is currently pursuing the Ph.D. degree. From 2017 to 2019, he was an alternative to military service with a real estate company in South Korea, where he was responsible for website development and maintenance. His research interests include machine learning, cyberspace, cyber warfare, and military science.

SUBONG JUNG received the B.S. degree in electronic engineering from the Naval Academy, Gyeongsangnam-do, Republic of Korea, in 1993, and the M.S. degree in industrial engineering from Kyung Hee University, Suwon, Republic of Korea, in 2001. He is currently a Manager with the Defense Future Technology Research Institute, LIG Systems, Seoul, Republic of Korea. His research interests include cyber warfare, decision-making, and information protection.

MYUNGKIL AHN received the B.S. degree in information and communication engineering from Chungnam National University, Daejeon, Republic of Korea, in 1997, the M.S. degree in computer engineering from Sogang University, Seoul, Republic of Korea, in 2003, and the Ph.D. degree in electrical and electronics engineering from Chung-Ang University, Seoul, in 2021. She is currently a Principal Researcher with the Cyber Technology Center, Agency for Defense Development, Seoul. Her research interests include computer security and cyberwarfare modeling and simulation.

DONGHWA KIM received the B.S. and M.S. degrees from the School of Electrical Engineering, Korea University, Seoul, Republic of Korea, in 2004 and 2007, respectively. He is currently pursuing the Ph.D. degree in computer engineering with Sejong University. He is a Senior Researcher with the Cyber Technology Center, Agency for Defense Development, Seoul. His research interests include cybersecurity training systems, M&S systems, and cyber red/blue team automation.

JAEPIL YOUN received the B.S. degree in computational information processing from the Korea Army Academy at Yeongcheon (KAAY), Republic of Korea, in 2008, the M.S. degree in cybersecurity from Ajou University, Suwon, Republic of Korea, in 2017, and the Ph.D. degree in computer engineering from Sejong University, Seoul, Republic of Korea, in 2023. From 2018 to 2020, he was a Researcher with the Agency for Defense Development (ADD), Republic of Korea. From 2021 to 2023, he was an Officer for cyber operations planning and cyber operations training at the Army Cyber Operations Center (ACOC). He currently conducts research at the Joint Forces Military University (JFMU), where he studies advancements in defense policy, military strategy, defense planning, and joint coalition operations. His research interests include cyber intelligence surveillance and reconnaissance (ISR) and cybersecurity.

DONGKYOO SHIN received the B.S. degree in computer science from Seoul National University, South Korea, in 1986, the M.S. degree in computer science from Illinois Institute of Technology, Chicago, IL, USA, in 1992, and the Ph.D. degree in computer science from Texas A&M University, College Station, TX, USA, in 1997. He is currently a Professor with the Department of Computer Engineering, Sejong University, South Korea. From 1986 to 1991, he was with Korea Institute of Defense Analyses, where he developed database application software. From 1997 to 1998, he was a Principal Researcher with the Multimedia Research Institute, Hyundai Electronics Company, South Korea. His research interests include machine learning, ubiquitous computing, biosignal data processing, and information security.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 December 2023, accepted 19 December 2023, date of publication 1 January 2024,
date of current version 16 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349022

RESEARCH ARTICLE

A Systematic Analysis of Enhancing Cyber Security Using Deep Learning for Cyber Physical Systems

SHIVANI GABA[1], ISHAN BUDHIRAJA[1], (Member, IEEE), VIMAL KUMAR[1], (Member, IEEE), SHESHIKALA MARTHA[2], (Member, IEEE), JEBREEL KHURMI[3], AKANSHA SINGH[1], (Member, IEEE), KRISHNA KANT SINGH[4], S. S. ASKAR[5], AND MOHAMED ABOUHAWWASH[6,7]

[1]School of Computer Science Engineering and Technology, Bennett University, Greater Noida, Uttar Pradesh 201310, India
[2]School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana 506371, India
[3]Department of Computer Science, College of Technology, Jazan University, Jazan 45142, Saudi Arabia
[4]Delhi Technical Campus, Greater Noida 201306, India
[5]Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia
[6]Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
[7]Department of Computational Mathematics, Science and Engineering, College of Engineering, Michigan State University, East Lansing, MI 48825, USA

Corresponding author: Akansha Singh (akansha1.singh@bennett.edu.in)

This project is funded by King Saud University, Riyadh, Saudi Arabia. Researchers Supporting Project number (RSP2024R167).

ABSTRACT In this current era, cyber-physical systems (CPSs) have gained concentrated consideration in various fields because of their emergent applications. Though the robust dependence on communication networks creates cyber-physical systems susceptible to deliberated cyber related attacks and detecting these cyber-attacks are the most challenging task. There is the interaction among the components of the cyber and physical worlds, so CPS security needs a distinct approach from past security concerns. Deep learning (DL) distributes better performance than machine learning (ML) due to its layered architecture and the efficient algorithm for extracting prominent information from training data. So, the deep learning models are taken into consideration quickly for detecting cyber-attacks in cyber physical systems. As numerous attack detection methods have been proposed by various authors for enforcing CPS security, this paper reviews and analyzes multiple ways of attack detection presented for CPS using deep learning. We will be putting the excellent potential for detecting cyber-attacks for CPS concerning deep learning modules. The admirable performance is attained partly as highly quality datasets are eagerly obtainable for the use of the public. Moreover, various challenges and research inclinations are also discussed in impending research.

INDEX TERMS Cybersecurity, cyberattacks, cyber physical systems (CPSs), deep learning (DL), attack detection.

I. INTRODUCTION

As there is a fast growth of technology in various communication networks and the field of computer science leads, cyber-physical systems (CPS) are rising widely in both areas, such as academia and industries. The cyber-physical systems are measured and supervised by computer-based algorithms, which are combined with networks and users. The cyber-physical systems comprise interacting network units with physical and computational devices. The applications of CPS are making a disproportionate impact on businesses, such as in industrial sectors, healthcare, and manufacturing.

As soon as the Internet of Things (IoT) initiates, various devices with security susceptibilities are connected to cyber-physical systems, resulting in multiple attacks. It has been observed in past years that the incidents of CPS attacks have increased after the Stuxnet attack back in 2010 [1]. If cyber-physical systems attacks are not perceived and reduced rapidly, they can cause massive consequences such

The associate editor coordinating the review of this manuscript and approving it for publication was Engang Tian.

as damage to equipment, financial losses, and public safety. So the security of CPS is one of the vital paradigms for this. But securing cyber-physical systems is also a challenging task due to its heterogeneity of components, complex interactions among cyber-physical systems, and the attack surface's complexities [2]. It is observed that an intruder can randomly interrupt the dynamism of systems or encourage agitations to cyber-physical systems deprived of the security of various strategies of hardware or software, which leads to substantial social victims or the lives of humans [3], [4], [5], [6], [7], [8], [9]. If cyberattacks are perceived and positioned quickly, the loss to overall systems will be measured within the acceptable time limit. Much of the existing literature on the detection of attacks is dependent on centralized architectures [10], [11], [12], [13].The attack detection schemes are usually categorized into knowledge-based and data-driven approaches [14]. The residual generation method is one of the representative detection strategies in many knowledge-based systems [15], [16], [17]. Usually, residual is intended by comparison of measurements of sensors and systematic model of the system. Afterward, it is equated with the static or time-variant threshold for determining whether it is an attack or not. In the case of data-driven methods, deep learning approach and heuristic algorithms are used for building models of cyber-physical systems [18], [19]. If this does not follow these associations, then the attack is assumed. Apart from centralized systems, many kinds of distributed systems appear nowadays. The main challenge of designing a distributed attack detection method is monitoring cyber-physical systems without adequate information. Most cyber-physical systems lack various cyber security mechanisms, such as message authentication, which results in numerous challenges for detecting data injection attacks [20]. The absence of worldwide encryption, mainly on systems engaging in dated technologies, makes it exciting to secure in contradiction of eavesdropping attacks. So, it is required to refer replay attacks. According to the report on the global cyber-physical system market and data bridge market research, the historical market and forecast CAGR is 7.8%. The traffic in global cyber-physical systems is expected to account for USD 12,356.23 million by 2028. This increase in traffic increases the burden on the CPS systems as the market increases. To overcome this problem, the researchers of both academia and industry explored this market, and as a result, the various privacy preservation methods are explored.

A. PROBLEM FORMULATION

Although there are various advantages of cyber-physical systems, these systems are susceptible to numerous cyber or physical security threats, attacks, and challenges. This occurs due to its non-homogeneous nature and dependency on sensitive and private data. This kind of planned or accidental acquaintance with these systems leads to terrible effects, which results in complex security measures. Though this leads to the undesirable overhead of networks. So the security measures of a cyber-physical system are required to formulate. Figure 1 represents the review methodology of this paper. It represents the searching process and reviewing results. The authors have read the various papers for collecting the noticeable information and deliberate the cyber physical systems, fault and failures, cyber security standards, and various challenges.

B. WHY DEEP LEARNING FOR CYBER PHYSICAL SYSTEMS (CPS)

Deep Learning (DL) [21], [22], [23] gives better results as compared to machine learning (ML). In case of passable data, deep learning models provide the best results. Even deep learning models are applied for solving cyber-physical system cybersecurity issues compared to other fields. It is also experiential that various deep learning models are anticipated in current publications for detecting cyber-physical systems' cyber-attacks. The main is not only the way to describe the difficulty of cyber-attack detections on cyber-physical systems; the main complexity arises when superimposing cyber security over cyber-physical systems [24]. Various authors have not had a detailed discussion on applying deep learning methods for detecting cyber-attacks contrary to cyber-physical systems. The brief survey was given by authors [25] with a four-step framework that uses deep learning methods for detecting cyber-physical systems cyber-attacks. The biggest concern nowadays is the security of CPS. Deep Learning approaches are precisely intended to handle large datasets compared to small datasets with numerous features. These methods can approximate any function as deep learning has a rich class of models. All these methods are appropriate in cyber-physical systems due to the following reasons:

- Information gathered from CPSs is commonly high layered as information from countless physical sensors and cyber sensors.
- A steady development of information because of upgrades and openness to novel susceptibilities are there.
- The models should be continually refreshed with novel information to represent the drifting of the framework and further vector assaults.

1) DEEP LEARNING WITH CPS

Deep learning has emerged as a powerful technique for handling the complexities of Cyber Physical Systems (CPS). It has been applied to various CPS applications such as anomaly detection, fault diagnosis, control, and optimization.

Deep learning algorithms such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Deep Belief Networks (DBN) have been used for CPS applications. CNNs have been used for image and signal processing tasks in CPS, while RNNs have been used for time-series data analysis in CPS. DBNs have been used for fault diagnosis and anomaly detection in CPS.

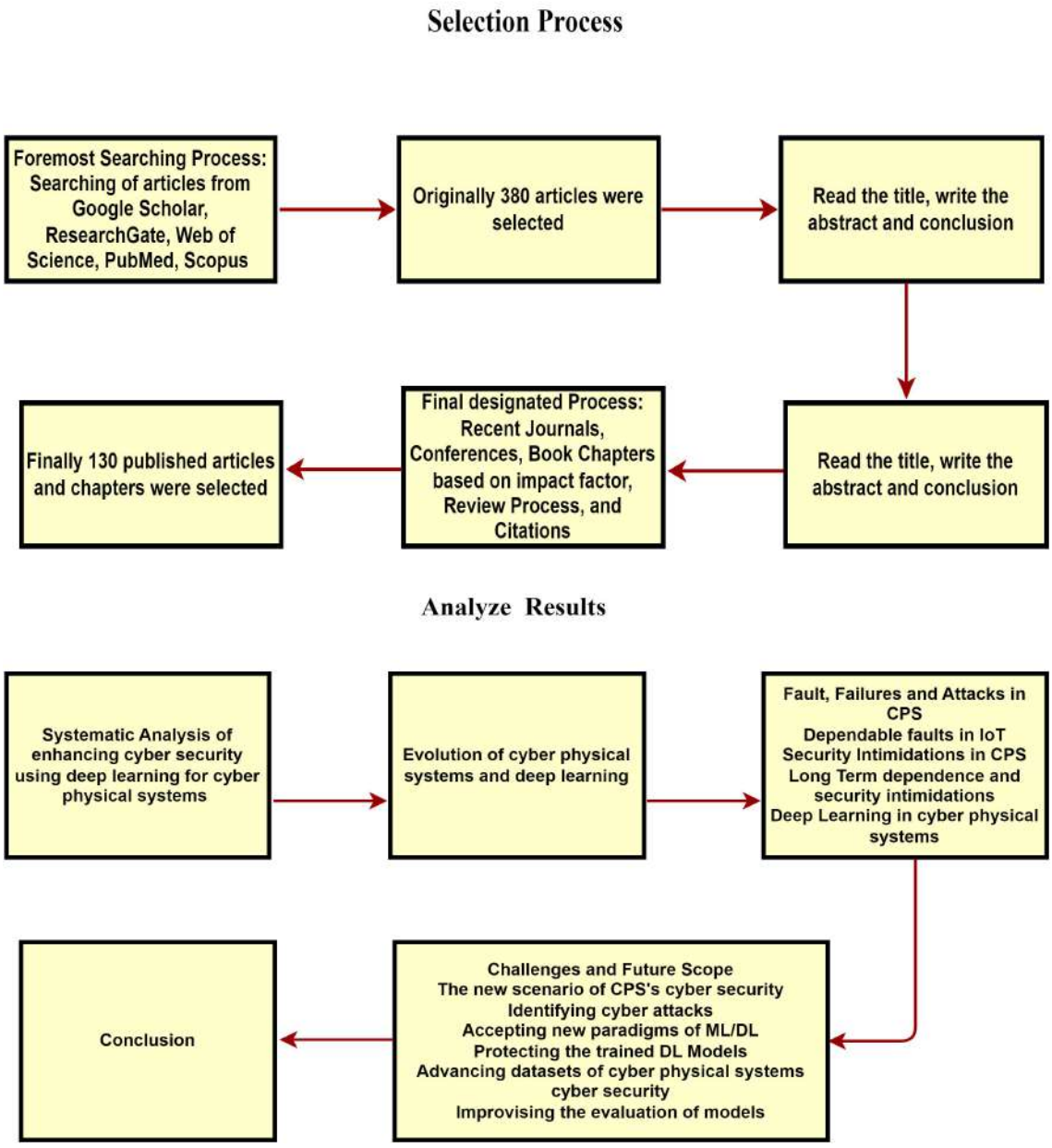


FIGURE 1. Methodology for selecting and analysing the survey.

Deep learning models require large amounts of data for training, and CPS data is often limited and expensive to collect. Transfer learning techniques have been applied to leverage pre-trained models and overcome this challenge. Additionally, the security of CPS can also be enhanced using deep learning techniques, such as using autoencoders for intrusion detection, and generative adversarial networks (GANs) for generating adversarial examples to improve the robustness of CPS. Overall, deep learning has shown promising results in various CPS applications and is expected to play a significant role in advancing the state-of-the-art in CPS.

C. QUANTUM LEARNING WITH CPS

Quantum machine learning is an emerging field that combines quantum computing and machine learning techniques to solve complex problems. However, quantum computing technology is still in its early stages of development, and its practical applications in the field of cybersecurity and CPS are still largely theoretical.

One of the potential advantages of quantum machine learning for CPS security is its ability to perform complex calculations faster than classical computing, which could potentially speed up the detection and response to cyber attacks. However, the development of quantum machine

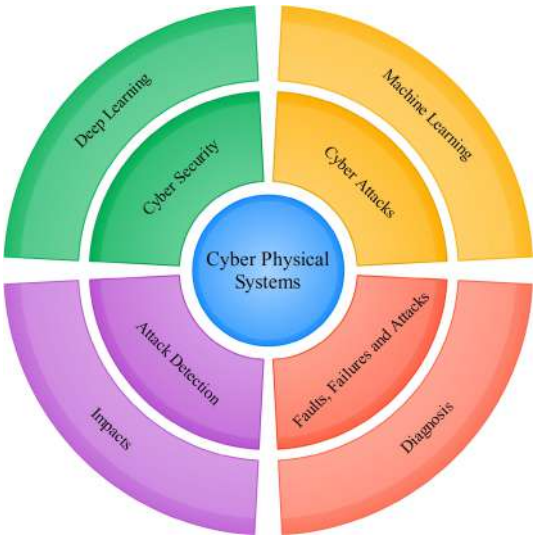


FIGURE 2. Broad division of concepts discussed in the paper.

learning algorithms and their integration into CPS systems is still a topic of ongoing research. Quantum learning with CPS is a promising area of research, but its practical applications in the field of cybersecurity and CPS are still largely speculative, and much work is needed to develop and test quantum machine learning algorithms for real-world CPS systems.

D. DEEP LEARNING AND QUANTUM LEARNING WITH CPS

Deep learning and quantum learning are two areas of research that can have potential applications in Cyber-Physical Systems (CPS).

Deep learning involves training deep neural networks to perform complex tasks, such as image and speech recognition, natural language processing, and even autonomous decision-making. In CPS, deep learning can be used to analyze large volumes of data generated by sensors and devices in real-time, detect anomalies and potential threats, and make accurate and timely decisions to ensure the safety and security of the system.

Quantum learning, on the other hand, uses the principles of quantum mechanics to process and analyze data. It involves the use of quantum algorithms and quantum computers to solve problems that are computationally infeasible using classical computers. In CPS, quantum learning can be used to optimize the performance of the system, reduce energy consumption, and enhance security by developing quantum-resistant encryption algorithms.

While both deep learning and quantum learning have potential applications in CPS, they are still in the early stages of development and require further research to fully understand their capabilities and limitations in this domain.

E. CONTRIBUTIONS

In this paper, we undertake an extensive investigation into the application of deep learning for cyber-attack detection within cyber-physical systems (CPS). Our contributions encompass:

- Here in this paper, authors have performed an exhaustive survey of contemporary methods and techniques for cyber-attack detection in CPS, harnessing the capabilities of deep learning.
- The authors have introduced a rigorous methodological framework that serves as the cornerstone of the research. This framework not only positions our work within the current landscape but also facilitates the systematic analysis and evaluation of recent developments in this domain.
- A comprehensive examination of reliability failures and security threats, specifically tailored to the various layers of CPS architecture.
- The authors have delved into the realm of solutions with meticulous attention to technical intricacies. The discussions provide in-depth insights into the implementation of security measures, considering factors such as encryption algorithms, anomaly detection thresholds, and real-time monitoring mechanisms.
- In an alignment with the core theme, the authors have engaged in a technical discourse surrounding challenges and future trends. This includes embracing novel paradigms in machine learning (ML) and deep learning (DL), devising techniques to safeguard trained DL models from adversarial attacks, advancing the construction of CPS cybersecurity datasets with a focus on data diversity and volume, and enhancing the technical rigor of model evaluation methodologies.

F. MOTIVATIONS

As soon as the intelligent computing systems introduce predictable intelligence towards the issues of cyber, so the researchers are more inclined to use intelligent computing for secure computations, as there are various challenges for detecting attacks also. The question is whether computation spectacle can help improve security concerns. The security of Cyber physical systems is a significant concern, and that's why it is mandated to study the safety of cyber physical systems. So an analysis of cyber security of the cyber physical system is required, and it is presented in this work.The taxonomy of paper is shown in Figure 3.

G. ORGANIZATION

The rest of the paper is categorized into various subsections: Section II describes the literature review of evolution of cyber physical systems and deep learning. Section III discusses the Mathematical Modeling Framework for Enhancing Cyber Security in Cyber-Physical Systems using Deep Learning. Section IV describes the fault, failures and attacks in cyber physical systems. Open issues for securing CPS are described in Section V.Challenges and Future Scope is described in Section VI. Finally paper is concluded in Section VII.

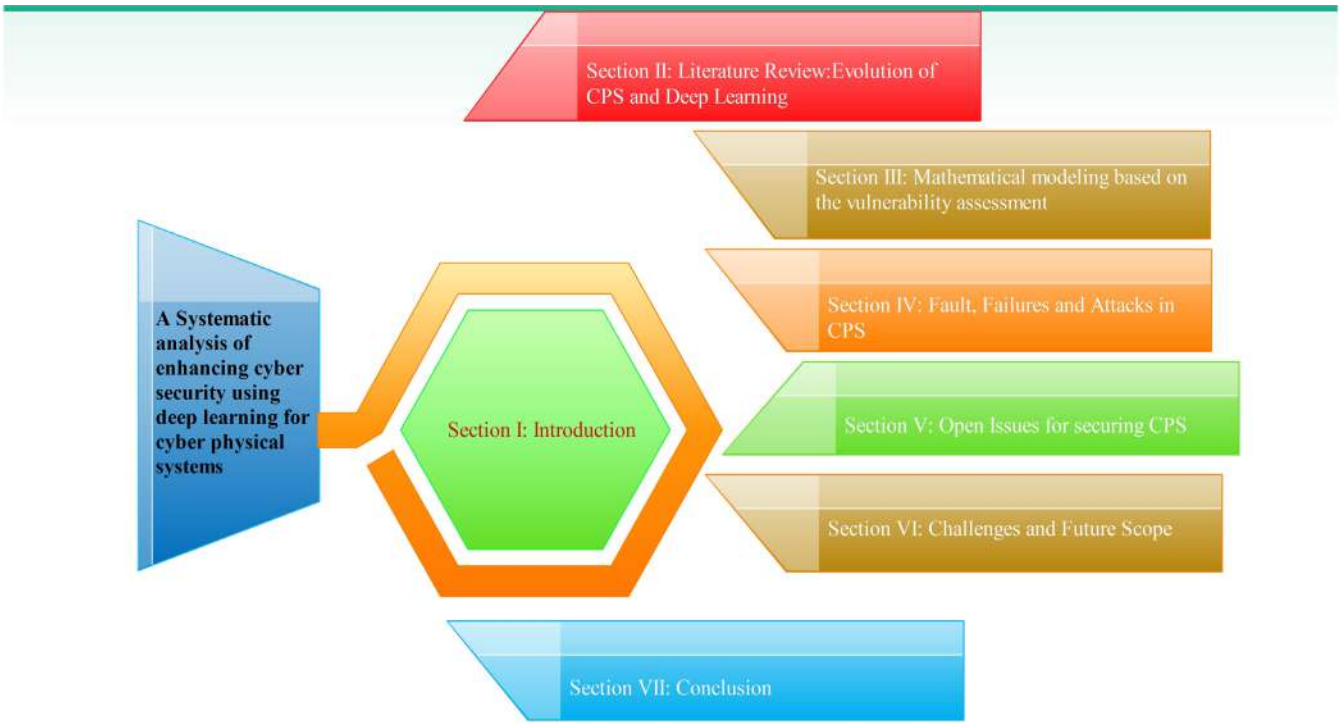


FIGURE 3. Organization of paper.

II. LITERATURE REVIEW: EVOLUTION OF CYBER PHYSICAL SYSTEMS AND DEEP LEARNING

Cyber-Physical System (CPS) is the coordination of computers with existing frameworks. The embedded computer screen, the actual control cycles, the feedback loops, and the physical approaches also influence calculations. Cyber-Physical System is near to convergence, with no association of the physical and the cyber world as a conceptual motivation. It consolidates designing representations and strategies from mechanical, ecological, typical, electrical, biomedical, compound, aeronautical, and modern designing with the models and techniques for software engineering. As the expressions ''the internet'' and ''cyber-physical system'' originate via a similar root, ''computer science,'' which is authored by Norbert Wiener [5], an American statistician who enormously affected the advancement of control frameworks theory, would be more precise. Wiener spearheaded innovation for the programmed pointing and shooting of hostile airplane weapons. Albeit the components he utilized didn't include computerized PCs, the standards contained are like those pre-owned nowadays in computer-based criticism controller frameworks [26]. The control rationale is a calculation, though one has done via simple circuits and mechanical portions, and consequently, computer science is the combination of actual cycles, analysis, and correspondence. The similitude is adept for control frameworks.

CPS is here and is mistaken for ''online protection,'' which concerns the secrecy, uprightness, and accessibility of information and has no characteristic association with actual cycles. The expression ''network protection'' along these lines is about the security of the internet and is subsequently, by implication, associated with computer science. CPS includes many testings security and protection concerns, yet these are in no way, shape, or form the main worries.

It is an innovation in that intelligence associates our actual world with our data world. Cyber Physical Systems is more essential and solid than these as it doesn't directly reference either execution draws near or precise applications like ''Industry'' in Industry 4.0. It centers as a substitute to the principal scholarly issue of adjoining the designing customs of the digital and an actual universe. One could discuss a CPS hypothesis like the ''direct frameworks hypothesis.'' CPS has turned out to be a common factor in critical infrastructure because of its massive influence and commercial assistance [6]. The growing reliance of crucial infrastructure on cyber-based skills has turned them susceptible to cyber-assaults like interference, auxiliary, and exclusion of data from the communique networks [7], [8], [9].' Therefore, the sanctuary of cyber-physical systems has become a perilous concern. A brief history of computer systems and cyber physical systems is illustrated in Figure 3.

Deep Learning (DL) has acquired huge consideration in previous years. It has worked on the state-of-art execution of numerous claims, remembering applications related to security for basic designs, like interruption identification, malware discovery, access control, and peculiarity recognition and orders [6]. DL was presented in the late twentieth era, which was begun with the investigation of Artificial Neural

FIGURE 4. Taxonomy of survey.

Networks (ANNs). Deep Neural Networks (DNN) comprise a set of layers that gain proficiency with a progression of hidden portrayals progressively [27], [28]. Higher-level descriptions contain enhanced parts of information tests that are helpful for segregation and stifle unessential highlights. Deep Learning models have worked on the cutting-edge execution in various assignments [10], [11]. The summary of related works of various methods and applications are shown in Table 1.

Figure 4 delineates the general idea of cyber-physical systems and the IoT for cyber physical systems. It displays current cyber-physical systems, how elements could be separated from such frameworks, conceivable deep learning models, and the benefits of utilizing deep learning [29]. Furthermore, the information gathered from existing digital frameworks is ordinarily high layered. Deep Learning models are explicitly intended to manage high layered information. Different attributes of CPS incorporate, proceed with the development of data, information float, and openness to new framework dangers. This way, it is crucial to assemble deep learning-based sanctuary models which are versatile and extendible with the information float, nonstop disclosure of new framework dangers and weaknesses [12].

This idea of ''Generalization'' is one significant issue for constructing security-based requests in cyber-physical systems as creating AI models for one situation is almost difficult to use, experiencing the same thing even in a similar setting. In this manner, it is a quintessence to zero in on speculation that deep learning models utilized in such applications are ordinarily high layered. Deep Learning models are explicitly intended to manage high-layered information. Different attributes of CPS incorporate, proceed

with the development of data, information float, and openness to new framework dangers. This way, it is crucial to assemble deep learning-based security models which are versatile and extensible by the information float, nonstop disclosure of new framework dangers and weaknesses [12]. This idea of ''Generalization'' is one significant issue for constructing security-based requests in cyber-physical systems as creating AI models for one situation is almost difficult to use, experiencing the same thing even in a similar setting. It is illustrative to zero in on speculation of deep learning models utilized in such applications [30], [31].

III. MATHEMATICAL MODELING BASED ON THE VULNERABILITY ASSESSMENT

The mathematical modeling framework for enhancing cyber security in Cyber-Physical Systems (CPS) using deep learning, based on the vulnerability assessment are stated below and explained in figure 5.

1) **Problem Formulation:** Minimize the objective function $J(\Theta)$, representing the cost or vulnerability.
2) **System Representation:** Define the CPS system as $CPS = \{C_1, C_2, \ldots, C_n\}$. Enumerate vulnerabilities as $Vulnerabilities = \{V_1, V_2, \ldots, V_m\}$.
3) **Threat Modeling:** Define a Threat Vector $T = [T_1, T_2, \ldots, T_k]$ representing potential threats.
4) **Deep Learning Integration:** Integrate deep learning models to process system information. Define the model's output as f_Θ(Input).
5) **Data Requirements:** Specify the dataset $D = \{(Input_1, Label_1), \ldots, (Input_N, Label_N)\}$ for model training.
6) **Mathematical Equations:** Develop equations to quantify vulnerability levels.

$$\text{Vulnerability Level} = g(\text{Threat Vector, Deep Learning Output})$$

7) **Quantification of Vulnerabilities:** Assign a vulnerability score based on the vulnerability level.

$$\text{Vulnerability Score} = h(\text{Vulnerability Level})$$

8) **Validation and Verification:** Establish validation metrics to evaluate model performance.
$ValidationMetric = ValidationFunction(ModelOutput, GroundTruth)$
9) **Sensitivity Analysis:** Assess model sensitivity to parameter changes.
$Sensitivity = \frac{\partial J}{\partial \Theta}$
10) **Limitations and Assumptions:** Clearly state any assumptions and limitations in the model.

$$\text{Assumption}_i : \ldots$$
$$\text{Limitation}_j : \ldots$$

11) **Comparative Analysis:** Develop metrics for comparing the model against other approaches.

$$\text{Comparison Metric} = \text{Compare(Model, Other Models)}$$

12) **Implications and Recommendations:** Discuss the implications of the findings. Provide recommendations for practical applications.

IV. FAULT, FAILURES AND ATTACKS IN CYBER PHYSICAL SYSTEMS

A failure is an occurrence that arises when an organization diverges as of its planned performance. The failure establishes because of its inadvertent state. The origin of a fault might be internal or external. The internal faults occur due to their physical nature (such as brokerage of the component connector), and faults occur due to their design (software or hardware-related bugs) [55]. Peripheral faults (External) initiate from the environmental cause like noise. Faults may be categorized into permanent and temporary faults. However, a temporary fault occurs for short time span. It may create an error, and this may lead to perpetual failure. Similarly, Physical faults and inputs can be temporary or it can be permanent, whereas the design faults are constantly permanent. The faults which could not be analytically imitated are usually known as irregular faults. This kind of fault can be led to soft errors.

The cyber-physical systems/Internet of Things (CPS/IoT) infrastructure is shown in the figure. Faults might arise at diverse layers of architecture, such as the physical layer or control layer, respectively [13]. The physical layer is susceptible to interruption, direct interference, or demolition of physical items. The network layer can make the connection of devices. The monitors and controllers in the control layer are susceptible to environmental uncertainties and handling of extents and control signals [55], [56]. The collection of information can be done by the information layer and is mainly vulnerable to issues related to secrecy and integrity.

A. DEPENDABLE FAULTS IN IoT

Internet of Things tends to communicate failures primarily due to its extent and heterogeneousness. Previously, traditional cyber-physical systems used to ignore or remove such failures by validating and verifying the design. Though IoT is involved in technology, it is growing in size with time. The subsequent faults might arise per cyber-physical systems layer:

- **Physical Layer: - Intrusion:** Interference of a signal. The quantity of associated devices and the radiation rises, affecting measurements of sensors, conveyed communications, or control indications.
- **Network Layer: - Collision of Messages:** In correspondence to intrusion, the quantity of interactive devices may activate communique failures such as crashes or overloading of the net- ' work. - Violation of Protocol:

TABLE 1. Summary of different methods and applications in the context of Deep Learning and CPS by various authors, along with challenges in ML algorithms.

Methods	Deep Learning	Application	Cyber Physical Systems	Reference	Challenges in ML Algorithms
Classification based techniques, Clustering based techniques, Statistical, anomaly recognition approaches	No	Cyber Interruption Detection, Detection of Fraud, etc	No	[32]	Limited labeled data, imbalanced datasets, model interpretability
Program Analysis	No	Commodity Internet of Things	Related but not fully covered	[33]	Scalability, real-time processing
Physical properties	No	Cyber Physical Systems	Yes	[34]	Sensor noise, environmental variability
Deep learning	Yes	Cyber Interruption Detection, Detection of Fraud	No	[35]	Model complexity, computational resources
Attack Based Tree, Model-based technique	No	Cyber Physical Systems (focus on SCADA)	Yes	[36]	Security of control systems, attack detection
Deep learning	Yes	Cyber Physical Systems	Yes	[37]	Scalability, real-time processing
Knowledge-Based technique, Behaviour-Based Interruption Recognition system	No	Cyber Physical Systems	Yes	[38]	Knowledge representation, anomaly detection
Interruption Detection system, Machine learning	No	Cyber Physical Systems	Yes	[39]	False positives, adaptive adversaries
-	-	Smart home IoT	Related but not fully covered	[40]	Privacy, device heterogeneity
Plant models based technique, Noise-based detection, State estimation based technique	No	Cyber Physical Systems	Yes	[41]	Model accuracy, noise robustness
Deep learning	Yes	Internet of Things	No	[42]	Energy efficiency, resource constraints

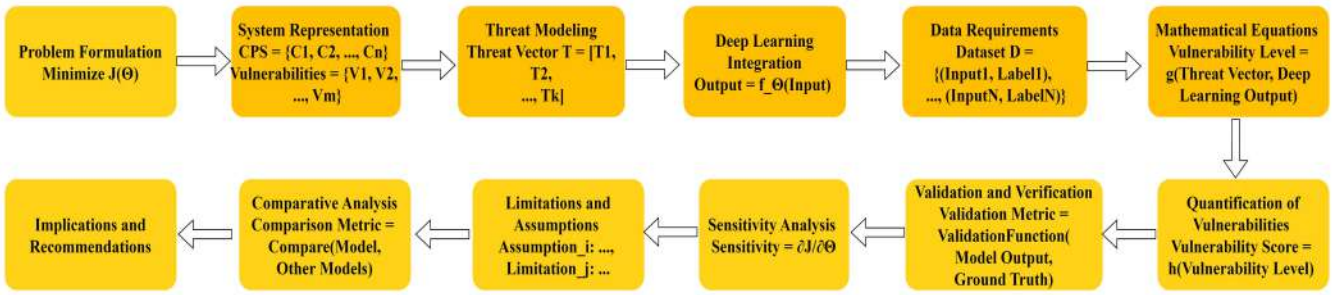


FIGURE 5. Mathematical modeling framework for enhancing cyber security in cyber-physical systems using deep learning.

The protocol violation occurs due to incorrect message content.

- **Control Layer: - Deadline Miss:** Delayed in the response of control signal. The Control loops has to survey the restraints related to timing of a cyber-physical system application. - Misusage: Sending erroneous inputs to a component
- **Information Layer: - Inaccessibility:** Lost data instigated by a skill apprise. The things might be linked, detached, or updated in the Internet of Things.

B. SECURITY INTIMIDATIONS IN CYBER-PHYSICAL SYSTEMS

Security has been one of the biggest concerns in computer networks to identify susceptibilities and avoid malicious attacks on the devices. Whereas in cyber-physical systems, more and more susceptibilities arise in the physical area and the indeterminate behavior of the physical atmosphere. The categorization of attacks applied per cyber-physical systems layer is given below:

- **Physical Layer: - Information Leakage:** Stealing perilous information from various devices such as private keys - Denial of Service: Manipulating various parameters for performing DoS attacks.
- **Network Layer: - Jamming:** Overloading the communication protocol by introducing false traffic. - Collision: Manipulation of timing, the power which leads to collision of data or violation of communication protocol. - Routing misdirects: Manipulating the routing mechanism leads to collision of data, flooding of data, and discriminating promoting of facts [57], [58].
- **Control Layer: - Desynchronizing:** Violating the timing or manipulation of clocks. This could lead to denial of service and leakage of information.
- **Information Layer: - Eavesdropping:** Stealing or sniffing of information. It is one of the biggest intimidations associated with confidentiality. Furthermore, data could also be deployed to accomplish various attacks. The potential intimidations and penalties could be stated in sanctuary intimidation models for cyber-physical systems.

TABLE 2. A systematic analysis of enhancing cyber security using deep learning for cyber physical systems" vs. existing survey papers.

Paper Reference	Scope and Focus	Methodological Approach	Technical Depth	Contributions	Originality
[43]	Smart Grid attacks, vulnerabilities, detection and defences	Review, analysis, and synthesis	Moderate	Identify key challenges, proposed solutions	Focused on existing attacks and defenses
[44]	IoT intrusion detection methods	Review, analysis, and synthesis	Moderate	Classification and detection techniques	Explores existing methods
[45]	ML techniques in CPS cyber security	Review, analysis, and synthesis	Moderate	Comparative analysis of methods	Emphasis on existing ML techniques
[46]	Survey on deep learning-based attack detection in CPS cybersecurity	Analytical survey of existing literature, summarizing key techniques and challenges	High technical depth, covers a wide range of deep learning techniques in the context of CPS security	Comprehensive analysis and insights into the state-of-the-art in deep learning -based attack detection for CPS cybersecurity	Original in presenting a holistic view of deep learning in CPS attack detection
[47]	Review of security analysis in CPS using machine learning	Literature review and analysis of machine learning applications in CPS security	Moderate technical depth, focuses on summarizing existing research in the domain	Provides a comprehensive review of security analysis in CPS using machine learning techniques	Original in presenting a consolidated overview of ML applications in CPS security
[25]	Survey on the generalization of deep learning for CPS security	Analytical survey of deep learning generalization techniques in CPS security	Moderate technical depth, emphasizing generalization aspects	Offers insights into the challenges and opportunities in applying deep learning for CPS security with a focus on generalization	Original in exploring generalization aspects in deep learning for CPS security
[48]	Survey on resilient machine learning for networked CPS	Analytical survey of machine learning security in the context of networked CPS	High technical depth, covers a range of resilient machine learning techniques	Provides a comprehensive survey on securing machine learning in networked CPS environments	Original in addressing resilience challenges in machine learning for networked CPS
[49]	Survey on deep learning-based anomaly detection in CPS	Analytical survey of progress, challenges, and opportunities in deep learning-based anomaly detection	High technical depth, explores various deep learning-based approaches	Offers a comprehensive overview of the progress and potential in deep learning-based anomaly detection for CPS	Original in presenting a state-of-the-art survey on anomaly detection in CPS using deep learning
[50]	Federated deep learning for intrusion detection in industrial cyber–physical systems	DeepFed: Federated deep learning	High	Intrusion detection, Federated learning	Novel approach in applying federated learning to industrial CPS
[51]	Attack graph model for cyber-physical power systems using hybrid deep learning	Hybrid deep learning approach	High	Attack graph modeling, Smart Grid security	Integration of deep learning into attack graph modeling for power systems
[52]	Real-time stability assessment in smart cyber-physical grids: a deep learning approach	Deep learning for real-time stability assessment	Moderate	Stability assessment in smart grids	Application of deep learning to real-time stability analysis in smart grids
[53]	Blockchain-based deep learning approach for cybersecurity in next-generation industrial cyber-physical systems	Blockchain and deep learning integration	High	Cybersecurity, Blockchain, Industrial CPS	Unique combination of blockchain and deep learning for enhanced cybersecurity
[54]	Deep learning-based DDoS-attack detection for cyber–physical system over 5G network	Deep learning for DDoS attack detection	High	DDoS attack detection, 5G networks	Application of deep learning for DDoS detection in 5G- enabled CPS
Our Paper	CPS security using DL	Systematic analysis	High	Innovative solutions, key challenges, and insights	Emphasis on original insights

C. LONG-TERM DEPENDENCE AND SECURITY INTIMIDATIONS

The Internet of things and cyber-physical systems will endure deviations over time, particularly when imperiled by a long operational period. Following features of the change might cause faults such as changes in the environment, functional changes, and changes in technology.The catagories of attacks implied on different CPS systems are mentioned below:

- **Physical Layer:**At this layer there is material decay and environmental effects issues at this layer and this voilates the environment.
- **Network Layer:**It overloads the information by putting false traffic on the network. Due to this the communication through protocol also voilates.
- **Control Layer:**It disturbs the timing or manipulates the clocks this leads to the aging effects and uncertainity effects.

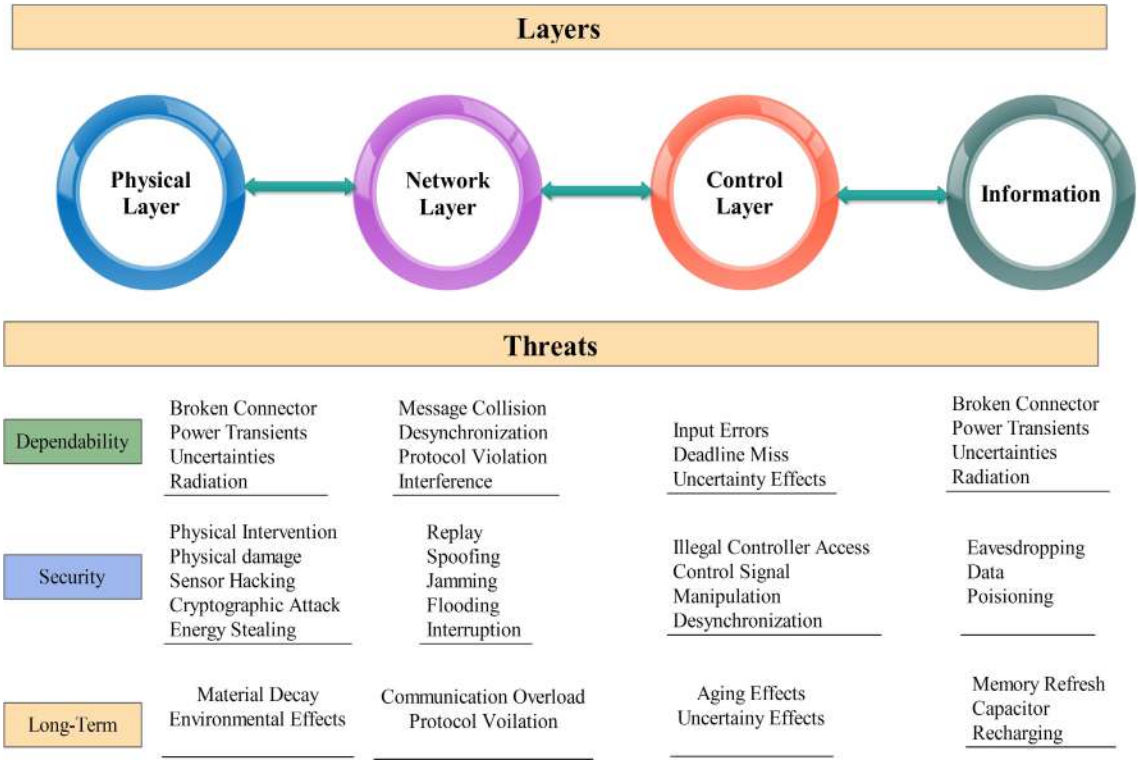


FIGURE 6. Reliability failures and sanctuary intimidations with reference to cyber physical systems layers.

TABLE 3. Threat models for different CPS layers.

Layers	Physical Layer	Sensor/Actuator Layer	Communic-ation Layer	Control Layer	Information Layer	Integration level Layer
Attacks	Manufacturer, Designer, External Attacker	Manufacturer, Designer, External Attacker	Manufacturer, Designer, External Attacker	Manufacturer, Designer, External Attacker	Manufacturer, Designer, External Attacker	Manufacturer, Designer, External Attacker
Methodology	Physical Interference	Hacking, Control Access, Information Influences	Replay, Sybil, Congestion, Implosion, Deceiving	Eavesdropping, Control Access	Eavesdropping	All conceivable control & Communication assaults
Payloads	Denial of Service, Aging Consistency	Energy Stealing, Denial of Service, Data Leakage, Desynchro-nization	Energy Thieving, Denial of Service, Information Leakage, Desynchro-nization	Leakage of Information, Denial of Service, Desynch-ronization	Leakage of Information	Stealing of energy, Denial of Service, Leakage of Information, Desynchron-ization

- **Information Layer:**The biggest itimidation concerned with confidentiality is stealing of information. The refershing of memory, recharging is theissues relaed to this layer.

D. DEEP LEARNING IN CYBER-PHYSICAL SYSTEMS

Here we discuss how deep learning can be applied in CPS. So, the introduction of deep learning is required for how it is used in security-related applications such as CPSs. Nowadays, DL is gaining huge focus in data science to enhance performance in various applications [12]. Deep Learning Algorithms contain hierarchical architectures with many layers in which higher-level features are explained in standings of lower-level features capable for the extraction of features and concepts from underlying data [14]. These architectures can produce outstanding results in applications like cyber-physical systems security [12], [15]. Figure 5 presents various applications of DL for CPS. The deep architectures are formed of various hidden layers [4]. Deep Learning methods can represent additional abstract illustrations of information due to the multi-level architecture.

Deep Learning models have revealed better generalization competence in many practical applications than shallow ANNs.

There are some major fields where deep learning has been effectively applied in cyber-physical systems for security-related determinations such as detection of anomaly, detection of malware and threat hunting, susceptibility recognition, interruption detection, prevention of blackouts, assaults, and destructions in CPSs.

E. CYBER ATTACKS

In current years, there was a hike in the proportion of cyber attacks aiming cyber physical systems with distressing significances. As per recent studies [86], [97], cyber physical systems are susceptible to malicious code injection attacks [66] and code reuse attacks [76]in addition with false data injection attacks [77], zero-control data attacks [83]. These kinds of attacks can lead to black out targeting cyber physical system's industrial devices and systems as shown in Table 4.

V. OPEN ISSUES FOR SECURING CPS

There are several open issues and research directions related to securing Cyber Physical Systems (CPS) using Deep Learning (DL). Some of the key areas of focus include:

- **Data Collection and Preparation:** CPS typically generate vast amounts of data that are relevant to the security of the system [31], [55]. However, collecting and preparing this data for use in DL models can be challenging, particularly when the data is highly heterogeneous and distributed across multiple sources [105].
- **Model Selection and Development:** There is a need to identify the most appropriate DL models for securing CPS and to develop these models so that they can be effectively applied to real-world scenarios [106]. This includes choosing the right type of model, such as CNNs or RNNs, and optimizing the model's architecture and parameters to improve its performance.
- Integration with Other Security Measures: DL models need to be integrated with other sanctuary procedures to ensure that they are effective in detecting and mitigating cyber threats [107], [108]. This may include integrating DL models with intrusion detection systems, firewalls, or access control systems, or incorporating additional data sources such as log data or network traffic data to improve the accuracy of the models.
- **Scalability and Real-Time Processing:** CPS generate huge quantities of data in real-time, which makes it challenging to use DL models to detect and answer to cyber intimidations in real-time [109], [110]. There is a need for DL models that are able to scale to handle large amounts of data and that can be implemented in real-time to detect and reply to cyber intimidations in real-time.
- **Explainability and Trustworthiness:** One of the challenges of using DL models for security purposes is the lack of transparency and interpretability of the models. There is a need to develop DL models that are more transparent and interpretable, so that security experts and decision-makers can understand the basis for the models' predictions and decisions [55], [56].
- **Adversarial Robustness:** CPS are often targeted by sophisticated cyber-attackers who use techniques such as adversarial machine learning to evade detection. There is a need for DL models that are robust to these attacks and that can continue to operate effectively even in the presence of adversarial inputs [111].

These are some of the key areas of focus for securing CPS using DL, and there is a growing body of research aimed at addressing these challenges. By developing DL models that are effective in detecting and mitigating cyber threats, and by integrating these models with other security measures, it is possible to improve the sanctuary of CPS and reduce the risk of cyber-attacks.

A. RESEARCH DIRECTIONS IN SECURING CPS USING DL

There are several open issues and research directions for securing CPS using DL techniques. Some of these include:

- **Development of robust and accurate DL-based intrusion detection systems for CPS:** This involves the usage of DL methods such as CNNs and RNNs to detect and classify various types of cyber-attacks in CPS.
- **Improving the interpretability of DL-based CPS security models:** Currently, one of the main limitations of deep learning models is their absence of interpretability, making it hard to comprehend how they attain at their decisions. Research is needed to make DL models more transparent and interpretable [57].
- **Anomaly detection in CPS using unsupervised DL techniques:** Unsupervised DL techniques such as Autoencoders and Variational Autoencoders (VAEs) could be utilized to detect anomalies in CPS by learning the normal behavior of the system and identifying nonconformities from this normal behavior [112].
- **Adversarial attacks on DL-based CPS security models:** Adversarial attacks are a major concern in DL, and they pose a threat to the security of CPS systems. Research is needed to develop defense mechanisms against these attacks and to enhance the robustness of DL-based security models [113].
- **Integration of DL with other security techniques:** DL-based security models can be combined with other security techniques such as firewall, detection of intrusion and anticipation systems, and encryption to create a more comprehensive and effective security system for CPS [114].
- **Handling large and complex data in CPS using DL:** CPS systems generate large amounts of data, and this data is often complex and unstructured. Research is needed to develop DL models that can handle this data effectively and efficiently [58], [115].

TABLE 4. Different CPS system with different types of anamolies.

CPS System	Existing Work	Type of Anomalies								
		Attacks					Faults			
		DoS	MITM	Packet Injection	Malware	FALSE Control Signals	Sensor Layer	Network Layer	Control System	Manually Created
Industrial Control System	[37]	Yes	Yes	X	No	No	No	No	No	No
	[59]	No	Yes	No	No	Yes	No	No	No	No
	[60]	No	No	No	No	No	No	No	No	X
	[61]	No	No	No	No	No	Yes	No	No	No
	[62]	No	No	No	No	No	Yes	No	No	Yes
	[63]	No	Yes	No	No	Yes	No	No	No	No
	[64]	Yes	No	Yes	No	Yes	No	No	No	No
	[65]	No	Yes	No	No	Yes	No	No	No	No
	[66]	No	No	No	No	No	No	Yes	No	No
	[67]	No	No	No	No	No	Yes	No	No	No
	[68]	No	No	No	No	No	Yes	No	No	No
	[69]	No	No	No	No	No	Yes	No	No	Yes
	[70]	No	No	No	No	No	Yes	No	No	Yes
	[71]	No	No	No	Yes	No	Yes	No	No	Yes
	[72]	Yes	No	Yes	Yes	X	No	No	No	Yes
	[73]	No	No	Yes	No	No	No	No	No	
	[74]	No	Yes	No	No	Yes	No	No	No	No
Smart Grid and ITS	[75]	No	No	No	No	No	No	No	No	Yes
	[76]	Yes	X	Yes	Yes	Yes	No	No	No	Yes
	[77]	No	Yes	No	No	No	No	No	No	Yes
	[78]	No	Yes	No	No	No	No	No	No	Yes
	[79]	No	Yes	No	No	No	No	No	No	Yes
	[80]	No	Yes	No	No	No	No	No	No	Yes
	[81]	No	Yes	No	No	No	No	No	No	Yes
	[82]	No	No	No	No	No	Yes	No	No	No
	[83]	No	Yes	No	No	No	No	No	No	Yes
	[84]	No	No	No	No	No	No	No	No	No
	[85]	No	Yes	No	No	No	Yes	No	No	Yes
	[86]	No	No	Yes	No	Yes	No	No	No	Yes
	[20]	Yes	No	Yes	No	Yes	No	No	No	Yes
	[87]	No	No	Yes	No	No	X	No	No	Yes
	[88]	Yes	Yes	Yes	No	Yes	No	No	No	No
	[89]	No	Yes	Yes	No	No	No	No	No	Yes
Aerial System	[90]	No	No	No	No	Yes	No	No	No	No
	[91]	No	No	No	No	Yes	No	No	No	No
	[92]	No	No	X	Yes	No	No	No	No	No
	[93]	No	No	No	No	No	No	Yes	No	No
	[94]	No	No	No	No	No	No	No	Yes	Yes
	[95]	No	No	No	No	No	No	No	No	Yes
	[96]	No	No	No	No	No	No	No	No	Yes

Note*: X belongs to Not Clear but inferred to be Yes

VI. CHALLENGES AND FUTURE SCOPE

The major potential fields are shown in the figure where the research areas may arise. The seven steps of the research methodology are already shown in the figure. The research literature can be improvised with our research methodology, and with the help of this, the comparative analysis can be done appropriately. The further challenges can be categorized into different directions, which are:

A. THE NEW SCENARIOS OF CYBER-PHYSICAL SYSTEM'S CYBERSECURITY

The various articles analyzed communication networks in the scenarios of cyber-physical systems [25], [103]. Most of the survey papers have examined the methods of cyber-physical systems in the intelligent grids or the water treatments of plants described in the 13/27 survey papers. It is the emergent way to apply deep learning in the current industry. Deep Learning is used for detecting the faults and defects in the industrial sector of complex items [97]. But these were not considered as there were no issues of cyber security covered in this. The cyber-attacks and threats usually exist in the cloud server where design models are stored. We suggest that the blockchain will be analyzed more in a broader way in collaboration with cyber-physical systems, and the variety and development of cyber-physical systems scenarios will lead to intense analysis of cyber security [116], [117], [118].

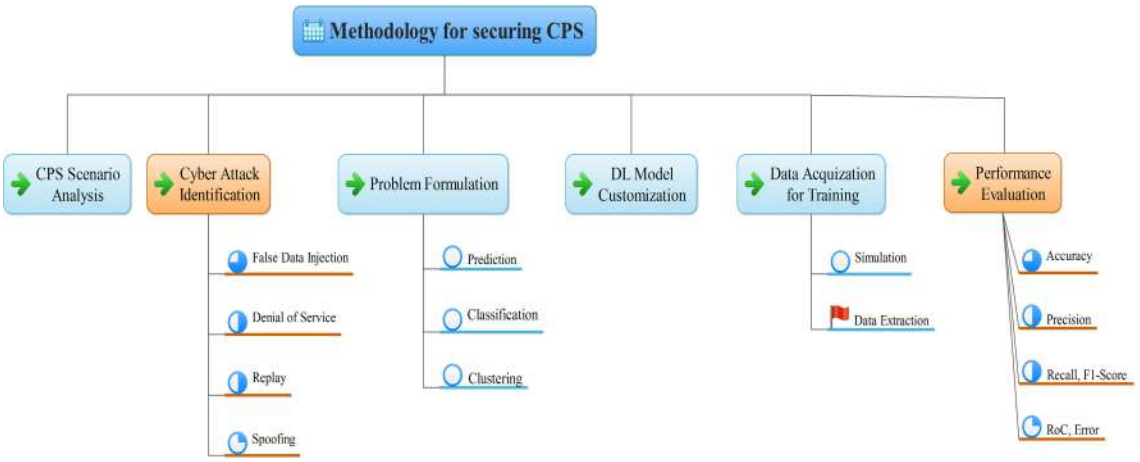


FIGURE 7. The deep learning driven methodology for security of cyber physical systems.

TABLE 5. Real cyber physical systems attacks.

Country	Nature of Attack	Type	Target	Motives	Date
USA	Slammer Worm Sensors Failure	Malware-DoS Accident	Ohio Nuke Plant Network [98] Taum Sauk Hydroelectric Failure of Power Station [99]	Criminal N/A	Jan 25,2003 14 Dec 2005
	Installed Software Update	Undefined Software	Georgia Nuclear Power Shutdown of Plant [100]	Unclear	Mar 7, 2008
	Reconnaissance	Undefined Software Programs	US Electricity Grid [101]	Political	Apr 8, 2009
	Backdoor	Unauthorised Access	Springfield Pumping Station [97]	Criminal	Nov 8, 2011
	Physical Breach	Unauthorised Access	Georgia Water Treatment Plant [102]	Criminal	Apr 26, 2013
Iran	Stuxnet [103]	Worm	Iranian nuclear facilities	Political	Nov,2007
	Stuxnet-2	worm	power plant and another industry	Political	25 Dec 2012
	DDoS	Disruptive	Iranian Infrastructure and communications companies	Political	03 Oct 2012
	Computer Virus	Malware	Iranian key oil facilities	Political	23 Apr 2012
Saudi Arabia	Shamoon-1	Malware	Saudi infrastructure in the energy industry	Religio-Political	15 Aug 2012
	Shamoon-2	Malware	Saudi government computers and targets	Religio-Political	17 Nov 2016
	Shamoon-3	Malware	Tasnee and other petrochemical firms, National Industrialization Company, Sadara Chemical Company	Religio-Political	23 Jan 2017
Qatar	Shamoon	Malware	Qatar's RasGas	Political	30 Aug 2012
United Arab Emirates	Trojan Laziok	Malware	UAE energy sector	Political	Jan-Feb 2015
Australia	Remote Access	Unauthorised Access	Maroochy Water Breach [73]	Criminal	March, 2000
Canada	Security Breach	Exploited Vulnerability	Telvent Company [104]	Criminal	Sept 10, 2012

B. IDENTIFYING CYBER ATTACKS

Most of the survey papers have analyzed the false data injection attacks. Recognizing surreptitious untruthful data injection attacks is challenging as a considerable amount of noise is being formed in the cyber-physical systems, and there is a deficiency in the mechanisms of cyber security for authenticating the devices and messages which are transmitted over the network. Some categories of false injection attacks depend on the information of invaders [119], [120]. As no such advanced information is needed to initiate a denial of services attacks, individually logged packets are required for replay attacks, scanned tools for penetrating attacks, and automated tools for fuzzy attacks. However, the cyber security of cyber-physical systems is a vast area compared to cyber-attacks in contradiction to cyber-physical systems. Detection of cyber-attacks that are initiated in cyberspace and infiltrate the physical domain is a challenging task [22], [121]. We assume that emergent cyber-attacks will head the defense devices, but the risk could be moderated via the data-driven approach.

C. ACCEPTING NEW PARADIGMS OF MACHINE LEARNING/DEEP LEARNING

Usually, all analyzed papers follow conventional machine learning standards, includes supervised and unsupervised learning. Around 4 papers inspected problems of regression, 3 papers are related to problems of clustering, and others are based on problems of classification. The directing usage of supervised learning reflects the value of using well-labeled data [21], [122], [123]. Particularly, network packets were labeled as usual or attack traffic, and the kinds of attacks were distinguished. This dependence on labeled data is limited to the broader acceptance of machine learning or deep learning methods. We suggest that the researchers and authors use new machine learning/deep learning paradigms [124]. It includes reinforcement and self-supervised learning to improve the explainability of the model. We suggest self-supervised learning flourishes in the cyber-physical system's domain as deep learning models suffer from deprived explainability. We are expectant about predicting that the deep learning models will be further reasonable when new tools and techniques are conceived and utilized [125].

D. PROTECTING THE TRAINED DEEP LEARNING MODELS

No survey papers are measured for defending the trained deep learning models, contrary to numerous attacks. In contrast, we highlight the significance of protecting the trained deep learning models due to the computational expenditures for introducing the deep learning models [126], [127]. The attackers can acquire adequate data to imitate a machine learning/deep learning model by generating many inquiries and conglomerating the outcomes. The removed data can be utilized to construct a mirroring model for the assailant to find conceivable avoidance assaults. We emphatically advocate that cyber defense be led quickly because of the ignorance of adversarial assaults in the cyber physical systems situations.

E. ADVANCING DATASETS OF CYBER PHYSICAL SYSTEMS CYBERSECURITY

Between the reviewed papers, datasets gathered in the area ruled the simulation with a proportion of 14:6. Simulated information was explored in the 2 cyber physical systems situations – shrewd matrices and vehicular organizations [128]. Five papers utilizing field information picked the Smack dataset, two reports the CICIDS2017 dataset, and the other diverse datasets [129], [130], [131].

Additionally, new datasets will constantly be essential and appreciated. In a perfect world, the new datasets are publicly released field information gathered from physical testbeds. A few cyber physical system testbeds are proposed to work with recognizing cyber assaults [24].The new pattern of expanding interest in building cyber physical system testbeds may help specialists to gather superior-grade assault and defense information [132], [133]. The new datasets are enormous enough to take advantage of deep learning models' power, and both new and old cyber assaults should be incorporated because cyber assaults advance rapidly. If naming information is tested, sequentially isolating the assaults from the typical traffic is a practical thought. Falsely mixing the information passages addressing assaults into a bunch of ordinary traffic records should be kept away from because the basic information increase strategy doesn't consider practicality, going after groupings, and potential connections changes. To help the headway of exploration and information, we emphatically energize more high-quality datasets increasingly to be made accessible to the local area [134], [135].

F. IMPROVISING THE EVALUATION OF MODELS

Standard execution measurements were utilized in the vast majority of the reviewed papers. Misleading up-sides were examined, precision and fault rate. This is demonstrated by authors [65] that it is fundamentally further hard to distinguish the seldom-happened assaults than the normal ones determined by the Bayesian regulations [136], [137].

Moreover, time is essential in ongoing investigations since each prepared machine learning or deep learning model's presentation will unavoidably corrupt over the long run. When the cyber develops quickly, the models prepared with old information will battle with identifying new assaults. A period rot metric was proposed in to assess a prepared model's presentation misfortune. By concentrating on the time rot, we will want to choose when the model should be retrained. We want to see future work like about cyber physical systems and cyber assaults. When top-to-bottom information is created and acquired, we might hope to relieve the risk of cyber-physical systems' cyber assaults.

VII. CONCLUSION

This review gives an ongoing perspective on recognizing cyber-attacks in the cyber physical systems.In particular, an inclusive perception is obtained through analyzing the cyber-physical systems situations, recognizing cybersecurity issues, interpreting the exploration issue to the machine learning/deep learning space, developing the deep learning model, planning datasets, and lastly, assessing the model. The Cyber attacks endure as a constant and conspicuous danger to the safety and betterment of cyber-physical systems. The work shows extraordinary potential to take advantage of cyber physical system's cyber information through deep learning models as a result of their promising demonstrations. We distinguished favorable examination issues, incorporating blockchain, identifying cutting-edge, steady dangers, taking on new machine learning and deep learning standards, avoiding adversarial and attacks of model extraction, enhancing datasets, and utilizing different execution measurements. We are hopeful and sure that the examination in this field will thrive.

REFERENCES

[1] N. Falliere, L. O. Murchu, and E. Chien, "W32. stuxnet dossier," Secur. Response, Symantec Corp., Cupertino, CA, USA, White Paper, Version 1.4, Feb. 2011, p. 29, vol. 5, no. 6.

[2] A. Humayed and B. Luo, "Cyber-physical security for smart cars: Taxonomy of vulnerabilities, threats, and attacks," in *Proc. ACM/IEEE 6th Int. Conf. Cyber-Phys. Syst.*, Seattle, WA, USA, Apr. 2015, pp. 252–253.

[3] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *Proc. Int. Conf. Crit. Infrastruct. Protection*. Boston, MA, USA: Springer, 2007, pp. 73–82.

[4] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Secur. Privacy*, vol. 9, no. 3, pp. 49–51, May 2011.

[5] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, Feb. 2011.

[6] C.-H. Lee, B.-K. Chen, N.-M. Chen, and C.-W. Liu, "Lessons learned from the blackout accident at a nuclear power plant in Taiwan," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2726–2733, Oct. 2010.

[7] J. P. Conti, "The day the samba stopped [power blackouts]," *Eng. Technol.*, vol. 5, no. 4, pp. 46–47, Mar. 2010.

[8] Y. Liu and S. Hu, "Cyberthreat analysis and detection for energy theft in social networking of smart homes," *IEEE Trans. Computat. Social Syst.*, vol. 2, no. 4, pp. 148–158, Dec. 2015.

[9] Y. Liu and S. Hu, "Smart home scheduling and cybersecurity: Fundamentals," in *Smart Cities and Homes*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 191–217.

[10] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3864–3872, Sep. 2016.

[11] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.

[12] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1652–1656, Oct. 2015.

[13] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Inform.*, vol. 13, no. 2, pp. 411–423, Sep. 2016.

[14] K. Paridari, N. O'Mahony, A. E.-D. Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration," *Proc. IEEE*, vol. 106, no. 1, pp. 113–128, Jan. 2018.

[15] A.-Y. Lu and G.-H. Yang, "Secure state estimation for cyber-physical systems under sparse sensor attacks via a switched Luenberger observer," *Inf. Sci.*, vol. 417, pp. 454–464, Nov. 2017.

[16] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[17] A. J. Gallo, M. S. Turan, P. Nahata, F. Boem, T. Parisini, and G. Ferrari-Trecate, "Distributed cyber-attack detection in the secondary control of DC microgrids," in *Proc. Eur. Control Conf. (ECC)*, Limassol, Cyprus, Jun. 2018, pp. 344–349.

[18] S. Altaf, A. Al-Anbuky, and H. GholamHosseini, "Fault diagnosis in a distributed motor network using artificial neural network," in *Proc. Int. Symp. Power Electron., Electr. Drives, Autom. Motion*, Ischia, Italy, Jun. 2014, pp. 190–197.

[19] B. M. Sanandaji, E. Bitar, K. Poolla, and T. L. Vincent, "An abrupt change detection heuristic with applications to cyber data attacks on power systems," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 5056–5061.

[20] M. Russo, M. Labonne, A. Olivereau, and M. Rmayti, "Anomaly detection in Vehicle-to-Infrastructure communications," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–6.

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[22] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for EMG-based human-machine interaction: A review," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021.

[23] L. Kuwatly, M. Sraj, Z. Al Masri, and H. Artail, "A dynamic honeypot design for intrusion detection," in *Proc. IEEE/ACS Int. Conf. Pervasive Services (ICPS)*, Beirut, Lebanon, Jul. 2004, pp. 95–104.

[24] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–29, Apr. 2014.

[25] C. S. Wickramasinghe, D. L. Marino, K. Amarasinghe, and M. Manic, "Generalization of deep learning for cyber-physical system security: A survey," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 745–751.

[26] S. Gaba, H. Khan, K. J. Almalki, A. Jabbari, I. Budhiraja, V. Kumar, A. Singh, K. K. Singh, S. S. Askar, and M. Abouhawwash, "Holochain: An agent-centric distributed hash table security in smart IoT applications," *IEEE Access*, vol. 11, pp. 81205–81223, 2023.

[27] A. Barnawi, S. Gaba, A. Alphy, A. Jabbari, I. Budhiraja, V. Kumar, and N. Kumar, "A systematic analysis of deep learning methods and potential attacks in Internet-of-things surfaces," *Neural Comput. Appl.*, vol. 35, no. 25, pp. 18293–18308, Sep. 2023.

[28] H. Sharma, N. Kumar, I. Budhiraja, and A. Barnawi, "Secrecy rate maximization in THz-aided heterogeneous networks: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13490–13505, Oct. 2023.

[29] V. Vishnoi, P. Consul, I. Budhiraja, S. Gupta, and N. Kumar, "Deep reinforcement learning based energy consumption minimization for intelligent reflecting surfaces assisted D2D users underlaying UAV network," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, May 2023, pp. 1–6.

[30] V. Vishnoi, I. Budhiraja, S. Ishan, and N. Kumar, "A deep reinforcement learning scheme for sum rate and fairness maximization among D2D pairs underlaying cellular network with NOMA," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13506–13522, 2023, doi: 10.1109/TVT.2023.3276647.

[31] S. Singh, A. Bhardwaj, I. Budhiraja, U. Gupta, and I. Gupta, "Cloud-based architecture for effective surveillance and diagnosis of COVID-19," in *Convergence of Cloud With AI for Big Data Analytics: Foundations and Innovation*. Hoboken, NJ, USA: Wiley, 2023, pp. 69–88.

[32] L. Cheng, K. Tian, and D. Yao, "Orpheus: Enforcing cyber-physical execution semantics to defend against data-oriented attacks," in *Proc. 33rd Annu. Comput. Secur. Appl. Conf.*, Orlando, FL, USA, Dec. 2017, pp. 315–326.

[33] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv:1901.03407*.

[34] R. Heartfield, G. Loukas, S. Budimir, A. Bezemskij, J. R. J. Fontaine, A. Filippoupolitis, and E. Roesch, "A taxonomy of cyber-physical threats and impact in the smart home," *Comput. Secur.*, vol. 78, pp. 398–428, Sep. 2018.

[35] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[36] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, Jul. 2019.

[37] P. Schneider and K. Böttinger, "High-performance unsupervised anomaly detection for cyber-physical system networks," in *Proc. Workshop Cyber-Phys. Syst. Secur. PrivaCy*, Jan. 2018, pp. 1–12.

[38] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2923–2960, 4th Quart., 2018.

[39] E. M. Veith, L. Fischer, M. Tröschel, and A. Nieße, "Analyzing cyber-physical systems from the perspective of artificial intelligence," in *Proc. Int. Conf. Artif. Intell., Robot. Control*, Dec. 2019, pp. 85–95.

[40] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–36, Jun. 2022.

[41] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Comput. Surveys*, vol. 46, no. 4, pp. 1–29, Apr. 2014.

[42] S. Nazir, S. Patel, and D. Patel, "Assessing and augmenting SCADA cyber security: A survey of techniques," *Comput. Secur.*, vol. 70, pp. 436–454, Sep. 2017.

[43] B. Siciliano, A. G. Scaglione, and L. Galluccio, "A survey of cyber-physical attacks and defenses in the smart grid," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3231–3240, Jun. 2020.

[44] I. Ndiaye, J. G. Nijlla, and I. Balogun, "A survey of intrusion detection in Internet of Things," *IEEE Access*, vol. 9, pp. 73900–73917, 2021.

[45] L. Yu, H. Wu, Z. Liu, Y. Li, and W. Zhao, "A review of machine learning methods in cybersecurity," *IEEE Access*, vol. 8, pp. 135695–135718, 2020.
[46] J. Zhang, L. Pan, Q.-L. Han, C. Chen, S. Wen, and Y. Xiang, "Deep learning based attack detection for cyber-physical system cybersecurity: A survey," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 377–391, Mar. 2022.
[47] A. A. Jamal, A.-A. M. Majid, A. Konev, T. Kosachenko, and A. Shelupanov, "A review on security analysis of cyber physical systems using machine learning," *Mater. Today*, vol. 80, pp. 2302–2306, Jan. 2023.
[48] F. O. Olowononi, D. B. Rawat, and C. Liu, "Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for CPS," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 524–552, 1st Quart., 2021.
[49] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–36, Jun. 2022.
[50] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5615–5624, Aug. 2021.
[51] A. Presekal, A. Ştefanov, V. S. Rajkumar, and P. Palensky, "Attack graph model for cyber-physical power systems using hybrid deep learning," *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 4007–4020, 2023, doi: 10.1109/TSG.2023.3237011.
[52] F. Darbandi, A. Jafari, H. Karimipour, A. Dehghantanha, F. Derakhshan, and K. Raymond Choo, "Real-time stability assessment in smart cyber-physical grids: A deep learning approach," *IET Smart Grid*, vol. 3, no. 4, pp. 454–461, Aug. 2020.
[53] S. Rathore and J. H. Park, "A blockchain-based deep learning approach for cyber security in next generation industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5522–5532, Aug. 2021.
[54] B. Hussain, Q. Du, B. Sun, and Z. Han, "Deep learning-based DDoS-attack detection for cyber–physical system over 5G network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 860–870, Feb. 2021.
[55] A. M. Aslam, R. Chaudhary, A. Bhardwaj, I. Budhiraja, N. Kumar, and S. Zeadally, "Metaverse for 6G and beyond: The next revolution and deployment challenges," *IEEE Internet Things Mag.*, vol. 6, no. 1, pp. 32–39, Mar. 2023.
[56] M. Gupta, B. Gupta, A. Jabbari, I. Budhiraja, D. Garg, K. Kotecha, and C. Iwendi, "A novel computer assisted genomic test method to detect breast cancer in reduced cost and time using ensemble technique," *Human-Centric Comput. Inf. Sci.*, vol. 13, no. 18, pp. 1–16, Feb. 2023, doi: 10.22967/HCIS.2023.13.008.
[57] A. Bhardwaj, I. Budhiraja, and U. Gupta, *Cloud-Based Architecture for Effective Surveillance and Diagnosis of COVID-19*. Hoboken, NJ, USA: Wiley, 2023.
[58] A. Bhardwaj, U. Gupta, I. Budhiraja, and R. Chaudhary, "Container-based migration technique for fog computing architecture," in *Proc. Int. Conf. Adv. Technol. (ICONAT)*, Jan. 2023, pp. 1–6.
[59] M. Kravchik and A. Shabtai, "Detecting cyber attacks in industrial control systems using convolutional neural networks," in *Proc. Workshop Cyber-Phys. Syst. Secur. PrivaCy*, Jan. 2018, pp. 72–83.
[60] Z. Zohrevand, U. Glässer, M. A. Tayebi, H. Y. Shahir, M. Shirmaleki, and A. Y. Shahir, "Deep learning based forecasting of critical infrastructure data," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Singapore, Nov. 2017, pp. 1129–1138.
[61] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Jul. 2019, pp. 2828–2837.
[62] B. Eiteneuer, N. Hranisavljevic, and O. Niggemann, "Dimensionality reduction and anomaly detection for CPPS data using autoencoder," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Melbourne, VIC, Australia, Feb. 2019, pp. 1286–1292.
[63] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *Proc. IEEE 18th Int. Symp. High Assurance Syst. Eng. (HASE)*, Singapore, Jan. 2017, pp. 140–145.
[64] C. Feng, T. Li, and D. Chana, "Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks," in *Proc. 47th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, Singapore, Jun. 2017, pp. 261–272.
[65] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, New Orleans, LA, USA, Nov. 2017, pp. 1058–1065.
[66] P. Ferrari, S. Rinaldi, E. Sisinni, F. Colombo, F. Ghelfi, D. Maffei, and M. Malara, "Performance evaluation of full-cloud and edge-cloud architectures for industrial IoT anomaly detection based on deep learning," in *Proc. II Workshop Metrol. Ind. IoT (MetroInd&IoT)*, Jun. 2019, pp. 420–425.
[67] A. Legrand, B. Niepceron, A. Cournier, and H. Trannois, "Study of autoencoder neural networks for anomaly detection in connected buildings," in *Proc. IEEE Global Conf. Internet Things (GCIoT)*, Naples, Italy, Dec. 2018, pp. 1–5.
[68] Z. Wu, Y. Guo, W. Lin, S. Yu, and Y. Ji, "A weighted deep representation learning model for imbalanced fault diagnosis in cyber-physical systems," *Sensors*, vol. 18, no. 4, p. 1096, Apr. 2018.
[69] Z. Li, J. Li, Y. Wang, and K. Wang, "A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment," *Int. J. Adv. Manuf. Technol.*, vol. 103, nos. 1–4, pp. 499–510, Jul. 2019.
[70] B. Lindemann, F. Fesenmayr, N. Jazdi, and M. Weyrich, "Anomaly detection in discrete manufacturing using self-learning approaches," *Proc. CIRP*, vol. 79, pp. 313–318, Jan. 2019.
[71] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, Oct. 2019.
[72] H. A. Khan, N. Sehatbakhsh, L. N. Nguyen, M. Prvulovic, and A. Zajić, "Malware detection in embedded systems using neural network model for electromagnetic side-channel signals," *J. Hardw. Syst. Secur.*, vol. 3, no. 4, pp. 305–318, Dec. 2019.
[73] Y.-J. Xiao, W.-Y. Xu, Z.-H. Jia, Z.-R. Ma, and D.-L. Qi, "NIPAD: A non-invasive power-based anomaly detection scheme for programmable logic controllers," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 4, pp. 519–534, Apr. 2017.
[74] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "Mad-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," in *Proc. Int. Conf. Artif. Neural Netw.*, Springer, 2019, pp. 703–716.
[75] N. L. Tasfi, W. A. Higashino, K. Grolinger, and M. A. M. Capretz, "Deep neural networks with confidence sampling for electrical anomaly detection," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jun. 2017, pp. 1038–1045.
[76] Y. Zhang, V. V. G. Krishnan, J. Pi, K. Kaur, A. Srivastava, A. Hahn, and S. Suresh, "Cyber physical security analytics for transactive energy systems," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 931–941, Mar. 2020.
[77] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, and X. Duan, "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4401–4410, Jul. 2019.
[78] Q. Deng and J. Sun, "False data injection attack detection in a power grid using RNN," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Washington, DC, USA, Oct. 2018, pp. 5983–5988.
[79] X. Niu, J. Li, J. Sun, and K. Tomsovic, "Dynamic detection of false data injection attack in smart grid using deep learning," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, Feb. 2019, pp. 1–6.
[80] H. Wang, J. Ruan, Z. Ma, B. Zhou, X. Fu, and G. Cao, "Deep learning aided interval state prediction for improving cyber security in energy Internet," *Energy*, vol. 174, pp. 1292–1304, May 2019.
[81] S. Basumallik, R. Ma, and S. Eftekharnejad, "Packet-data anomaly detection in PMU-based state estimator using convolutional neural network," *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 690–702, May 2019.
[82] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Appl. Energy*, vol. 211, pp. 1123–1135, Feb. 2018.
[83] Y. Wang, D. Chen, C. Zhang, X. Chen, B. Huang, and X. Cheng, "Wide and recurrent neural networks for detection of false data injection in smart grids," in *Proc. 14th Int. Conf., Wireless Algorithms, Syst., Appl. (WASA)*. Honolulu, HI, USA: Springer, 2019, pp. 335–345.

[84] E. Khanapuri, T. Chintalapati, R. Sharma, and R. Gerdes, "Learning-based adversarial agent detection and identification in cyber physical systems applied to autonomous vehicular platoon," in *Proc. IEEE/ACM 5th Int. Workshop Softw. Eng. Smart Cyber-Phys. Syst. (SEsCPS)*, Montreal, QC, Canada, May 2019, pp. 39–45.

[85] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1264–1276, Mar. 2020.

[86] A. Taylor, S. Leblanc, and N. Japkowicz, "Anomaly detection in automobile control network data with long short-term memory networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Montreal, QC, Canada, Oct. 2016, pp. 130–139.

[87] T. Kieu, B. Yang, and C. S. Jensen, "Outlier detection for multidimensional time series using deep neural networks," in *Proc. 19th IEEE Int. Conf. Mobile Data Manage. (MDM)*, Aalborg, Denmark, Jun. 2018, pp. 125–134.

[88] K. Zhu, Z. Chen, Y. Peng, and L. Zhang, "Mobile edge assisted literal multi-dimensional anomaly detection of in-vehicle network using LSTM," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4275–4284, May 2019.

[89] C. Jichici, B. Groza, and P.-S. Murvay, "Examining the use of neural networks for intrusion detection in controller area networks," in *Proc. Int. Conf. Secur. Inf. Technol. Commun.*, Springer, 2018, pp. 109–125.

[90] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, London, U.K., Jul. 2018, pp. 387–395.

[91] S. Tariq, S. Lee, Y. Shin, M. S. Lee, O. Jung, D. Chung, and S. S. Woo, "Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Anchorage, AK, USA, Jul. 2019, pp. 2123–2133.

[92] O. M. Ezeme, Q. H. Mahmoud, and A. Azim, "DReAM: Deep recursive attentive model for anomaly detection in kernel events," *IEEE Access*, vol. 7, pp. 18860–18870, 2019.

[93] L. Gunn, P. Smet, E. Arbon, and M. D. McDonnell, "Anomaly detection in satellite communications systems using LSTM networks," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, Canberra, ACT, Australia, Nov. 2018, pp. 1–6.

[94] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using recurrent neural networks (RNN)," in *Proc. Integr. Commun. Navigat. Surveill. (ICNS)*, Herndon, VA, USA, Apr. 2016, p. 5C2-1.

[95] E. Habler and A. Shabtai, "Using LSTM encoder-decoder algorithm for detecting anomalous ADS-B messages," *Comput. Secur.*, vol. 78, pp. 155–173, Sep. 2018.

[96] O. M. Ezeme, M. Lescisin, Q. H. Mahmoud, and A. Azim, "DeepAnom: An ensemble deep framework for anomaly detection in system processes," in *Proc. 32nd Can. Conf. Artif. Intell., Adv. Artif. Intell. (Canadian AI)*. Kingston, ON, Canada: Springer, May 2019, pp. 549–555.

[97] V. L. Do, L. Fillatre, I. Nikiforov, and P. Willett, "Security of SCADA systems against cyber–physical attacks," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 5, pp. 28–45, May 2017.

[98] K. Poulsen, "Slammer worm crashed Ohio nuke plant network," *Secur. Focus*, vol. 19, 2003.

[99] J. D. Rogers and C. M. Watkins, "Overview of the Taum Sauk pumped storage power plant upper reservoir failure, Reynolds County, MO," in *Proc. 6th Int. Conf. Case Histories Geotechnical Eng.*, Arlington, VA, USA, 2008, pp. 1–13.

[100] T. FoxBrewster, "Ukraine claims hackers caused Christmas power outage," *Forbes Secur.*, 2016.

[101] S. Gorman, "Electricity grid in us penetrated by spies," *Wall Street J.*, vol. 8, no. 8, 2009.

[102] M. J. Credeur, "FBI probes Georgia water plant break-in on terror concern," Bloomberg, 2013.

[103] J. Slay and M. Miller, *Lessons Learned From the Maroochy Water Breach*. Boston, MA, USA: Springer, 2008.

[104] F. Y. Rashid. (2012). *Telvent Hit by Sophisticated Cyber-Attack, SCADA Admin Tool Compromised*. [Online]. Available: http://www.securityweek.com/telvent-hit-sophisticated-cyber-attack-scada-admin-tool-compromised

[105] M. A. Almaiah, F. Hajjej, A. Ali, M. F. Pasha, and O. Almomani, "A novel hybrid trustworthy decentralized authentication and data preservation model for digital healthcare IoT based CPS," *Sensors*, vol. 22, no. 4, p. 1448, Feb. 2022.

[106] R. F. Mansour, "Artificial intelligence based optimization with deep learning model for blockchain enabled intrusion detection in CPS environment," *Sci. Rep.*, vol. 12, no. 1, p. 12937, Jul. 2022.

[107] Z. A. Sheikh, Y. Singh, P. K. Singh, and K. Z. Ghafoor, "Intelligent and secure framework for critical infrastructure (CPS): Current trends, challenges, and future scope," *Comput. Commun.*, vol. 193, pp. 302–331, Sep. 2022.

[108] D. M. Sharma and S. K. Shandilya, "Attack detection based on machine learning techniques to safe and secure for CPS—A review," in *Proc. Int. Conf. IoT, Intell. Comput. Secur., Select (IICS)*, pp. 273–286, Springer, 2023.

[109] A. Albasir, K. Naik, and R. Manzano, "Toward improving the security of IoT and CPS devices: An AI approach," *Digit. Threats: Res. Pract.*, vol. 4, no. 2, pp. 1–30, Jun. 2023.

[110] G. Epiphaniou, M. Hammoudeh, H. Yuan, C. Maple, and U. Ani, "Digital twins in cyber effects modelling of IoT/CPS points of low resilience," *Simul. Model. Pract. Theory*, vol. 125, May 2023, Art. no. 102744.

[111] A. Albasir, K. Naik, and R. Manzano, "Toward improving the security of IoT and CPS devices: An AI approach," *Digit. Threats, Res. Pract.*, vol. 4, no. 2, pp. 1–30, Jun. 2023.

[112] A. Aggarwal, S. Gaba, S. Nagpal, and A. Arya, "A comparative analysis among task scheduling for grouped and ungrouped grid application," in *Proc. CEUR Workshop, Int. Conf. Emerg. Technol., AI, IoT, CPS Sci. Technol. Appl.* Chandigarh, India: NITTTR, Sep. 2021, pp. 1–5.

[113] A. Aggarwal, S. Gaba, S. Nagpal, and B. Vig, "Bio-inspired routing in VANET," in *Cloud and IoT-Based Vehicular Ad Hoc Networks*, 2021, pp. 199–220.

[114] D. Aggarwal and S. Gaba, "A comparative study: Reviewing performance of routing protocols in mobile ad-hoc network," *Vol*, vol. 4, no. 8, pp. 528–532, Jun. 2018.

[115] I. Budhiraja et al., "A comprehensive review on variants of SARS-CoVs-2: Challenges, solutions and open issues," *Comput. Commun.*, vol. 197, pp. 34–51, 2023.

[116] H. Khan, I. Budhiraja, S. A. Wahaj, M. Z. Alam, S. T. Siddiqui, and M. I. Alam, "IoT and blockchain integration challenges," in *Proc. IEEE Int. Conf. Current Develop. Eng. Technol. (CCET)*, Dec. 2022, pp. 1–5.

[117] P. Rani, V. Kumar, I. Budhiraja, A. Rathi, and S. Kukreja, "Deploying electronic voting system use-case on Ethereum public blockchain," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2022, pp. 1–6.

[118] P. Consul, I. Budhiraja, R. Chaudhary, and D. Garg, "FLBCPS: Federated learning based secured computation offloading in blockchain-assisted cyber-physical systems," in *Proc. IEEE/ACM 15th Int. Conf. Utility Cloud Comput. (UCC)*, Dec. 2022, pp. 412–417.

[119] R. Nijhawan, M. Juneja, N. Kaur, A. Yadav, and I. Budhiraja, "Automated deep learning based approach for albinism detection," in *Proc. Int. Conf. Recent Trends Image Process. Pattern Recognit.* Kingsvilla, TX, USA: Springer, 2022, pp. 272–281.

[120] A. Aggarwal, S. Gaba, S. Chawla, and A. Arya, "Recognition of alphanumeric patterns using backpropagation algorithm for design and implementation with ANN," *Int. J. Secur. Privacy Pervasive Comput.*, vol. 14, no. 1, pp. 1–11, Feb. 2022.

[121] S. Gaba, A. Aggarwal, and S. Nagpal, "Role of machine learning for ad hoc networks," in *Cloud and IoT-Based Vehicular Ad Hoc Networks*. Hoboken, NJ, USA: Wiley, 2021, pp. 269–291.

[122] A. Aggarwal, S. Gaba, J. Kumar, and S. Nagpal, "Blockchain and autonomous vehicles: Architecture, security and challenges," in *Proc. 5th Int. Conf. Comput. Intell. Commun. Technol. (CCICT)*, Jul. 2022, pp. 332–338.

[123] S. Gaba, S. Nagpal, and A. Aggarwal, "A comparative study of convolutional neural networks for plant phenology recognition," in *Advanced Sensing in Image Processing and IoT*. Boca Raton, FL, USA: CRC Press, 2022, pp. 109–136.

[124] A. Barnawi, I. Budhiraja, K. Kumar, N. Kumar, B. Alzahrani, A. Almansour, and A. Noor, "A comprehensive review on landmine detection using deep learning techniques in 5G environment: Open issues and challenges," *Neural Comput. Appl.*, vol. 34, no. 24, pp. 21657–21676, Dec. 2022.

[125] S. Gaba, D. Kumar, S. Nagpal, and A. Aggarwal, "A quick analysis on cyber physical systems for sustainable development," *Grenze Int. J. Eng. Technol.*, vol. 8, no. 1, pp. 621–627, 2022.

[126] P. Consul, I. Budhiraja, R. Chaudhary, and N. Kumar, "Security reassessing in UAV-assisted cyber-physical systems based on federated learning," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2022, pp. 61–65.
[127] A. Verma, P. Bhattacharya, I. Budhiraja, A. K. Gupta, and S. Tanwar, "Fusion of federated learning and 6G in internet-of-medical-things: Architecture, case study and emerging directions," in *Proc. 4th Int. Conf. Futuristic Trends Netw. Comput. Technol.* Ahmedabad, India: Springer, Jul. 2022, pp. 229–242.
[128] P. Arpaia, C. Manna, and G. Montenero, "Ant-search strategy based on likelihood trail intensity modification for multiple-fault diagnosis in sensor networks," *IEEE Sensors J.*, vol. 13, no. 1, pp. 148–158, Jan. 2013.
[129] I. Budhiraja, N. Kumar, H. Sharma, M. Elhoseny, Y. Lakys, and J. J. P. C. Rodrigues, "Latency-energy tradeoff in connected autonomous vehicles: A deep reinforcement learning scheme," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 13296–13308, 2023, doi: 10.1109/TITS.2022.3215523.
[130] A. Barnawi, N. Kumar, I. Budhiraja, K. Kumar, A. Almansour, and B. Alzahrani, "Deep reinforcement learning based trajectory optimization for magnetometer-mounted UAV to landmine detection," *Comput. Commun.*, vol. 195, pp. 441–450, Nov. 2022.
[131] Deepanshi, I. Budhiraja, D. Garg, N. Kumar, and R. Sharma, "A comprehensive review on variants of SARS-CoVs-2: Challenges, solutions and open issues," *Comput. Commun.*, vol. 197, pp. 34–51, Jan. 2023.
[132] S. Gaba, S. Nagpal, A. Aggarwal, S. Kumar, and P. Singh, "A modified approach for accuracy enhancement in intruder detection with optimally certain features," in *Proc. 3rd Mobile Radio Commun. 5G Netw. (MRCN)*. Kurukshetra, India: Springer, 2023, pp. 149–157.
[133] S. Gaba, I. Buddhiraja, V. Kumar, and A. Makkar, "Federated learning based secured computational offloading in cyber-physical IoST systems," in *Proc. Int. Conf. Recent Trends Image Process. Pattern Recognit.* Kingsvilla, TX, USA: Springer, 2022, pp. 344–355.
[134] S. Gaba, S. Nagpal, A. Aggarwal, R. Kumar, and S. Kumar, "An analysis of Internet of Things (IoT) malwares and detection based on static and dynamic techniques," in *Proc. 7th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Nov. 2022, pp. 24–29.
[135] P. Singh, G. Bathla, D. Panwar, A. Aggarwal, and S. Gaba, "Performance evaluation of genetic algorithm and flower pollination algorithm for scheduling tasks in cloud computing," in *Proc. Int. Conf. Signal Process. Integr. Netw.* Noida, India: Springer, 2022, pp. 139–154.
[136] H. Sharma, I. Budhiraja, P. Consul, N. Kumar, D. Garg, L. Zhao, and L. Liu, "Federated learning based energy efficient scheme for MEC with NOMA underlaying UAV," in *Proc. 5th Int. ACM Mobicom Workshop Drone Assist. Wireless Commun. 5G Beyond*, Oct. 2022, pp. 73–78.
[137] P. Consul, I. Budhiraja, D. Garg, and A. Bindle, "Power allocation scheme based on DRL for CF massive MIMO network with UAV," in *Proc. Innov. Inf. Commun. Technol. (ICIICT)*. Thailand: Springer, 2022, pp. 33–43.

SHIVANI GABA received the B.Tech. and M.Tech. degrees from Kurukshetra University, in 2015 and 2017, respectively. She is currently an Educator, a Researcher, and a Philanthropist. She is also a Microsoft Technology Associate (MTA) and a Microsoft Office Specialist (MOS) Certified. She is also a Research Scholar with the School of Computer Science and Engineering, Bennett University, Greater Noida. She has presented and published abundant papers and chapters in national/international conferences and journals. Her research interests include AI, blockchain, deep learning, and cyber-attacks.

ISHAN BUDHIRAJA (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Uttar Pradesh Technical University, Lucknow, India, in 2008, the M.Tech. degree in electronics and communication engineering from Maharishi Dayanand University, Rohtak, Haryana, in 2012, and the Ph.D. degree in computer science engineering from the Thapar Institute of Engineering & Technology, Patiala, India, in 2021. He was a Research Associate on the Project Energy Management of Smart Home Using Cloud Infrastructure-A Utility Perspective, funded by CSIR, New Delhi, India. Some of his research findings are published in top-cited journals, such as IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE INTERNET OF THINGS JOURNAL, *IEEE Wireless Communication Magazine*, and IEEE SYSTEMS JOURNAL, and various international top-tiered conferences, such as IEEE GLOBECOM, IEEE ICC, IEEE WCMC, ACM, and IEEE Infocom. His research interests include device-to-device communications, the Internet of Things, non-orthogonal multiple access, femtocells, deep reinforcement learning, and microstrip patch antenna.

VIMAL KUMAR (Member, IEEE) received the M.Tech. and Ph.D. degrees from MNNIT Allahabad, Prayagraj, Uttar Pradesh, India. He is currently an Assistant Professor in SCSET with Bennett University (Times of India Group), Greater Noida. He has more than 17 years of teaching and research experience. His past research work is more focused on multipath mobile computing in heterogeneous networks for multi-interface enabled smart mobile devices to improve the QoE of end users. He has published 17 research papers in various reputed SCI/Scopus/WoS/ESCI indexed journals and conferences and his papers are awarded best papers in conferences. He is currently passionate about innovations in blockchain use cases and use of IoT devices in various domains. He is also a member of Internet Society.

SHESHIKALA MARTHA (Member, IEEE) received the Ph.D. degree from K. L. University. She has been a Professor and the Head of SR University, since 2012, and having total experience of more than 18 years. She has published more than 50 publications in reputed research journals. Her research interests include data mining, machine learning, deep learning, and cyber-physical systems.

JEBREEL KHURMI received the B.S. degree in Computer Engineering and Networking from Jazan University, Jazan, Kingdom of Saudi Arabia, the M.S. degree in Computer Science Networks and Telecommunications from University of Missouri–Kansas City, MO, USA. Currently, he is a Lecturer with Jazan College of Technology, Saudi Arabia. His research interests are the IoT, Smart Applications, and Sensors Enhancements.

AKANSHA SINGH (Member, IEEE) received the B.Tech. and M.Tech. degrees in computer science and the Ph.D. degree in image processing and machine learning from IIT Roorkee. She is also a Professor with the School of Computer Science and Engineering, Bennett University, Greater Noida, India. She has also undertaken government funded project as a principal investigator. Her research interests include image processing, remote sensing, the IoT, and machine learning. She has served as an associate editor and a guest editor for several journals.

KRISHNA KANT SINGH received the B.Tech., M.Tech., M.S., and Ph.D. degrees in image processing and machine learning from IIT Roorkee. He is currently working as the Director with Delhi Technical Campus, Greater Noida, UP, India. He has wide teaching and research experience. He has authored more than 116 research articles in Scopus and SCIE indexed journals of repute. He has also authored 25 technical books. He is an Associate Editor of *Journal of Intelligent and Fuzzy Systems* (SCIE Indexed) and IEEE Access (SCIE Indexed) and a Guest Editor of *Open Computer Science* and *Wireless Personal Communications*. He is serving as a member of Editorial Board for *Applied Computing and Geoscience* (Elsevier).

S. S. ASKAR received the B.Sc. degree in mathematics and the M.Sc. degree in applied mathematics from Mansoura University, Egypt, in 1998 and 2004, respectively, and the Ph.D. degree in operation research from Cranfield University, U.K., in 2011. He has been an Associate Professor with Mansoura University, since 2016. He has joined King Saud University, in 2012, where he is currently a Professor with the Department of Statistics and Operation Research. His main research interests include game theory and its applications that include mathematical economy, dynamical systems, and network analysis.

MOHAMED ABOUHAWWASH received the B.Sc. and M.Sc. degrees in statistics and computer science from Mansoura University, Mansoura, Egypt, in 2005 and 2011, respectively, and the joint Ph.D. degree in statistics and computer science from Michigan State University, East Lansing, MI, USA, and Mansoura University, Egypt, in 2015. Currently, he holds significant academic positions at Distinguished Institutions, including Computational Mathematics, Science, and Engineering (CMSE), Biomedical Engineering (BME), and Radiology, Institute for Quantitative Health Science and Engineering (IQ), Michigan State University. Additionally, he serves as an Associate Professor at the Department of Mathematics, Faculty of Science, Mansoura University. During 2018, he dedicated to advancing knowledge transcends geographical boundaries, as evidenced by his role as a Visiting Scholar at the Department of Mathematics and Statistics, Faculty of Science, Thompson Rivers University, Kamloops, BC, Canada. He is a Distinguished Researcher and an Academician, widely recognized for his outstanding contributions to the fields of computational intelligence, machine learning, and image reconstruction. With an illustrious career, he has published over 160 papers in esteemed journals, including notable publications like IEEE Transactions on Evolutionary Computation, IEEE Transactions on Medical Imaging, IEEE Transactions on Emerging Topics in Computational Intelligence, Artificial Intelligence Review, Expert Systems with Applications, Swarm and Evolutionary Computation, Knowledge-Based Systems, and Applied Soft Computing. In addition to his prolific research output, he has showcased his expertise by authoring several edited books published by reputable academic publishers such as *Springer, Wiley, Taylor, and Francis*. His impact on the academic community is further amplified through his editorial board service in numerous prestigious journals and conferences. Throughout his illustrious career, he has received recognition for his academic excellence, notably being honoured with the best master's and Ph.D. Thesis Awards from Mansoura University in 2012 and 2018, respectively.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 July 2023, accepted 28 July 2023, date of publication 7 August 2023, date of current version 16 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3303205

RESEARCH ARTICLE

A Comparative Analysis of Industrial Cybersecurity Standards

FATIHA DJEBBAR[1], (Member, IEEE), AND KIM NORDSTRÖM[2]
[1]Department of Engineering Science, Högskolan Väst, 46153 Trollhättan, Sweden
[2]Cybersecurity Product Compliance Group, 10392 Stockholm, Sweden

Corresponding author: Fatiha Djebbar (fatiha.djebbar@hv.se)

ABSTRACT Cybersecurity standards provide a structured approach to manage and assess cybersecurity risks. They are the primary source for security requirements and controls used by organizations to reduce the likelihood and the impact of cybersecurity attacks. However, the large number of available cybersecurity standards and frameworks make the selection of the right security standards for a specific system challenging. The absence of a comprehensive comparison overlap across these standards further increases the difficulty of the selection process. In situations where new business needs dictate to comply or implement additional security standard, there may be a risk of duplicating existing security requirements and controls between the standards resulting in unnecessary added cost and workload. To optimize the performance and cost benefits of compliance efforts to standards, it is important to analyze cybersecurity standards and identify the overlapping security controls and requirements. In this work, we conduct a comparative study to identify possible overlaps and discrepancies between three security standards: ETSI EN 303 645 v2.1.1 for consumer devices connected to the internet, ISA/IEC 62443-3-3:2019 for industrial automation and control systems, and ISO/IEC 27001:2022 for information security management systems. The standards were carefully chosen for their broad adoption and acceptance by the international community. We intentionally selected standards with different areas of focus to illustrate the significant overlaps that can exist despite being designed for different environments. Our objective is to help organizations select the most suitable security controls for their specific needs and to simplify and clarify the compliance process. Our findings show a significant overlap among the three selected standards. This information can help organizations gain a comprehensive understanding of common security requirements and controls, enabling them to streamline their compliance efforts by eliminating duplicated work especially when meeting the requirements of multiple standards.

INDEX TERMS Cybersecurity, security controls, security standards, cybersecurity concepts, threats, security requirements.

I. INTRODUCTION

Embracing emerging technologies have resulted in remarquable added capabilities, values and experiences. However, these new technologies have been consistent target of diverse threat actors, each driven by different motivations and capabilities [1]. To fully benefit from the competitive advantage of these technologies, cybersecurity is currently a top priority and a major theme in industrial sectors and consumers market. Statistics showed that in 93% of cases, an external attacker can breach an organization's network perimeter and gain access to local network resources [2]. Cybersecurity standards and frameworks provide guidelines and best practices for organizations to follow to enhance their overall security posture. Implementing cybersecurity frameworks also helps businesses to comply with relevant regulations and laws [3]. The chair of multiple committees in the recognized European Telecommunications Standards Institute (ETSI), affirms that "Cybersecurity standards are critical to the collective effort to prevent attacks in the first place and reduce the

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero.

effectiveness of successful incursions'' [4]. Therefore, various standard organizations have taken a proactive approach to develop, best practices, guidelines, and other resources to assist organizations in securing their data and systems. This has led to broad collaboration on the creation and implementation of cybersecurity standards among organizations such as: the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), the International Society of Automation (ISA), ETSI, the International Telecommunication Union - Telecommunication (ITU-T), European Union Agency for Network and Information Security (ENISA). Furthermore, there have been recent updates and releases of several regulations. The EU Cybersecurity Act (CSA) was enacted on April 17, 2019 (Regulation (EU) 2019/881) [5] to strengthen the mandate of the EU cybersecurity. This act granted ENISA a permanent mandate to address cybersecurity threats and establish an EU-wide cybersecurity certification framework to enhance the security of connected products, Internet of Things (IoT) devices as well as critical infrastructure through such certificates. This framework incorporates security features in the initial stages of their technical design and development. The EU Network and Information Security (NIS) directive was adopted in 2016 (EU 2016/1148) [6] and was the first piece of EU-wide cybersecurity legislation. The updated NIS 2 Directive [7] include improved cybersecurity risk management and new reporting obligations across sectors such as digital infrastructure. The scope of the Radio Equipment Directive (RED) 2014/53/EU [8] has been updated in February 2022 to include cybersecurity requirements for radio products which will become mandatory in August 2024 through a Delegated Act on Internet-connected radio equipment. The General Data Protection Regulation (GDPR) [9] was entered into force in May 2018 and established security requirements for data protection to safeguard EU citizens. Other regulation proposals, such as the Artificial Intelligence Act, the Data Act, and the Cybersecurity Resilience Act, aim to address risks and establish rules regarding the use of data generated by connected products, protecting consumers and businesses who use digital components in products or software. Various industrial sectors, such as road vehicles, industrial automation and control systems, information security management systems, and consumer devices connected to the Internet, have shown significant activity in developing standards that specifically address their specific security needs. Notable examples include cybersecurity standards like ISO/SAE 21434 [10], ETSI EN 303 645 [11], ISA/IEC 62443 [12], and ISO/IEC 27001 [13]. These standards and regulations promote the development and implementation of security requirements to ensure the protection of organizations, critical infrastructures, and consumers' products.

Disconcerted by the substantial number of cybersecurity standards, this study aims at identifying and reviewing commonly adopted cybersecurity standards. The goal is to understand their security control objectives to uncover overlapping requirements, and contradictions. The results of this study can assist organizations, cybersecurity professionals, academics, and researchers in understanding the current state of the art and in selecting the best standards for their needs, balancing performance and cost-effectiveness. Furthermore, the objective of this study is to identify any existing gaps within the selected standards and address challenges arising from overlapping requirements and controls, irrespective of their specific application context. As a contribution, this paper aims to fulfil the following objectives:

1) To conduct a comprehensive review of commonly adopted cybersecurity standards, and present a literature review on the current state of the art.
2) To identify prevalent domain-specific cybersecurity standards that form a strong basis to mitigate cybersecurity threats.
3) To identify the overlap and gaps in security requirements and controls between the studied standards with the aim of avoiding redundant efforts when complying with multiple standards.
4) To identify and discuss the challenges related to the creation and compliance to multiple security standards.

As for the remaining part of the paper, section II presents an overview and motivation for this study while the background and existing research work are presented in section III. Section IV provides a formal classification for security standards and section V explains the research methodology used in this study followed by section VI which presents reviews on the analyzed standards and the mapping outcomes. Section VII discusses the findings, while section VIII highlights the challenges associated to the implementation of these standards. Finally, section IX concludes the paper and proposes future research work.

II. OVERVIEWS AND MOTIVATION

The rapid pace adoption of digital technology is leading to the creation of new business models and market opportunities. As the volume of interconnected products and services rises, the importance of cybersecurity also grows in tandem with the expanding digitization and connectivity [9], [14], [15], [16]. To effectively combat the growing risk of cybercrime, it is essential to integrate systematic and well-structured cybersecurity measures into a comprehensive strategy that encompasses individuals, processes, and technology. This entails, in part, adopting appropriate standards and frameworks to ensure a robust defense against cyber threats. ISO defines a standard as ''a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines, or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context'' [18]. Standards have special significance in the domain of cybersecurity addressing confidentiality, integrity, and availability of data [19]. They are collections of best practices created by experts to protect organizations from cyber threats and help improve their cybersecurity posture by protecting their most valuable assets at an effective spending. These

best practices emphasize the importance of implementing a comprehensive security program that includes a range of controls to protect organizational assets. These security controls are generally organized into five categories: Identify-Detect-Respond-Protect-Recover (IDRPR) [20], [21]. By organizing security controls into these categories, organizations can better understand the specific areas they need to focus on to build a robust security program. The approach allows organizations to effectively and efficiently manage specific cybersecurity risks to data and systems.

There exists a large number of security standards. For instance, ISO/IEC 27000 series alone encompasses over 60 standards that address a broad spectrum of information security concerns. This proliferation and diversification in security standards can be confusing, and in most of the cases complex to cybersecurity practitioners and organizations. The requirements for cybersecurity are distributed across numerous standards, resulting in a fragmentation issue. This can lead to the implementation of redundant or conflicting security controls when an organization must comply with multiple standards.

This work is driven by the belief that proper alignment of security controls with an organization's business needs, goals, and objectives is crucial for ensuring the effective security of their endpoint devices, data, networks, and critical infrastructure. Although standards are the primary structured source for security controls and requirements that protect organizations and systems from cyber threats, other sources of protection also exist, such as frameworks, guidelines, and legislation. Table 1 provides definitions, examples, authoritative level and scope for these additional sources of protection.

While it is very important for organizations to implement a cybersecurity standard to safeguard their valuable assets and digital space, so is the selection of the appropriate set of security controls to be implemented. In some business areas such as e-commerce, it is obligatory to comply with governmental or commercial regulatory standards. In other areas, standards adoption is voluntary or may be required in the near future. In the case where there is a need to comply with more than one standard, it can be confusing, time consuming and financially overwhelming if these standards are in part overlapping. This situation can occur especially when a new environment is added to the organization. For example, a manufacturing organization is expanding to include e-commerce. Initially, this organization had to comply with ISA/IEC 62443 [12] for example and it will need also to comply with the Payment Card Industry Data Security Standard (PCI-DSS) [22], which encompasses a set of security standards applicable to any organization handling payment card information to maintain the security and trustworthiness of the payment card industry. Even though these standards may differ based on their scope, they may include in part, similar security controls objectives. Identifying these security controls will help organizations remove overlapping controls and streamline their cyber defence mechanisms. Thus, simplifying the process of compliance and reducing the implementation time and cost for of the whole standards. Additionally, contradictory security controls objectives in standards are equally important to identify to avoid inconsistent security enforcement. An analysis of commonly adopted security standards is therefor imposed in order to expose forms of similarities and possible contradictions in security standards. This study also identifies and discusses open issues and challenges based on a mapping process to selected standards. Discussing and evaluating individual standards is outside the scope of this study, however, future research may consider individual discussions and evaluations of specific standards identified as well.

III. BACKGROUND AND RELATED WORK

A. BACKGROUND

A key responsibility of cybersecurity is to ensure the confidentiality, integrity, and availability of data and systems [23]. This can be achieved, in part, by implementing a suitable set of controls, policies, processes as well as organizational structures that support a systematic mitigation of cyber risks. Cybersecurity will continue to pose a significant challenge in the years ahead. The implementation of best practices in organizations is greatly supported by the use of standards [24]. These documents serve as a set of regulations that specify how organizations should carry out their operations and processes. Security standards are often embraced because they are proved to be effective in providing well-structured security requirements and controls. They provide a multitude of benefits that justify the time and financial resources required to produce and apply them. A raising number of manufacturers and vendors are using these standards in order to produce and sell standards-compliant products and services. Governments and businesses increasingly mandate the implementation of security standards as well. According to a recent survey conducted by Gartner, Inc. [25], 75% of organizations are actively seeking security vendor consolidation in 2022, which marks a significant increase from 29% in 2020. The requirement for secure integration and compatibility of ICT systems using technical standards is increasingly necessary. This is especially relevant in open markets where individuals have the ability to combine equipment and services from various providers, resulting in cost-saving benefits for organizations. The rapid growth of IoT devices, cyber-physical systems, and algorithm-controlled embedded systems like autonomous vehicles and digital twins is also contributing to this need [10]. Cloud computing relies heavily on standardization of hardware, software, and the services they run to ensure interoperability [26]. However, as cloud computing expands, connected systems will be exposed to new and evolving cybersecurity threats. In response, a growing number of organizations are participating and contributing to the development of cybersecurity standards. This has resulted in a significant increase in the number of standards. This trend is expected to continue, necessitating the development of new standards in the future.

TABLE 1. External Security Requirements and Control Sources.

Source	Objective	Selected sources	Owner	Focus area
Standard	Insights into security controls recommendations meant to establish Minimum Security Requirements (MSR) that ensure systems, applications and processes are designed and operated to include appropriate cybersecurity and privacy protections.	ISO/IEC 27001:2022 [13]	ISO and IEC	Addresses cybersecurity requirements.
		ISO/IEC 27002:2022 [27]	ISO and IEC	Addresses cybersecurity controls.
		ISA/IEC 62443-3-3:2019 [28]	ISA and IEC	Addresses Network and system security for Industrial Automation and Control Systems.
		PCI-DSS- The Payment Card Industry Data Security Standard [22]	PCI Security Standards Council – USA	Focus on protecting consumer financial information when stored electronically.
		ISO/SAE 21434 [10]	ISO and the Society of Automotive Engineers (SAE)	focuses on the cybersecurity risks inherent in the design and development of car electronics.
		ETSI EN 303 645 [11]	ETSI	Focus on security and data protection provisions for consumer IoT devices.
Framework	Security best practices, methods, and guidelines that organizations can embrace to get the best results for implementing a successful program.	NIST 800-37 [29]	National Institute of Standards and Technology (NIST) – USA	Provides guidelines for applying the (Risk Management Framework) RMF to information systems and organizations.
		ISO/IEC 29100 [30]	ISO and IEC	Provides high-level framework for protection of personally identifiable information within information and communication technology systems.
		COBIT- Control Objectives for Information Technology	The Information Systems Audit and Control Association (ISACA) [31].	focuses on IT security, governance, and management in organizations that want to improve product quality and, at the same time, adhere to enhanced security best practices.
		CMMC- Cybersecurity Maturity Model Certification [32]	Department of Defense (DoD)-USA	Focus on normalizing and standardizing cybersecurity preparedness across the federal governments defense industrial base (DIB).
		TARA: Threat Assessment and Remediation Analysis [33]	Jackson E. Wym. The MITTRE Corporation	Identifying and assessing cyber vulnerabilities and selecting effective countermeasures to mitigate them.
Guideline	Recommended practices that are based on industry-recognized secure practices. They lack the level of consensus and formality associated with standards.	IoT code of practice [34]	Australian Cybersecurity Center	Provides code of Practice for IoT Security for manufacturers, with guidance for consumers on smart devices at home.
		OWASP- Open Web Application Security Project [35].	Open Web Application Security Project Foundation	Focus on web security, application security and vulnerability assessment.
		NIST 800-53 [36]	NIST	Focus on security and privacy controls for information systems and organizations.
		VDI/VDE. 2182 [37]	VDI/VDE- VDI (The Association of German Engineers)	Identifying and assessing cyber vulnerabilities and selecting effective describes how specific measures can be implemented to guarantee the IT security of automated machines and plant.
Legislation	These are the highest levels of documentation in relation to cybersecurity from which other documents are created. It can incorporate security controls and standards. It is mandated by a government body, and required by law, to be complied with.	GDPR [9]	European Parliament and Council of the European Union (EU)	Focus on data protection and privacy in the European Economic Area.
		HIPAA- Health Insurance Portability and Accountability act [14]	Department of Health and Human Services (HSS)- USA.	Focus on the security and privacy of sensitive health information.
		UNECE WP29 [38]	Inland Transport Committee (ITC) of the United Nations Economic Commission for Europe (UNECE).	Focus on protecting road vehicles and road users from cybersecurity threats.
		NIS2 EU directive [7]	European Parliament and Council of the EU.	Focus on improving Member State cybersecurity capabilities, developing cybersecurity risk management in the internal market and encouraging information sharing.
		RED 2014/53/EU [8]	European Parliament and Council of the EU.	Focus on establishing a regulatory framework for radio equipment, setting essential requirements for safety and health, electromagnetic compatibility (EMC) and radio spectrum efficiency.
		Cybersecurity act [5]	European Parliament and Council of the EU.	Aims to achieve a high level of cybersecurity, cyber resilience, and trust in the EU.
		New Zealand privacy act [39]	New Zealand	Promotes and protect individual privacy.

B. RELATED WORK

In this section, we present a survey of various research works on cybersecurity standards. These studies generally emphasize the scope of applicability of different standards, the challenges, and the evolution of the taxonomy of the field. The authors in [16] report the results of a questionnaire among industry sectors and found two standards that are most applied in industry: ISO/IEC 27000-series, and the Common Criteria ISO/IEC 15408 for Information security, cybersecurity and privacy protection [17]. They also

provide a valuable table of standards that are used for specific sectors of industry. While they provide survey results of commonly used standards, they do not contrast or compare these standards. The work presented in [15] surveyed and compared commonly used standards for creating secure software applications. The authors suggest that many standards might not cover all the security requirements for secure software development when used individually. Instead, a process for creating secure software relies on implementing more than one standard, particularly to comply with regulations or obtain certification for a secure software application. Authors in [40] reviewed the development of design notations, models, and languages that can be applied to describing the IoT security and privacy requirements. The authors also discussed possible risk assessment methods and how they can be incorporated in the IoT applications and systems. The authors explained why it is important to integrate privacy in the early stage of system development. Their study shows that while most of the research articles analyze security in some way, they seldom investigate data privacy. In this survey, the authors emphasized the potential challenges and opportunities for proactive design tools that support IoT privacy. Moreover, the authors identified six research challenges related to privacy in IoT systems and their implications for the IoT research community about how to address these challenges. In [41], the authors analyzed multiple authoritative cybersecurity standards, manuals, handbooks, and literary works to present the unanimous meaning and construct of the term cyber threat. The author's work reveals that although cyber threat definitions are mostly consistent, most of them lack the inclusion of disinformation in their list/glossary of cyber threats. Hence, they conducted an in-depth comparative analysis of disinformation and its similar nature and characteristics with the prevailing and existing cyber threats. They, therefore, argue for its recommendation as an official and actual cyber threat. The authors recommend a taxonomy correction and hope that it influences future policies and regulations in combating disinformation and its propaganda In [42], the authors reviewed some of the most common industrial security standards. In total, they reviewed five standards: ISA/IEC 62443, ISO/IEC 27000 series, ISO/IEC 15408, VDI/VDE 2182, and NIST SP 800-82. It has been concluded that standards are not always one-size-fits-all. The applicability and implementation of security standards in the industrial domain may differ significantly depending on the size of the organization. Some of the mentioned standards are more applicable for larger organizations, making it more challenging for smaller organizations to implement them. This issue often results in smaller industrial organizations hiring external cybersecurity personnel that do not understand the attributes and characteristics of the domain. To help organizations adopt the cybersecurity standard or framework that best fits their cybersecurity requirements, authors in [43] reviewed published papers in the academic database to extract commonly used industrial systems cybersecurity standards. The findings of their study highlighted the comprehensive coverage of both technical and organizational best practice measures in ISA/IEC 62443. The authors in [44], discussed cybersecurity strategies and challenges in standardization and government policies with close attention to the Cybersecurity Incident Management Framework (CIMF). The authors have also provided recommendations for effective cyber defense and cybersecurity. The standards PCI DSS and ISO 17799 are reviewed and compared in [45]. The study has concluded that although both standards have similar objectives, they differ significantly in terms of scope. ISO 17799 is applicable to all types of organizations, regardless of their size and type; however, PCI is applicable for a limited range of information systems, and its implication costs depend on the maturity of the systems and the security processes and controls within a system.

While previous research have greatly advanced our understanding of security standards adoption and implementation. There remain gaps in addressing the issue of streamlining compliance efforts. Through the identification of similarities between standards, organizations can eliminate redundant work and simplify the compliance process. This, will reduce both the implementation time and the cost associated with meeting the full set of standards. The objective of this research is to provide a comprehensive evaluation of widely adopted security standards in key industry sectors demonstrating the benefits of recognizing the similarities between them.

IV. STANDARDS CLASSIFICATION

To better manage and understand the large number of cybersecurity standards that currently exist, formal classification schemes have been proposed [46], [47]. Standards can generally be categorized into regulatory, best practice (industrial), or regional as elaborated next. A full view of standards classification is depicted in Figure 1.

A. REGULATORY STANDARDS

There are two main recognized types of regulatory standards [48]:

1) DE JURE STANDARDS

De jure standards refer to standards that are established by law. They are often established by industry groups, a government body or internationally or nationally recognized standards bodies. The development process often involves negotiations between parties with different interests in the standard and these standards are often critically assessed before being approved. Each such standard is ratified through the corresponding organization's official procedures and before approval. De jure standards reflect a state of affairs that is in accordance with law and non-compliance with the standard may therefore be officially sanctioned [48]. Within the European Union, standards organisations like ETSI [11], the European Committee for Standardization (CEN) and the

European Committee for Electrotechnical Standardization (CENELEC) [49] have been a key factor in the creation of a single European market that is governed by harmonized standards [3], which we define next.

a: EU HARMONIZED STANDARDS

Harmonized standards provide the technical details to meet the essential requirements of a specific legal act within the European Union. They apply in all EU countries and replace any conflicting national standards [50]. When harmonized standards are used and applied in a correct way, they give a presumption of conformity that legal requirements are fulfilled. By implementing a harmonized standard, manufacturers and service providers can therefore demonstrate that their services or products comply with relevant EU legislation. Only harmonised standards referred to and published in the Official Journal of the European Union (OJEU) [51] are valid.

2) DE FACTO STANDARDS

De facto standards are those which have been widely accepted as the best standard for their purpose (e.g. ETSI EN 303 645) [48]. Such standards are also referred to as market-driven standards. This is often because they have a proven track record for efficiency and reliability. A De facto standard that become accepted by an industry are also known as industry standards or professional standards. They can also be formalized and turned into de jure standards with the approval of an official standards organization,

B. INDUSTRIAL STANDARDS

Many of these standards must be purchased [52], some may be downloaded for free of charge [11]. Paid standards often offer more comprehensive details and specifications. However, legal and financial obligations need to be considered by organizations when opting for such standards. Furthermore, standards can be viewed as vertical or horizontal standards as explained next (Figure 1).

- Vertical standards: apply to a particular industry, for example: PCI DSS which is specific to the ''payment Card Industry Data Security''.
- Horizontal standards: are generic, they have broad scope (e.g., ISO/IEC 27001) and are adopted by multiple industries, including automotive, banking, manufacturing and service providers.

C. REGION-BASED STANDARDS

In addition to the regulatory and industrial classification of standards, there exist also a classification based on the region or country where the standard is developed or adopted. Region-based standards can be developed by national, international or regional standardization organizations as shown in Figure 1. Classifying standards by region ensures that they meet the specific needs and requirements of a given country or region.

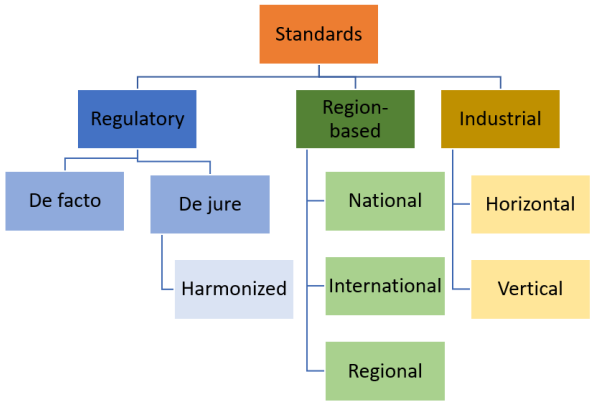


FIGURE 1. Organizing cybersecurity Standards.

- International standards are developed by international organizations such as ISO and IEC which can be adopted by countries worldwide.
- Regional standards are created by regional organizations such as the European Union (EU) and can be adopted by countries within that specific region.
- National standards are developed by a specific country such as ANSI/CTA-2088-A in the United States and the Minimum Cybersecurity Standard (MCSS) for UK.

Standards can vary in their content based on their purpose and the regulations and requirements of the region or country in which they are developed. Despite this, a standard can still belong to multiple categories. For instance, NIST 800-82 is initially a US national standard, but it has attained international recognition due to its widespread adoption. Additionally, it is also classified as an horizontal industrial standard. Similarly, ETSI EN 303 645, which was originally a European standard (regional), has gained international recognition and transformed into an international standard due to its extensive adoption. Figure 2 provides an illustrative example of these classifications. The aforementioned standards categorization, often result in security practitioners not paying enough attention to differences between organizations and their unique situational security requirements [43], [53].

This classification of scalability considerations influences the implementation of security controls, which may differ in common or unique form based on factors such as the organization's size, complexity, the importance of the information system's mission, and the organization's control scope.

V. METHODOLOGY

The overall goal for the mapping is to be as specific as possible, leaning towards under-mapping versus over-mapping. In this study, the general approach entails identifying all the elements encompassed by a control in a particular standard and then determining if a corresponding control in the compared standard articulates the exact same concept [54]. In order to accomplish this objective, we will employ the teleological interpretation method, which holds great significance within

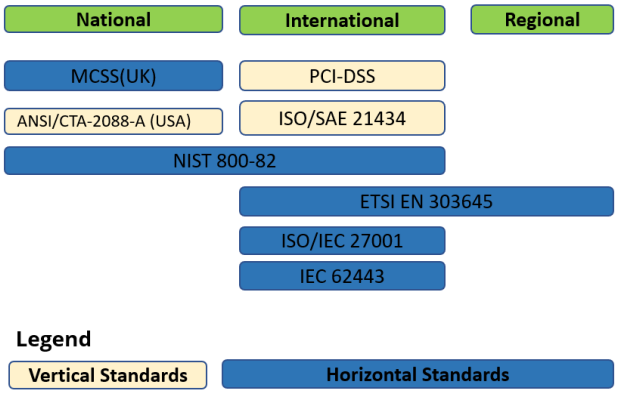


FIGURE 2. Industrial Security Standards: A classification example under region-based criteria.

the legal domain. Teleology comes from two Greek words: telos, meaning ''end, purpose or goal'', and logos, meaning ''explanation or reason'' [55]. Teleology is hence a method of explaining something through its function or purpose, rather than the thing itself. Both European national constitutional courts and the European Court of Human Rights utilize this method when justifying the interpretation of a legal rule in a concrete case. They maintain that such an interpretation can be justified by considering the goal (telos) that the rule is intended to realize [56]. As control objectives are intended to meet specific security goals outlined by a particular standard, the application of teleological interpretation is a valid approach for determining the meaning of a control. Hence, in this work, the requirements and the controls have been interpreted, compared and mapped according to their wording as well as their purpose or goal. More precisely, if the wording of the two controls are the same, they are matched with the relationship ''**Equivalent**''. If the controls have not identical wording but achieve the same purpose or goal, the type of the relationship between two defensive countermeasures is further analysed and the relationship is considered as ''**Related**''. As an example:

1– ISO/IEC 27001:2022 requirement 8.24 ''Use of cryptography'' is ***Equivalent*** to ISA/IEC 62443-3-3:2019 requirement 8.5 SR 4.3 ''Use of cryptography''.

2– ISO/IEC 27001:2022 requirement 8.21 ''Networks security'' is ***Related*** to ETSI EN 303 645 requirement 5.6-1 ''All unused network and logical interfaces shall be disabled''.

VI. MAPPING ISO/IEC 27001:2022 TO ISA/IEC 62443-3-3 AND ETSI EN 303645

In this section, we, first, present a comprehensive overview of the selected security standards ETSI EN 303 645 v2.1.1 [11], ISO/IEC 27001:2022 [13] and ISA/IEC 62443-3-3:2019 [28]. Subsequently, we perform a mapping analysis to uncover any similarities and disparities in the security requirements among the standards, providing a comprehensive examination of our findings. For this comparative analysis, we have mapped both ISA/IEC 62443-3-3 and ETSI EN 303 645 to ISO 27001:2022, a widely recognized security standard that serves as a reference for many organizations. Considering its extensive acceptance, the decision to use ISO 27001:2022 as the baseline for this comparison was a reasonable and expected choice.

The mapping process encompasses all the security controls outlined in ISO/IEC 27001:2022. Each control is thoroughly examined and evaluated, then the teleological interpretation method is applied to determine if a corresponding control exists in the standards being compared. If the security control encompasses multiple sub-controls (Figure 3), they are also included in the mapping. To accommodate the extensive number of security controls in each of the analyzed standards, the mapping tables in Appendix IX (Tables 5 and 6) solely display the controls that demonstrate alignment between the standards. Controls that lack a corresponding entry are excluded from these tables.

A. OVERVIEW OF THE SELECTED STANDARDS

The choice of the aforementioned security standards was made deliberately and thoughtfully, to showcase that despite their distinct application environments, there are still potential similarities among them. In addition, these standards are widely accepted, produced by various standardization bodies, and regarded as the best practices in their specific domains. They encompass a comprehensive set of cybersecurity controls for Information Security Management Systems (ISMS) [52], industrial systems [12], and IoT consumers [11] and are relevant to a range of environments, both horizontal and vertical. In the following sections, a more in-depth examination of each selected standard will be provided.

1) ISO/IEC 27001:2022

The ISO/IEC 27001:2022 standard outlines security controls for setting up, implementing, maintaining, and continually enhancing an Information Security Management System (ISMS). This includes administrative aspects of cybersecurity, such as security policies, as well as the human factors involved in privacy protection. A comprehensive list of all controls can be found in ISO/IEC 27001:2022 Annex A. ISO/IEC 27001:2022 is part of the ISO 27000 series, and is widely adopted by various countries and industries [52]. It can serve as a reference for identifying and implementing security controls in an ISMS, or as a source of guidance for creating industry-specific cybersecurity controls.

ISO/IEC 27001:2022 is the most recent update made by ISO, incorporating 93 high level controls (Figure 3) integrated into four distinct areas in terms of organizational, people, physical, and technology as presented in Figure 4. Each of these area controls must be addressed to respond to the challenges associated with ISMS cybersecurity.

This new version supersedes ISO 27001:2013, which comprised 114 controls across 14 categories, and introduces enhanced requirements and controls to address privacy protection, as well as the impact of technological advancements

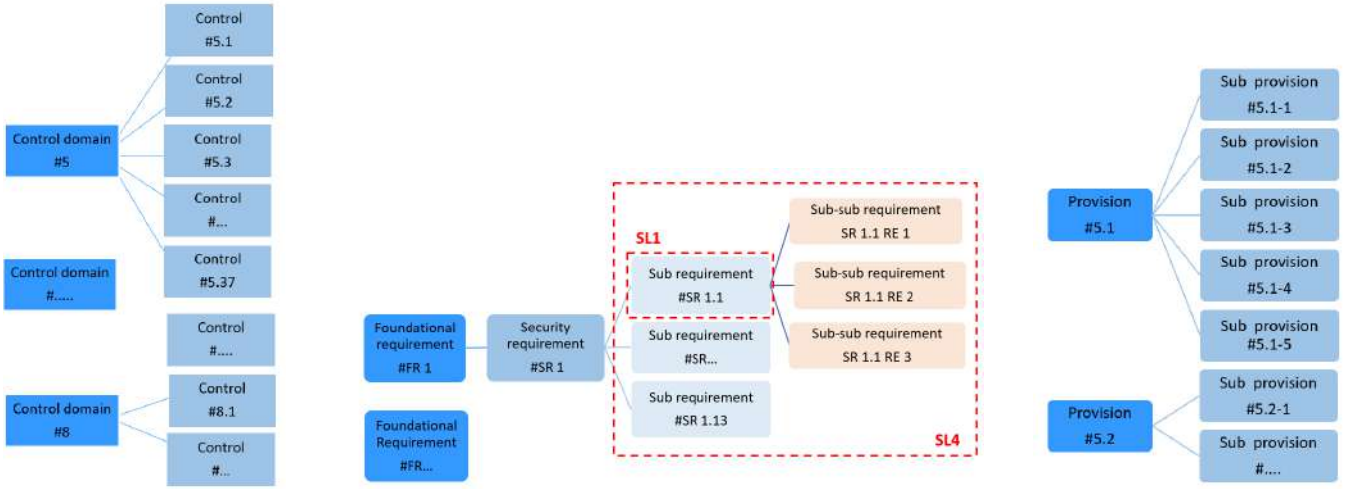


FIGURE 3. Security controls and requirements hierarchy of ISO/IEC 27001:2022 [13] (a) , ISA/IEC 62443-3-3 [28](b) and ETSI EN 301 645 [11](c). The red dashed squares is used to illustrate the security requirements (SRs) in a foundational requirement (FR) that could be included in a SL1 and SL4).

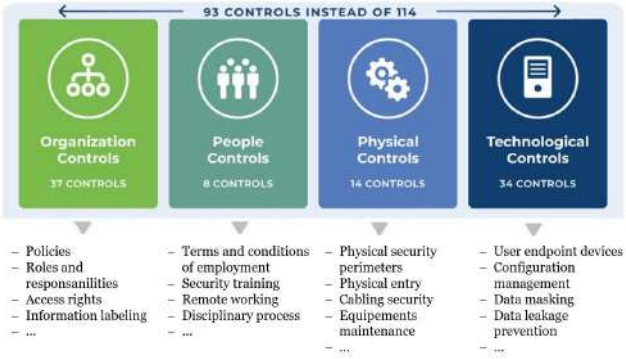


FIGURE 4. Controls areas in ISO 27001:2022.

TABLE 2. Security levels (SLs) in ISA/IEC 62443 [12].

Security Level	Description
SL0	No specific requirements or security protection.
SL1	Protection against casual or coincidental violation.
SL2	Protection against intentional violation using simple means with low resources, generic skills and low motivation.
SL3	Protection against intentional violation using sophisticated means with moderate resources, system-specific skills and moderate motivation.
SL4	Protection against intentional violation using sophisticated means with extended resources, system specific skills and high motivation

and evolving industrial practices. These changes reflect current security challenges in relation to modern risks and their associated controls.

2) ISA/IEC 62443-3-3:2019

The International Society of Automation (ISA) and The International Electro-technical Commission (IEC) jointly developed a flexible framework to address and mitigate current and future security vulnerabilities in industrial automation and control systems (IACSs). ISA/IEC 62443 includes detailed technical control system requirements (SRs) and requirement enhancements (RE) for Industrial Automation and Control Systems (IACSs) related to seven foundational requirements (FRs) (Figure 3), which define the requirements for control system capability security levels (SLs) and their components [12]. The industrial control system architecture should according to the standard be split into segments of zones and conduits, where the segmentation is an outcome of a security risk assessment. A zone is a collection of assets that have common security requirements. Conduits on the other hand is a logical grouping of communication channels between two or more zones. To achieve the desired security level and an acceptable level of risk for their network and components, organizations have the option to select from five different security levels, namely SL0 to SL4 as described in Table 2. As the security level increases, the number of necessary security controls also increases.

ISA/IEC 62443 standard consists of 12 standards arranged into 4 packages that address various aspects or levels of IACS security, including system availability, protection of the plant, and time-critical system response [12] enforced by various access control and network security requirements. For the purpose of limiting the extent of this study, we concentrate on the ISA/IEC 62443-3-3 standard, which provides specific documentation for system security requirements and security levels. It is deemed as a crucial standard within the ISA/IEC 62443 framework. The complete rundown of the security requirements are detailed in the standard document.

3) ETSI EN 303 645 v2.1.1

In 2020, ETSI introduced the standard ETSI EN 303 645 [11] with the objective of establishing high-level security and data protection provisions for consumer Internet of Things (IoT) devices connected to network infrastructure. This standard targets all parties that are involved in manufacturing and developing products and appliances that work based on the Internet of Things technology. The standard consists of 13 high-level recommendations that encompass 68 provisions, of which 33 are mandatory and the remaining are recommendations, applicable to general horizontal or sector-specific security requirements. The comprehensive listing of the provisions is accessible in the standard document [11]. Essentially, ETSI EN 303 645 places a strong emphasis on the protection of consumer data, the security of IoT devices and the protection of consumer's privacy. The standard has become a widely recognized reference for securing IoT devices globally and is utilized in various cybersecurity certification programs. As the first globally applicable cybersecurity standard for consumer IoT devices, ETSI EN 303645 is suitable for a diverse range of consumer products and is a demonstration of security best practice through voluntary industry compliance.

B. MAPPING ETSI EN 303 645 TO ISO/IEC 27001:2022

The mapping analysis, including ISO/IEC 27001:2022 controls and ETSI EN 303645 v.2.1.1 high-level and low-level provisions, shows that all ETSI EN 303 645 requirements can be aligned with ISO/IEC 27001:2022. This result is plausible as IoT consumer products can be considered as information technology devices. Therefore, it can be safely concluded that, to some extent, implementing ISO/IEC 27001:2022 can also fulfill the requirements of ETSI EN 303 645. Nevertheless, the study also shows that 64 out of the 93 ISO/IEC 27001 controls do not have a corresponding provision in ETSI EN 303 645, found particularly within the category of organizational controls which focuses on organizational leadership and employment aspects. This discrepancy can be justified as these requirements are typically not relevant to individuals, for instance:

- ISO/IEC 27001 controls ranging from 5.2 to 5.13: ensure that security policies are written and reviewed in accordance with the organization's information security practices and establish a framework for adequately implementing and maintaining these practices. These controls are directed towards organizations and do not apply to individuals.
- ISO/IEC 27001:2022 controls from 6.2 to 6.6: focus on defining the employment and termination conditions for organizational employees, and are viewed as a logical gap because they are crucial for employees but have no relevance for individuals in a personal capacity.
- ISO/IEC 27001:2022 controls 7.1 to 7.12: outline physical access controls and are not applicable to IoT environment. It is also expected that a device intended for personal use would not require physical access controls.
- ISO/IEC 27001:2022 controls 8.29 to 8.31: pertain to technological controls for security testing and monitoring and reviewing activities related to outsourced system development, but do not apply to personal devices.

The full mapping result of this comparison is displayed in Appendix IX (Table 5).

C. MAPPING ISA/IEC 62443-3-3:2019 TO ISO/IEC 27001:2022

The comparison between ISO/IEC 27001:2022 to ISA/IEC 62443-3-3, as depicted in Appendix IX (Table 6), reveals that while there are a large overlap between the two standards, we also found several gaps (see Table 3). Some of the omissions in ISA/IEC 62443-3-3 standard may be addressed in other parts of the ISA/IEC 62443 standards series. For instance, the security policy controls in ISO 27001:2022, have not been addressed in ISA/IEC 62443-3-3, but they are covered in ISA/IEC 62443-2-1. This suggests that the ISA/IEC 62443 standard series is designed to be complementary, with each part addressing different aspects of ICS security and filling in any gaps left by other parts. Other gaps can be justified as follows:

- ISA/IEC 62443-3-3 Req 6.4 and 6.5: Wireless connections and wireless endpoints devices are similar to other types of network connections but wireless devices can require a different set of security controls. Requirements related to wireless connectivity also differ to some extent between ISA/IEC 62443-3-3 and ISO/IEC 27001:2022. The requirements for wireless industry automation components based on ISA/IEC 62443-3-3 note the importance on strict use control measures where the focus is on identifying unauthorized wireless devices. In ISO/IEC 27001:2022 on the other hand is highlighting the challenge in controlling wireless network perimeter and procedures for configuration of wireless network devices. Radio coverage adjustments is here mentioned as a control for segregation of wireless networks. Requirements in ISA/IEC 62443-3-3 related to configuration of portable and mobile devices are more strict and indicate automatic enforcement of configurable usage restrictions.
- ISA/IEC 62443-3-3 Req 6.6: it covers requirements for mobile code technologies and indicate for example the need for capabilities to prevent execution of mobile code as well as restricting transfer of mobile code to/from devices. A similar requirement is not defined in ISO/IEC 27001:2022.
- ISA/IEC 62443 Req 9.4: The ISA/IEC 62443 series standards has introduced the concept of security zones, where a zone is a group of logical or physical assets that share common security requirements. Security controls can be defined both for zone boundaries and controls that are valid within a specific zone. ISA/IEC 62443-3-3 also include requirements for zone boundary protection. An

equivalent control system that would provide capabilities to monitor and control communications and connections between system boundaries is not included in ISO/IEC 27001:2022. Segregation of networks with the purpose to split the network into security boundaries and control the network perimeter of each domain using e.g. gateways is defined in ISO/IEC 27001:2022, but it is not analogous the concept of zones in ISA/IEC 62443-3-3.
– ISA/IEC 62443 Req 9.5: prohibits all general purpose person-to-person communications which is an example of a industry automation specific requirement. From an industry control system perspective it is essential to prohibit the usage of the industrial automation system for the purpose of private communication, since this could potentially be an attack vector to exploit vulnerabilities in a factory environment. It is understandable that a corresponding requirement is not included in ISO/IEC 27001:2022 due to the fact that the scope is different.

VII. DISCUSSION OF THE MAPPING RESULTS

The objective of this study was to analyse the similarities and differences between the security controls of three well-established industrial cybersecurity standards: ISO/IEC 27001, ISA/IEC 62443-3-3, and ETSI EN 303 645. The study also aims at identifying strengths and weaknesses of each of the mentioned standards. Although the mapping analysis revealed some gaps between the standards as illustrated in Table 4, it can be reasonably argued that there are numerous common, generic cybersecurity requirements (Figure 5) that are valid and applicable to various industries and ICT environments. It also showed that all the three analyzed standards encompass a collection of generic requirements that can enhance an organization's cybersecurity posture. In order to provide further insight into the results of the mapping study, we aligned the security controls of each standard to one of the cybersecurity functions, as defined in ISO/IEC 27001:2022, ISO/IEC TS 27110 [21] and the NIST cybersecurity Framework (CSF) [57]. These standards categorize cybersecurity functions, referred to as cybersecurity concepts, into five categories such as: Identify, Protect, Detect, Respond, and Recover. By doing so, one can determine which areas of system security each standard prioritizes. The strength or weaknesses of a specific area are demonstrated by the number of security controls created for each concept. The analysis indicates that ISO 27001:2022 has a more comprehensive set of controls for each cybersecurity concept with a total of 125 controls compared to 113 in ISA/IEC 62443-3-3:2019 and 72 in ETSI EN 303 645 V2.1.1 (as depicted in Figure 6), suggesting its superiority compared to the other two standards. Next we will elaborate on the distinctive characteristics of each standard.

A. ISO/IEC 27001:2022

ISO/IEC27001:2022 views cybersecurity as a combination of requirements and controls related to organization, people, process, and technology as highlighted in Table 4. The

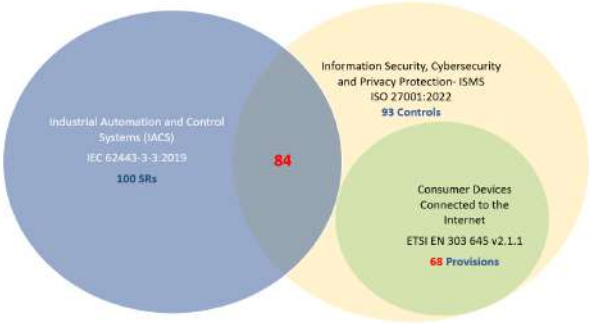


FIGURE 5. Security standards coverage.

study revealed that ISO/IEC 27001:2022 emphasized human resource security with controls for employment, termination, and changes of employment, applying to both employees and contractors, a feature lacking in the other two standards. The findings also indicate that ISO/IEC 27001:2022 had a clear advantage over the other two standards in facilitating and simplifying the mapping process. All ISO/IEC27001:2022 requirements are written at a high-level and do not include any low-level requirements. However, ETSI EN 303 645 and ISA/IEC 62443-3-3 were more challenging to map as each control encompasses additional sub-controls that required careful examination (Figure 3), sometimes leading to ambiguity and confusion. For instance, ETSI EN 303 645's provision "no default passwords 5.1-1" includes additional low-level provisions for authentication mechanisms. Figure 6 illustrates how ISO/IEC 27001:2022 has been updated to include a more comprehensive coverage of cybersecurity concepts of 125 controls. All of the standards contain a greater number of controls dedicated to the protection of the system, compared to the other cybersecurity concepts. In particular, ISO/IEC 27001:2022 supersedes both standards with controls that are crucial to identify the risk, respond to, and recover from attacks. The emphasis on risk identification highlights the standard's increased focus on preventing attacks and minimizing the costs associated with mitigation. Furthermore, the ISO standard places greater emphasis on implementing measures to respond to and recover from a cyber attack, which demonstrates its commitment to promoting system resilience and facilitating a rapid return to normal operations in the event of an attack.

B. ETSI EN 303 645 V2.1.1

The ETSI EN 303 645 standard provides baseline security provision for consumer IoT focusing on data protection and consumer privacy. Since the devices addressed by this standard are intended for personal use, the focus is primarily on protection measures and risk identification with very limited controls to detect, respond and recover from attacks (Figure 6). Furthermore, unlike the ISO/IEC 27001:2022 standard, it does not address people and physical controls as they are not applicable to ETSI standard scope (Table 4). From the mapping analysis presented in

TABLE 3. Unmapped ISA/IEC 62443-3-3:2019 requirements.

Requirement identifier	Requirement name	Description
6.4	SR 2.2 – Wireless use control	The control system shall provide the capability to authorize, monitor and enforce usage restrictions for wireless connectivity to the control system according to commonly accepted security industry practices.
6.4.3.1	SR 2.2 RE 1 – Identify and report unauthorized wireless devices.	The control system shall provide the capability to identify and report unauthorized wireless devices transmitting within the control system physical environment.
6.5	SR 2.3 – Use control for portable and mobile devices.	The control system shall provide the capability to automatically enforce configurable usage restrictions.
6.5.3.1	SR 2.3 RE 1 – Enforcement of security status of portable and mobile devices.	The control system shall provide the capability to verify that portable or mobile devices attempting to connect to a zone comply with the security requirements of that zone.
6.6	SR 2.4 – Mobile code.	The control system shall provide the capability to enforce usage restrictions for mobile code technologies based on the potential to cause damage to the control system.
6.6.3.1	SR 2.4 RE 1 – Mobile code integrity check	The control system shall provide the capability to verify integrity of the mobile code before allowing code execution.
6.12	SR 2.10 – Response to audit processing failures	The control system shall provide the capability to alert personnel and prevent the loss of essential services and functions in the event of an audit processing failure. The control system shall provide the capability to support appropriate actions in response to an audit processing failure according to commonly accepted industry practices and recommendations.
9.4	SR 5.2 – Zone boundary protection.	The control system shall provide the capability to monitor and control communications at zone boundaries to enforce the compartmentalization defined in the risk-based zones and conduits model.
9.4.3.1	SR 5.2 RE 1 – Deny by default, allow by exception.	The control system shall provide the capability to deny network traffic by default and allow network traffic by exception (also termed deny all, permit by exception).
9.4.3.2	SR 5.2 RE 2 – Island mode	The control system shall provide the capability to prevent any communication through the control system boundary (also termed island mode). NOTE Examples of when this capability may be used include where a security violation and/or breach has been detected within the control system, or an attack is occurring at the enterprise level.
9.4.3.3	SR 5.2 RE 3 – Fail close	The control system shall provide the capability to prevent any communication through the control system boundary when there is an operational failure of the boundary protection mechanisms (also termed fail close). This 'fail close' functionality shall be designed such that it does not interfere with the operation of a SIS or other safety-related functions.
9.5	SR 5.3 – General purpose person-to-person communication restrictions.	The control system shall provide the capability to prevent general purpose person-to-person messages from being received from users or systems external to the control system.
9.5.3.1	SR 5.3 RE 1 – Prohibit all general-purpose person-to-person communications	The control system shall provide the capability to prevent both transmission and receipt of general-purpose person-to-person messages.
11.8.3.1	SR 7.6 RE 1 – Machine-readable reporting of current security settings.	The control system shall provide the capability to generate a report listing the currently deployed security settings in a machine-readable format.

TABLE 4. Control domains coverage.

Control domains	ISMS ISO/IEC 27001:2022	IACS ISA/IEC 62443-3-3:2019	Consumer IoT ETSI EN 303 645 v2.1.1
Organizational requirements	✓		✓
People requirements	✓		
Physical requirements	✓	✓	
Technological requirements	✓	✓	✓

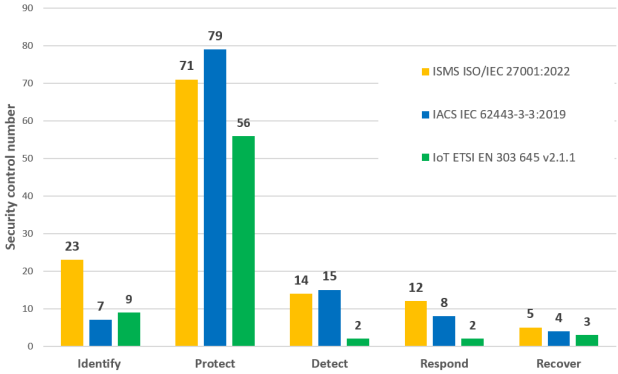


FIGURE 6. Security standards controls coverage.

Appendix IX (Table 5), it can be safely concluded that the organization and technology controls in the ISO/IEC 27001:2022 standard provide full coverage of the ETSI EN 303 645 standard. This is supported by the fact that all 68 ETSI EN provisions were successfully mapped to 29 ISO/IEC 27001 controls (Figure 5). Therefore, organizations can leverage the ISO/IEC 27001:2022 standard to effectively implement the security requirements outlined in the ETSI EN 303 645 standard for their consumer IoT devices. When using the ISO/IEC 27001:2022 standard to implement the security requirements of the ETSI EN 303 645 standard, it is important for organizations to consider that the ETSI standard covers devices without passwords, such as household appliances with limited computing power like coffee makers or refrigerators. This means that they have to implement controls that are appropriate and effective for these devices by prioritizing practical solutions over complex security measures like authentication and authorization. The objective is to provide practical household connectivity solutions that make everyday tasks more manageable, like remotely starting a washing machine or cooking utensil, prioritizing ease of use over extensive security measures.

TABLE 5. Mapping ETSI EN 303 645 to ISO 27001:2022.

ISO/IEC 27002:2022 control identifier	ISO/IEC 27002:2022 control name	ETSI EN 303 645 requirement identifier	ETSI EN 303 645 requirement name
5.1	Policies for information security	5.2-1	The manufacturer shall make a vulnerability disclosure policy publicly available.
5.14	Information transfer	5.5-1	The consumer IoT device shall use best practice cryptography to communicate securely.
		5.5-6	Critical security parameters should be encrypted in transit, with such encryption appropriate to the properties of the technology, risk, and usage.
		5.8-1	The confidentiality of personal data transiting between a device and a service, especially associated services, should be protected, with best practice cryptography.
		5.8-2	The confidentiality of sensitive personal data communicated between the device and associated services shall be protected, with cryptography appropriate to the properties of the technology and usage.
5.15	Access control	5.1-3	Authentication mechanisms used to authenticate users against a device shall use best practice cryptography, appropriate to the properties of the technology, risk and usage.
		5.6-8	The device should include a hardware-level access control mechanism for memory.
5.17	Authentication information	5.1-1	Where passwords are used and, in any state, other than the factory default, all consumer IoT device passwords shall be unique per device or defined by the user.
		5.1-2	Where pre-installed unique per device passwords are used, these shall be generated with a mechanism that reduces the risk of automated attacks against a class or type of device.
		5.1-3	Authentication mechanisms used to authenticate users against a device shall use best practice cryptography, appropriate to the properties of the technology, risk, and usage.
		5.1-4	Where a user can authenticate against a device, the device shall provide to the user or an administrator a simple mechanism to change the authentication value used.
5.29	Information security during disruption	5.9-1	Resilience should be built into consumer IoT devices and services, taking into account the possibility of outages of data networks and power.
		5.9-2	Consumer IoT devices should remain operating and locally functional in the case of a loss of network access and should recover cleanly in the case of restoration of a loss of power.
		5.9-3	The consumer IoT device should connect to networks in an expected, operational and stable state and in an orderly fashion, taking the capability of the infrastructure into consideration.
5.30	ICT readiness for business continuity	5.9-1	Resilience should be built into consumer IoT devices and services, considering the possibility of outages of data networks and power.
		5.9-2	Consumer IoT devices should remain operating and locally functional in the case of a loss of network access and should recover cleanly in the case of restoration of a loss of power.
		5.9-3	The consumer IoT device should connect to networks in an expected, operational and stable state and in an orderly fashion, taking the capability of the infrastructure into consideration.
5.34	Privacy and protection of PII	5.8-1	The confidentiality of personal data transiting between a device and a service, especially associated services, should be protected, with best practice cryptography.
		5.8-2	The confidentiality of sensitive personal data communicated between the device and associated services shall be protected, with cryptography appropriate to the properties of the technology and usage.
		5.8-3	All external sensing capabilities of the device shall be documented in an accessible way that is clear and transparent for the user.
		5.11-1	The user shall be provided with functionality such that user data can be erased from the device in a simple manner.
		5.11-2	The consumer should be provided with functionality on the device such that personal data can be removed from associated services in a simple manner.
		5.11-3	Users should be given clear instructions on how to delete their personal data.
		5.11-4	Users should be provided with clear confirmation that personal data has been deleted from services, devices and applications.
		6-1	The manufacturer shall provide consumers with clear and transparent information about what personal data is processed, how it is being used, by whom, and for what purposes, for each device and service.
		6-2	Where personal data is processed based on consumers' consent, this consent shall be obtained in a valid way.
		6-3	Consumers who gave consent for the processing of their personal data shall have the capability to withdraw it at any time.
		6-4	If telemetry data is collected from consumer IoT devices and services, the processing of personal data should be kept to the minimum necessary for the intended functionality.
		6-5	If telemetry data is collected from consumer IoT devices and services, consumers shall be provided with information on what telemetry data is collected, how it is being used, by whom, and for what purposes.
7.13	Equipment maintenance	5.12-1	Installation and maintenance of consumer IoT should involve minimal decisions by the user and should follow security best practice on usability.
7.14	Secure disposal or reuse of equipment	5.11-1	The user shall be provided with functionality such that user data can be erased from the device in a simple manner.
		5.11-2	The consumer should be provided with functionality on the device such that personal data can be removed from associated services in a simple manner.
		5.11-3	Users should be given clear instructions on how to delete their personal data.
		5.11-4	Users should be provided with clear confirmation that personal data has been deleted from services, devices, and applications.
8.5	Secure authentication	5.1-3	Authentication mechanisms used to authenticate users against a device shall use best practice cryptography, appropriate to the properties of the technology, risk, and usage.
		5.1-5	When the device is not a constrained device, it shall have a mechanism available which makes brute-force attacks on authentication mechanisms via network interfaces impracticable.
8.8	Management of technical vulnerabilities	5.2-2	Disclosed vulnerabilities should be acted on in a timely manner.
		5.2-3	Manufacturers should continually monitor for, identify, and rectify security vulnerabilities within products and services.

TABLE 5. *(Continued.)* **Mapping ETSI EN 303 645 to ISO 27001:2022.**

8.9	Configuration management	5.6-3	Device hardware should not unnecessarily expose physical interfaces to attack.
		5.6-4	Where a debug interface is physically accessible, it shall be disabled in software.
		5.6-5	The manufacturer should only enable software services that are used or required for the intended use or operation of the device.
		5.12-2	The manufacturer should provide users with guidance on how to securely set up their device.
		5.12-3	The manufacturer should provide users with guidance on how to check whether their device is securely set up.
8.10	Information deletion	5.11-1	The user shall be provided with functionality such that user data can be erased from the device in a simple manner.
		5.11-2	The consumer should be provided with functionality on the device such that personal data can be removed from associated services in a simple manner.
		5.11-3	Users should be given clear instructions on how to delete their personal data.
		5.11-4	Users should be provided with clear confirmation that personal data has been deleted from services, devices, and applications.
8.12	Data leakage prevention	5.4-1	Sensitive security parameters in persistent storage shall be stored securely by the device.
		5.5-1	The consumer IoT device shall use best practice cryptography to communicate securely.
		5.5-2	The consumer IoT device should use reviewed or evaluated implementations to deliver network and security functionalities, particularly in the field of cryptography.
8.14	Redundancy of information processing facilities	5.9-1	Resilience should be built in to consumer IoT devices and services, taking into account the possibility of outages of data networks and power.
8.15	Logging	5.10-1	If telemetry data is collected from consumer IoT devices and services, such as usage and measurement data, it should be examined for security anomalies.
		6-3	If telemetry data is collected from consumer IoT devices and services, the processing of personal data should be kept to the minimum necessary for the intended functionality.
		6-4	If telemetry data is collected from consumer IoT devices and services, consumers shall be provided with information on what telemetry data is collected, how it is being used, by whom, and for what purposes.
8.16	Monitoring activities	5.7-2	If an unauthorized change is detected to the software, the device should alert the user and/or administrator to the issue and should not connect to wider networks than those necessary to perform the alerting function.
		5.10-1	If telemetry data is collected from consumer IoT devices and services, such as usage and measurement data, it should be examined for security anomalies.
		6-3	If telemetry data is collected from consumer IoT devices and services, the processing of personal data should be kept to the minimum necessary for the intended functionality.
		6-4	If telemetry data is collected from consumer IoT devices and services, consumers shall be provided with information on what telemetry data is collected, how it is being used, by whom, and for what purposes.
8.19	Installation of software on operational systems	5.3-1	All software components in consumer IoT devices should be securely updateable.
		5.3-2	When the device is not a constrained device, it shall have an update mechanism for the secure installation of updates.
		5.3-3	An update shall be simple for the user to apply.
		5.3-4	Automatic mechanisms should be used for software updates.
		5.3-5	The device should check after initialization, and then periodically, whether security updates are available.
		5.3-6	If the device supports automatic updates and/or update notifications, these should be enabled in the initialized state and configurable so that the user can enable, disable, or postpone installation of security updates and/or update notifications.
		5.3-7	The device shall use best practice cryptography to facilitate secure update mechanisms.
		5.3-8	Security updates shall be timely.
		5.3-9	The device should verify the authenticity and integrity of software updates.
		5.3-10	Where updates are delivered over a network interface, the device shall verify the authenticity and integrity of each update via a trust relationship.
		5.3-11	The manufacturer should inform the user in a recognizable and apparent manner that a security update is required together with information on the risks mitigated by that update.
		5.3-12	The device should notify the user when the application of a software update will disrupt the basic functioning of the device.
		5.3-13	The manufacturer shall publish, in an accessible way that is clear and transparent to the user, the defined support period.
		5.3-14	For constrained devices that cannot have their software updated, the rationale for the absence of software updates, the period and method of hardware replacement support and a defined support period should be published by the manufacturer in an accessible way that is clear and transparent to the user.
		5.3-15	For constrained devices that cannot have their software updated, the product should be isolable and the hardware replaceable.
		5.3-16	The model designation of the consumer IoT device shall be clearly recognizable, either by labelling on the device or via a physical interface.
		5.7-1	The consumer IoT device should verify its software using secure boot mechanisms.
		5.7-2	If an unauthorized change is detected to the software, the device should alert the user and/or administrator to the issue and should not connect to wider networks than those necessary to perform the alerting function.
		5.12-1	Installation and maintenance of consumer IoT should involve minimal decisions by the user and should follow security best practice on usability.
8.20	Networks security	5.6-1	All unused network and logical interfaces shall be disabled.
		5.6-2	In the initialized state, the network interfaces of the device shall minimize the unauthenticated disclosure of security-relevant information.
8.24	Use of cryptography	5.1-3	Authentication mechanisms used to authenticate users against a device shall use best practice cryptography, appropriate to the properties of the technology, risk, and usage.
		5.3-7	The device shall use best practice cryptography to facilitate secure update mechanisms.
		5.4-1	Sensitive security parameters in persistent storage shall be stored securely by the device.
		5.5-1	The consumer IoT device shall use best practice cryptography to communicate securely.

TABLE 5. *(Continued.)* **Mapping ETSI EN 303 645 to ISO 27001:2022.**

		5.5-2	The consumer IoT device should use reviewed or evaluated implementations to deliver network and security functionalities, particularly in the field of cryptography.
		5.5-3	Cryptographic algorithms and primitives should be updateable.
		5.5-6	Critical security parameters should be encrypted in transit, with such encryption appropriate to the properties of the technology, risk, and usage.
		5.8-1	The confidentiality of personal data transiting between a device and a service, especially associated services, should be protected, with best practice cryptography.
		5.8-2	The confidentiality of sensitive personal data communicated between the device and associated services shall be protected, with cryptography appropriate to the properties of the technology and usage.
8.25	Secure development life cycle	5.6-9	The manufacturer should follow secure development processes for software deployed on the device.
8.26	Application security requirements	5.5-1	The consumer IoT device shall use best practice cryptography to communicate securely.
		5.5-2	The consumer IoT device should use reviewed or evaluated implementations to deliver network and security functionalities, particularly in the field of cryptography.
		5.13-1	The consumer IoT device software shall validate data input via user interfaces or transferred via Application Programming Interfaces (APIs) or between networks in services and devices.
8.27	Secure system architecture and engineering principles	5.4-1	Sensitive security parameters in persistent storage shall be stored securely by the device.
		5.4-2	Where a hard-coded unique per device identity is used in a device for security purposes, it shall be implemented in such a way that it resists tampering by means such as physical, electrical or software.
		5.4-4	Any critical security parameters used for integrity and authenticity checks of software updates and for protection of communication with associated services in device software shall be unique per device and shall be produced with a mechanism that reduces the risk of automated attacks against classes of devices.
		5.5-3	Cryptographic algorithms and primitives should be updateable.
		5.5-4	Access to device functionality via a network interface in the initialized state should only be possible after authentication on that interface.
		5.5-5	Device functionality that allows security-relevant changes in configuration via a network interface shall only be accessible after authentication.
		5.5-6	Critical security parameters should be encrypted in transit, with such encryption appropriate to the properties of the technology, risk, and usage.
		5.5-7	The consumer IoT device shall protect the confidentiality of critical security parameters that are communicated via remotely accessible network interfaces.
		5.5-8	The manufacturer shall follow secure management processes for critical security parameters that relate to the device.
		5.6-1	All unused network and logical interfaces shall be disabled.
		5.6-2	In the initialized state, the network interfaces of the device shall minimize the unauthenticated disclosure of security-relevant information.
		5.6-7	Software should run with least necessary privileges, taking account of both security and functionality.
8.28	Secure coding	5.4-3	Hard-coded critical security parameters in device software source code shall not be used.
		5.6-4	Where a debug interface is physically accessible, it shall be disabled in software.
		5.6-6	Code should be minimized to the functionality necessary for the service/device to operate.
		5.13-1	The consumer IoT device software shall validate data input via user interfaces or transferred via Application Programming Interfaces (APIs) or between networks in services and devices.
8.32	Change management	5.7-2	If an unauthorized change is detected to the software, the device should alert the user and/or administrator to the issue and should not connect to wider networks than those necessary to perform the alerting function.

C. ISA/IEC 62443-3-3:2019

The concept of a risk based segmented architecture with zones, conducts and security levels differentiates the ISA/IEC 62443-3-3 from the other standards we have analysed in this study. This approach allows organizations to apply security controls based on the level of acceptable risk and protection needed. Lower security levels such as SL0 or SL1 may suffice for non-critical industrial environments, while higher security levels such as SL3 or SL4 are essential for high-risk or critical systems. Despite these particularities, Appendix IX (Table 6) testifies on the large overlap between ISO 27001 and ISA/IEC 62443. In fact, out of 100 requirements from ISA/IEC 62443-3-3, 84 have been mapped to equivalent or related controls in ISO/IEC 27001 (Figure 5). The unmapped requirements as shown in Table 3 indicates a number of requirement enhancements used in SL3 or SL4 that are relevant and important for high-level security systems such as critical systems. Therefore it might in some cases be justified to implement a set of baseline cybersecurity requirements defined in ISO/IEC 27001 in a non-critical industrial automation environment. In a high risk or critical industrial environments additional system level requirements designed to protect against intentional violations needs to be considered. ISA/IEC 62443-3-3 places greater emphasis on technical protection measures with a total of 79 protective controls compared to 71 in ISO 27001:2022. Additionally, ISA/IEC 62443-3-3 focuses on controls to detect attacks, but places less importance on controls for pre- and post-attacks. This direction has also been followed in the other two standards. It is important to note that all controls in ISA/IEC 62443-3-3 are physical or technological requirements as shown in Table 4, as this standard is intended for system requirements. Organizational and people controls are addressed in other parts of the ISA/IEC 62443 standard package.

VIII. CHALLENGES

The evolving nature of the cybersecurity area, characterized by the emergence of new threats and vulnerabilities, makes unrealistic to establish a permanent and steady level of system security over time. Instead cybersecurity is optimized to a level business leaders define, balancing the limited resources available to the acceptable risk appetite. Complying to a cybersecurity standard can partially manage cybersecurity challenges, attacks opportunities and cyber risks. However, not all risks can be mitigated through standards and frameworks. Given the cross-functional nature of cybersecurity, the development and implementation of effective security standards and frameworks present additional challenges that

TABLE 6. Mapping ISA/IEC 62443-3-3:2019 to ISO/IEC 27002:2022.

ISO/IEC 27002:2022 control identifier	ISO/IEC 27002:2022 control name	ISA/IEC 62443-3-3:2019 requirement identifier	ISA/IEC 62443-3-3:2019 requirement name
5.3	Segregation of duties	5.3	SR 1.1 – Human user identification and authentication
		6.3	SR 2.1 – Authorization enforcement
		6.3.3.1	SR 2.1 RE 1 – Authorization enforcement for all users
5.9	Inventory of information and other associated assets	11.10	SR 7.8 – Control system component inventory
5.14	Information transfer	8.3	SR 4.1 – Information confidentiality
		8.3.3.1	SR 4.1 RE 1 – Protection of confidentiality at rest or in transit via untrusted networks
		8.3.3.2	SR 4.1 RE 2 – Protection of confidentiality across zone boundaries
5.15	Access control	5.3	SR 1.1 – Human user identification and authentication
		5.3.3.1	SR 1.1 RE 1 – Unique identification and authentication
		5.3.3.2	SR 1.1 RE 2 – Multifactor authentication for untrusted networks
		5.3.3.3	SR 1.1 RE 3 – Multifactor authentication for all networks
		5.4	SR 1.2 – Software process and device identification and authentication
		5.4.3.1	SR 1.2 RE 1 – Unique identification and authentication
5.16	Identity management	5.6	SR 1.4 – Identifier management
		5.7	SR 1.5 – Authenticator management
		5.7.3.1	SR 1.5 RE 1 – Hardware security for software process identity credentials
		5.8	SR 1.6 – Wireless access management
		5.8.3.1	SR 1.6 RE 1 – Unique identification and authentication
5.17	Authentication information	5.9	SR 1.7 – Strength of password-based authentication
		5.9.3.1	SR 1.7 RE 1 – Password generation and lifetime restrictions for human users
		5.9.3.2	SR 1.7 RE 2 – Password lifetime restrictions for all users
5.18	Access rights	5.5	SR 1.3 – Account management
		5.5.3.1	SR 1.3 RE 1 – Unified account management
		6.3	SR 2.1 – Authorization enforcement
		6.3.3.1	SR 2.1 RE 1 – Authorization enforcement for all users
		6.3.3.2	SR 2.1 RE 2 – Permission mapping to roles
		6.3.3.4	SR 2.1 RE 4 – Dual approval
5.28	Collection of evidence	6.10	SR 2.8 – Auditable events
		6.10.3.1	SR 2.8 RE 1 – Centrally managed, system-wide audit trail
5.29	Information security during disruption	7.8	SR 3.6 – Deterministic output
		11.3	SR 7.1 – Denial of service protection
		11.3.3.1	SR 7.1 RE 1 – Manage communication loads
		11.3.3.2	SR 7.1 RE 2 – Limit DoS effects to other systems or networks
		11.4	SR 7.2 – Resources management
5.30	ICT readiness for business continuity	11.5	SR 7.3 – Control system backup
		11.5.3.1	SR 7.3 RE 1 – Backup verification
		11.5.3.2	SR 7.3 RE 2 – Backup automation
		11.6	SR 7.4 – Control system recovery and reconstitution
		11.7	SR 7.5 – Emergency power
5.33	Protection of records	7.11	SR 3.9 – Protection of audit information
		8.3	SR 4.1 – Information confidentiality
		8.3.3.1	SR 4.1 RE 1 – Protection of confidentiality at rest or in transit via untrusted networks
		8.3.3.2	SR 4.1 RE 2 – Protection of confidentiality across zone boundaries
5.34	Privacy and protection of PII	8.3	SR 4.1 – Information confidentiality
		8.3.3.1	SR 4.1 RE 1 – Protection of confidentiality at rest or in transit via untrusted networks
		8.3.3.2	SR 4.1 RE 2 – Protection of confidentiality across zone boundaries.
6.6	Confidentiality or non-disclosure agreements	8.3	SR 4.1 – Information confidentiality
6.7	Remote working	5.15	SR 1.13 – Access via untrusted networks
		5.15.3.1	SR 1.13 RE 1 – Explicit access request approval
7.7	Clear desk and clear screen	8.3	SR 4.1 – Information confidentiality
7.10	Storage media	8.3	SR 4.1 – Information confidentiality
7.11	Supporting utilities	11.7	SR 7.5 – Emergency power
8.1	User endpoint devices	6.7	SR 2.5 – Session lock
8.2	Privileged access rights	5.3	SR 1.1 – Human user identification and authentication
		5.4	SR 1.2 – Software process and device identification and authentication
8.4	Access to source code	7.6	SR 3.4 – Software and information integrity
		7.6.3.1	SR 3.4 RE 1 – Automated notification about integrity violations
8.5	Secure authentication	5.3	SR 1.1 – Human user identification and authentication
		5.3.3.2	SR 1.1 RE 2 – Multifactor authentication for untrusted networks
		5.3.3.3	SR 1.1 RE 3 – Multifactor authentication for all networks
		5.9	SR 1.7 – Strength of password-based authentication
		5.9.3.1	SR 1.7 RE 1 – Password generation and lifetime restrictions for human users
		5.9.3.2	SR 1.7 RE 2 – Password lifetime restrictions for all users
		5.10	SR 1.8 – Public key infrastructure (PKI) certificates
		5.11	SR 1.9 – Strength of public key authentication
		5.11.3.1	SR 1.9 RE 1 – Hardware security for public key authentication
		5.12	SR 1.10 – Authenticator feedback
		5.13	SR 1.11 – Unsuccessful login attempts
8.6	Capacity management	6.11	SR 2.9 – Audit storage capacity
		6.11.3.1	SR 2.9 RE 1 – Warn when audit record storage capacity threshold reached
8.9	Configuration management	11.10	SR 7.8 – Control system component inventory
8.10	Information deletion	8.4	SR 4.2 – Information persistence
		8.4.3.1	SR 4.2 RE 1 – Purging of shared memory resources
8.11	Data masking	8.3	SR 4.1 – Information confidentiality

TABLE 6. *(Continued.)* **Mapping ISA/IEC 62443-3-3:2019 to ISO/IEC 27002:2022.**

8.12	Data leakage prevention	8.3	SR 4.1 – Information confidentiality
8.13	Information backup	11.5	SR 7.3 – Control system backup
		11.5.3.1	SR 7.3 RE 1 – Backup verification
		11.5.3.2	SR 7.3 RE 2 – Backup automation
8.14	Redundancy of information processing facilities	11.4	SR 7.2 – Resource management
8.15	Logging	6.10	SR 2.8 – Auditable events
		6.14	SR 2.12 – Non-repudiation
		6.14.3.1	SR 2.12 RE 1 – Non-repudiation for all users
		7.4	SR 3.2 – Malicious code protection
		7.4.3.1	SR 3.2 RE 1 – Malicious code protection on entry and exit points
		7.4.3.2	SR 3.2 RE 2 – Central management and reporting for malicious code protection
		10.3	SR 6.1 – Audit log accessibility
8.16	Monitoring activities	10.4	SR 6.2 – Continuous monitoring
8.17	Clock synchronization	6.13	SR 2.11 – Timestamps
		6.13.3.1	SR 2.11 RE 1 – Internal time synchronization
		6.13.3.2	SR 2.11 RE 2 – Protection of time source integrity
8.18	Use of privileged utility programs	6.3	SR 2.1 – Authorization enforcement
		6.3.3.3	SR 2.1 RE 3 – Supervisor override
8.19	Installation of software on operational systems	7.6	SR 3.4 – Software and information integrity
8.20	Networks security	11.8	SR 7.6 – Network and security configuration settings
8.21	Security of network services	11.8	SR 7.6 – Network and security configuration settings
8.22	Segregation of networks	9.3	SR 5.1 – Network segmentation
		9.3.3.1	SR 5.1 RE 1 – Physical network segmentation
		9.3.3.3	SR 5.1 RE 3 – Logical and physical isolation of critical networks
8.23	Web filtering	9.5	SR 5.3 – General purpose person-to-person communication restrictions
8.24	Use of cryptography	7.3	SR 3.1 – Communication integrity
		7.3.3.1	SR 3.1 RE 1 – Cryptographic integrity protection
		8.5	SR 4.3 – Use of cryptography
8.26	Application security requirements	7.8	SR 3.6 – Deterministic output
		7.9	SR 3.7 – Error handling
		9.6	SR 5.4 – Application partitioning
8.27	Secure system architecture and engineering principles	6.8	SR 2.6 – Remote session termination
		6.9	SR 2.7 – Concurrent session control
		7.10	SR 3.8 – Session integrity
		7.10.3.1	SR 3.8 RE 1 – Invalidation of session IDs after session termination
		7.10.3.2	SR 3.8 RE 2 – Unique session ID generation
		7.10.3.3	SR 3.8 RE 3 – Randomness of session Ids
		11.9	SR 7.7 – Least functionality
8.28	Secure coding	7.6	SR 3.4 – Software and information integrity
		7.7	SR 3.5 – Input validation
8.29	Security testing in development and acceptance	7.5	SR 3.3 – Security functionality verification
		7.5.3.1	SR 3.3 RE 1 – Automated mechanisms for security functionality verification
		7.5.3.2	SR 3.3 RE 2 – Security functionality verification during normal operation

demand close coordination among multiple stakeholders. Selecting a framework or standard can be challenging, considering the excess of security standards, the resulting security controls fragmentation and the complexity of implementing the standards across different domains.

When organizations are mandated to comply with several standards, they may end up implementing redundant or conflicting security controls. In order to overcome this challenge, organizations can focus on identifying duplicated controls to simplify the process and minimize expenses. However, mapping controls between standards can be a difficult task because controls are written in various ways, with some being written at a high-level, while others have low-levels requirements and some may even contain ambiguous requirements that require careful examination. Another challenge or common mistake is addressing cybersecurity on a system-by-system basis. Consequently, the security perspective of the entire system, including its intended use, operational environment, and characteristics, should be evaluated from end-to-end. This approach is recommended by the ISA/IEC 62443 standard for establishing an industrial automation and control system security (IACSs) program. However, implementing a security management program for IACSs based on the ISA/IEC 62443 framework can turn out to be a time consuming exercise. The wide-ranging management system encompassing policies, procedures, and personnel utilizing the IACSs in addition to the IACS itself. It is important to emphasize that industrial automation and control systems are employed across various industries, and it is essential to acknowledge that not all industrial systems and applications should be classified as critical. In fact it is not unusual to use commercial off-the-shelf (COTS) components and consumer products in an industrial environment. In a critical systems these kind of products may not be robust enough from a cybersecurity perspective, but in a non-critical industrial automation setup they might be appropriate to use. Ultimately, cybersecurity remains the art of tolerating imperfection. Despite organizations' best efforts to implement cybersecurity measures, there is always a possibility of vulnerabilities, breaches, and other security incidents. Cybersecurity professionals must constantly adapt and respond to new and emerging threats, and prioritize their efforts based on the level of risk and available resources. In this regard, a framework or standard can be a valuable tool to assess risks, implement mitigation controls, and work in a structured way.

IX. CONCLUSION AND FUTURE WORK

The realm of cybersecurity encompasses a wide range of standards at various levels, including national, international, regional, and industry-specific. These standards can often

be overly generic, complex and hard to follow, neglecting the fact that each organization has its own distinct security needs based on its size and business type. In this study, we performed a comparative analysis between the security requirements and controls across three widely adopted standards, namely ISA/IEC 62443-3-3:2019 which addresses network and system requirements, ISO/IEC 27001:2022 deals with information security management systems and ETSI EN 303 645 v2.1.1 serves as a baseline standard for consumer IoT products. The findings of our study suggest that despite being designed for distinct environments and scopes, these standards exhibit significant similarities in their security requirements and controls. Notably, ISO/IEC 27001:2022 fully encompasses the security provisions outlined in ETSI EN 303645, while it largely covers ISA/IEC 62443-3-3 requirements. The observed gaps between the standards is attributed to the specificity of ETSI 303 645 in providing provisions for devices with limited computing capabilities that do not require complex security solutions, such as those without passwords. In contrast, ISA/IEC 62443-3-3 includes security requirements for critical industrial systems, which demand unique security considerations, resulting in differing security requirements compared to the other two standards. Our study also revealed that ISO 27001:2022 provides controls covering organization, physical, technology, and people security requirements. ETSI focuses on provisions for organization and technology security, while ISA/IEC 62443:2019 places emphasis on physical and technology security requirements. Additionally, the findings show that while all three standards prioritize protection controls, only ISO27001:2022 emphasizes the need for cyber resilience. The standard provides measures for responding to and restoring systems and operations after an attack, which is not adequately covered by the other two standards. Our work holds practical future prospects. By identifying and addressing overlaps and gaps in industrial standards security controls, we can streamline compliance efforts for organizations facing the challenge of adhering to multiple standards simultaneously. This streamlining can save valuable resources, reduce redundancy, and improve overall efficiency in cybersecurity implementation. Moreover, it can promote consistency across different standards, fostering a more integrated and effective cybersecurity framework. Since this case study involves three environment-specific standards, we will expand our efforts in the future to include additional well-established security standards to evaluate potential overlaps. Our goal is to find out a more comprehensive standard that can contribute in addressing the fragmentation issue and reduce the additional cost and effort required when complying with multiple security standards.

APPENDIX

See Tables 5 and 6.

REFERENCES

[1] P. K. Joshi. (2023). *Governance, Risk Management, and Compliance in the Cybersecurity Framework*. Accessed: Jul. 7, 2023. [Online]. Available: https://www.eccouncil.org/cybersecurity-exchange/whitepaper/governance-risk-and-compliance/

[2] C Brooks. (2022). *Alarming Cyber Statistics For Mid-Year 2022 That You Need to Know*. Accessed: Jun. 26, 2023. [Online]. Available: https://www.forbes.com/

[3] ENISA. (2020). *Standards*. Accessed: Jun. 26, 2023. [Online]. Available: https://www.enisa.europa.eu/topics/standards

[4] Alex Leadbeater. *Interview With Alex Leadbeater, Chair of TC Cyber at ETSI*. Accessed: Jun. 26, 2023. [Online]. Available: https://cybersecurity-magazine.com/interview-with-alex-leadbeater-chair-of-tc-cyber-at-etsi/

[5] European Union. (2019).*The EU Cybersecurity Act*. Accessed: Jul. 1, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2019/881/oj

[6] European Union. (2016). *The EU Network and Information Security (NIS) Directive*. Accessed: Jun. 27, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2016/1148/oj

[7] European Union. (2022). *NIS 2 Directive*. Directive (EU) 2022/2555. Accessed: Jun. 27, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2022/2555/oj

[8] European Parliament and Council of the EU. *On the Harmonisation of the Laws of the Member States Relating to the Making Available on the Market of Radio Equipment and Repealing Directive*. Accessed: Jul. 2, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0053&from=E

[9] European Union. (2016). *General Data Protection Regulation*. Regulation (EU) 2016/679. Accessed: Jul. 1, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[10] ISO/SAE. *Road Vehicles—Cybersecurity Engineering*, Standard ISO/SAE 21434, 2021. [Online]. Available: https://www.iso.org/standard/70918.html

[11] ETSI. (2020). *Cybersecurity for Consumer Internet of Things*. Accessed: Jul. 2, 2023. ETSI EN 303 645v02. [Online]. Available: https://www.etsi.org/deliver/etsi_en/303600_303699/303645/02.01.01_60/en_303645v020101p.pdf

[12] ISA/IEC. *Security of Industrial Automation and Control Systems (IACS)-IEC*, Standard ISA/IEC 62443, 2019. [Online]. Available: https://isagca.org/isa-iec-62443-standards

[13] ISO/IEC. *Information Security Management Systems*, Standard ISO/IEC 27001, 2022. [Online]. Available: https://www.iso.org/isoiec-27001-information-security.html

[14] The U.S. Department of Health and Human Services (HHS). (1996). *Health Information Privacy*. Health Insurance Portability and Accountability Act (HIPAA). Accessed: Jul. 7, 2023. [Online]. Available: https://www.hhs.gov/hipaa/for-individuals/index.html

[15] A. Ramirez, A. Aiello, and S. J. Lincke, "A survey and comparison of secure software development standards," in *Proc. 13th CMI Conf. Cybersecurity Privacy (CMI)-Digit. Transformation-Potentials Challenges*, Nov. 2020, pp. 1–6.

[16] L. Shan, B. Sangchoolie, P. Folkesson, J. Vinter, E. Schoitsch, and C. Loiseaux, "A survey on the application of safety, security, and privacy standards for dependable systems," in *Proc. 15th Eur. Dependable Comput. Conf. (EDCC)*, Sep. 2019, pp. 71–72.

[17] *Information Security, Cybersecurity and Privacy Protection—Evaluation Criteria for IT Security*, Standard ISO/IEC 15408-1, 2022. [Online]. Available: https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html

[18] ISO. *Consumers and Standards: Partnership for a Better World*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.iso.org/sites/ConsumersStandards/6_review_questions.html

[19] Fortinet. *CIA Triad*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.fortinet.com/resources/cyberglossary/cia-triad

[20] G. Mutune. *Top Cybersecurity Frameworks*. Accessed: Jul. 3, 2023. [Online]. Available: https://cyberexperts.com/cybersecurity-frameworks/#2_NIST_Cybersecurity_Framework3

[21] ISO/IEC. *Information Technology, Cybersecurity and Privacy Protection—Cybersecurity Framework Development Guidelines*, Standard ISO/IEC TS 27110, 2021. [Online]. Available: https://www.iso.org/standard/72435.html

[22] (2022). *PCI-Security Standards Council, PCI DSS: V4.0*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.pcisecuritystandards.org/document_library/

[23] Office of Information Security. *Confidentiality, Integrity, and Availability: The CIA Triad*. Accessed: Jul. 1, 2023. [Online]. Available: https://informationsecurity.wustl.edu/items/confidentiality-integrity-and-availability-the-cia-triad/

[24] U.S. IT Governance. *Cybersecurity Standards and Frameworks*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.itgovernanceusa.com/cybersecurity-standards
[25] Gartner. (2022). *Top Trends in Cybersecurity 2022—Vendor Consolidation*. Accessed: Jun. 25, 2023. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2022-09-12-gartner-survey-shows-seventy-five-percent-of-organizations-are-pursuing-security-vendor-consolidation-in-2022
[26] NIST. *The NIST Cloud Federation Reference Architecture*, Standard NIST-SP 500-332, 2020. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-332.pdf
[27] *Information Security, Cybersecurity and Privacy Protection—Information Security Controls*, Standard ISO/IEC 27002, 2022. [Online]. Available: https://www.iso.org/standard/75652.html
[28] *Industrial Communication Networks—Network and System Security*, Standard ISA/IEC 62443-3-3, 2019. [Online]. Available: https://www.nen.nl/en/nen-en-iec-62443-3-3-2019-en-258484
[29] *Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy*, Standard NIST 800-37r2, 2018. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf
[30] *Information Technology—Security Techniques—Privacy Framework*, Standard ISO/IEC 29100, 2020. [Online]. Available: https://www.sis.se/api/document/preview/80022590
[31] ISACA. (2019). *COBIT—Control Objectives for Information Technology*. COBIT 5 Framework. Accessed: Jul. 1, 2023. [Online]. Available: https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004KoCDEA0
[32] OUSD(A&S) and United States DoD. *Cybersecurity Maturity Model Certification (CMMC 2.0)*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.acq.osd.mil/cmmc/
[33] J. E. Wynn. *Threat Assessment and Remediation Analysis (TARA)*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.mitre.org/news-insights/publication/threat-assessment-and-remediation-analysis-tara
[34] Australian Cybersecurty Center. *IoT Code of Practice: Guidance for Manufacturers*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.cyber.gov.au/acsc/view-all-content/publications/iot-code-practice-guidance-manufacturers
[35] Open Web Application Security Project (OWASP) Foundation. (2021). *OWASP Application Security Verification*. Accessed: Jul. 1, 2023. [Online]. Available: https://github.com/OWASP/ASVS/raw/v4.0.3/4.0/OWASP%20Application%20Security%20Verification%20Standard%204.0.3-en.pdf
[36] *Security and Privacy Controls for Information Systems and Organizations*. Standard NIST.SP.800-53r5, 2020. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf
[37] Verband der Elektrotechnik, Elektronik und Informationstechnik. (2020). *IT-Security for Industrial Automation—Recommendations for the Implementation of Security Properties for Components, Systems, and Equipment*. VDI/VDE 2182. Accessed: Jul. 1, 2023. [Online]. Available: https://www.vdi.de/en/home/vdi-standards/details/vdivde-2182-blatt-4-it-security-for-industrial-automation-recommendations-for-the-implementation-of-security-properties-for-components-systems-and-equipment
[38] The United Nations Economic Commission for Europe (UNECE). (2000). *World Forum for Harmonization of Vehicle Regulations*. UNECE WP29. Accessed: Jul. 1, 2023. [Online]. Available: https://unece.org/transport/vehicle-regulations/world-forum-harmonization-vehicle-regulations-wp29
[39] New Zealand. (2020). *Privacy Act*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html
[40] N. M. Karie, N. M. Sahri, W. Yang, C. Valli, and V. R. Kebande, "A review of security standards and frameworks for IoT-based smart environments," *IEEE Access*, vol. 9, pp. 121975–121995, 2021.
[41] K. M. Caramancion, Y. Li, E. Dubois, and E. S. Jung, "The missing case of disinformation from the cybersecurity risk continuum: A comparative assessment of disinformation with other cyber threats," *Data*, vol. 7, no. 4, p. 49, Apr. 2022.
[42] C. Shearon, "The new standard for cybersecurity," in *Proc. Pan Pacific Microelectron. Symp. (Pan Pacific)*, 2020, pp. 1–9.
[43] P. Wagner, G. Hansch, C. Konrad, K.-H. John, J. Bauer, and J. Franke, "Applicability of security standards for operational technology by SMEs and large enterprises," in *Proc. 25th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, vol. 1, Sep. 2020, pp. 1544–1551.
[44] H. Taherdoost, "Understanding cybersecurity frameworks and information security standards—A review and comprehensive overview," *Electronics*, vol. 11, no. 14, p. 2181, Jul. 2022.
[45] J. Srinivas, A. K. Das, and N. Kumar, "Government regulations in cybersecurity: Framework, standards and recommendations," *Future Gener. Comput. Syst.*, vol. 92, pp. 178–188, Mar. 2019.
[46] ENISA, "Standardization in support of the cybersecurity certification," Eur. Union Agency Cybersecur., Greece, Dec. 2019.
[47] European Commission. *Internal Market, Industry, Entrepreneurship and SMEs: Harmonised Standards*. Accessed: Jul. 1, 2023. [Online]. Available: https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en
[48] T. Carpenter, "9—Electronic publishing standards," in *Academic and Professional Publishing*. U.K.: Chandos Publishing, 2012, pp. 215–241.
[49] CEN-CENELEC. *The European Committee for Standardization and the European Committee for Electrotechnical Standardization*. Accessed: Jun. 25, 2023. [Online]. Available: https://www.cencenelec.eu/
[50] European Commission. *Harmonised Standards*. Accessed: Jul. 7, 2023. [Online]. Available: https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en
[51] European Union. *Official Journal of the European Union (OJEU)*. Accessed: Jul. 6, 2023. [Online]. Available: https://eur-lex.europa.eu/homepage.html
[52] B. Shojaie, H. Federrath, and I. Saberi, "The effects of cultural dimensions on the development of an ISMS based on the ISO 27001," in *Proc. 10th Int. Conf. Availability, Rel. Secur.*, Aug. 2015, pp. 159–167.
[53] M. Siponen and R. Willison, "Information security management standards: Problems and solutions," *Inf. Manag.*, vol. 46, no. 5, pp. 267–270, Jun. 2009.
[54] Center for Internet Security. *CIS Critical Security Controls Version 8*. Accessed on: Jun. 25, 2023. [Online]. Available: https://www.cisecurity.org/controls/v8#v8-mappings
[55] E. T. Feteris, "The pragma-dialectical analysis and evaluation of teleological argumentation in a legal context," *Argumentation*, vol. 22, no. 4, pp. 489–506, Nov. 2008.
[56] O. Pollicino, "Legal reasoning of the court of justice in the context of the principle of equality between judicial activism and self-restraint," *German Law J.*, vol. 5, no. 3, p. 289, 2004. [Online]. Available: http://www.germanlawjournal.com/index.php?pageID=11&artID=402
[57] (2018). *NIST Cybersecurity Framework*. NIST CSF 1.1. Accessed: Mar. 22, 2023. [Online]. Available: https://www.nist.gov/cyberframework/online-learning/five-functions

FATIHA DJEBBAR (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from the University of Quebec, Canada, and the Ph.D. degree in signal and image processing from the University of Bretagne Occidental, Brest, France. She is currently a Senior lecturer with Högskolan Väst, Sweden. Prior to this role, she was a cybersecurity product compliance specialist in Sweden. Her general research interests include network security, the IoT security, information security, digital forensics, and cybersecurity, in particular cybersecurity risk assessment, privacy preserving techniques, and cyber physical system protection.

KIM NORDSTRÖM received the B.Sc. degree in computer science from the Arcada University of Applied Sciences, Helsinki, Finland, the M.Sc. degree in business administration from Åbo Akademi Univerisity, Turku, Finland, and the master's degree in law from the University of Turku, Finland. He is currently a cybersecurity product compliance specialist in Sweden. He holds CISA and CISM CRISC certificates in cybersecurity.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 October 2022, accepted 16 November 2022, date of publication 18 November 2022,
date of current version 28 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223440

RESEARCH ARTICLE

CyberTOMP: A Novel Systematic Framework to Manage Asset-Focused Cybersecurity From Tactical and Operational Levels

MANUEL DOMÍNGUEZ-DORADO[1], JAVIER CARMONA-MURILLO[2], DAVID CORTÉS-POLO[3], AND FRANCISCO J. RODRÍGUEZ-PÉREZ[2]

[1]Department of Information Systems and Digital Toolkit, Public Business Entity Red.es., 28020 Madrid, Spain
[2]Department of Computing and Telematics Engineering, Universidad de Extremadura, 10003 Cáceres, Spain
[3]Department of Signal Theory and Communications and Telematics Systems and Computing, Rey Juan Carlos University, Móstoles, 28933 Madrid, Spain

Corresponding author: Manuel Domínguez-Dorado (manuel.dominguez@red.es)

This work was supported in part by Project TED2021-131699B-I00 and Project MCIN/AEI/10.13039/501100011033; in part by the European Union NextGenerationEU''/The Recovery, Transformation and Resilience Plan (PRTR); and in part by the Regional Government of Extremadura, Spain, under Grant GR21097.

ABSTRACT Currently different reference models are used to manage cybersecurity, although practically none are applicable ''as is'' to lower levels as they do not detail specific procedural aspects for them. However, they urge organizations to develop a methodological foundation to manage cybersecurity at those levels. Although they allow organizations to adhere to a recognized standard at the strategic level, this advantage vanishes when organizations must define specific low-level procedures, allowing the appearance of inconsistency at tactical and operational levels between departments of the same organization or between organizations. The design of these elements with the required holism and homogeneity is difficult, and this is why generic processes focused on getting certified regarding a standard are usually originated, but they are insufficient to obtain effective cybersecurity because they are not focused on dealing with real cyber threats. Because of the great responsibility of lower levels to achieve effective cybersecurity, this lack of methodological definition makes it difficult to adapt cybersecurity to the highly dynamic cyber context with the required holism and strategic alignment. Our proposal provides CyberTOMP, a process for managing cybersecurity at lower levels, as well as a set of methodological elements that support it. The novelty of these contributions is that they complement the strategic standard selected by the organization, providing it with a set of procedural elements ready to be used out of the box, contributing those aspects required by high-level frameworks to manage cybersecurity at lower levels, for which there is no alternative with a managerial approach.

INDEX TERMS Business asset, cybersecurity management, cybersecurity metrics, cyber threats, CyberTOMP, holistic cybersecurity, strategic alignment, tactical and operational cybersecurity, unity of action.

I. INTRODUCTION

Currently, various approaches to the security aspects of the digital world coexist. These strategies correspond to different organizations' digital evolution stages from decades ago to the present. Over time, the organizations' degree of digitization has increased, causing their most relevant assets at those moments to have been affected by a different threat context and, therefore, have required a specific risk analysis and a particular way of dealing with them. Depending on the specific stage, we can use an information technologies *(IT)* security approach [1], [2], an information security approach [3], [4], [5] or a cybersecurity approach [6], [7] among the main ones.

A. EVOLUTION TOWARDS A CYBERSECURITY APPROACH

Around the decades of the fifties and sixties, under an IT security approach, the most important organizations asset was the technology itself; this was a time when the cost of the first

The associate editor coordinating the review of this manuscript and approving it for publication was Sedat Akleylek.

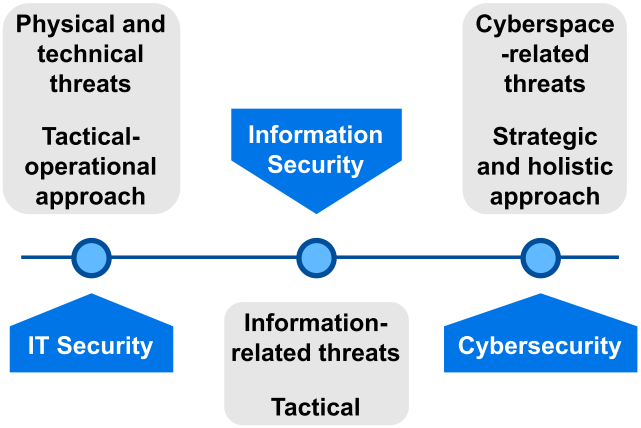


FIGURE 1. From IT security to Cybersecurity. Moving from a single-departmental approach to an organization-wide approach.

FIGURE 2. Cybersecurity checkpoints agenda at different levels during a four-years strategy. The tactical and operational levels must deal with the greatest variations of the cyber threats context. These variations are often hidden to higher levels due to the observation of variables that do not correctly reflect variations in the short and medium term.

mainframes constituted a large investment. The associated risks were mainly circumscribed to the technical and physical spheres and were addressed by most technical departments within the organizations. As information systems evolved, the value provided by the information increased, transforming it into a highly valued asset and forcing organizations to adapt their strategies towards an information security approach. Different departments that owned that information began to be involved in managing and handling the risks associated with it. They started to understand the threats that could affect the information and, by extension, the normal development of their own activities.

This paradigm has been prevailing for many years and is still used as the main approach in many organizations today. However, with the irruption of cyberspace, the information security approach has become insufficient. Cyberspace, understood as a set of interconnected information systems through communication networks in which people and entities interact and accomplish their activities, has unique characteristics: high dynamism; it is a common playing field where each organization controls only part of it; it has a high dependency on third parties; it requires the focus to be placed not so much or not only on information, but also on the continuity of business processes/assets; there is a need for cyber resilience, etc.

Parallel to the massive adoption of cyberspace, a set of specific threats has emerged that can potentially affect the capability of organizations to develop their activities, interact with third parties, and even preserve their image, reputation, and the trust vested in them. To deal with this evolution (fig. 1), with an increasing cyber threat context, the only approach to properly manage the current cyber risks and cyber threats is cybersecurity, mistakenly understood as information security synonymous on many occasions [8], [9]. This is not only because of cyberspace features but also because the greater digital dependency of organizations on cyberspace has brought to light new vital organizational assets, affected by cyber threats, which cannot be analyzed easily by employing an information security approach [10]: reputation, trust placed by third parties, people's physical integrity, supply chains, the organization's capabilities, Internet of Things *(IoT)* specific threats [11], etc.

Cybersecurity requires unity of action from the whole organization, leadership from strategic levels [12] and a high degree of holism [13], from its conception to its practical application, focusing on business assets [14]. It demands a proactive attitude that takes into account the response and recovery from cyber incidents as well as business continuity [15], aspects that must be managed throughout the entire life cycle, carefully considering the critical success factors to achieve effective cybersecurity [16].

B. RESPONSIBILITY OF TACTICAL AND OPERATIONAL LEVELS IN CYBERSECURITY

The main standards and reference models used for cybersecurity provide guidelines for its evaluation, although this is a high-level evaluation. This implies that variations in the state of cybersecurity can only be measured at the strategic level in the medium/long term. In scopes other than cybersecurity, assessing within such periodicity might be acceptable if the context is not very changing and significant corrective or adaptive actions are not frequently required. Under these circumstances, high-level assessments and corrections may be sufficient to maintain the state of the organization aligned with strategic goals.

However, this does not occur in the field of cybersecurity. Cyberspace and its associated cyber threat context evolve very dynamically, intensely, and frequently. For this reason, most corrective or adaptive actions, as well as the measurement of their effects, must be carried out in the medium/short term, that is, at tactical and operational levels within the organization. Thus, a large part of the responsibility for preserving the cybersecurity state aligned with an organization's cybersecurity strategy falls on them, who are also responsible for maintaining the unity of action and the holistic approach required by cybersecurity. Accomplishing these requirements from lower levels that are distributed

throughout the organization in several departments and areas that usually operate as silos and have different chains of command is very difficult.

Regrettably, the aforementioned standards and frameworks do not supply these levels, out of the box, with detailed methodological elements to help them manage and evaluate cybersecurity; neither do they provide standardized mechanisms to maintain the strategic alignment nor to quickly detect new cyber threats and nimbly apply the necessary actions to deal with them (fig. 2). Consequently, it cannot be taken for granted that these levels have the necessary mechanisms to carry out this work for the mere fact that the organization has adhered to a high-level standard in the strategic sphere.

C. CONTRIBUTIONS OF OUR WORK

From the current state-of-the-art, which we detail in later sections, needs are identified in the frameworks commonly used to manage cybersecurity. They are defined at a strategic, level and almost all urge organizations to develop a methodological base to be used in cybersecurity management at lower levels so that the cybersecurity strategy can be broken down and transferred correctly to the whole organization. As explained in the previous paragraphs, and we will expand on it in the article, we understand that the responsibility of these levels in the management of cybersecurity is relevant, but it encounters a series of challenges derived, on the one hand, from these aspects not covered by high-level frameworks and on the other hand by the structural rigidity of many organizations. Using any of the existing high-level frameworks, organizations can adhere to a widely recognized standard at the strategic level. But by having to define their own cybersecurity management process and procedures for the lower levels of the organization, this advantage, in a way, vanishes, inducing inconsistency between different organizations or even within different departments and functional areas of the same organization at tactical and operational levels.

Defining these elements is not always simple; it is almost never homogeneous and seldom consider cyber threats, but simply organizational aspects. On more occasions than is recommended, the difficulty in developing methodological elements for the tactical and operational levels leads to generic processes and procedures that are sufficient to obtain a certification with respect to the selected strategic framework, but insufficient to obtain effective cybersecurity.

Our work provides CyberTOMP as a means of managing cybersecurity at the tactical and operational levels, as well as a set of methodological elements, knowledge bases and concepts on which it is based. They are designed to complement the standard selected by the organization in the strategic sphere, providing it with a set of processes and procedures ready to be used out of the box. They contribute aspects required by the methodological guidelines of the high-level framework and by the organization to manage cybersecurity at tactical and operational level, levels for which there is no alternative with a managerial approach. Our proposal constitutes a procedural and methodological solution and not a technical one. Specifically, our proposal supplies lower levels with:

- Mechanisms to manage cybersecurity at tactical and operational levels, regardless of the higher-level standard or framework adopted by the organization, are thus a complement and not a disruptive element.
- A set of techniques and metrics focused on business assets to quantitatively and homogeneously assess cybersecurity, at different levels and degrees of aggregation.
- A homogeneous set of expected cybersecurity outcomes that arises from the analysis and combination of well-recognized international sources.
- The capability to maintain alignment with the cybersecurity strategy, under a holistic approach, from the tactical and operational levels, engaging all functional areas involved in the process.
- Procedures to incorporate the dynamic variations of the real cyber threats context, in an agile way, into cybersecurity daily grinds.

D. ORGANIZATION OF THIS DOCUMENT

The remainder of this work is organized as follows: in section II, the aspects found in the current state of the art that must be overcome to achieve effective cybersecurity management at low levels of the organization, are identified; in section III the methodological elements, knowledge bases and concepts developed in our proposal as support for the practical application of cybersecurity management at tactical and operational levels, are described; the section IV defines and describes in detail the CyberTOMP, our core contribution that, based on the rest of the elements detailed in section III, allows the organization to manage cybersecurity at tactical and operational levels; in this section recommendations and guidelines for its practical application are proposed as well.

II. STATE OF THE ART AND PROBLEM STATEMENT

From a theoretical perspective, the adoption of a cybersecurity approach does not have apparent complexity. However, based on the current standards commonly used for cybersecurity at a strategic level, there are different aspects that hinder its practical adoption in organizations when it is applied from lower levels, especially considering the differentiating characteristics of cybersecurity with respect to previous approaches and the need to change the way it is addressed [17]. In the following subsections we identify the current problems that our proposal addresses.

A. LACK OF HIGH-LEVEL STANDARDS THAT PROVIDE PROCEDURAL ELEMENTS FOR TACTICAL AND OPERATIONAL LEVELS

There are many frameworks and standards that can be useful, in certain cases, to manage cybersecurity [18], which sometimes makes it difficult to choose one and implement it in organizations [19]. A large number of them, such

as Capability Maturiry Model Integration *(CMMI)* [20], [21], [22] or Information Technology Infrastructure Library *(ITIL)* [23], [24] are generic and applicable to multiple spheres. When applied to cybersecurity, they can contribute to managing it. Some even contain elements related to security in the digital field [25]. However, they are, in no case, specific models for cybersecurity, so their advantages are very limited in this regard [26], in addition to being defined at a very high level [27].

Other frameworks and standards are focused on information security management, not on cybersecurity, for instance, the ISO 27000 family of standards [28], [29], the Model of Indicators for the Improvement of Cyber Resilience *(IMC)* [30], [31] or even the Spanish National Security Scheme *(ENS)* [32], [33], [34], [35]. They are commonly used to address cybersecurity, although they are based on or bear a clear perspective of information security and do not properly cover the specific aspects of the cybernetic context; therefore, they do not allow, *per se*, meeting the requirements of a cybersecurity model.

To conclude, there are other works, such as the one developed by MITRE in the Adversarial Tactics, Techniques and Common Knowledge matrix *(ATT&CK©)* [36], [37] (used in various works on threat intelligence [38], [39]), the Critical Security Controls for Effective Cyber Defense *(CSC)* [40], [41] from the Center for Internet Security *(CIS)*, even with its shortcomings [42], the Open Web Application Security Project *(OWASP)* Top 10 project [43], [44], the Community Defense Model *(CDM)* [45] from the CIS, that aligns the CSC to cover the threats documented by MITRE, helping to implement the mitigations that it proposes [46] or those known as nine D's of cybersecurity described in [47] (so called because they are recommendations that all begin with this letter). All of them are sets of recommendations, good practices and specific tools for cybersecurity, which are very useful but disconnected from a comprehensive framework that covers all organizations' levels.

Among the analyzed models, the Framework for Improving Critical Infrastructure Cybersecurity [48], [49], from the National Institute of Standards and Technology *(NIST)* stands out. It is a complete framework for cybersecurity that is accompanied by the SP-800 series of guides [50] (where guide SP-800-53 [51] can be especially highlighted), which provides the organization with high levels of cyber resilience under a cybersecurity approach. This framework in conjunction with the Cybersecurity Maturity Model *(CMM)* [52], [53] also allows the evaluation of third parties that must be part of the organization's supply chain. There are other less common models as, for example, the one developed in [54], [55] which focuses on the managerial aspects of cybersecurity to protect critical infrastructure. It is defined at a very high level of abstraction and does not provide procedural elements for direct application. However, it provides a modern view that cybersecurity is not only related to technical domains but also involves the whole organization.

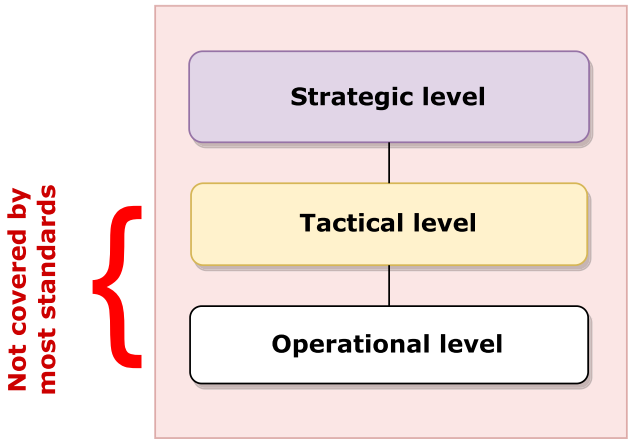


FIGURE 3. It is necessary to provide the tactical and operational levels with homogeneous methodological tools for cybersecurity management.

There are published works that focus on cybersecurity very applied to specific and particular cases. A deeper literature review and an analysis of the body of knowledge in the field of cybersecurity can be found in [56], [57], [58], [59], [60], and [61], for general cases and also specific ones. They generally follow technical approaches that do not address organizational cybersecurity from a procedural perspective. But it is also important to study the problem from the managerial point of view within the current standards and new contributions such as the one we will describe in this paper.

Nevertheless, none of these frameworks or initiatives, and even the NIST framework, includes a detailed methodological description of how cybersecurity should be managed at the organization's tactical/operational levels. This means that none of them are applicable without being complemented, since cybersecurity must be administered on many occasions from these levels (fig. 3). It is the responsibility of each organization to design the set of processes and procedures indicated by these frameworks for their lower levels.

By not including specific standardized guidelines, the tactical/operational application of these models can be completely different between organizations, between areas within the same organization, or it cannot even take place.

There are several factors why an organization could choose to use them even though they are not fully defined options to address cybersecurity at all levels of the organization: because they are certifiable standards that allow positioning against competitors, because they are widespread and finding workers trained in them is easier, because they are required by third parties to access contracts, or because they are mandatory rules according to the legal framework surrounding the organization. For these reasons, replacing these frameworks in the organization is not always an option, but they should be complemented to provide them with what they lack. They should be provided with methodological elements that apply at the

lowest levels to address the deficiencies in this area. Hence, it is necessary to provide tactical and operational levels with homogeneous cybersecurity management mechanisms that allow them to adapt to the cyber threat context and maintain alignment with the strategic cybersecurity objectives.

In [62], a use case in Portugal for the implementation of information security actions in a group of SMEs was explained in detail. Some aspects of this work are similar to those adopted in our proposal: a set of information security controls from a recognized standard, which have been grouped into different groups of controls to respond to different needs. Subsequently, the characterization of each control depends on the type of organization and other aspects.

However, this very well-prepared work has, in our opinion, some limitations. It is based on the ISO 27001 standard, a standard for information security and not for cybersecurity. At the procedural level, it does not detail the elements of management, processes and procedures used at tactical and operational levels to coordinate the efforts of the organization's workforce. This is most likely because their destination is small and medium-sized companies, where this distinction between levels makes perhaps less sense.

Paraphrasing the conclusions of the authors of this work: *However, ISO-27001:2013 is a single tool for achieving the project goal and it can be seen as a limitation in this study. In that sense, other best practices and frameworks should be addressed, implemented, and compared.*

In our work, we present a wider solution based on several standards and initiatives specific to cybersecurity and not information security. It also contributes the required processes, procedures and metrics to be used out of the box that can be applied to tactical and operational levels.

B. LACK OF MECHANISMS TO PROVIDE HOLISM FROM LOWER LEVELS

Cybersecurity requires something that, until now, none of the previous approaches related to digital security required [63]: a holistic approach, promotion from the strategic levels to the whole organization, unity of action to address cybersecurity risks, and proactive mindset and focus on cyber incident response and recovery tasks.

Since a large part of the initiative in cybersecurity must be driven at tactical and operational levels, the interdepartmental coordination required to provide a holistic approach must also be addressed from these levels.

Notwithstanding, the areas or units that compose these levels do not have direct visibility, communication, and coordination between them, and usually work under different chains of command in isolated silos. Habitual conflict escalation mechanisms are useful for inter-area communication in specific situations, but not for managing the daily grinds at lower levels. Under these circumstances, it is difficult for lower levels to achieve the coordination, unity of action, and holistic and proactive vision required by cybersecurity (fig. 4).

This situation is amplified when the organization is more distributed in silos. In any event, this communication is

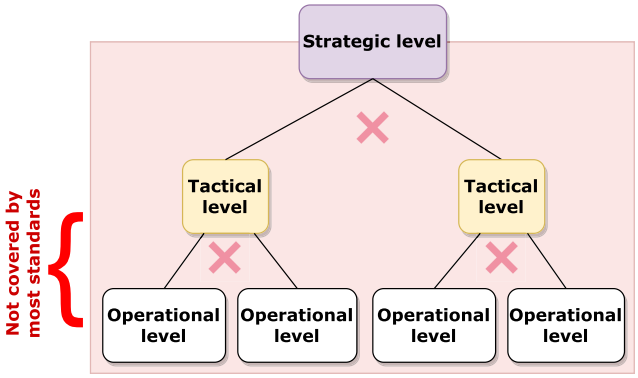


FIGURE 4. The distribution of the organization in silos hinders a fluent communication and collaboration between functional units and the achievement of the holism and unity of action required by a cybersecurity approach.

fundamental because people from different functional areas of the organization must agree on the actions they have to implement, on the metrics that will affect them, on the weight and responsibility that each one will have with respect to the cybersecurity of business assets, and so on. This should not be done independently but jointly, coordinated, taking advantage of existing synergies and forming a team.

For these reasons, it is necessary to provide these levels with tools that ensure that they can design and execute joint cybersecurity actions proactively, quickly, with holistic vision and unity of action; avoiding the appearance of conflicts despite the distribution of teammates among several functional areas.

C. LACK OF HOMOGENEOUS CYBERSECURITY EVALUATION CRITERIA

What has not been measured cannot be improved. This statement, extrapolated to cybersecurity, implies the need to evaluate the effectiveness of cybersecurity controls [64] and safeguards, from a holistic and multidisciplinary perspective, and offer a shared vision of the organization's cybersecurity posture.

When people from different functional areas collaborate to ensure the cybersecurity status of business assets and meet strategic cybersecurity objectives, there is a need to measure progress [65] because this allows continuous decision-making at different levels [66], [67]. But current standards and frameworks define neither measurement mechanisms nor assessment criteria that can be used by tactical and operational levels to fit this need, aspects with which all the parties should agree, and that allow focusing on solutions and not on resolving the differences around the assessment process itself. Otherwise, several discrepancies and conflicts will tend to arise between the areas co-responsible for cybersecurity, which prevents having a clear vision of their real cybersecurity state.

When different organization units, follow non-identical assessment criteria to evaluate the same element

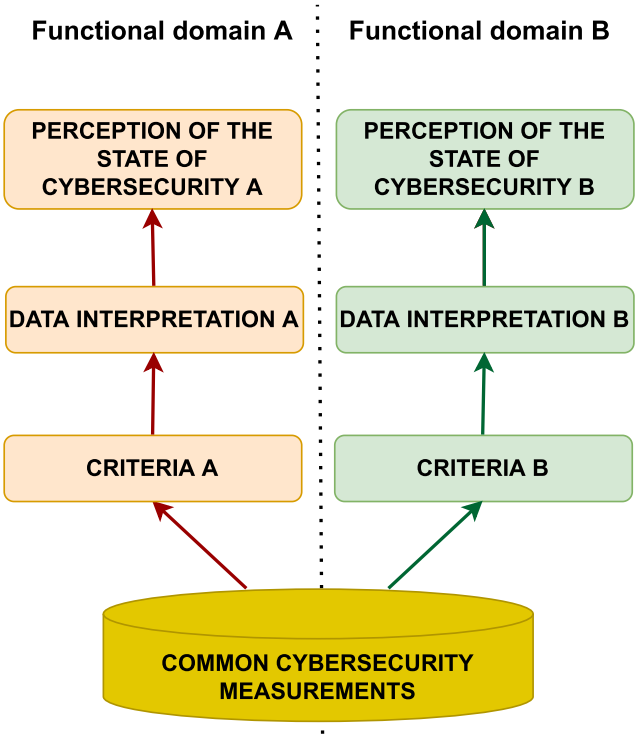


FIGURE 5. Silos in organizations frequently imply the existence of different criteria and disjointed interpretations of the real state of cybersecurity, even when the same data is valued. A common standard should be defined for the evaluation of cybersecurity at these levels.

(cybersecurity in this case), it is likely that none of these evaluations coincide with the rest (fig. 5) unless they share a common vision, which is a common way of interpreting the measurements, leading to a lack of coordination in cybersecurity due to different perceptions. For these reasons, it is necessary to have standardized and homogeneous tools that provide a common shared measurement of the performance and state of cybersecurity at these levels, and also allow quantitative evaluation of the effectiveness of the implemented actions for decision-making in the short and middle terms.

III. TOOLKIT TO SUPPORT CYBERSECURITY MANAGEMENT FROM TACTICAL-OPERATIONAL LEVELS

After a review of models and initiatives commonly used to manage cybersecurity, we designed a proposal that combines the existing elements that may be useful for the purpose of our work with other specific elements designed in our study that complete it to address all the needs identified in Section II. We have always tried that our solution consists of an evolution or a combination of fundamentals already consolidated and accepted, and not of a theoretically excellent proposal but difficult to run in practice by any organization. In addition, special emphasis has been placed on keeping the solution limited to management at lower levels (tactical/operational), assuming that the organization will have specific frameworks for managing at higher levels (strategic/tactical), although

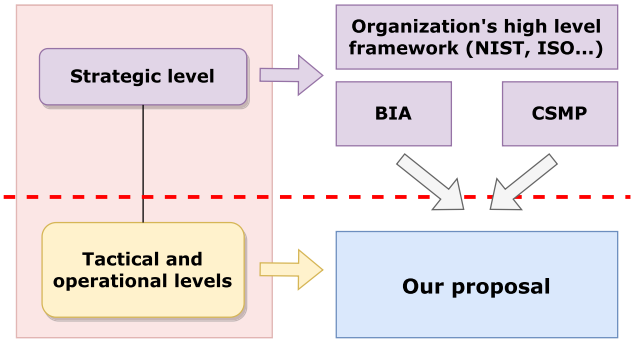


FIGURE 6. BIA and CSMP, both slightly modified, connect the organization's strategic framework to our proposal for tactical and operational levels.

perhaps they may not be appropriate "as is" for cybersecurity management, as explained in Section II.

In the following paragraphs, every decision and auxiliary solution that makes up our proposal will be discussed, justifying the reasons for it.

A. CONNECTING OUR PROPOSAL WITH THE CORPORATE STRATEGY

In our proposal, we chose to minimize the dependence on the high-level framework used at the strategic level to ensure its applicability in different organizations while guaranteeing that it serves as a cybersecurity management tool at tactical and operational levels of the organization and maintain alignment with the corporate strategy from these levels. However, a method is needed to connect and align the activity of lower levels towards the strategy. For this, we propose to use two elements present in almost any medium-sized organization, regardless of the regulatory framework to which they have adhered: the Business Impact Analysis *(BIA)* and the Cybersecurity Master Plan *(CSMP)*, or the set of cybersecurity projects, if applicable, that come from the application of the framework used at strategic levels (fig. 6).

1) BIA REQUIREMENTS FOR ASSET FOCUS AND BUSINESS CONTINUITY

The concept of business continuity refers to the ability of an organization to identify threats that can become disruptive events that affect its activity, and plan the response and recovery in advance to guarantee the normal development of business activities [68], [69]. The greater this capacity, the more resilient is the company.

It is not a new concept, nor is it solely focused on cybersecurity. An entity could be affected by multiple events; some recent events such as the lock-down suffered by the COVID-19 pandemic, but also natural disasters, labor conflicts, lack of qualified workers, events linked to information security, or cybersecurity incidents.

The requirements for cybersecurity are in many ways similar to the requirements for ensuring business continuity: holistic view; impulse from the strategic level to the entire

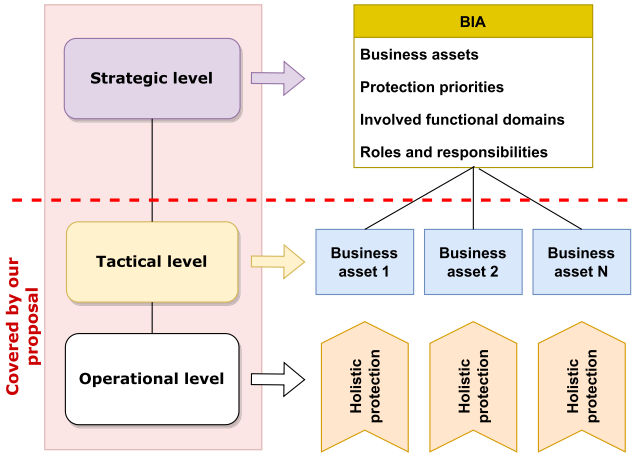


FIGURE 7. Using the BIA to connect the strategic level to the lower ones provides this proposal with the capability of integrating cybersecurity-related business continuity requirements and a focus on the business assets in the daily cybersecurity grinds.

organization; unity of action in crisis management; proactive approach; development of plans to respond and recover in the face of different situations and actions that reduce the impact when crises break out. Therefore, with organizations making massive use of cyberspace and with a great dependence on this medium, cybersecurity, correctly put into practice, contributes significantly to business continuity in crisis situations caused by cybersecurity incidents [70].

In their business continuity management, it is common for organizations to carry out the BIA [71], generating a document in which the organization details aspects such as the critical business processes, the assets on which these processes depend, the criticality of each one, the maximum tolerable interruption times, or the tolerable recovery times. The BIA is, therefore, a strategic declaration of intent coming from the highest level of the organization, where it is evaluated and indicated which assets to protect (and recover, where appropriate) and with what intensity, to ensure that the impact of a crisis on the overall business is as small as possible. It is also common for BIA to define roles, responsibilities, strategies, communication mechanisms, etc. for all areas, and for cybersecurity.

Our proposal provides mechanisms that allow organizations to align cybersecurity with business continuity requirements, as the maximum expression of the organization's survival needs. In particular, at tactical and operational levels, which are often the executors of recovery actions. However, business continuity associated with cybersecurity, expressed as a whole, is difficult to understand at operational and tactical levels. It is too broad and difficult to manage and, therefore, difficult to understand, communicate, and plan at those levels. For this reason, the first decision in our proposal is the application of the "divide and conquer" paradigm to have a smaller and more manageable scope at such levels. In addition, it is more understandable, allowing greater cohesion between the multidisciplinary and holistic operational team in charge of its cybersecurity and continuity.

Since the BIA identifies and prioritizes the business assets that support the organization's activity, we propose focusing cybersecurity efforts on them [72] and assign them as a basic unit at the tactical and operational levels for their cyber protection, understanding that this element is sufficiently manageable at these levels.

Each organization develops a BIA according to its needs, although it is common for a BIA to include information relevant to the business. Nevertheless, to provide it with the utility intended in this work, the BIA must include at least:

- Identification of business assets.
- Functional areas responsible for business assets and those that depend on their results.
- Continuity strategies for different crisis scenarios.
- The parameters in which business assets can be discontinued without generating a disproportionate impact, and therefore, the levels of this discontinuity acceptable to the organization.
- The impact on the business in the event of a discontinuity that extends beyond the parameters considered acceptable by the organization.
- A map of high-level dependencies between the different business assets.
- Based on the above, prioritization that reflects the protection required by business assets. On a scale of three values, LOW, MEDIUM, and HIGH.

In this way in our proposal, the BIA becomes one of the two points of interconnection between the strategic area of the organization and the rest of the lower levels (fig. 7). This provides the following four main strengths for cybersecurity:

- This allows for a more manageable and understandable scope for lower levels of the organization.
- Allows maintaining the focus on the business asset and its derivative assets.
- It allows the integration of business continuity strategies related to cybersecurity in daily activity.
- It allows the incorporation of the risk-based approach (related to business continuity) [73], [74] so that business cyber continuity risk requirements can be introduced in the tactical and operational cybersecurity management cycle.

2) CSMP REQUIREMENTS FOR A STRATEGIC ALIGNMENT

CSMP is a tool commonly used by cybersecurity managers to orchestrate all the needs and context of cybersecurity in a portfolio of cybersecurity programs and projects aligned with the needs of the organization. In this way, the cybersecurity effort and the necessary budget are focused on achieving the organization's strategic cybersecurity objectives and, by extension, the company's business goals.

The design of CSMP includes systematic phases so that it covers all aspects of cybersecurity in an integral way, which allows focusing and optimizing resources to achieve the interests of the company in this area. It includes, among many other aspects, cybersecurity guidelines; strategic

cybersecurity objectives; the definition of high-level cybersecurity controls and safeguards; the definition of cybersecurity architecture, covering all areas where cybersecurity is applicable; the definition of roles, responsibilities, processes, and procedures; the quantification of expenses and investments in cybersecurity, and the high-level planning of cybersecurity actions/projects. This allows an incremental development of the cybersecurity strategy and the achievement of short, medium and long-term goals. From all of the above, which represents a high-level comprehensive plan for cybersecurity management throughout the organization, we would like to emphasize that it is in this CSMP that the framework and regulatory framework related to cybersecurity are defined and the cybersecurity projects required by the organization, as well as the strategic cybersecurity objectives and the specific objectives of each designed project.

Theoretically, CSMP is an optimal tool for providing cybersecurity with a comprehensive vision. However, and this is relevant, during the preparation of this plan, the strategic framework that the organization will use for the direction and management of cybersecurity must be defined, as well as the associated processes and procedures. But if the execution of the CSMP depends on any of the main existing frameworks ''as is'', the problem described in the section II resurfaces, since practically all of the high-level frameworks and standards do not provide methodological tools applicable to tactical and operational levels and focus mainly on the strategic levels; so that even with a CSMP, organizations must develop their processes and procedures to manage cybersecurity at the tactical and operational level. Most of these high-level frameworks indicate that this methodological base should be developed. And this is precisely what our proposal provides. Our proposal can be used to complete the methodological guidelines of high-level frameworks and can be included in the CSMP to be used in cybersecurity management at the tactical and operational levels of the organization.

In our solution, the use of CSMP is proposed as a second point of connection with the strategic level of the organization (fig. 8). To do this, CSMP projects, or cybersecurity projects in the event that there is no properly defined CSMP, must meet certain requirements:

- Every business assets must have their own project in the CSMP. A project may cover more than one asset if its cybersecurity objectives coincide with others.
- These projects must be defined at a high level and specify the objective, but not detail the tactical/operational actions, so that rolling wave planning can be carried out [79] at lower levels as information from the context analysis becomes available. The planning of CSMP projects is therefore simplified.
- The objectives of the indicated projects must be defined based on the cybersecurity metrics and indicators described in our proposal, as developed later in this section.

Building the CSMP as described in our proposal provides four main benefits:

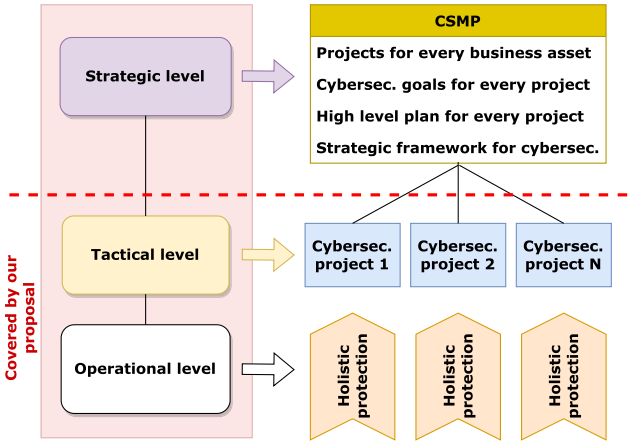


FIGURE 8. Using the CSMP to connect the strategic level to the lower ones provides this proposal with the capability of integrating cybersecurity risks and cybersecurity strategic goals in low levels' activities.

- It allows for more manageable and understandable cybersecurity projects for lower levels of the organization.
- Allows maintaining focus on strategic objectives for business assets and their derivative assets.
- It allows alignment towards the cybersecurity strategy in the daily activity of its management from the lower levels.
- It allows the incorporation of the risk-based approach (related to cybersecurity) [75], [76], [77], [78], so that cybersecurity risks requirements can be introduced in the tactical and operational cybersecurity management cycle.

B. CYBERSECURITY FUNCTIONS FOR BUSINESS ASSETS

With the use of BIA and CSMP as described in our proposal, a multidisciplinary operational team in charge of the cybersecurity of a certain business asset would have a manageable scope. Even so, in our work we propose to make this scope even more manageable to further increase its understanding and facilitate the evaluation of its cybersecurity state. Among the frameworks reviewed in Section II, the most complete and focused on cybersecurity is the NIST cybersecurity framework, which organizes different cybersecurity safeguards in a tree-like manner, very useful, in continuous security functions, categories, and subcategories. The functions provide a high-level strategic view of the cybersecurity risk management process life cycle and their subsequent breakdown into categories, and sub-categories brings this strategic view closer to the tactical and operational levels:

1) **Identify**. This function enables a greater understanding of organization's context to focus and prioritize its efforts in accordance with the risk management strategy and its needs.
2) **Protect**. The purpose is to develop and implement appropriate safeguards and controls to ensure the delivery of critical services. This is the basis for the

subsequent limitation or containment of the impact of a possible cybersecurity incident.

3) **Detect**. The purpose is to develop and implement appropriate activities to identify the occurrence of a cybersecurity event.
4) **Respond**. The purpose purpose is to develop and implement appropriate activities to take action regarding a detected cybersecurity incident. It allows, among other aspects, containing the impact of cybersecurity incidents.
5) **Recover**. Its purpose is to develop and implement appropriate activities to maintain resilience plans and recover any capacity or service affected by a cybersecurity incident. Allows the recovery of the usual activities of the organization.

This functional classification is easily understandable and, following it, a tactical/operational team could focus on different aspects of the cybersecurity of the business asset, which could also be evaluated separately. The identification of specific responsibilities of each functional area of cybersecurity is facilitated and favors the creation of specialized operational subgroups in each of the functions, categories or subcategories. In addition, the ''Response'' and ''Recovery'' functions are closely linked to business continuity and cyber resilience, so they fit very well in cybersecurity focused on business assets from the BIA, as indicated in our proposal.

The subcategories (expected outcomes) and categories defined within the NIST framework [48] contribute hierarchically to the achievement of the objectives of each function on which they depend. Each is traceable to the most relevant regulatory frameworks and initiatives, such as CIS CSC, NIST SP 800-53, ISO 27001, which facilitates coexistence with these standards.

Therefore, we have considered it convenient to reuse this classification in functions, categories, and subcategories in our proposal. The NIST framework will not be used in most strategic aspects in order four our proposal to remain independent of the higher level regulatory framework used in the organization: NIST, CMMI, ISO 27001, ENS, etc.

In the rest of our proposal, it is considered that any activity carried out by tactical and operational teams for the cybersecurity of a business asset must be included in one of the defined cybersecurity functions or in its derived hierarchy.

C. UNIFIED LIST OF EXPECTED OUTCOMES FOR THE CYBERSECURITY OF BUSINESS ASSETS

The finest grain level of the NIST classification is a subcategory. In that model they are also called ''expected outcomes'' which is very appropriate because it reflects that these subcategories are the goals, which are achieved with the operational implementation of the corresponding controls and safeguards. In our proposal, we reuse the NIST definition of ''expected outcomes'' since implicitly this denomination is a proactive requirement for the teams in charge of executing cybersecurity actions, an aspect that we consider essential for modern cybersecurity.

However, the expected outcomes from the NIST framework are not the only source of relevant information clearly focused on cybersecurity, and being a fairly broad set, it is true that it is not updated very frequently. There are other sources that are either updated more frequently or simply supplement NIST's set of expected outcomes. For example, in [36], MITRE identifies cyberattacks observed in the real world and the tactics, techniques, and procedures followed by cyber attackers to carry them out: the *modus operandi*. The main mitigation actions for each case are solso defined. In [40], the CIS details the most critical cybersecurity controls that should be implemented in any organization. For this, it uses what it calls the ''Implementation Group'' (IG), numbered from 1 to 3. IGs are a way to identify groups of controls that need to be implemented together to address existing threats. IG1 controls, once implemented, allow for dealing with a wide variety of cyber threats. The IG2 controls include those from IG1, and the IG3 controls include all. Consequently, depending on the context of the organization and the protection needs it requires, it must implement IG1, IG2, or IG3 controls. IG3 is the most complete and allows for a higher level of cybersecurity against the most complex threats (it also includes the most complex and costly controls). The CIS itself, in [45], calculates the level of coverage of the threats identified by MITRE after the implementation of the different IGs, ranging from 77% of threats in the worst case by implementing IG1 to 95% in the best case, implementing IG3; a relevant coverage in any of the cases. Finally, in [47], a series of recommendations are defined, which are applicable to any cybersecurity scenario and can be very useful for minimizing exposure to cyber threats: the nine D's of cybersecurity.

As expected outcomes will determine what cybersecurity actions operational teams need to take, we consider it essential in our proposal to have an expanded list of expected outcomes that brings together not only information from the NIST framework but also from the cited sources. That is why we have approached this task by thoroughly analyzing these sources and integrating them into a Unified List of Expected Outcomes (ULEO) that:

- Retains the same classification of functions, categories, and subcategories as NIST.
- Groups the expected outcomes in the same implementation groups defined by the CIS, with the same meaning.
- Expands the focus and number of original expected outcomes from the NIST model, including inputs from other complementary or more up-to-date sources.
- Maintains alignment with the work of MITRE, so that the application of each IG allows addressing a certain percentage of cyber threats observed in the real world.

When building the ULEO we have been especially careful in the process of integrating controls from other cybersecurity initiatives, to ensure that this range of threat coverage is not altered downwards. In all cases, stricter controls than those proposed by the NIST have been added or replaced by more extensive controls, but in no case the controls were

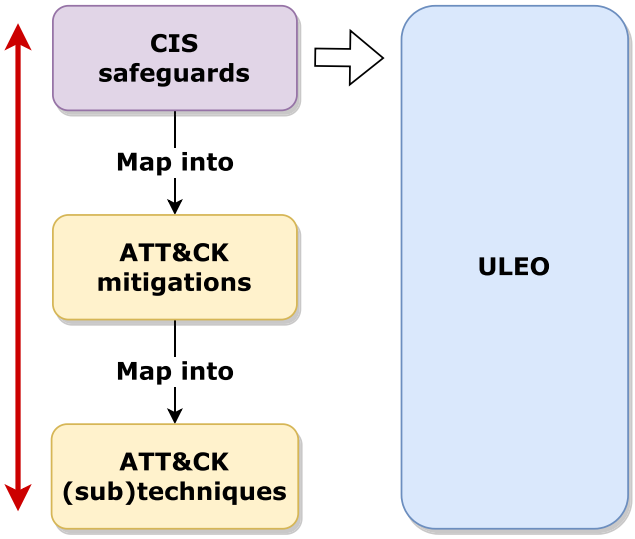


FIGURE 9. Our proposal indirectly incorporates the mitigations and TTPs of MITRE to the ULEO through the inclusion of the corresponding CIS safeguards.

relaxed, which is the reason why these ranges of coverage can be ensured. Therefore, the proposed method maintains or improves the coverage percentages calculated by the CIS in [45].

The following subsections define ULEO and describe the process followed for its analysis and construction.

1) PHASE I. FUSION OF MITRE RECOMMENDATIONS WITH CIS CONTROLS AND NIST SUBCATEGORIES. CREATION OF INITIAL ULEO

The starting point for the construction of ULEO in our proposal is the complete set of functions, categories, and subcategories defined in the NIST framework.

Our proposal does not directly include the mitigations identified by MITRE to address the cyberattacks documented in the ATT&CK©matrix. In [45], the CIS does an excellent job analyzing in depth which of its controls and safeguards allow the implementation of the necessary mitigations to face the Tactics, Techniques and Procedures *(TTPs)* employed in the cyberattacks documented by MITRE. These requirements were grouped into each of the three IGs used in our study. Thus, in our proposal we take advantage of this effort by including the CSCs from CIS which also allows us to indirectly include the needs and requirements identified by MITRE (fig. 9).

In [80], the CIS performed a comparative analysis of the equivalence between the expected outcomes from NIST and CIS CSCs. In our proposal we have taken this initial comparative analysis as a basis, which does not merge elements but rather identifies them, to make the first combination of the expected outcomes of the NIST and CIS CSCs, as follows:

1) Cases where a CIS control or safeguard does not have a related NIST subcategory. In this case, we have that control or safeguard to the list, considering that it complements the NIST model itself, covering cases that it did not consider.
2) Cases where a CIS control or safeguard further defines and completes a similar subcategory within the NIST framework. In this case, we replaced the NIST subcategory with CIS control or safeguard that addresses the same problem, but with greater completeness.
3) Cases in which CIS control or safeguard is defined in less detail and completes a similar subcategory within the NIST framework. In this case, we have maintained the NIST subcategory, ignoring CIS controls or safeguards that address the same problem but with less completeness than NIST.
4) Cases in which CIS controls or safeguards equivalently define a similar subcategory within the NIST framework. In this case, we chose to maintain the NIST subcategory as it addresses the same problem under equal conditions. Choosing an equivalent CIS control or safeguard would not have added or subtracted anything.
5) Cases in which a CIS control or safeguard partially defines a NIST subcategory and vice versa; that is, both NIST and CIS address the same problem, but neither of them does so completely, rather they intersect. In this case, we included both the NIST subcategory and the CIS control or safeguard because both offer a better response to the same problem than either of the two separately.
6) Cases in which a NIST subcategory does not have an equivalent CIS control or safeguard; that is, it is something that only exists within the NIST framework and not within the CIS framework. In this case, we maintained this NIST subcategory because we understand that it provides a security plus.

The previous combination was carried out by analyzing each control, safeguard, and expected outcome, one by one, to identify, after an analysis of the textual description of each item, to which NIST function, category, and subcategory it belonged. In addition, to determine the implementation group it should be placed in. The result of this process is the first version of ULEO.

2) PHASE II. INCORPORATION OF THE NINE D's OF CYBERSECURITY TO THE ULEO

The nine D's of cybersecurity are textual recommendations that lack a classification system. Therefore, in the first place, we have provided each of them with a code that can be shown in Table 1, similar to the functions, categories, and subcategories of the NIST or the controls and safeguards of the CIS in their respective models. We assimilate each of them at the level of a subcategory or expected outcome.

Subsequently, the textual descriptions of each of them were analyzed in the same way that was done with the CSCs of CIS, to identify which function or category of cybersecurity they contribute to. The nine D's of cybersecurity were systematically analyzed with respect to the controls,

TABLE 1. Identifiers assignment for the nine D's of cybersecurity.

ID	Description
9D-1	Deter attacks
9D-2	Detect attacks
9D-3	Drive up difficulty
9D-4	Differentiate protections
9D-5	Dig beneath the threat
9D-6	Diffuse protection throughout the payload
9D-7	Distract with decoys
9D-8	Divert attackers to other targets
9D-9	Depth of defense

safeguards, and subcategories of the initial ULEO previously generated, so that:

1) Cases in which a D does not have a related subcategory in the initial ULEO. We choose to add such D considering that it complements the set.
2) Cases in which a D defines a subcategory of the initial ULEO in a more detailed and complete manner. We decided to replace it with that D which addresses the same problem, but with greater completeness.
3) Cases in which a D defines a subcategory of the initial ULEO in a less detailed or complete manner. We choose to retain this subcategory and not include this D because it addresses the same problem in less depth or detail.
4) Cases in which a D defines a subcategory of the initial ULEO with the same level of detail and depth. We choose to retain this subcategory because they address the same problem under equal conditions. Choosing an equivalent D does not add or subtract anything.
5) Cases in which a D partially defines the same case as a subcategory of the initial ULEO and vice versa, that is, both cases address the same problem, but neither of them does so completely, rather they intersect. In this case, we included both the previously existing subcategory in the initial ULEO and the corresponding D because both offer a better answer to the same problem than either of them separately.
6) Cases in which a subcategory of the initial ULEO does not have an equivalent D, that is, it is something that exists only in the initial ULEO and not in [47]. In this case, we maintained this subcategory because we understood it provides a plus of security.

After this combination, we finished the inclusion of all the intended information in the ULEO: expected outcomes from NIST, controls and safeguards from CIS, the nine D's of cybersecurity, and, indirectly, mitigations from MITRE.

3) PHASE III. FILTERING AND GENERATION OF THE FINAL ULEO

After the two previous phases, the resulting ULEO contained redundant expected outcomes, whose only difference was the application in different IGs, an example of which is shown in Table 2. To remediate this redundancy, we performed a cleaning process consisting of consolidating these redundancies into a single expected outcome, leaving a single appearance that will apply to these IGs. In Table 3 the result of redundancy removal for the case presented in Table 2, can be shown.

TABLE 2. Example of redundant expected outcomes that apply to different IGs.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.AM	ID.AM-1	✓	✓	✓
Identify	ID.AM	ID.AM-1		✓	✓
Identify	ID.AM	ID.AM-1			✓

TABLE 3. Example of redundancy reduction.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.AM	ID.AM-1	✓	✓	✓

The final ULEO was obtained by repeating this process. It incorporates a total of 169 expected outcomes organized in the same functions and categories used by the NIST framework, but keeping traceability to MITRE mitigations while including information from the nine D's of cybersecurity and the CIS CSCs. In Appendix V, Tables 4 to 26 show the ULEO for each function and category. The expected outcomes are referenced by their code, being those that begin with 'CSC' those from the set of CSCs from CIS; those that start with '9D' those corresponding to the nine Ds of cybersecurity as indicated in the Table 1 and the rest, the original of the NIST framework.

4) ULEO BENEFITS

The ULEO we have built provides several advantages to the solution we propose:

- It classifies the expected outcomes into three IGs, following the same approach that the CIS uses for its critical controls. In practice, this allows to obtain three different sets of expected outcomes applicable to three different scenarios where the cybersecurity needs are LOW, MEDIUM, or HIGH.
- As it has been built, it incorporates the best recommendations of the NIST, the CIS, and the 9 D's of cybersecurity, eliminating the existing redundancies between them. It also brings together the best of each approach: security functions (and their division into categories and subcategories), IGs, etc. Moreover, based on the unified list of expected results of the NIST Cybersecurity Framework, not only cybersecurity controls are

considered in our proposal, but also the main controls related to privacy, closely linked, as detailed in [81].
- The expected outcomes of each implementation group allow for effective cyber defense against the TTPs documented by MITRE (and associated cyber threats).
- Its hierarchical arrangement allows the state of cybersecurity to be evaluated with different granularity and to easily identify which aspects must be improved to achieve the expected outcomes.
- Although our proposal should not be understood as a cyber-incident management process, it helps to deal with cyber-incidents by facilitating to the organization to acquire the skills and elements necessary for it, as a consequence of the implementation of the expected results of the functions ''Detect'' and ''Respond'' of the ULEO.
- The mere use of ULEO makes it possible to reduce the risks related to cybersecurity and business continuity by facilitating the organization to acquire the necessary skills and elements for it, as a consequence of the implementation of the expected results of the ''Identify'' and ''Recover'' functions. In addition, the ULEO has been built in such a way that there is a direct mapping from it to the mitigations defined by MITRE to face the most important real cyber threats.

D. CYBER SECURITY DOMAINS

As mentioned throughout this work, many organizations manage their cybersecurity using information security regulatory frameworks. For this reason, it is likely that they have not assimilated the need for participation in many of the functional areas whose involvement is required for cybersecurity. This is a clear mistake that must be corrected if organizations intend to deal with cyber threats using a cybersecurity approach, so it is necessary to change this trend and adopt a much broader and more integrated vision.

To help with this purpose, in our proposal we use the main cybersecurity domains of [82], because it is the most complete work and at the same time focused on cybersecurity of the sources that we have analyzed. To the previous ones, we added an additional domain related to corporate communication, marketing and institutional relations, which we consider essential to face the emerging cyberattacks in the last two years, with an impact on the supply chain and on the image and reputation of the organization; and because it is a necessary area to achieve some of the cybersecurity expected outcomes of the ULEO. In our work we will understand the domains of cybersecurity as the functional areas of an organization with responsibilities in cybersecurity. The complete list of functional areas of cybersecurity included in our proposal can be found in Table 27 (Appendix V), with the following scope:

- **FA1**. In charge of IoT device security.
- **FA2**. Active defense, vulnerability management, threat hunting, SIEM operation, cybersecurity operations center activities, or incident response [83].
- **FA3.** Prepare human resources regarding cybersecurity threats through continuous training and its reinforcement, as well as the design and execution of practical cybersecurity exercises [84].
- **FA4**. In charge of the analysis of internal and external threats, the exchange of threat intelligence with third parties or the preparation and incorporation of Indicators Of Compromise *(IOCs)*.
- **FA5**. With tasks related to the surveillance of applicable regulations and their incorporation into cybersecurity. In addition, the monitoring of different performance indicators, and the establishment of strategies, policies, standards, processes, procedures or corporate instructions.
- **FA6**. Focused on risk treatment, business continuity management, crisis management, establishing the organization's position regarding cyber risks, insurance contracting, risk registration, auditing, defining groups of risk management, or defining those responsible and owners of the processes and assets [85].
- **FA7**. Responsible for cybersecurity risk analysis, vulnerability scanning, supply chain risk identification and analysis, asset inventory, risk monitoring, and penetration testing of infrastructure, people, or systems of information, among others.
- **FA8**.With the mission of leading the secure software development cycle, continuous integration and deployment, user experience security, software quality, API security, identification of information flows in information systems, management of the free software used, or the static or dynamic analysis of the code.
- **FA9**. In charge of the management, development, implementation, and verification of compliance with the standards and regulations defined at the corporate level for cybersecurity: CIS controls, MITRE matrix, NIST framework for the improvement of cybersecurity of critical infrastructures, or the family of standards ISO27000 [19].
- **FA10**. With activities such as management, definition, implementation, operation, prevention, etc., in relation to cryptography, key and certificate management, encryption standards, security engineering, access controls with or without multiple authentication factors, single sign-on, privileged access management, identity management, identity federation, cloud security, container security, endpoint security, data protection and prevention of data leakage, network design to prevent distributed denial of service attacks, development and secure configuration of systems, patch and update management or the establishment of secure reference configurations.
- **FA11**. To promote study, education, and training, attendance at conferences, or participation in related professional groups, training, or certification.
- **FA12**. Specific activities include internal and external corporate communication, social networks

management, marketing, or the establishment and maintenance of institutional relationships with interested third parties with whom the organization maintains some type of contact.

E. AGGREGATED CYBERSECURITY ASSESSMENT

Cybersecurity assessment, especially in environments involving different functional areas, is often problematic because of its ambiguity, different interpretations, or different interests. However, having a unified, realistic and unbiased view of the state of cybersecurity is essential. Based on what was previously discussed in this study, our proposal defines the necessary aspects to provide a shared vision of cybersecurity.

1) IG IDENTIFICATION

In our work, we have elaborated on the ULEO in such a way that it allows a direct association between the protection priority indicated in the BIA for each business asset and different IGs. The correspondence between the priority established in the BIA and the IGs that should be applied to the asset can be shown in Table 28 (Appendix V), in such a way that, to provide cybersecurity to a business asset cataloged with LOW priority, actions must be put in place to achieve all the expected outcomes of the IG1 implementation group. For the assets cataloged with MEDIUM and HIGH priorities, those of the IG2 and IG3 groups, respectively. These groups and their associated actions are homogeneous for all business assets in the organization.

2) RELATIVE WEIGHT OF EACH SECURITY FUNCTION

The hierarchical structure embedded in the ULEO allows us to infer the weight of each cybersecurity function (fig. 10) for each IG with respect to the global cybersecurity of the business asset. These weights can be calculated as a percentage (or normalized between 0.00 and 1.00). In our proposal we calculated the weights of each security function for IG1, IG2 and IG3. These weights have been rounded to the second decimal place and are shown in table 29, Table 30 and Table 31 (Appendix V), respectively, where:

- F, represents the continuous cybersecurity function.
- N_c, represents the number of categories that the function F includes for the corresponding IG.
- W_f, represents the relative weight of the F function with respect to the global cybersecurity value of the asset.

3) RELATIVE WEIGHT OF EACH CATEGORY AND EXPECTED OUTCOME

For the same reasons expressed in the previous point, the ULEO allows determining the weight of each category, for each IG, with respect to each cybersecurity function, as well as the weight of each expected outcome with respect to its category. In our proposal, we calculated the weights of each category and expected outcomes, as shown in Appendix C. The weights corresponding to 'Identify' categories and expected outcomes can be seen in Tables 32 to 34; those related to 'Protect' categories and expected outcomes in Tables 35 to 37; values related to 'Detect' sub-items are shown in Tables 38 to 40; the weights of categories and expected outcomes belonging to 'Respond' are in Tables 41 to 43, and those corresponding to the 'Recover' function are shown in Tables 44 to 46. In all cases:

- C, represents the category.
- N_o, represents the number of expected outcomes of that category.
- W_c, represents the relative weight of C category with respect to its function (rounded to the second decimal place).
- W_o, represents the relative weight of each expected outcome with respect to its category.

A visual description of category weights for functions 'Identify', 'Protect', 'Detect', 'Respond' and 'Recover' is shown in figs. 11, 12, 13, 14 and 15, respectively.

The previous calculations allow a tree-like set of weights to be calculated in an aggregated way for the cybersecurity posture of the business asset in relation to its criticality. At all levels, expected outcome, category, function, or global.

4) DISCRETE LEVELS OF IMPLEMENTATION

It is convenient to define unambiguous values to establish the achievement/implementation status of each expected outcome. This issue is a common source of discrepancies and conflicts in organizations, either because each functional area has different perspectives on implementation status or because they do not have the ability to adequately measure at such a detailed level. Therefore, in our proposal, we have chosen to use Discrete Levels of Implementation *(DLIs)*, as standardized values to communicate the status of implementation of the cybersecurity actions that allows obtaining the expected outcomes (fig. 16). In our study these are the only possible values for expressing the state of progress in the implementation of each action related to an expected outcome.

Because they are not subject to interpretation and have the same meaning regardless of the functional area, action or expected outcome in question, *DLIs* are a good communication mechanism that avoids conflicts between functional areas and provides the same and shared perception of cybersecurity status.

5) ASSET BREAKDOWN

The main element of this proposal is the business asset, understanding that this unit is sufficiently small to be addressed at lower levels without too many problems. However, there may be situations where it is necessary to break down such business assets into secondary assets, for example, because it is easier to take care of cybersecurity in this way or because it facilitates the distribution of tasks between different operational groups of the same functional area or different functional areas. If necessary, the asset can be broken down as many times as necessary, following the guidelines designed

FIGURE 10. Relative weights of each cybersecurity function and the three IGs.

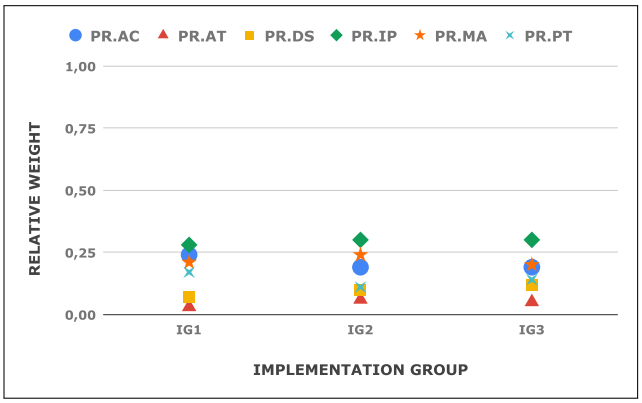


FIGURE 12. Relative weights of every category in 'Protect' function and the three IGs.

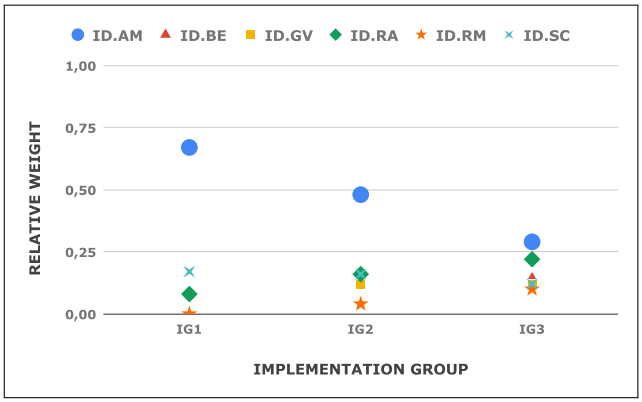


FIGURE 11. Relative weights of every category in 'Identify' function and the three IGs.

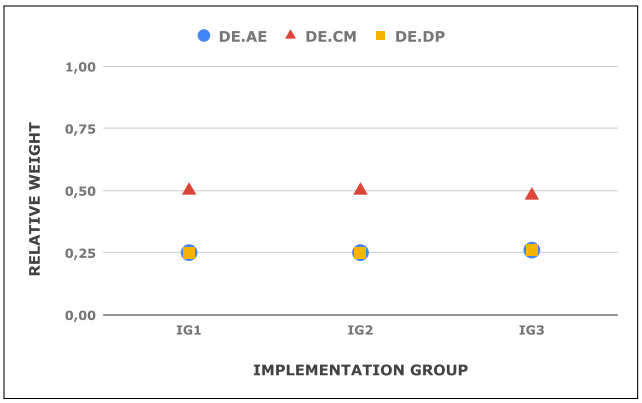


FIGURE 13. Relative weights of every category in 'Detect' function and the three IGs.

in our proposal. Bearing in mind that L represents the level of the asset, with L0 being the business asset and increasing to L1, L2... as the assets are broken down into more manageable assets:

- Each asset that is broken down must be broken down into elements that constitute an independent whole by themselves, as shown in equation 1.

$$Asset(L) \Rightarrow \bigcap_{i=1}^{n} Asset(L+1)_i = 0 \tag{1}$$

- The sub-assets in which an asset is broken down must represent the total of the asset on which they depend. In other words, the total top-level asset has been broken down into the sub-assets that make it up, as shown in equation 2.

$$Asset(L) = \sum_{i=1}^{n} Asset(L+1)_i \tag{2}$$

- Each sub-asset must have a weight (ω), as a reflection of its contribution to the higher-level asset, consisting of a normalized value between 0.00 and 1.00, equivalent to a percentage between 0% and 100% of the parent asset, respectively, as shown in equation 3.

$$Asset(L) = \sum_{i=1}^{n} \omega_i \cdot Asset(L+1)_i \tag{3}$$

subject to the following restriction (equation 4)

$$\sum_{i=1}^{n} \omega_i = 1, \forall \omega \in \mathbb{R}, \omega \subset [0,1] \tag{4}$$

- The implementation group corresponding to the parent asset will apply to all its sub-assets, as specified in equation 5.

$$IG(Asset(L+1)) = IG(Asset(L)) \tag{5}$$

Likewise, there are two types of assets/sub-assets: those that have been broken down into sub-assets, which we call 'inner assets', and those that have not been broken down into sub-assets, which we call 'leaf assets'. It is important to understand this distinction which is necessary for an aggregate evaluation of asset cybersecurity.

Figure 17 shows an example of a properly performed breakdown of a fictitious business asset at three levels. The weights and number of sub-actives in the figure are invented and placed like this for merely didactic purposes. However, it

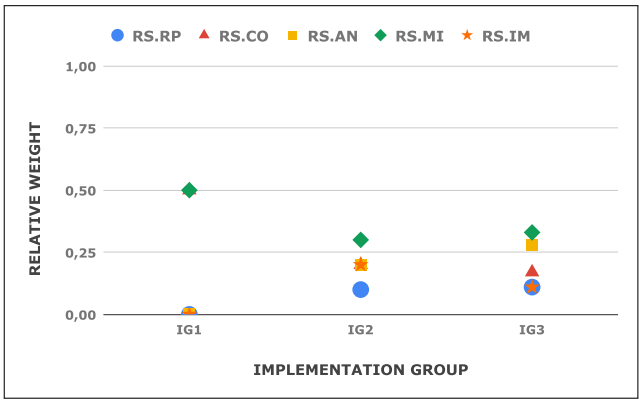


FIGURE 14. Relative weights of every category in 'Respond' function and the three IGs.

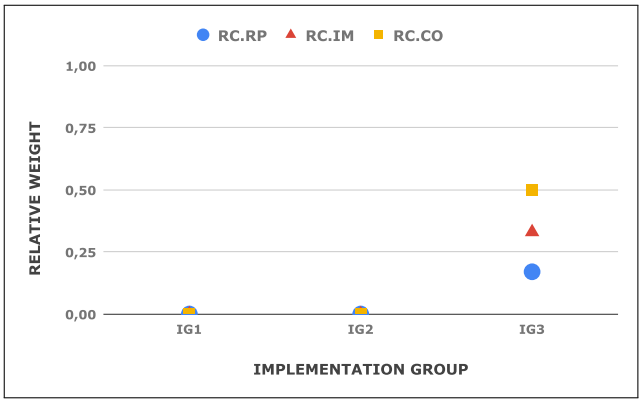


FIGURE 15. Relative weights of every category in 'Recover' function and the three IGs.

is necessary, as can be seen in the figure, that the sum of the weights of the sub-assets into which an asset has been broken down, is 1.00 in all cases. The figure also shows in different colors the inner assets (blue) and the leaf assets (yellow).

6) ASSET's CYBERSECURITY IDEAL STATE AND ASSET's CYBERSECURITY EXPECTED STATE

The Asset's Cybersecurity Ideal State *(ACIS)* will always be 1.00, which is achieved when a DLI of 1.00 has been reached for all the expected outcomes that correspond to it according to the applicable IG. It is important to understand this nuance, since the same level of implementation for the same expected outcomes that for an asset could represent an *ACIS*, for another asset it could represent a state of, for example, 0.54 (so not ideal), simply because a different implementation group applies to it.

The Asset's Cybersecurity Expected State *(ACES)*, will be determined by the organization as a cybersecurity objective, referring to a specific value of one, several, or all cybersecurity functions, categories, or expected outcomes. This expected state could result from any combination of DLIs applied to any applicable set of expected outcomes, which allows reaching that value. Understand this distinction.

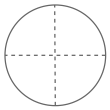
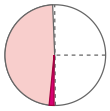
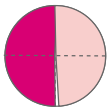
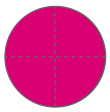
COVERAGE	DLI	EXPLANATION
	0.00	None of the necessary actions have been implemented to obtain the expected outcome.
	0.33	Some of the actions necessary to obtain the expected outcome have been implemented, but less than half.
	0.66	Half of the actions necessary to obtain the expected outcome, or more, have been implemented.
	1.00	All the necessary actions have been implemented to obtain the expected outcome.

FIGURE 16. Discrete levels of implementation (DLIs). black shows the minimum coverage required to be qualified as the corresponding DLI. Pink shows the maximum coverage (together with the black portion) before hopping to the next DLI.

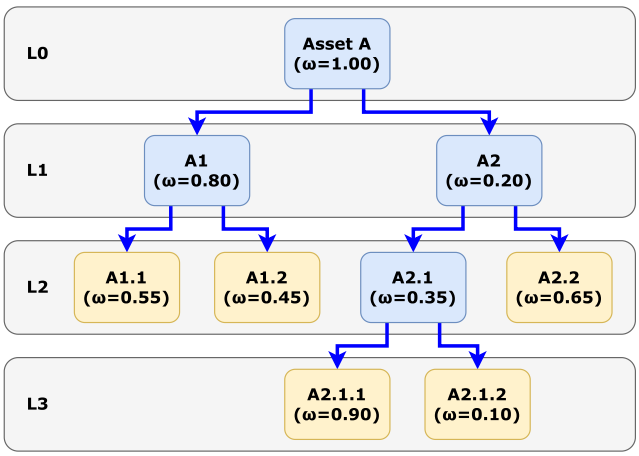


FIGURE 17. Example of a correct asset breakdown.

Although there is only one option to achieve an *ACIS* (the one described in the previous paragraph), to achieve an *ACES*, there may be multiple possible combinations on which a selection process will have to be carried out; this is covered in Section IV.

7) COMPUTING THE ASSETS' CYBERSECURITY STATUS

The defined structure and weights calculated in our proposal allow the evaluation of the cybersecurity status of an asset by adding information in a bottom-up process. The formulas that we have designed in our solution are easy to implement in any programming language or dashboard solution. Its tree-like structure facilitates the implementation of navigation through the organization, assets, sub-assets, functions, categories, and expected outcomes, to detect deficiencies in cases in which

the state of cybersecurity is not the expected or planned at any of these levels.

In the case of a leaf asset, the evaluation is performed as follows:

- **First step.** It consists of assigning to each expected outcome that applies the DLI that best reflects the status of the implementation of the associated actions. Thus, this information can be propagated upwards, starting by calculating the Category's Cybersecurity State (CCS_i) of each cybersecurity categories of the model of our proposal (equation 6).

$$CCS_i = \sum_{j=1}^{n} Wo_{ij} \cdot DLI_{ij} \tag{6}$$

  That is, the weighted sum of the discrete level of implementation of each expected outcome included in the category is calculated, based on its relative weight with respect to this category.
- **Second step**. Once the CCS_i values are known for all categories, the metrics can continue to be propagated upwards to calculate the Function's Cybersecurity State (FCS_i) of each cybersecurity function of the model of our proposal (equation 7).

$$FCS_i = \sum_{j=1}^{n} Wc_{ij} \cdot CCS_{ij} \tag{7}$$

  That is, the weighted sum of the cybersecurity status of each category of the function is calculated, considering its relative weight with respect to this function.
- **Third step**. And finally, having already calculated the FCS_i values for each function, we can calculate, going higher, the Asset's Cybersecurity Status (ACS_i) for each evaluated leaf asset (equation 8).

$$ACS_t = \sum_{j=1}^{n} Wf_{tj} \cdot FCS_{tj} \tag{8}$$

  This formula calculates the weighted sum of the cybersecurity status of each function applied to the asset, considering its relative weight with respect to its global cybersecurity. The t sub-index means that the ACS value is computed at a given moment, and subsequent measurements can throw different values.

In the case of inner assets, the calculation is based on previous knowledge of the ACS_i value of each sub-asset using the technique explained in the previous steps. Once these values are known, this information can be added, and the value of ACS_i for the inner asset can be calculated as follows (equation 9):

$$ACS_t = \sum_{j=1}^{n} Wsa_{tj} \cdot ACSsa_{tj} \tag{9}$$

where $ACSsa_{tj}$ is the ACS_{tj} value calculated independently for each sub-asset and Wsa_i is the relative weight of that sub-asset. In other words, the weighted sum of the cybersecurity status of each sub-asset is calculated while considering its relative weight with respect to the parent asset.

Because of the possibility of having different ACS_t values depending on the moment when the measurement is taken, our proposal allows computing the behavior of the ACS value over the time ($ACSev$), as shown in equation 10.

$$ACSev = \frac{t\sum_{i=1}^{t} t_i ACS_i - \sum_{i=1}^{t} t_i \sum_{i=1}^{t} ACS_i}{t\sum_{i=1}^{t} t_i^2 - (\sum_{i=1}^{t} t_i)^2} \tag{10}$$

$ACSev$ will take values from 0.00 to 1.00, because it is an additive time series. Values close to 1.00 indicate that the $ACIS$ for that asset will be achieved quickly, whereas values close to 0.00 predict ACS for that asset increases slowly and, therefore, it will take longer to achieve its $ACIS$.

8) COMPUTING THE ORGANIZATION's CYBERSECURITY STATUS

Although our proposal does not intend to address the strategic area, thanks to this, it is possible to evaluate the Organization's Cybersecurity Status (OCS) by continuing with bottom-up aggregation, in a similar way to what was explained in the previous section.

If the organization has identified weights for business assets that comply with the provisions for asset breakdown, the OCS can be calculated as follows (equation 11):

$$OCS_t = \sum_{j=1}^{n} Wba_{tj} \cdot ACSba_{tj} \tag{11}$$

where:

- Wba_{tj} is the relative weight of each business asset of the organization.
- $ACSba_{tj}$ is the cybersecurity status of each business asset calculated as described in the previous section. The t subindex, again, means that the $ACSba$ value is computed at a given moment and subsequent measurements can throw different values.

The above formula calculates the weighted sum of the cybersecurity status of each business asset, using its relative weight with respect to the organization. As in the previous paragraphs, owing to the possibility of having different OCS_t values depending on the moment when the measurement is taken, our proposal allows the calculation of the behavior of the OCS value over time ($OCSev$), as shown in equation 12.

$$OCSev = \frac{t\sum_{i=1}^{t} t_i OCS_i - \sum_{i=1}^{t} t_i \sum_{i=1}^{t} OCS_i}{t\sum_{i=1}^{t} t_i^2 - (\sum_{i=1}^{t} t_i)^2} \tag{12}$$

$OCSev$ will take values from 0.00 to 1.00, because it is an additive time series. Values close to 1.00 indicate that the cybersecurity status for the organization will be achieved quickly, whereas values close to 0.00 predict the OCS increases slowly and, therefore, it will take longer to achieve the expected cybersecurity status.

IV. CYBERSECURITY TACTICAL AND OPERATIONAL MANAGEMENT PROCESS

A. OVERVIEW

To articulate all the elements defined in Section III and that in this way our proposal constitutes a systematic mechanism, we have developed a Cybersecurity Tactical and Operational Management Process (CyberTOMP).

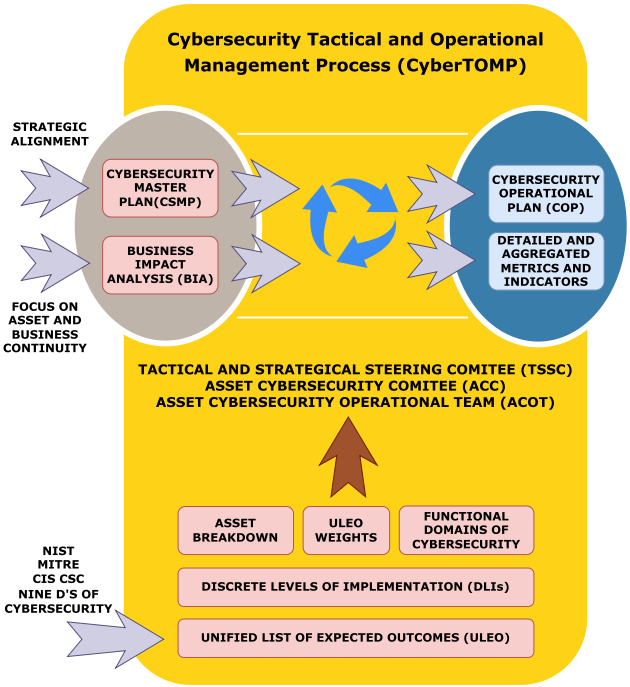


FIGURE 18. CyberTOMP high-level view.

Fig. 18 shows a coarse-grained view of the process, with the main inputs, outputs, and involved elements. The high-level objective of this process is to facilitate cybersecurity management by focusing on a business asset in each case. For this to be possible, the process, which will be discussed in the following sections, will be based on the organization's CSMP and BIA. This, together with the requirements expressed in Section III, provides the necessary alignment with the strategic objectives of the organization, both in terms of cybersecurity and business continuity, as well as a focus on business assets.

As a result of the application of CyberTOMP, a specific Operational Cybersecurity Plan *(COP)* is obtained for the business asset whose cybersecurity is being managed, as well as a set of metrics and indicators detailed and addable upwards. Both results, agreed upon by all functional areas involved in cyber defense/cyber protection of business assets. CyberTOMP facilitates the application of change management techniques [86] by following an inclusive and progressive approach.

The process that we developed achieves the necessary cooperation between all the functional areas of the organization in cybersecurity matters through three multidisciplinary bodies that participate at different times:

- **The Tactical-Strategic Steering Committee *(TSSC)*.** An interdepartmental multidisciplinary committee composed of members of the organization's steering committee, who preferably, participated in both the preparation of the CSMP and the BIA. With initial inclusion, if necessary, of tactical personnel.
- **The Asset's Cybersecurity Committee *(ACC)*.** An interdepartmental multidisciplinary committee made up of all intermediate positions with responsibilities at a tactical level for the business asset to be protected. With sporadic participation, if necessary, of operational personnel.
- **The Asset's Cybersecurity Operational Team *(ACOT)*.** An interdepartmental multidisciplinary team made up of all positions in the organization with responsibilities at the operational level, as well as external personnel incorporated into the organization belonging to service providers, who regularly participate in the daily work of the organization. In both cases, when these tasks are related to the business asset to be protected.

Each of these bodies must include people from all areas of knowledge of the organization that must participate in the cybersecurity of the business asset. In this way, these will be the bodies that facilitate the unity of action and holistic approach. Their participation in the process will be in increasing order, with the TSSC being the body that has to use the least effort in the process and the ACOT being the one that has to make the most.

At a greater level of detail, CyberTOMP includes five phases, that are similar to those commonly accepted for project management [87], with some modifications in the final phase becasue, although considering that the protection of assets emanates from projects defined in the CSMP, it is an ongoing task. These phases are: Initiating, Planning, Execution, Monitoring and Controlling, and Continuous Improvement, each containing a series of clear steps, as presented in fig. 19, which shows CyberTOMP's detailed view.

These phases, as well as the activities included in them, their peculiarities, and their explanations are detailed in the following sections with the intention of serving as a guide for their practical application in any organization. We believe this level of detail is necessary because precisely what our work tries to solve is the lack of procedural elements to manage cybersecurity at the tactical and operational levels.

B. INITIATING

This initial phase of the process is focused on:

- Ensure that cybersecurity management focuses on business assets, using those identified in the BIA.
- Ensure strategic alignment by assigning requirements derived from the BIA as well as tasks, objectives, and high-level requirements from different projects defined in the CSMP.

- Ensure that the required holism is provided to protect the business asset on a daily basis.
- Ensure that guidelines are provided to achieve shared leadership and co-governance in cybersecurity management for each business asset.

These elements have a marked strategic nature, are defined at a high level, and are presumably endowed with greater stability over time. The 'Initiation' phase consists of two main activities as detailed below.

1) DEFINE INITIAL ACC

In this activity (fig. 20), the TSSC analyzes the information contained in both the CSMP and BIA to determine the following:

- The business assets identified in the BIA and their high-level cybersecurity and continuity needs, including the potential needs for actions to respond to cybersecurity incidents and/or to recover from unavailability with regard to cybersecurity.
- The projects defined at a high level in the CSMP for each of the assets established in the BIA, their objectives, and their actions at a high level.
- Based on the above, the functional areas of the organization that should be involved in the cybersecurity of each business asset established in the BIA.
- People, at a tactical level, identified in each of these areas.

This group of individuals identified by the TSSC will form the initial ACC. If the TSSC deems it necessary, it may consult those people directly to determine more accurately whether other people not considered should also be part of the initial ACC. The initial ACC should include, for each person, high-level reasons why that person should be part of the ACC and high-level expectations for the cybersecurity of the business asset from their functional area.

As a guideline for this step, the set of cybersecurity functional domains identified in Section III can be used, which provides a fairly detailed representation of the functional areas involved in cybersecurity. The TSSC will define as many ACCs as business assets need cyber protection.

2) DEFINE INITIAL CYBERSECURITY ASSIGNMENT

In this step (fig. 21), based on the analysis of the BIA and CSMP, the TSSC will prepare a high-level list of cybersecurity and continuity needs and objectives (in relation to cybersecurity) for the business asset and will formalize a cybersecurity assignment for the asset, which will be delivered to the people who form the initial ACC. The needs and objectives will be extracted from the cybersecurity projects included in the CSMP and will be expressed in the form of high-level *ACES*, preferably as requirements on the metrics ACS_i or FCS_i of the asset indicated in the assignment. For example, the objectives of the business asset cybersecurity assignment can be:

- Increasing the ACS_i a 10%.
- Increasing the FCS_i, for the 'Respond' function, a 12%.
- Keeping the ACS_i at the current 75% relative to the current threat context.
- Keeping the ACS_i after a change in prioritization of business assets in the BIA.
- Keeping the ACS_i after a remodeling of the organizational structure.
- Assessing the ACS_i.
- Achieving the *ACIS*.

Or similar objectives. The cybersecurity assignment for the asset includes the indicated goals, the group of people that will form the initial ACC, the written statement of the assignment, and each area or functional unit represented. For practical reasons, it may be more agile to carry out this delivery through a joint meeting where the details of the assignment can be explained. Finally, the assignment must reach all the members of the initial ACC in a more formal way.

The assignment will include a period for the ACC to refine, adjust, and complete it after a more detailed analysis at the tactical level as a step prior to its final formalization.

The TSSC will carry out as many cybersecurity assignments as business assets need cyber protection.

C. PLANNING

This phase of the process is intended to delve into the details of the actions that must be undertaken to achieve the objectives requested in the assignment. For this, a series of iterative activities is carried out until the granularity that allows:

- Breaking down the business assets if it is considered necessary for a better distribution of tasks, greater control, or in general, to facilitate the management of the work to be carried out at tactical and operational levels.
- Identifying and distributing the scope of actions among different areas of knowledge represented in the ACC.
- Providing context to the cybersecurity needs of the assignment and adapting the actions that must be undertaken to the reality of the moment in the cyber field, from a multidisciplinary and holistic approach.
- Agreeing on the distribution of cybersecurity metrics and indicators.
- Updating the initial cybersecurity assignment, completing it with the aspects considered necessary.

In this phase, the ACC deals with planning in two stages that allow:

- Having a tactical-strategic planning, with a minimum participation of the TSSC.
- Having a later tactical-operational planning, more detailed, without the participation of the TSSC, and with the growing involvement of the operational teams.

The 'Planning' phase is consists of eight activities, which are detailed below.

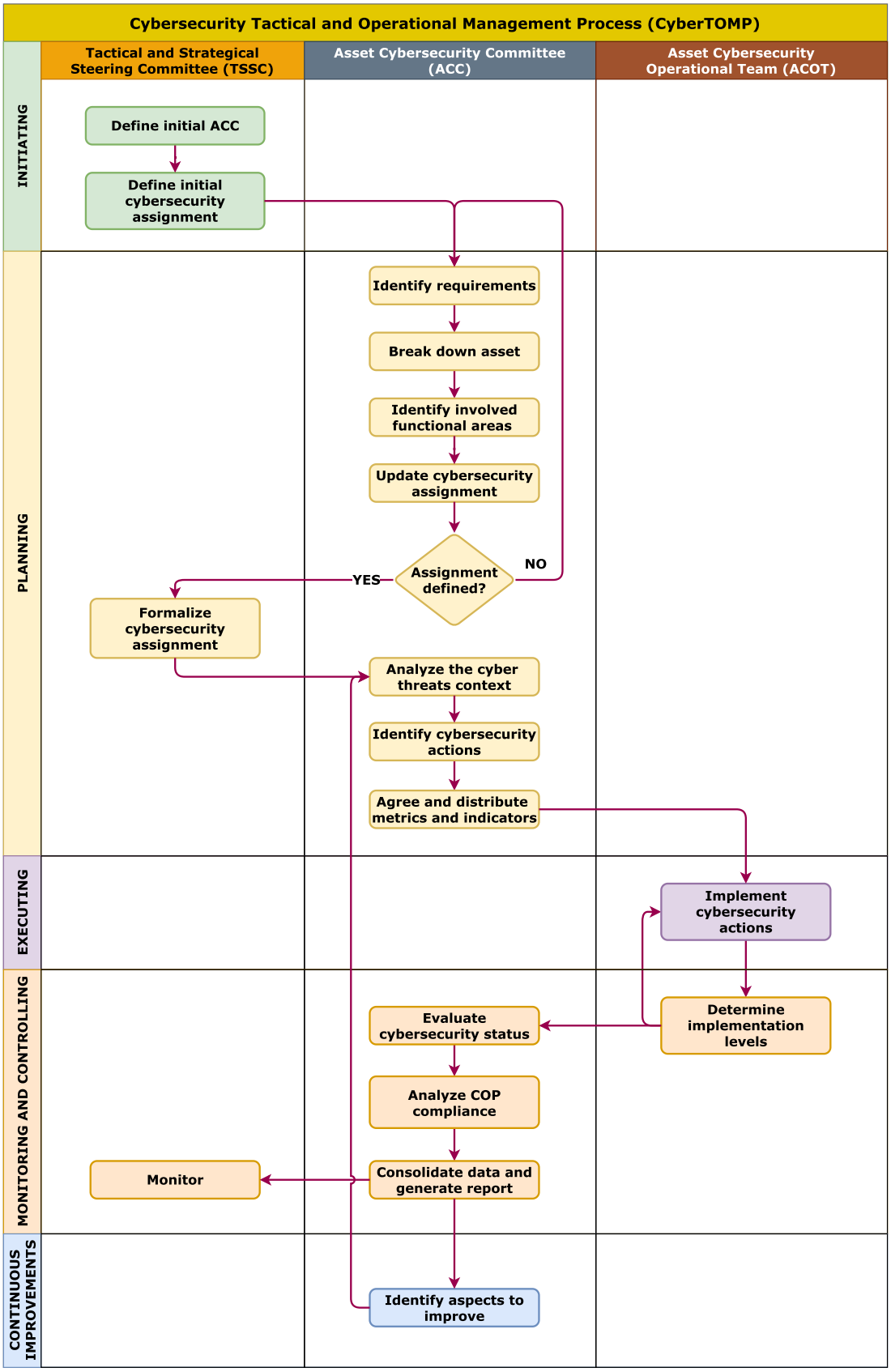


FIGURE 19. Detailed CyberTOMP steps and activities.

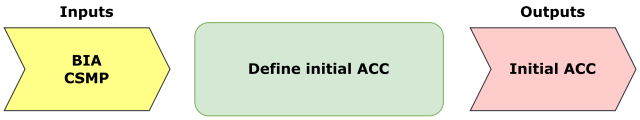


FIGURE 20. Inputs and outputs of 'Define initial ACC' activity.

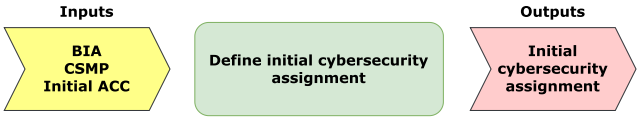


FIGURE 21. Inputs and outputs of 'Define initial cybersecurity assignment' activity'.

1) IDENTIFY REQUIREMENTS

In this activity (fig. 22), the ACC in the cybersecurity assignment for the asset will receive the priority corresponding to it, as the organization has assigned to that asset in the BIA. Accordingly, ACC will be able to directly identify the corresponding IG from the ULEO defined in this study, as described in Section III. Because each IG determines the expected outcomes for each existing function and category, the ACC will know all the expected outcomes whose implementation would allow the business asset to reach the *ACIS*. This value will be used as a reference for the maximum cybersecurity with which the asset must be provided.

The ACC must analyze the objectives (the *ACES*) set by the TSSC in the cybersecurity assignment and determine the categories or expected outcomes of the ULEO that will need to be taken into consideration to achieve that objective without going deeper into the specific actions that involve each of them. The ACC will add this additional detail to the cybersecurity assignment and update the *ACES* to reflect on what was identified.

This step begins with tactical-strategic planning of the actions required for the cybersecurity of the business asset.

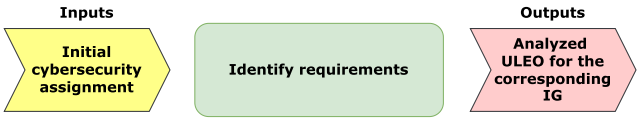


FIGURE 22. Inputs and outputs of 'Identify requirements' activity.

2) BREAK DOWN ASSET

If greater ease of management or understanding is needed, the ACC may break down the asset (fig. 23) into others of smaller caliber. The breakdown mechanism is presented in detail in Section III. Each sub-asset generated in this process is managed by the same ACC within the same assignment.

This subdivision allows different members of the ACC to focus more (although coordinated) on some of the broken-down sub-assets. It can also facilitate the assignment of activities between different areas or operational groups with greater

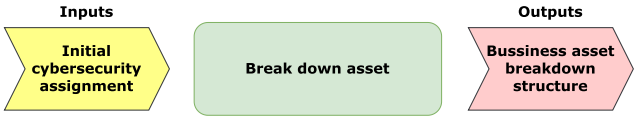


FIGURE 23. Inputs and outputs of 'Break down asset' activity.

specialization in specific tasks, without losing alignment with the proposed objective from the strategic level.

3) IDENTIFY INVOLVED FUNCTIONAL AREAS

It is likely that after the analysis of the requirements and the possible breakdown of assets into smaller ones, the need to incorporate some additional functional areas that must participate in the cybersecurity of the asset will be detected. If this is the case, the ACC will include tactical managers of such functional areas in CyberTOMP (fig. 24). The functional areas described in Section III are clear candidates.

FIGURE 24. Inputs and outputs of 'Identify functional areas involved' activity.

4) UPDATE CYBERSECURITY ASSIGNMENT

The ACC updates the cybersecurity assignment for the business asset (fig. 25) by documenting the identified requirements, the expected outcomes that must be considered to achieve the objectives, the new functional areas identified that must participate in the cybersecurity of the asset, the estimated breakdown of the business asset, and the agreed weights for all. In short, it should provide a more complete vision of cybersecurity assignment and provide the necessary justifications for it.

Once the assignment has been updated, it will be analyzed whether it can be considered complete and final, in which case the ACC will request formal approval from the TSSC. Otherwise, the process iterates, returning to the ''Identify requirements'' step.

An assignment cannot be considered complete if new functional areas are added to the process. If this happens, to prevent this inclusion from being merely cosmetic and ultimately causing tensions due to the assumption of non-agreed responsibilities, it will be necessary to iterate again (from the first step of 'Planning' phase) so that these functional areas can participate in all the steps prior to the final definition of the cybersecurity assignment.

5) FORMALIZE CYBERSECURITY ASSIGNMENT

TSSC analyzes the updated cybersecurity assignment for the asset submitted by ACC. It will evaluate its content, its convenience and feasibility, and the existence of the necessary

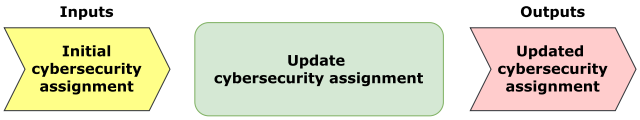


FIGURE 25. Inputs and outputs of 'Update cybersecurity assignment' activity.

consensus to provide holism and unity of action. It will approve the assignment (fig. 26) by signing it, the TSSC as a whole, the Chief Information Security Officer *(CISO)*, or the Chief Executive Officer *(CEO)*. It sends it to all members of the ACC as a final cybersecurity assignment for the protection of the business asset.

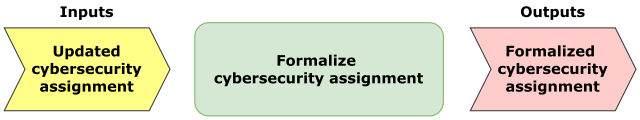


FIGURE 26. Inputs and outputs of 'Formalize cybersecurity assignment' activity.

This step ends the tactical-strategic planning of the actions required for cybersecurity of the business asset.

6) ANALYZE THE CYBER THREATS CONTEXT

In this phase, the ACC, supported by members of the ACOT, if necessary, will analyze the organization's cybersecurity context (fig. 27) in detail. In addition to the cyber threat context, in relation to business assets that they have been commissioned to protect. From both internal and external perspectives.

In this phase, renewed knowledge is acquired regarding the evolution of threats to the business in the cyber context. To express this in more detail, the cybersecurity status of a business asset can be altered simply because the context has changed, new threats have emerged, or there are exceptional situations that involve variations in the exposure level to different cybersecurity risks.

From this point is when the tactical-operational levels use their creativity, skills, and effort to cushion the enormous fluctuations in the cyber context and thus contribute, from the lower levels, to the strategic objectives of cybersecurity and the maintenance of the long-term corporate strategy.

This step is extremely important because allows a later definition of the form ('how') in which different cybersecurity actions must be implemented to ensure the achievement of the expected outcomes.

As a result of this step, it will be documented how low-level assets are impacted by the internal and external cyber context.

In this activity, in the event that it is a second or later iteration, the improvement opportunities identified in the continuous improvement phase of CyberTOMP will also be considered.

This step begins with tactical-operational planning of the actions required for the cybersecurity of the business asset.

FIGURE 27. Inputs and outputs of 'Analyze the cyber threats context' activity.

7) IDENTIFY CYBERSECURITY ACTIONS

In this activity, it is important to understand that expected outcomes are called that way precisely because they are the results that will presumably be obtained by carrying out different actions. Actions defined in greater detail in the textual description of each expected outcome.

For example, the CIS safeguard 'CS-11.1 Establish and Maintain a Data Recovery Process' would be the expected outcome, whereas the actions defined by the CIS for that safeguard would be those that allow it to be achieved: 'Establish and maintain a data recovery process. In the process, address the scope of data recovery activities, recovery prioritization, and the security of backup data. Review and update documentation annually, or when significant enterprise changes occur that could impact this Safeguard'. Only when everything described for that safeguard is done, it can be indicated that it is fully implemented.

As explained in the previous sections, there is only one way to obtain the *ACIS*, but there are many combinations to obtain the *ACES*. Therefore, both *ACC* and *ACOT* must analyze the different existing options that allow reaching the required *ACES*.

In this activity, the ACC will take the approved cybersecurity assignment, where the expected outcomes for which specific actions must be designed have already been identified, as well as the analysis carried out in the cyber threat context. (fig. 28). For each, the *ACC* will analyze the details of its description:

- For ULEO subcategories from the NIST cybersecurity framework, they should review the relevant description [48] in the framework itself or in the associated guides [50], [51].
- For the subcategories included in the ULEO and coming from the CIS, the relevant description [40] in the list of CSCs can be reviewed.
- For the subcategories incorporated into the ULEO and coming from the nine D's of cybersecurity, they should consult the description of each D [47] described in the original work.

The objective of this activity is to identify the potential list of cybersecurity actions that would address the cyber threat

FIGURE 28. Inputs and outputs of 'Identify cybersecurity actions' activity.

context to achieve the goals included in the cybersecurity assignment.

8) AGREE AND DISTRIBUTE METRICS AND INDICATORS

In this activity, the *ACC* and *ACOT* will reach a consensus (fig. 29) to select the expected outcomes and the actions that lead to them, among those identified, in a way that optimizes resources, management is facilitated, the workload and responsibilities of the different participating functional areas are reasonably distributed, existing technologies or knowledge can be reused; conflicts are minimized, etc.

With the above, each functional area of the *ACOT* will have the expected outcomes and the associated tasks that they have to undertake from their scope, the description of such tasks, the roles and responsibilities, metrics and weights, planning of the actions and milestones, their dependencies, and the periods to evaluate the progress. All this, as a whole, will constitute the Cybersecurity Operational Plan *(COP)* for the asset accompanied by the corresponding metrics and indicators. This plan will be fully aligned with the corresponding cybersecurity assignment mandated by the *TSSC* and, by extension, with the BIA and associated CSMP project.

The *ACC* defines a minimum DLI for each expected outcome, which must allow the achievement of what is required by the *TSSC* in the cybersecurity assignment for the asset. In this way, each person from the *ACOT* will know the target level of implementation for the actions that correspond to them. This step ends the tactical-operational planning of the actions required for cybersecurity of the business asset.

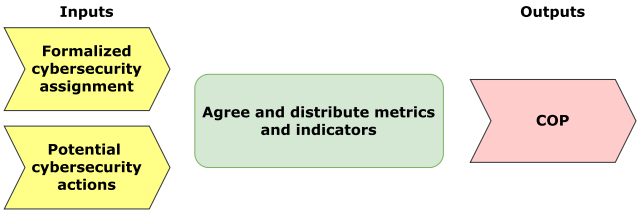


FIGURE 29. Inputs and outputs of 'Agree and distribute metrics and indicators' activity.

D. EXECUTING

The objective of this phase effectively implement the actions planned in the *COP*.

1) IMPLEMENT CYBERSECURITY ACTIONS

In this activity, the *ACOT* will be the team in charge of implementing the specific measures to achieve the expected outcomes that have been assigned (fig. 30), so that the micromanagement of these actions can be carried out in a decentralized manner in each *ACOT* functional area once the *ACC* has already agreed on the set of precise actions.

In practice, this step allows the performance of short-term tasks in a semi-autonomous and self-organized manner, ultimately contributing to the organization's cybersecurity and business continuity objectives (in relation to cybersecurity).

The different members of the *ACOT* can be helped, especially in the more technical functional areas, by the different existing guides, such as, for example, [33], [50] o [46].

FIGURE 30. Inputs and outputs of 'Implement cybersecurity actions' activity.

E. MONITORING AND CONTROL

This phase is focused on evaluating the cybersecurity status of business assets in relation to the cybersecurity assignment ordered by the *TSSC* and the corresponding *COP* generated in previous phases, to build valuable information so that the different levels of the organization can clearly understand the cybersecurity situation of the asset, with the necessary detail, and make decisions in this regard.

The evaluation of the state of cybersecurity will be carried out at three levels: operational, tactical, and strategic, which will be carried out with different frequencies, the most frequent being the operational evaluation, followed by the tactical one and the least frequent, the strategic evaluation, for a correct assessment of the impact of the actions as well as the new needs in the short, medium, and long term, respectively.

1) DETERMINE IMPLEMENTATION LEVELS

In this activity, with the periodicity indicated by the *ACC*, each member of the *ACOT* establishes the current *NDI* for each expected outcome that has been assigned (fig. 31), as indicated in Section III. In this way, the *ACC* will have the *NDI* for all expected outcomes included in the *COP* of the asset.

Together with this information, the *ACOT* will succinctly detail difficulties, synergies, proposals arising during the course of the work, or unexpected situations or situations not initially analyzed, if they exist. This will be performed individually for each expected outcome.

Progress information, together with the relevant information that allows its contextualization, will be included in an Operational Cybersecurity Report *(OCR)*, which can be as complex or simple as the organization requires.

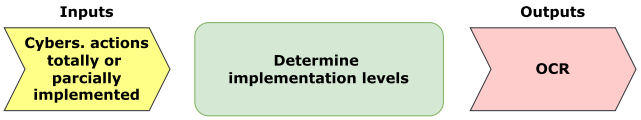


FIGURE 31. Inputs and outputs of 'Determine implementation levels' activity.

2) EVALUATE CYBERSECURITY STATUS

In this activity, with the agreed frequency, the *ACC* will receive the *OCRs* sent by the *ACOT* and proceed to evaluate the cybersecurity of the asset (fig. 32) using the *DLIs* contained in that report. They will do it following what is specified in Section III, taking into account the relative weights and calculating, for the business asset, the values CCS_i, FCS_i and ACS_i, so that at the end, the information aggregation and construction process will have, for each asset and sub-asset into which the business asset has been broken down:

- The status of achievement of each expected outcome.
- The cybersecurity status with respect to each category.
- The cybersecurity status with respect to each function.
- The cybersecurity status of the business asset.

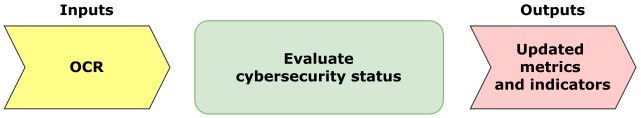


FIGURE 32. Inputs and outputs of 'Evaluate cybersecurity status' activity.

3) ANALYZE COP COMPLIANCE

In this activity, with the frequency that has been agreed upon for the tactical evaluation of cybersecurity, the *ACC* will analyze the current state and evolution of the different metrics and indicators associated with the cybersecurity assignment (fig. 33), calculated and aggregated in the previous step using the different *OCRs* that the *ACOT* has been sending to it and that have not yet been jointly analyzed or compared with the *COP* forecasts. It is recommended that this activity coincide with the last release of *OCR* by *ACOT* in order to have the most up-to-date view possible.

In addition, it will use the relevant information provided by the *ACOT* in the *OCRs* to contextualize possible deviations from what was planned and understand the circumstances that may have caused such deviations or the synergies and opportunities that may exist. All of this will be included in the Tactical Cybersecurity Report *(TCR)*.

Finally, the *ACC* updates, if it exists, the organization's cybersecurity dashboard with the current CCS_i, FCS_i, and ACS_i values.

4) CONSOLIDATE DATA AND GENERATE REPORT

In this activity, with the periodicity required by the *TSSC*, the *ACC* will analyze the degree of achievement of what is required in the cybersecurity assignment for the business asset, using such an assignment as a source and also the information of the different *TCRs*. It is recommended that this task is carried out coinciding with the generation of the last *TRC* to obtain the most up-to-date and recent view. With all this, it will generate a Strategic Cybersecurity Report *(SCR)* that will broadly identify the advances or delays and their main causes, as well as evolutionary data and tactical decisions taken or planned, if appropriate, in a very executive way (fig.34).

The *ACC* will report the status to the *TSSC*, forwarding that report.

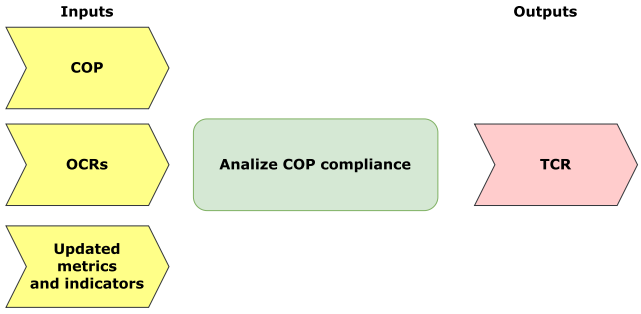


FIGURE 33. Inputs and outputs of 'Analyze COP compliance' activity.

5) MONITORING

The *TSSC* receives, with the required frequency, the last *SCR* regarding cybersecurity assignment for the protection of the business asset. With this information and that of the rest of the cybersecurity assignments they have assigned, they can, if desired, calculate the *OCS* value, taking into account the weights that could have been defined at a strategic level for each business asset.

FIGURE 34. Inputs and outputs of 'Consolidate data and generate report' activity.

The TSSC will use this monitoring information (fig. 35) to modify or update the cybersecurity assignment for strategic decision-makers in general or to generate additional strategic information that it deems necessary. This aspect is not addressed in detail in CyberTOMP, whose main scope is the tactical and operational levels.

F. CONTINUOUS IMPROVEMENT

The purpose of this phase is to identify the margins for improvement in different aspects, which can later be used as a basis for designing and executing additional actions in cybersecurity.

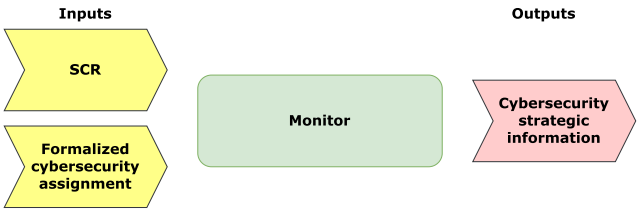


FIGURE 35. Inputs and outputs of 'Monitor' activity.

1) IDENTIFY ASPECTS TO IMPROVE

In this activity (fig. 36), the *ACC* will analyze the information from the *TCR*, paying attention not so much to possible deviations, but to the relevant information provided by the different members of the *ACOT*, which may include identified synergies, barriers found, opportunities, difficulties, and so on. The improvements likely to be identified in this activity are, without being an exhaustive list:

- New mechanisms for better coordination between functional areas.
- New mechanisms for better coordination and communication in the *ACC*.
- The need to search for alternatives for the implementation of operational actions that have been more complex or costly to implement in practice than initially planned.
- The use of tools that allow greater agility in work.
- The possibility of including common elements that suppose an optimization of costs and effort.
- The need to reinforce the operational work with new staff.
- Others of a similar nature.

This identification must be the result of a joint debate within the *ACC* and must not focus on the search for solutions, an aspect that is dealt with in the new analysis of the context, but on the identification and documentation of improvement opportunities.

Once this activity is done, the process must iterate again from the activity "Analyze the cyber threats context". Thus, CyberTOMP allows design of a new modified *COP* to include new cybersecurity actions to improve the detected weaknesses and adapt to the dynamic cyber threat context.

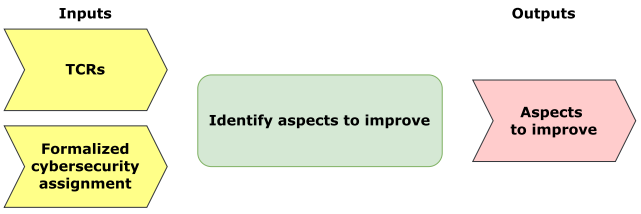


FIGURE 36. Inputs and outputs of 'Identify aspects to improve' activity.

G. PERIODICITY AND END OF THE PROCESS

CyberTOMP only ends when the *TSSC* carries out a new cybersecurity assignment for the same business asset or when it is decided from by strategic sphere of the organization. Otherwise, CyberTOMP will continue even if the *ACES* or *ACIS* has been reached. This is because, as has been commented on throughout this document, that state can change simply because the context changes. For example:

- If the context of cyberspace varies significantly and controls currently in place for the cybersecurity of the asset no longer have the same validity.

TABLE 4. ULEO for 'Identify' function and 'Assets Management' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.AM	CSC-1.1	✓	✓	✓
Identify	ID.AM	CSC-12.4		✓	✓
Identify	ID.AM	CSC-14.1	✓	✓	✓
Identify	ID.AM	CSC-2.2	✓	✓	✓
Identify	ID.AM	CSC-3.1	✓	✓	✓
Identify	ID.AM	CSC-3.2	✓	✓	✓
Identify	ID.AM	CSC-3.6	✓	✓	✓
Identify	ID.AM	CSC-3.7		✓	✓
Identify	ID.AM	ID.AM-1	✓	✓	✓
Identify	ID.AM	ID.AM-2	✓	✓	✓
Identify	ID.AM	ID.AM-2		✓	✓
Identify	ID.AM	ID.AM-3		✓	✓

TABLE 5. ULEO for 'Identify' function and 'Business Environment' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.BE	9D-1		✓	✓
Identify	ID.BE	ID.BE-1			✓
Identify	ID.BE	ID.BE-2			✓
Identify	ID.BE	ID.BE-3			✓
Identify	ID.BE	ID.BE-4			✓
Identify	ID.BE	ID.BE-5			✓

TABLE 6. ULEO for 'Identify' function and 'Governance' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.GV	CSC-17.4		✓	✓
Identify	ID.GV	ID.GV-1	✓	✓	✓
Identify	ID.GV	ID.GV-2		✓	✓
Identify	ID.GV	ID.GV-3			✓
Identify	ID.GV	ID.GV-4			✓

TABLE 7. ULEO for 'Identify' function and 'Risk Assessment' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.RA	9D-1		✓	✓
Identify	ID.RA	CSC-18.2		✓	✓
Identify	ID.RA	CSC-18.5			✓
Identify	ID.RA	CSC-3.7		✓	✓
Identify	ID.RA	ID.RA-1	✓	✓	✓
Identify	ID.RA	ID.RA-2			✓
Identify	ID.RA	ID.RA-3			✓
Identify	ID.RA	ID.RA-4			✓
Identify	ID.RA	ID.RA-6			✓

TABLE 8. ULEO for 'Identify' function and 'Risk Management Strategy' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.RM	9D-8		✓	✓
Identify	ID.RM	ID.RM-1			✓
Identify	ID.RM	ID.RM-2			✓
Identify	ID.RM	ID.RM-3			✓

TABLE 9. ULEO for 'Identify' function and 'Supply Chain Risk Management' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Identify	ID.SC	ID.SC-1		✓	✓
Identify	ID.SC	ID.SC-2	✓	✓	✓
Identify	ID.SC	ID.SC-3		✓	✓
Identify	ID.SC	ID.SC-4			✓
Identify	ID.SC	ID.SC-5	✓	✓	✓

- If there are organizational changes that eliminate, add, or reorganize the functional areas or personnel associated with it.
- If the implemented solutions depend on formalized contracts with service providers that end.
- If the business asset is expanded or reduced with new functionalities or components.
- If employees leave the organization or move horizontally and are replaced by others with different skills or training, or they are not replaced.
- If there is a budget reduction that prevents the maintenance of cybersecurity measures implemented around the asset.

TABLE 10. ULEO for 'Protect' function and 'Identity Management and Access Control' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.AC	CSC-12.5		✓	✓
Protect	PR.AC	CSC-12.6		✓	✓
Protect	PR.AC	CSC-13.4		✓	✓
Protect	PR.AC	CSC-4.7	✓	✓	✓
Protect	PR.AC	CSC-5.2	✓	✓	✓
Protect	PR.AC	CSC-5.6		✓	✓
Protect	PR.AC	CSC-6.8			✓
Protect	PR.AC	PR.AC-1	✓	✓	✓
Protect	PR.AC	PR.AC-2			✓
Protect	PR.AC	PR.AC-3		✓	✓
Protect	PR.AC	PR.AC-3	✓	✓	✓
Protect	PR.AC	PR.AC-4	✓	✓	✓
Protect	PR.AC	PR.AC-5	✓	✓	✓
Protect	PR.AC	PR.AC-6			✓
Protect	PR.AC	PR.AC-7	✓	✓	✓

TABLE 11. ULEO for 'Protect' function and 'Awareness and Training' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.AT	CSC-14.9		✓	✓
Protect	PR.AT	CSC-15.4		✓	✓
Protect	PR.AT	PR.AT-1	✓	✓	✓
Protect	PR.AT	PR.AT-2		✓	✓

TABLE 12. ULEO for 'Protect' function and 'Data Security' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.DS	9D-6			✓
Protect	PR.DS	CSC-3.4	✓	✓	✓
Protect	PR.DS	PR.DS-1		✓	✓
Protect	PR.DS	PR.DS-2		✓	✓
Protect	PR.DS	PR.DS-3	✓	✓	✓
Protect	PR.DS	PR.DS-4			✓
Protect	PR.DS	PR.DS-5			✓
Protect	PR.DS	PR.DS-6		✓	✓
Protect	PR.DS	PR.DS-7		✓	✓
Protect	PR.DS	PR.DS-8			✓

H. RECOMMENDATIONS FOR A CORRECT APPLICATION

Practical implementation of CyberTOMP can be facilitated or improved by applying a series of recommendations:

TABLE 13. ULEO for 'Protect' function and 'Information Protection Processes and Procedures' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.IP	9D-3		✓	✓
Protect	PR.IP	9D-5		✓	✓
Protect	PR.IP	9D-8		✓	✓
Protect	PR.IP	9D-9	✓	✓	✓
Protect	PR.IP	CSC-11.1	✓	✓	✓
Protect	PR.IP	CSC-16.1		✓	✓
Protect	PR.IP	CSC-16.14			✓
Protect	PR.IP	CSC-18.4			✓
Protect	PR.IP	CSC-2.5		✓	✓
Protect	PR.IP	CSC-2.6		✓	✓
Protect	PR.IP	CSC-2.7			✓
Protect	PR.IP	CSC-4.3	✓	✓	✓
Protect	PR.IP	PR.IP-1	✓	✓	✓
Protect	PR.IP	PR.IP-10		✓	✓
Protect	PR.IP	PR.IP-11	✓	✓	✓
Protect	PR.IP	PR.IP-12		✓	✓
Protect	PR.IP	PR.IP-2		✓	✓
Protect	PR.IP	PR.IP-3			✓
Protect	PR.IP	PR.IP-4	✓	✓	✓
Protect	PR.IP	PR.IP-5			✓
Protect	PR.IP	PR.IP-6	✓	✓	✓
Protect	PR.IP	PR.IP-7		✓	✓
Protect	PR.IP	PR.IP-8			✓
Protect	PR.IP	PR.IP-9	✓	✓	✓

- **Application of change management techniques**. In the development of our proposal, we understand the following circumstances concur:
  - A collaborative habit is required to reach consensus.
  - By employing three collegiate groups for decision-making, those roles that would normally have the possibility of making decisions individually may understand it as an attack on their competencies and present opposition to the changes.

  To facilitate both, we recommend the professional application of specific techniques for change management that ease the applicability of this proposal. For example, finding change agents to actively participate in the implementation. This change management approach should include training in soft skills that will equip participants with the ability to achieve win-win agreements.
- **The necessary role of CISO**. In light of what is stated in our solution, this could give the impression that the role of the CISO is diluted, becoming a point of potential conflict. It is recommended that the CISO have a relevant leadership role in the *TSSC*. Leadership, not necessarily hierarchical superiority. However, as the role with the most developed skills in cybersecurity, it should be the person responsible for ensuring the correct execution of CyberTOMP and who mediates in the case of conflicts or doubts.
- **Automation**. The use of tools to automate the calculation of metrics and indicators in the cybersecurity evaluation process can significantly facilitate the use of CyberTOMP and the generation of reports. All metrics and indicators have been defined in such a way that they can be easily automated and information can be provided

TABLE 14. ULEO for 'Protect' function and 'Maintenance' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.MA	9D-5		✓	✓
Protect	PR.MA	9D-9		✓	✓
Protect	PR.MA	CSC-12.1	✓	✓	✓
Protect	PR.MA	CSC-12.3		✓	✓
Protect	PR.MA	CSC-13.5		✓	✓
Protect	PR.MA	CSC-16.13			✓
Protect	PR.MA	CSC-18.3		✓	✓
Protect	PR.MA	CSC-4.2	✓	✓	✓
Protect	PR.MA	CSC-4.6	✓	✓	✓
Protect	PR.MA	CSC-4.8		✓	✓
Protect	PR.MA	CSC-4.9		✓	✓
Protect	PR.MA	CSC-7.3	✓	✓	✓
Protect	PR.MA	CSC-8.1	✓	✓	✓
Protect	PR.MA	CSC-8.10		✓	✓
Protect	PR.MA	CSC-8.3	✓	✓	✓
Protect	PR.MA	CSC-8.9		✓	✓
Protect	PR.MA	PR.MA-1			✓

TABLE 15. ULEO for 'Protect' function and 'Protective Technology' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Protect	PR.PT	9D-4		✓	✓
Protect	PR.PT	9D-7			✓
Protect	PR.PT	CSC-4.12			✓
Protect	PR.PT	CSC-4.4	✓	✓	✓
Protect	PR.PT	CSC-4.5	✓	✓	✓
Protect	PR.PT	CSC-9.5		✓	✓
Protect	PR.PT	PR.PT-1	✓	✓	✓
Protect	PR.PT	PR.PT-2	✓	✓	✓
Protect	PR.PT	PR.PT-3			✓
Protect	PR.PT	PR.PT-4			✓
Protect	PR.PT	PR.PT-5	✓	✓	✓

TABLE 16. ULEO for 'Detect' function and 'Anomalies and Events' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Detect	DE.AE	CSC-8.12			✓
Detect	DE.AE	DE.AE-1		✓	✓
Detect	DE.AE	DE.AE-2		✓	✓
Detect	DE.AE	DE.AE-3	✓	✓	✓
Detect	DE.AE	DE.AE-4			✓
Detect	DE.AE	DE.AE-5			✓

TABLE 17. ULEO for 'Detect' function and 'Security Continuous Monitoring' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Detect	DE.CM	CSC-13.1		✓	✓
Detect	DE.CM	CSC-13.5		✓	✓
Detect	DE.CM	CSC-3.14			✓
Detect	DE.CM	DE.CM-1		✓	✓
Detect	DE.CM	DE.CM-2			✓
Detect	DE.CM	DE.CM-3			✓
Detect	DE.CM	DE.CM-4	✓	✓	✓
Detect	DE.CM	DE.CM-5			✓
Detect	DE.CM	DE.CM-6			✓
Detect	DE.CM	DE.CM-7	✓	✓	✓
Detect	DE.CM	DE.CM-8		✓	✓

TABLE 18. ULEO for 'Detect' function and 'Detection Processes' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Detect	DE.DP	CSC-17.1	✓	✓	✓
Detect	DE.DP	CSC-17.4		✓	✓
Detect	DE.DP	CSC-17.5		✓	✓
Detect	DE.DP	DE.DP-2			✓
Detect	DE.DP	DE.DP-3			✓
Detect	DE.DP	DE.DP-5			✓

at all levels in almost real time, reducing the workload of the *ACC*.

- **Gradual implementation**. A progressive application is recommended, starting with a business asset that is relatively simple to manage and with few functional areas involved, and subsequently including others of greater complexity until this proposal is applied to all the business assets of the organization. The application to simpler cases in the first instance allows the refinement of the process, training of the team and obtaining good results that serve as a hook for the expansion of the solution.

TABLE 19. ULEO for 'Respond' function and 'Analysis' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Respond	RS.AN	CSC-17.9			✓
Respond	RS.AN	RS.AN-1		✓	✓
Respond	RS.AN	RS.AN-2			✓
Respond	RS.AN	RS.AN-3			✓
Respond	RS.AN	RS.AN-5		✓	✓

TABLE 20. ULEO for 'Respond' function and 'Communications' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Respond	RS.CO	CSC-17.4	✓	✓	✓
Respond	RS.CO	CSC-17.5		✓	✓
Respond	RS.CO	RS.CO-5			✓

TABLE 21. ULEO for 'Respond' function and 'Improvements' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Respond	RS.IM	RS.IM-1		✓	✓
Respond	RS.IM	RS.IM-2		✓	✓

TABLE 22. ULEO for 'Respond' function and 'Mitigation' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Respond	RS.MI	CSC-1.2	✓	✓	✓
Respond	RS.MI	CSC-4.10		✓	✓
Respond	RS.MI	CSC-7.7		✓	✓
Respond	RS.MI	RS.MI-1			✓
Respond	RS.MI	RS.MI-2			✓
Respond	RS.MI	RS.MI-3			✓

TABLE 23. ULEO for 'Respond' function and 'Response Planning' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Respond	RS.RP	CSC-17.6		✓	✓
Respond	RS.RP	RS.RP-1			✓

V. CONCLUSION

Tactical and operational levels are responsible for the practical implementation of cybersecurity. The standards used for

TABLE 24. ULEO for 'Recover' function and 'Communications' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Recover	RC.CO	RC.CO-1			✓
Recover	RC.CO	RC.CO-2			✓
Recover	RC.CO	RC.CO-3			✓

TABLE 25. ULEO for 'Recover' function and 'Improvements' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Recover	RC.IM	RC.IM-1			✓
Recover	RC.IM	RC.IM-2			✓

TABLE 26. ULEO for 'Recover' function and 'Recovery Planning' category.

Function	NIST Category	Unified expected outcome	IG1	IG2	IG3
Recover	RC.RP	RC.RP-1			✓

TABLE 27. Functional areas involved in cybersecurity, reused and improved in our proposal.

FA ID	Main cybersecurity responsibilities
FA1	Physical security
FA2	Security operations
FA3	User education
FA4	Threat intelligence
FA5	Governance
FA6	Enterprise risk management
FA7	Risk assessment
FA8	Application security
FA9	Frameworks and standards
FA10	Security architecture
FA11	Career development
FA12	Corporate communications

TABLE 28. Correspondence between cyberprotection priorities and IGs.

Cyberprotection priority (from BIA)	Corresponding IG
LOW	IG1
MEDIUM	IG2
HIGH	IG3

TABLE 29. Weights of cybersecurity functions for IG1.

F	N_c	W_f
Identify	4	0.27
Protect	6	0.40
Detect	3	0.20
Respond	2	0.13
Recover	0	0.00

TABLE 30. Weights of cybersecurity functions for IG2.

F	N_c	W_f
Identify	6	0.30
Protect	6	0.30
Detect	3	0.15
Respond	5	0.25
Recover	0	0.00

TABLE 31. Weights of cybersecurity functions for IG3.

F	N_c	W_f
Identify	6	0.26
Protect	6	0.26
Detect	3	0.13
Respond	5	0.22
Recover	3	0.13

TABLE 32. Weights for category 'Identify' and IG1.

C	N_o	W_c	W_o
ID.AM	8	0.67	0.125
ID.BE	0	0.00	0.00
ID.GV	1	0.08	1.00
ID.RA	1	0.08	1.00
ID.RM	0	0.00	0.00
ID.SC	2	0.17	0.50

TABLE 33. Weights for category 'Identify' and IG2.

C	N_o	W_c	W_o
ID.AM	12	0.48	$1/12$
ID.BE	1	0.04	1.00
ID.GV	3	0.12	$1/3$
ID.RA	4	0.16	0.25
ID.RM	1	0.04	1.00
ID.SC	4	0.16	0.25

cybersecurity encourage organizations to develop procedural elements for effective cybersecurity management at these levels, but do not provide such a procedural basis so that it can be used as is. This causes indeterminacy in how each organization manages cybersecurity at lower levels, often resulting in a lack of holism, strategic alignment, differing perceptions of the state of cybersecurity or difficulty quickly adapting to a changing cyber threat landscape.

TABLE 34. Weights for category 'Identify' and IG3.

C	N_o	W_c	W_o
ID.AM	12	0.29	$1/12$
ID.BE	6	0.15	$1/6$
ID.GV	5	0.12	0.20
ID.RA	9	0.22	$1/9$
ID.RM	4	0.10	0.25
ID.SC	5	0.12	0.20

TABLE 35. Weights for category 'Protect' and IG1.

C	N_o	W_c	W_o
PR.AC	7	0.24	$1/7$
PR.AT	1	0.03	1.00
PR.DS	2	0.07	0.50
PR.IP	8	0.28	0.125
PR.MA	6	0.21	$1/6$
PR.PT	5	0.17	0.20

TABLE 36. Weights for category 'Protect' and IG2.

C	N_o	W_c	W_o
PR.AC	12	0.19	$1/12$
PR.AT	4	0.06	0.25
PR.DS	6	0.10	$1/6$
PR.IP	18	0.30	$1/18$
PR.MA	15	0.24	$1/15$
PR.PT	7	0.11	$1/7$

TABLE 37. Weights for category 'Protect' and IG3.

C	N_o	W_c	W_o
PR.AC	15	0.19	$1/15$
PR.AT	4	0.05	0.25
PR.DS	10	0.12	0.10
PR.IP	24	0.30	$1/24$
PR.MA	17	0.20	$1/17$
PR.PT	11	0.14	$1/11$

TABLE 38. Weights for category 'Detect' and IG1.

C	N_o	W_c	W_o
DE.AE	1	0.25	1.00
DE.CM	2	0.50	0.50
DE.DP	1	0.25	1.00

TABLE 39. Weights for category 'Detect' and IG2.

C	N_o	W_c	W_o
DE.AE	3	0.25	$1/3$
DE.CM	6	0.50	$1/6$
DE.DP	3	0.25	$1/3$

TABLE 40. Weights for category 'Detect' and IG3.

C	N_o	W_c	W_o
DE.AE	6	0.26	$1/6$
DE.CM	11	0.48	$1/11$
DE.DP	6	0.26	$1/6$

TABLE 41. Weights for category 'Respond' and IG1.

C	N_o	W_c	W_o
RS.RP	0	0.00	0.00
RS.CO	1	0.50	1.00
RS.AN	0	0.00	0.00
RS.MI	1	0.50	1.00
RS.IM	0	0.00	0.00

TABLE 42. Weights for category 'Respond' and IG2.

C	N_o	W_c	W_o
RS.RP	1	0.10	1.00
RS.CO	2	0.20	0.50
RS.AN	2	0.20	0.50
RS.MI	3	0.30	$1/3$
RS.IM	2	0.20	0.50

TABLE 43. Weights for category 'Respond' and IG3.

C	N_o	W_c	W_o
RS.RP	2	0.11	0.50
RS.CO	3	0.17	$1/3$
RS.AN	5	0.28	0.20
RS.MI	6	0.33	$1/6$
RS.IM	2	0.11	0.50

TABLE 44. Weights for category 'Recover' and IG1.

C	N_o	W_c	W_o
RC.RP	0	0.00	0.00
RC.IM	0	0.00	0.00
RC.CO	0	0.00	0.00

Our proposal comprises a common set of expected cybersecurity results rooted in the most recognized cybersecurity standards and initiatives, as well as a set of metrics that allow a homogeneous evaluation of cybersecurity at different levels. This is orchestrated by CyberTOMP, a process for managing cybersecurity at tactical and operational levels.

Together, these elements complement the standard for cybersecurity used at a strategic level, regardless of what

TABLE 45. Weights for category 'Recover' and IG2.

C	N_o	W_c	W_o
RC.RP	0	0.00	0.00
RC.IM	0	0.00	0.00
RC.CO	0	0.00	0.00

TABLE 46. Weights for category 'Recover' and IG3.

C	N_o	W_c	W_o
RC.RP	1	0.17	1.00
RC.IM	2	0.33	0.50
RC.CO	3	0.50	$1/3$

this standard is, being able to be used as is, out of the box, for the holistic management of cybersecurity at all levels while maintaining alignment with the corporate cybersecurity strategy.

This proposal is being implemented in an entity in the Public Sector, a process that will provide the necessary feedback for its evolution and formal validation, results we hope to share with the scientific community in a future study.

APPENDIX A
ULEO TABLES

See Tables 4–26.

APPENDIX B
FUNCTIONAL AREAS INVOLVED IN CYBERSECURITY AND CORRESPONDENCE CYBERPROTECTION PRIORITIES - IGs

See Tables 27 and 28.

APPENDIX C
WEIGHTS OF EVERY CYBERSECURITY FUNCTION, CATEGORY AND EXPECTED OUTCOME

See Tables 29–46.

REFERENCES

[1] F. Y. Sattarova and T. H. Kim, "IT security review: Privacy, protection, access control, assurance and system security," *Int. J. Multimedia Ubiquitous Eng.*, vol. 2, no. 2, pp. 17–32, 2007.

[2] J. L. Fennelly, *Effective Physical Security*. Oxford, U.K.: Butterworth-Heinemann, 2016.

[3] M. E. Whitman and J. Herbert Mattord, *Management of Information Security*. Boston, MA, USA: Cengage Learning, 2013.

[4] R. von Solms, "Information security management: Why standards are important," *Inf. Manage. Comput. Secur.*, vol. 7, no. 1, pp. 50–58, Mar. 1999.

[5] M. E. Whitman and J. H. Mattord, *Principles of Information Security*. Boston, MA, USA: Cengage Learning, 2021.

[6] T. Chmielecki, P. Pacyna, P. Potrawka, N. Rapacz, R. Stankiewicz, and P. Wydrych, "Enterprise-oriented cybersecurity management," in *Proc. Ann. Comput. Sci. Inf. Syst.*, Sep. 2014, pp. 1–8.

[7] N. Kshetri, *Cybersecurity Management: An Organizational and Strategic Approach*. Toronto, ON, Canada: University of Toronto Press, 2021.

[8] R. von Solms and J. van Niekerk, "From information security to cyber security," *Comput. Secur.*, vol. 38, pp. 97–102, Oct. 2013, doi: 10.1016/j.cose.2013.04.004.

[9] R. Reid and J. Van Niekerk, "From information security to cyber security cultures," in *Proc. Inf. Secur. South Afr.*, Aug. 2014, pp. 1–7.

[10] J. V. D. Ham, "Toward a better understanding of 'Cybersecurity,'" *Digit. Threats, Res. Pract.*, vol. 2, no. 3, pp. 1–3, Sep. 2021.

[11] M. Roopak, G. Y. Tian, and J. Chambers, "An intrusion detection system against DDoS attacks in IoT networks," in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2020, pp. 562–567.

[12] R. A. Rothrock, J. Kaplan, and F. Van der Oord, "The board's role in managing cybersecurity risks," *MIT Sloan Manag. Rev.*, vol. 59, no. 2, pp. 12–15, 2018.

[13] K. T. Dean, "Cyber-security holism: A system of solutions for a distributed problem," Marine Corps Command and Staff College, Quantico, VA, USA, Tech. Rep. ADA601717, 2013.

[14] H. I. Kure and S. Islam, "Assets focus risk management framework for critical infrastructure cybersecurity risk management," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 4, no. 4, pp. 332–340, Dec. 2019.

[15] R. Phillips and B. Tanner, "Breaking down silos between business continuity and cyber security," *J. Bus. Continuity Emergency Planning*, vol. 12, no. 3, pp. 224–232, 2019.

[16] R. Rajan, N. P. Rana, N. Parameswar, S. Dhir, Sushil, and Y. K. Dwivedi, "Developing a modified total interpretive structural model (M-TISM) for organizational strategic cybersecurity management," *Technol. Forecasting Social Change*, vol. 170, Sep. 2021, Art. no. 120872.

[17] I. N. Fovino, "Cybersecurity, our digital anchor," Eur. Union, Luxembourg, Tech. Rep. JRC121051, 2020, doi: 10.2760/352218.

[18] D. Sulistyowati, F. Handayani, and Y. Suryanto, "Comparative analysis and design of cybersecurity maturity assessment methodology using NIST CSF, COBIT, ISO/IEC 27002 and PCI DSS," *Int. J. Informat. Visualizat.*, vol. 4, no. 4, p. 225, Dec. 2020.

[19] A. Bahuguna, R. K. Bisht, and J. Pande, "Roadmap amid chaos: Cyber security management for organisations," in *Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2018, pp. 1–6.

[20] R. Miñana. (2021). *?'Qué es Capability Maturity Model Integration? (CMMI)*. Accessed: Jul. 7, 2022. [Online]. Available: https://www2.deloitte.com/es/es/pages/technology/articles/que-es-cmmi-capability-maturity-modelintegration.html

[21] K. Balla, M. Tang, P. Mowat, M. Rasking, S. Chaobo, E. van Veenendaal, and Z. Hongbao, "Changes in CMMI 2.0 and how they can affect TMMi," TMMi Foundation, Bulverde, TX, USA, Tech. Rep., 2020.

[22] ISACA. *CMMI Adoption & Transition Guidance 2021*. Accessed: Jul. 7, 2022. [Online]. Available: https://cmmiinstitute.com/getattachment/5868888b-5f37-4715-bc8b-c43250ec0abc/attachment.aspx

[23] C. Agutter, "ITIL 4 essentials, second edition," IT Governance Publishing Ltd., Cambridge, U.K., Tech. Rep. 5524, 2020.

[24] *ITIL Foundation. ITIL 4 Edition. Glossary*, Axelos, London, U.K., 2019.

[25] R. Jašek, L. Králík, and M. Popelka, "ITIL and information security," in *Proc. AIP Conf.*, Helsinki, 2015, Art. no. 550020.

[26] E. R. Larrocha, G. Díaz, J. M. Minguet, M. Castro, and A. Vara, "Filling the gap of information security management inside ITIL: Proposals for posgraduate students," in *Proc. IEEE EDUCON Conf.*, Apr. 2010, pp. 907–912.

[27] J. Gillingham. (Aug. 2021). *An Introduction To Information Security Management in ITIL*. Accessed: Jul. 7, 2022. [Online] Available: https://www.invensislearning.com/blog/information-security-management/

[28] *UNE-ISO/IEC 27001. Tecnología de la Información. Técnicas de Seguridad. Sistema de Gestión de Seguridad de la Información (SGSI) Requisitos*. AENOR, Madrid, Spain, 2014

[29] *UNE-ISO/IEC 27002. Tecnología de la Información. Técnicas de Seguridad. Código de Prácticas Para Los Controles de Seguridad de la Información*, AENOR, Madrid, Spain, 2015.

[30] H. R. Suárez, J. D. P. Álvarez, and M. G. Hidalgo, "Ciber-resiliencia. Aproximación a un marco de medición," Nat. Inst. Commun. Technol. (INTECO), Tech. Rep., 2014.

[31] *IMC_01—Metodología de Evaluación de Indicadores Para Mejora de la Ciberresiliencia (IMC)*, Spanish Nat. Cybersecur. Inst. (INCIBE), 2020.

[32] G. D. España, "Real decreto 311/2022, de 3 de mayo, por el que SE regula el esquema nacional de seguridad," *Boletín Oficial del Estado*, vol. 106, pp. 61715–61804, May 2020.

[33] CCN. *Guías Esquema Nacional de Seguridad 2022*. Accessed: Jul. 7, 2022. [Online] Available: https://www.ccn-cert.cni.es/guias/guias-series-ccn-stic/800-guia-esquema-nacional-de-seguridad.html

[34] *Guía de Seguridad CCN-STIC-806. Esquema Nacional de Seguridad, Plan de Adecuación*, Centro Criptológico Nacional, Madrid, Spain, 2011.

[35] Centro Criptológico Nacional. (2021). *Adecuación al ENS y Seguimiento del Progreso*. Accessed: Jul. 7, 2022. [Online]. Available: https://www.ccn-cert.cni.es/gestion-de-incidentes/lucia/2-uncategorised/48-adecuacion-al-ens-y-seguimiento-del-progreso.html
[36] MITRE. *MITRE ATT&CK©, 2021*. Accessed: Jul. 7, 2022. [Online] Available: https://attack.mitre.org/
[37] E. S. Blake, A. Andy, P. M. Doug, C. N. Kathryn, G. P. Adam, and B. T. Cody, "MITRE ATT&CK©: Design and philosophy," MITRE, McLean, VA, USA, Tech. Rep., 2020
[38] R. Kwon, T. Ashley, J. Castleberry, and S. N. G. Gourisetti, "Cyber threat dictionary using MITRE ATT&CK matrix and NIST cybersecurity framework mapping," in *Proc. Resilience Week (RWS)*, Oct. 2020.
[39] W. Xiong, E. Legrand, and O. Åberg, and R. Lagerström, "Cyber security threat modeling based on the MITRE Enterprise ATT&CK matrix," *Softw. Syst. Model.*, vol. 21, pp. 157–177, Jun. 2021.
[40] *CIS Security Controls, Version 8*, CIS, East Greenbush, NY, USA, 2021
[41] B. Shamma, *Implementing CIS Critical Security Controls for Organizations on a Low-Budget*. Ann Arbor, MI, USA: ProQuest LLC, 2018.
[42] S. Gros, "A critical view on CIS controls," in *Proc. 16th Int. Conf. Telecommun. (ConTEL)*, Jun. 2021, pp. 122–128.
[43] OWASP. (2021). *OWASP TOP 10 Project*. Accessed: Jul. 7, 2022. [Online] Available: https://owasp.org/www-project-top-ten/
[44] M. Bach-Nutman, "Understanding the top 10 OWASP vulnerabilities," 2020, *arXiv:2012.09960*.
[45] *Center for Internet Security, CIS Community Defense Model, Version 2.0*, CIS, East Greenbush, NY, USA, 2021.
[46] MITRE. (2022). *MITRE ATT&CK© Enterprise Mitigations*. Accessed: Jul. 7, 2022. [Online] Available: https://attack.mitre.org/mitigations/enterprise/
[47] K. S. Wilson and M. A. Kiy, "Some fundamental cybersecurity concepts," *IEEE Access*, vol. 2, pp. 116–124, 2014.
[48] *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*, NIST, Gaithersburg, MD, USA, 2018
[49] Organización de los Estados Americanos y AWS, "Ciberseguridad, marco NIST. Un abordaje integral de la ciberseguridad," Org. Amer. States (OEA), USA, White Paper, 5th ed. OEA, 2019.
[50] NIST Computer Security Resource Center. *SP 800 Series, 2021*. Accessed: Jul. 7, 2022. [Online] Available: https://csrc.nist.gov/publications/sp800
[51] *NIST Special Publication 800–53. Revision 5. Security and Privacy Controls for Information Systems and Organizations*, NIST, Gaithersburg, MD, USA, 2020
[52] *Adquisition and sustainment, Cybersecurity Maturity Model Certification (CMMC) Model Overview. Version 2.0*, Office of the Under Secretary of Defense, Department of Defense, Richmond, VA, USA, 2021
[53] Office of the Under Secretary of Defense. (Dec. 2021). *Adquisition and Sustainment, CMMC 2.0 Spreadsheet and Mapping*. Accessed: Jul. 7, 2022. [Online] Available: https://www.acq.osd.mil/cmmc/docs/CMMCModel_V2_Mapping.xlsx
[54] T. Limba, T. Plėta, K. Agafonov, and M. Damkus, "Cyber security management model for critical infrastructure," *Entrepreneurship Sustainability Issues*, vol. 4, no. 4, pp. 559–573, 2017, doi: 10.9770/jesi.2017.4.4(12).
[55] M. Tvaronaviciene, T. Pleta, and S. D. Casa, "Cyber security management model for critical infrastructure protection," in *Proc. Int. Sci. Conf. Contemp. Issues Bus., Manag. Econ. Eng.*, 2021, pp. 133–139.
[56] K. Barbara, E. W. N. Bernroider, and R. Walser, "Evaluation of cybersecurity management controls and metrics of critical infrastructures: A literature review considering the NIST cybersecurity framework," in *Proc. Nordic Conf. Secure IT Syst.*, Cham, Switzerland: Springer, 2018, pp. 369–384.
[57] N. Tissir, S. El Kafhali, and N. Aboutabit, "Cybersecurity management in cloud computing: Semantic literature review and conceptual framework proposal," *J. Reliable Intell. Environments*, vol. 7, no. 2, pp. 69–84, Jun. 2021.
[58] L. Maximilian, E. Markl, and M. Aburaia, "Cybersecurity management for (industrial) Internet of Things-challenges and opportunities," *J. Inf. Technol. Softw. Eng.*, vol. 8, no. 5, pp. 1–9, 2018.
[59] S. Ali, "Cybersecurity management for distributed control system: Systematic approach," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 11, pp. 10091–10103, Nov. 2021.
[60] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, "Harnessing artificial intelligence capabilities to improve cybersecurity," *IEEE Access*, vol. 8, pp. 23817–23837, 2020.
[61] F. Alrimawi, L. Pasquale, and B. Nuseibeh, "On the automated management of security incidents in smart spaces," *IEEE Access*, vol. 7, pp. 111513–111527, 2019.
[62] M. Antunes, M. Maximiano, R. Gomes, and D. Pinto, "Information security and cybersecurity management: A case study with SMEs in Portugal," *J. Cybersecurity Privacy*, vol. 1, no. 2, pp. 219–238, Apr. 2021.
[63] M. S. Tisdale, "Architecting a cybersecurity management framework," *Issues Inf. Syst.*, vol. 17, no. 4, pp. 1–284, 2016.
[64] L. Axon, A. Erola, A. Janse van Rensburg, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Practitioners' views on cybersecurity control adoption and effectiveness," in *Proc. 16th Int. Conf. Availability, Rel. Secur.*, Aug. 2021, pp. 1–10.
[65] United States Government Accountability Office, "Critical infrastructure protection. Sector-specific agencies need better measure cybersecurity progress," U.S. Government Accountability Office (GAO), USA, Tech. Rep. GAO-16-79, 2015.
[66] T. Kissoon, "Optimum spending on cybersecurity measures," *Transforming Government, People, Process Policy*, vol. 14, no. 3, pp. 417–431, doi: 10.1108/TG-11-2019-0112.
[67] J. Breier and L. Hudec, "On selecting critical security controls," in *Proc. Int. Conf. Availability, Rel. Secur.*, Sep. 2013, pp. 582–588.
[68] P. Speight, "Business continuity," *J. Appl. Secur. Res.*, vol. 6, no. 4, pp. 529–554, 2011.
[69] B. Zawada, "The practical application of ISO 22301," *J. Bus. Continuity Emergency Planning*, vol. 8, no. 1, pp. 83–90, 2014.
[70] M. H. Bejarano, R. J. Rodriguez, and J. Merseguer, "A vision for improving business continuity through cyber-resilience mechanisms and frameworks," in *Proc. 16th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2021, pp. 1–5.
[71] R. L. Tammineedi, "Business continuity management: A standards-based approach," *Inf. Secur. J., A Global Perspective*, vol. 19, no. 1, pp. 36–50, Mar. 2010.
[72] M. Clark, J. Espinosa, and W. Delone, "Defending organizational assets: A preliminary framework for cybersecurity success and knowledge alignment," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2020, pp. 4283–4292.
[73] H. Kure, S. Islam, and M. Razzaque, "An integrated cyber security risk management approach for a cyber-physical system," *Appl. Sci.*, vol. 8, no. 6, p. 898, May 2018.
[74] A. Couce-Vieira, D. R. Insua, and A. Kosgodagan, "Assessing and forecasting cybersecurity impacts," *Decis. Anal.*, vol. 17, no. 4, pp. 356–374, Dec. 2020.
[75] Z. A. Collier and I. Linkov, and J. H. Lambert, "Four domains of cybersecurity: A risk-based systems approach to cyber decisions," *Environ. Syst. Decis.*, vol. 33, pp. 2194–5411, Nov. 2013.
[76] A. M. Rea-Guaman, J. Mejía, T. San Feliu, and J. A. Calvo-Manzano, "AVARCIBER: A framework for assessing cybersecurity risks," *Cluster Comput.*, vol. 23, no. 3, pp. 1827–1843, Sep. 2020.
[77] C. T. Harry and N. Gallagher, "An effects-centric approach to assessing cybersecurity risk," Center Int. Secur. Stud., Univ. Maryland, College Park, MD, USA, Tech. Rep. resrep20424, 2019.
[78] A. A. Ganin, P. Quach, M. Panwar, Z. A. Collier, J. M. Keisler, D. Marchese, and I. Linkov, "Multicriteria decision framework for cybersecurity risk assessment and management," *Risk Anal.*, vol. 40, no. 1, pp. 183–199, Jan. 2020.
[79] J. R. S. Cristóbal, "Complexity in project management," *Proc. Comput. Sci.*, vol. 121, pp. 762–766, Jan. 2017.
[80] CIS. (2021). *CIS Critical Security Controls V8 Mapping to NIST CSF*. Accessed: Jul. 7, 2022. [Online] Available: https://www.cisecurity.org/white-papers/cis-controlsv8-mapping-to-nist-csf/
[81] NIST. (2021). *Mappings: Cybersecurity Framework and Privacy Framework to Rev. 5*. Accessed: Sep. 23, 2022. [Online] Available: https://csrc.nist.gov/CSRC/media/Publications/sp/800-53/rev-5/final/documents/csf-pf-to-sp800-53r5-mappings.xlsx
[82] H. Jiang. (2021). *Cybersecurity Domain Map Ver 3.0*. Accessed: Jul. 7, 2022. [Online]. Available: https://www.linkedin.com/pulse/cybersecurity-domain-map-ver-30-henry-jiang/
[83] A. Ahmad, K. C. Desouza, S. B. Maynard, H. Naseer, and R. L. Baskerville, "How integration of cyber security management and incident response enables organizational learning," *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 8, pp. 939–953, Aug. 2020.
[84] N. Chowdhury and V. Gkioulos, "Cyber security training for critical infrastructure protection: A literature review," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100361.
[85] H. I. Kure, S. Islam, M. Ghazanfar, A. Raza, and M. Pasha, "Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 493–514, Jan. 2022.

[86] A. Zimmermann, *Gestión del Cambio Organizacional: Caminos y Herramientas*, 2nd ed. Quito: Ediciones Abya-Yala, 2000.
[87] *A guide to the Project Management Body of Knowledge. PMBoK Guide*. 7th ed., Project Management Institute, Newtown Square, PA, USA, 2021.

MANUEL DOMÍNGUEZ-DORADO received the B.Sc. and M.Sc. degrees in computer science from the University of Extemadura and the master's degree in cybersecurity management (CISO) from the International Institute for Global Security Studies. He worked as a Researcher with the University of Extemadura. Nowadays, he works as the Cybersecurity Manager of the Public Business Entity Red.es. His research interests include cybersecurity in organizations and in communications networks and cybersecurity management.

JAVIER CARMONA-MURILLO received the Ph.D. degree in computer science and communications from the University of Extremadura, Spain, in 2015. From 2005 to 2009, he was a Research and Teaching Assistant. Since 2009, he has been an Associate Professor with the Department of Computing and Telematics System Engineering, Universidad de Extremadura. During the past years, he has spent research periods with the Centre for Telecommunications Research, King's College London, U.K., and Aarhus University, Denmark. His current research interests include 5G networks, mobility management protocols, performance evaluation, and the quality of service support in future mobile networks.

DAVID CORTÉS-POLO received the degree in computer science from the University of Extremadura, Spain, and the Ph.D. degree in telematics from the University of Extremadura, in 2015. From 2011 to 2014, he worked as a Researcher and a Teaching Assistant with the University of Extremadura. From 2020 to 2022, he was an Associate Professor with the Department of Computing and Telematics System Engineering, Universidad de Extremadura. Since September 2022, he has been an Assistant Professor at King Juan Carlos University, Madrid. His research interests include IP-based mobility management protocols, performance evaluation, and network CDR analytics.

FRANCISCO J. RODRÍGUEZ-PÉREZ received the degree in computer science engineering and the Ph.D. degree from the University of Extremadura, Spain, in 2000 and 2015, respectively. His research interests include the design and implementation of algorithms and signaling techniques to improve reliability, performance, delay, computing load, and energy consumption, and other metrics of prioritized quality of service aware flows over multiprotocol label switching packet transport networks, the Internet of Things systems, wireless *ad-hoc* networks, and smart cities environments.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 June 2023, accepted 13 July 2023, date of publication 20 July 2023, date of current version 11 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297446

RESEARCH ARTICLE

Safety and Cybersecurity Assessment Techniques for Critical Industries: A Mapping Study

IEVGEN BABESHKO[1,2] AND FELICITA DI GIANDOMENICO[2]

[1]Computer Systems, Networks and Cybersecurity Department, National Aerospace University "Kharkiv Aviation Institute," 61070 Kharkiv, Ukraine
[2]Istituto di Scienza e Tecnologie dell'Informazione Alessandro Faedo-CNR, 56127 Pisa, Italy

Corresponding author: Ievgen Babeshko (ievgen.babeshko@isti.cnr.it)

ABSTRACT The paper presents a mapping study of safety and cybersecurity assessment techniques used in critical industries such as nuclear power plants, the oil and gas sector, autonomous vehicles, railways, etc., with particular emphasis on instrumentation and control systems (I&C). Modern I&Cs are complex electronic systems comprising thousands of components, therefore their reliability and safety when employed in critical application domains are challenging. With the development and integration of Industry 4.0 technologies such systems become more open for communication and flexible usage due to gradual interconnection with public networks and the Internet, but new cybersecurity and safety challenges are introduced. This paper states research questions and provides analysis results of recent relevant sources. Initially, 320 records (acquired between 2018 and 2022 inclusive) were identified. Later on, 187 studies were processed to check eligibility criteria. Overall, this mapping study includes 49 papers, after examining the pre-defined criteria and guidelines. The results of the analysis performed allow to systemize techniques being utilized in practice right now, as well as to identify trends of further techniques development. In fact, although the techniques used are not novel and most of them have been used for decades, our study shows that there are still some new trends in this field. In particular, the unified safety and cybersecurity assessment technique is a promising research direction, worth further investigation.

INDEX TERMS Safety, cybersecurity, assessment techniques, instrumentation and control systems.

I. INTRODUCTION

Safety and cybersecurity issues have always been among the top priorities in critical industries, but today they are becoming even more urgent. Assessment of modern critical instrumentation and control systems is a complicated process, principally due to the size (system consists of many components) and volatility (system perpetually evolves throughout lifecycle) problem. Cybersecurity contributes to safety and sometimes conflicts with it, but it is not always considered at all lifecycle stages together with safety. The results of the assessment are considerably dependent on metrics/techniques/assumptions chosen. Therefore, arranging an assessment process based on solid methodologies/techniques is of high importance, because there is a risk of safety underestimation or overestimation, with potential severe impact on the service delivered. With the focus on critical sectors, this paper considers the following domains: nuclear power plants, the oil and gas sector, autonomous vehicles, and railways.

The purpose of the work is to survey recent literature in order to develop a mapping study useful to understand:

- which 'classical' (described in standards or other normative documents) assessment techniques are used in recent primary studies;
- advancements of such "classical" techniques, to respond to needs posed by modern critical systems;
- application of specific techniques to assess different metrics/properties they were originally developed for (i.e. modification of reliability assessment techniques for cybersecurity assessment);
- combinations of techniques used;
- needs for additional research in the generalization of assessment techniques so as to provide a unified assessment approach.

The associate editor coordinating the review of this manuscript and approving it for publication was Diana Gratiela Berbecaru.

The paper is organized as follows. In Section II we analyze existing systematic literature reviews, surveys, and mapping studies on adjacent topics. A description of the approach used, as well as research questions, are provided in Section III. In section IV we present our analysis of the collected data and our results in response to the research questions. In Section V we list key findings. Section VI presents the threats to the validity. Finally, we make conclusions and outline future research directions in Section VII.

II. COMPARISON WITH RELATED WORKS

This study fills a gap in research on cybersecurity and safety assessment techniques: although several reviews exist, to the best of the authors' knowledge no previous work provides a comprehensive and up-to-date systematic mapping study that covers different critical domains. To facilitate comparison, related works are summarized in Table 1. For each work, the following information is presented:

- Reference;
- Year of publication;
- Application domain;
- The number of references included in the paper.

TABLE 1. Comparison with other systematic literature reviews, surveys, and mapping studies.

Ref.	Year	Domain	Number of references
[19]	2019	Nuclear	32
[24]	2021	Nuclear	52
[30]	2021	Critical Infrastructures	107
[41]	2020	Autonomous Vehicles	23

The review made in [19] discusses U.S. Nuclear Regulatory Commission (NRC)'s proposed vulnerability assessment methodology, as well as additions and changes that must be made to increase its efficacy. It mainly includes references to normative documents for the nuclear field, not research studies.

In [24], the focus is put on the identification of scientific papers discussing cybersecurity frameworks, standards, guidelines, best practices, and any additional cybersecurity protection measures for the nuclear domain. Safety issues are not covered, as well as cybersecurity and safety co-engineering were not addressed in this report.

Report [30] focuses on studies that combine Bayesian Networks and Graph Theory for safety and cybersecurity integrated assessment. Other techniques and their combinations are not covered.

In [41], blockchain-based methods are discussed for cybersecurity assurance in the autonomous vehicles domain.

III. REVIEW APPROACH

A. GENERAL INFORMATION

This study was performed according to guidelines on systematic literature reviews and surveys [59] and guidelines for conducting systematic mapping studies [60]. First of all, a set of research questions that our study aims to answer was formulated. These research questions address safety and cybersecurity techniques used, as well as their combinations and modifications, and are listed in Section III-B. From the research questions, we defined the research query and then the search strategy, as presented in Section III-C. We applied this search strategy to the following popular electronic databases:

- IEEE Explore (https://ieeexplore.ieee.org/);
- ScienceDirect (https://www.sciencedirect.com/);
- SpringerLink (https://link.springer.com/);
- Wiley (https://onlinelibrary.wiley.com/);
- MDPI (https://www.mdpi.com/).

After that, the selection process described in Section III-D was applied so as to identify the set of relevant primary studies that we analysed to answer the research questions. We present the results of our analysis in Section IV and Section V.

B. RESEARCH QUESTIONS

Implementation of deep and throughout safety assessment was a strong requirement for critical industries for a long time, but the essential rise of cyberattacks and malware targeted for this particular sector during the last 5 years has intensified the discussions around the convergence of safety and cybersecurity.

Traditional safety assessment approaches either did not focus on cybersecurity, leaving its issues to particular separate disciplines, or at most referred to generic cybersecurity approaches and guidelines which were not feasible to follow or implement.

To overcome the abovementioned challenges, traditional approaches were modified in different ways, so as to consider cybersecurity-related threats and make assessment more comprehensive. Such modifications could be the following:

- assessment techniques determine the impact of cybersecurity threats and vulnerabilities on system safety as an adjunct to 'traditional' hazards; an example of such an approach is Hazard Analysis and Risk Assessment (HARA) combined with Threat Analysis and Risk Assessment (TARA);
- adaptation of traditional dependability and safety assessment techniques to the cybersecurity domain; an example of such an approach is Intrusion Modes, Effects, and Criticality Analysis (IMECA), where the traditional Failure Modes, Effects, and Criticality Analysis (FMECA) approach is utilized for intrusion analysis;
- include combinations of several safety and cybersecurity assessment techniques.

Despite the variety of approaches safety and cybersecurity assessment for critical industries is still a challenge requiring further investigation.

The following research questions were formulated to attain such investigation:

- (RQ1) Which safety indicators (metrics) are considered during safety assessment?
- (RQ2) Which cybersecurity indicators (metrics) are considered during cybersecurity assessment?
- (RQ3) Which techniques (classical, modified, combinations) are used for safety assessment?
- (RQ4) Which techniques (classical, modified, combinations) are used for cybersecurity assessment?
- (RQ5) Which limitations are applied to techniques currently used?

C. SEARCH STRATEGY

The search string used for the selection of studies is presented in Table 2. Only studies published from 2018 through 2022 inclusive were considered.

TABLE 2. Search string.

({safety} < OR > {cybersecurity} < OR > {security}) < AND > ({assessment} < OR > {evaluation} < OR > {analysis}) < AND > ({nuclear} < OR > {oil} < OR > {vehicle} < OR > {transport} < OR > {railway} < OR > {automotive})

D. SELECTION PROCESS

The following inclusion and exclusion criteria (Table 3) were applied to the studies identified using the search string (figure 1).

TABLE 3. Inclusion and exclusion criteria.

Inclusion Criteria
• Papers published in journals or conference proceedings. • Studies presenting a modified technique or combination of several techniques and description of usage • Studies providing use cases to support the performed assessment or introducing a tool • Studies that are peer-reviewed
Exclusion Criteria
• Studies that are PhD thesis, published in workshop proceedings and book chapters. • Studies from fields different from safety and cybersecurity assessment in critical industries domains (nuclear, aerospace, maritime, oil and gas, railway, automotive) • Studies that do not provide clear evidence of the benefits obtained through a proposed modified technique (criteria of clearness: measurable results compared to unmodified technique(s)) • Multiple studies authored by the same researchers on the same/similar topic (in this case the more relevant source was chosen, i.e. journal paper had priority over conference proceeding, most recent one had priority over older ones) • Studies that are not written in English

To ensure quality assessment the following questions were addressed:

- Are claims clearly defined?
- Is it possible to reuse the presented assessment technique, its modification or a combination of techniques (is description detailed enough)?

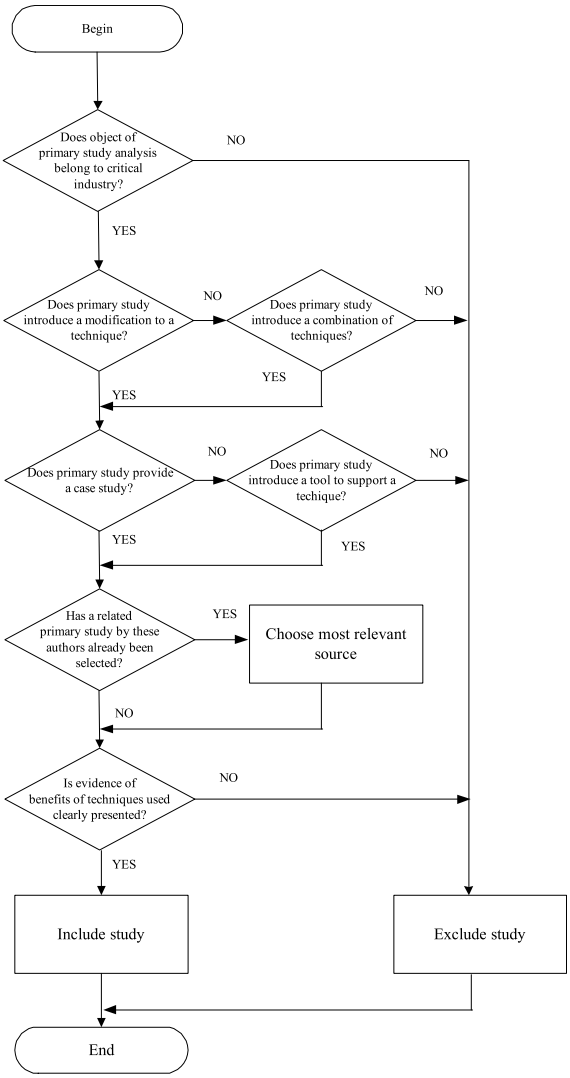


FIGURE 1. Flowchart of the selection process for each primary study.

Initially, 320 records (acquired between 2018 and 2022 inclusive) were identified according to the search string. After examining titles, abstracts and keywords, the number of records was reduced to 187 by excluding not relevant studies.

After the application of the selection process shown on Fig. 1, 49 papers were selected from a total number of 187.

IV. DATA ANALYSIS

A. DISTRIBUTION BY YEAR AND TYPE

The distribution of primary studies by years in the window 2018-2022 is shown in Fig. 2.

Most of the studies are journal papers as shown in Table 4 and Fig. 3, but conference proceedings were also analysed.

B. OVERVIEW OF THE ADOPTED TECHNIQUES AND ASSESSMENT METRICS

The performed research has shown that techniques listed in Table 5 below are typically used during the safety and/or cybersecurity assessment process.

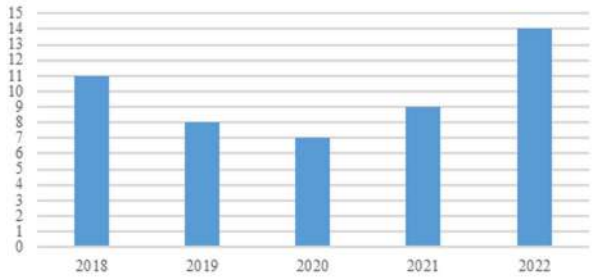


FIGURE 2. Distribution of primary studies by year.

TABLE 4. Year and type of primary studies.

Year	Type	List of References
2018	Conference	[26], [27], [50]
	Journal	[18], [21], [29], [32], [39], [40], [45], [51]
2019	Conference	[38]
	Journal	[1], [23], [33], [34], [35], [37], [47]
2020	Conference	-
	Journal	[12], [20], [22], [36], [46], [49], [52]
2021	Conference	[2], [4], [6], [7], [42], [48]
	Journal	[11], [28], [31]
2022	Conference	[3], [14]
	Journal	[5], [8], [9], [10], [13], [15], [16], [17], [25], [44], [43], [53]

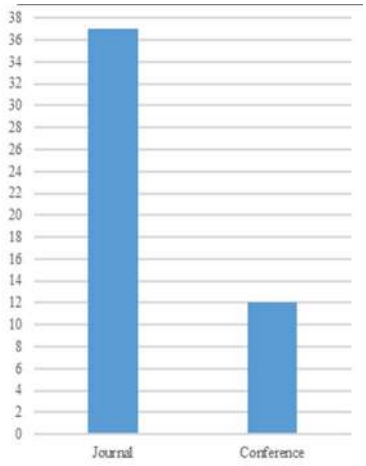


FIGURE 3. Distribution of primary studies by type.

We classified techniques listed in Table 5 by their focus (safety or cybersecurity) and analysis process (spreadsheet-based, scenario-based, tree-based, and model-based) and prepared a taxonomy shown in Fig. 4.

By spreadsheet-based process (Fig. 5) we mean an approach that gathers data into a single spreadsheet and the main deliverables (metrics, assessment results) are based on processing the spreadsheet data. A typical example of a spreadsheet-based process is a failure mode, effect, and diagnostic analysis (FMEDA), a systematic analysis technique to obtain subsystem/product level failure rates, failure modes, and diagnostic capability. The main purpose of FMEDA is to evaluate hardware architecture metrics and safety goal violations due to random hardware failures and provide sufficient information to improve safety gaps if the required hardware safety level is not fulfilled [54].

Another example of spreadsheet-based process is a risk-based inspection (RBI) which is well-established and used in the Oil& Gas and Chemical industries. This approach, along with risk-based maintenance, is described by API RP 581 [55], originally developed for application in the refining industry. The standard represents a correlation between maintenance activities and main events in the industries. RBI is also adapted and applied in many other sectors and inspection activities, allowing for the identification of failure mechanisms and rates based on equipment status.

Instead, tree-based techniques (Fig. 6) process graphical representation in the form of a tree. The classical example of a tree-based technique is a fault tree analysis (FTA) used for the reliability assessment of a system. FTA is a deductive

TABLE 5. Techniques used for safety and cybersecurity analysis.

Technique Id	Abbreviation	Technique Title
T1	FMEDA	Failure Modes, Effects, and Diagnostics Analysis
T2	FTA	Fault Tree Analysis
T3	BDMP	Boolean-driven Markov process
T4	HARA	Hazard Analysis and Risk Assessment
T5	RBD	Reliability Block Diagram
T6	RBI	Risk-based inspection
T7	ATA	Attack tree analysis
T8	PSA	Probabilistic safety assessment
T9	SM	Semi-Markov
T10	BN	Bayesian Networks
T11	MC	Monte-Carlo Simulation
T12	BA	Bowtie Analysis
T13	ETA	Event Tree Analysis
T14	FMEA / FMECA	Failure Modes and Effects Analysis / Failure Modes, Effects, and Criticality Analysis
T15	STAMP	Systems-Theoretic Accident Model and Process
T16	DVAG	Dynamic Vulnerability Assessment Graph
T17	MBAEM	Model-based Assurance Evidence Management.
T18	TARA	Threat Analysis and Risk Assessment
T19	IMECA	Intrusion Modes, Effects, and Criticality Analysis
T20	CRA	Cybersecurity Risk Assessment

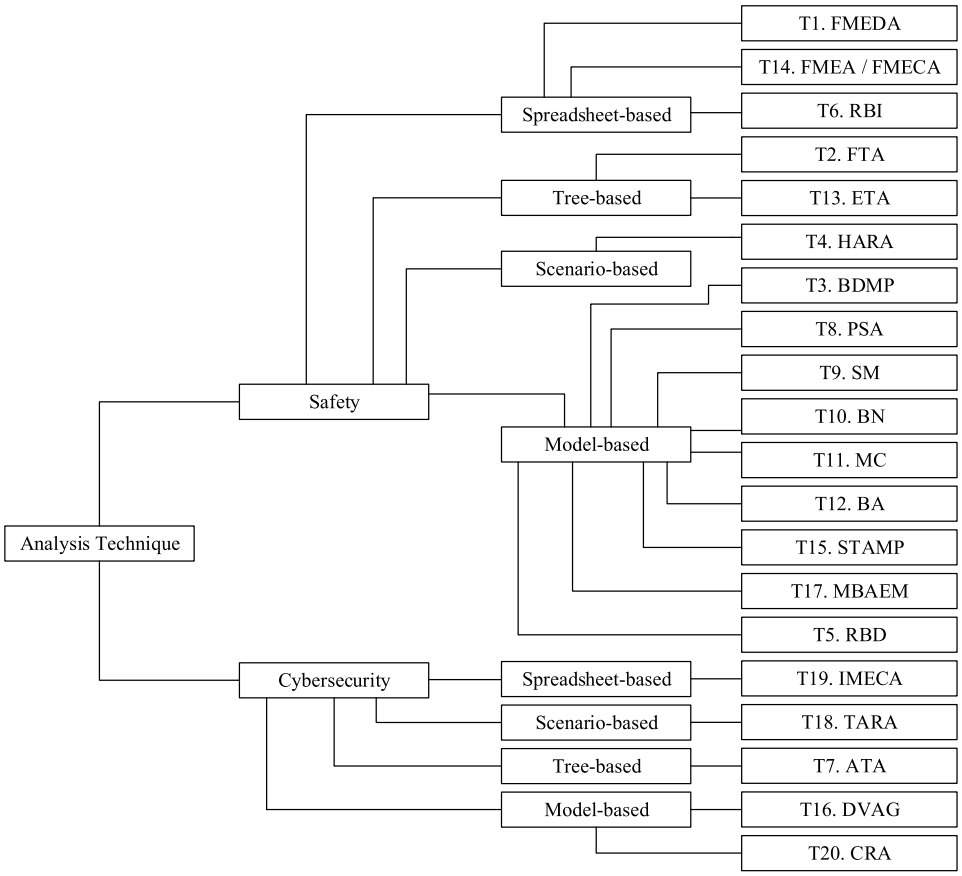


FIGURE 4. Taxonomy of analysis techniques.

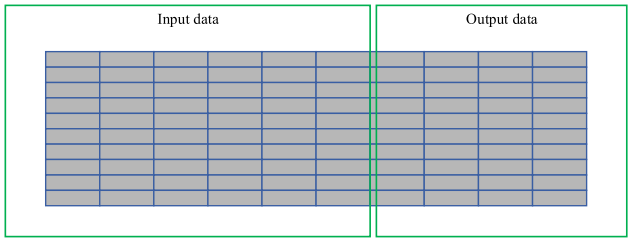


FIGURE 5. Spreadsheet-based technique.

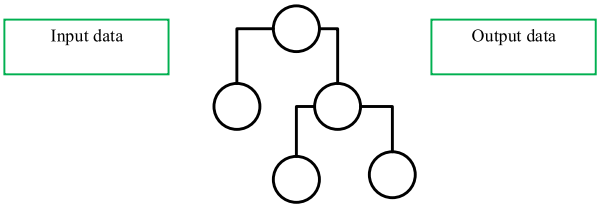


FIGURE 6. Tree-based technique.

process by means of which an undesirable event, called the top event, is postulated, and after that, the possible ways for this event to occur are systematically deduced. The deduction process is performed so that the fault tree embodies all

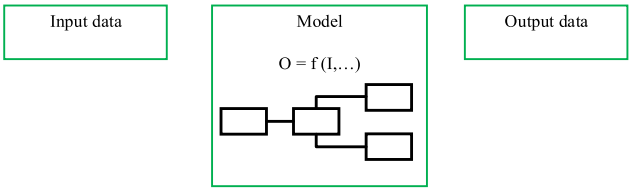


FIGURE 7. Model-based technique.

component failures (i.e., failure modes) that contribute to the occurrence of the top event. The fault tree itself is a graphical representation of the various combinations of failures that led to the occurrence of the top event [56]. In [45] it is proposed to apply FTA for cybersecurity assessment by using a model that integrates fault tree analysis, decision theory, and fuzzy theory to ascertain the current causes of cyberattack prevention failures and determine the vulnerability of a given cybersecurity system. Moreover, for cybersecurity assessment, another tree-based technique called attack tree analysis (ATA) is actively utilized [57].

By model-based techniques (Fig. 7) we mean approaches that perform an assessment using different models – graphs, equations etc.

For example, reliability block diagrams (RBD) represent sequences of system components and their connections. Each

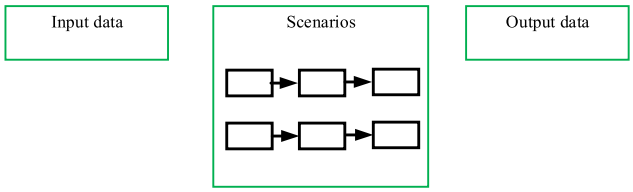


FIGURE 8. Scenario-based technique.

sequence consists of an input point and output point, several blocks representing system components, and the multiple paths from the input point to the output point that represent successful system operations, where an interruption of these paths may lead to the failure of the whole system. Therefore, an RBD model represents the static topology of I&C reliability, where the topology can be a serial, parallel or a combination of serial and parallel sections. Contrary to FTA, RBD models are success-oriented sequences that describe the function of a system by probabilistic means. Component blocks in an RBD are arranged to illustrate the proper combinations of working components that keep the entire system operational and, therefore, safe. Failure of a component can be represented by removing the component as well as its connections with other components from the sequence. When the number and position of failed components in the RBD model are such that there is no connection between the input and output point, the whole system fails.

Another example is a Bayesian network (BN) that represents a hypothesis of rationalizing from uncertain evidence to uncertain conclusions since it can perform the factorization of the collective distribution of variables, based on the conditional dependencies. BN helps to address uncertainty and incompleteness problems; thus, it is extensively applied in several domains. BNs are generally utilized for examining the hazards and vulnerabilities of networks, which are acyclic graphs that provide a quantitative and qualitative assessment of risks.

Model-based assurance evidence management (MBAEM) is another model-based technique that considers different activities for assurance evidence management, namely the determination of the evidence to provide, the possibility of reusing evidence, the collection of evidence information, tracing, evaluation, and change impact analysis of assurance evidence, and the use of the evidence for, e.g., compliance management and argumentation.

Finally, Probabilistic Safety Assessment (PSA) is the most common method to assess the risk of a nuclear power plant. It employs a graphical approach based on event and fault tree methods.

As for scenario-based processes (Fig. 8), they are typically based on profiling of scenarios collection obtained from different sources (accident, research, expert data, etc.) by focusing on safety and/or cybersecurity-relevant scenarios.

A typical example of a scenario-based process is HARA, where malfunctions and/or the functional insufficiencies are analyzed in terms of identification of both safety-relevant scenarios (known-safe and known-unsafe), as well as a set of unknown-unsafe scenarios, with further focus on required countermeasures.

The list of metrics is given in Table 6 and includes both safety and cybersecurity related metrics.

TABLE 6. Assessment metrics.

Metric Id	Metric Name
M1	SFF – Safe Failure Fraction
M2	SIL – Safety Integrity Level
M3	SPFM – Single-Point Fault Metric
M4	LFM – Latent (Multi-Point) Fault Metric
M5	PMHF – Probabilistic Metric for Hardware Failures
M6	PFH – Probability of Failure on Demand per Hour
M7	SL – Security Level
M8	ASIL – Automotive Safety Integrity Level
M9	Risk
M10	CDF – core damage frequency
M11	InTo-CSI – Intrusion Tolerance-based Cyber Security Index
M12	MTTC – Mean Time To Compromise
M13	CVSS – common vulnerability scoring system

SFF is a metric used to measure the likelihood of getting a dangerous failure that is not detected by diagnostics.

SIL is used to claim that all safety instrumented functions are operating satisfactorily under all stated conditions within a stated period of time.

SPFM is a hardware architectural metric used to show the sufficiency of safety mechanisms to prevent risk from single-point faults.

LFM is a hardware architectural metric used to show the sufficiency of safety mechanisms to prevent risk from latent faults.

PMHF is a probability of a safety goal violation caused by a random hardware failure.

PFH is the probability of dangerous failure that would prevent the system to be able to perform its safety function when required.

SL is a metric to measure how well a system component is protected from a certain level of threat and potential vulnerabilities.

ASIL is a risk classification metric.

CDF is a metric used to measure frequency and consequences considering initiating event frequency with system failure probabilities and fatalities (or environmental effects).

InTo-CSI is an index defined through relative comparison of two security states of the same system: a system without any cyber security controls, and a system with scrutiny controls.

The value of MTTC is the estimated figure of the time required for the valid attack assuming uniformly expended efforts.

CVSS is used to evaluate the severity of vulnerabilities, representing the virtual consequences on the vulnerable component in terms of confidentiality, integrity, and availability.

A considerable number of studies use Risk to represent the outputs of assessment technique utilization. In most cases, this metric represents the likelihood of the hazardous event and the severity of its consequences. Typical examples of what is used as a risk in the analysed studies are provided below. In [1], [4], [7], and [9] traditional risk priority number (RPN) is used, which is determined by three indicators: effect severity, occurrence probability, and detection difficulty. In [3] it is extended to cover risk interaction. In [5] risk assessment objectivity and accuracy are enhanced by the utilization of fuzzy confidence interval number (FCIN), generalized trapezoidal fuzzy numbers (GTrFN) evaluation model and the evaluation parameter sensitivity analysis. In [6] risk is calculated using the severity of the hazard, the exposure of that particular situation and the controllability of the system to mitigate hazardous situations. In [8] fairness risk is also considered separately from safety risk. In [10] special attention is given to considering assurance risks. In [12] risk is computed using potential risk impact due to vulnerabilities/attacks and the likelihood of the risk. In [13] risk includes attack cost, attack difficulty, and detected possibility.

C. USING PRIMARY STUDIES TO ANSWER RESEARCH QUESTIONS

To answer research questions (RQ1) and (RQ2), we have arranged the selected primary studies in the form of a table with the following columns (see Table 7):

- Reference;
- Techniques used (see Table 5);
- Metrics (see Table 6).

Based on the analysed studies, the resulting most popular techniques are listed in Table 8 below.

Therefore, the answer to RQ1 includes the following metrics: PFH, SFF, SIL, and ASIL are the most popular safety metrics. Also, in many studies, generic risk metric is used.

As for cybersecurity (RQ2), SL, MTTC, InTo-CSI, and CVSS scores are used as quantitative metrics. Just like with safety, the major part of studies considers generic risk metric more appropriate and comprehensive.

To answer research questions (RQ3) and (RQ4), we have arranged the list of selected primary studies in the form of a table with the following columns (see Table 10):

- Reference;
- Focus on safety;
- Focus on cybersecurity;
- Usage of several assessment techniques;
- Usage of modified assessment techniques;

TABLE 7. Techniques and metrics of primary studies.

Ref.	Techniques	Metrics
[1]	T14, T2	M9
[2]	T1, T2	M3, M4, M5
[3]	T14	M9
[4]	T14	M9
[5]	T14	M9
[6]	T4	M9
[7]	T14	M9
[8]	T14	M9
[9]	T14	M9
[10]	T17	M9
[11]	T18, T4	M8
[12]	T18	M9
[13]	T20	M9
[14]	T19	M9
[15]	T10	M13
[16]	T7	M9
[17]	T1,T14	M9
[18]	T11	M9
[20]	T8, T2, T13	M9, M10
[21]	T13	M11, M12
[22]	T9, T2, T13	M9
[23]	T2, T8, T13	M9
[25]	T16	M9
[26]	T8	M7, M9
[27]	T2	M9
[28]	T6	M9, M6
[29]	T10	M9
[31]	T15	M9
[32]	T10	M9
[33]	T10	M9
[34]	T16	M9
[35]	T10	M9
[36]	T10	M9
[37]	T2, T4, T14	M5, M6
[38]	T9, T10	M9
[39]	T12, T7	M9
[40]	T14	M9
[42]	T4, T18	M9
[43]	T2, T14	M9
[44]	T1	M1, M2, M3, M4, M5, M8
[45]	T2	M9
[46]	T2, T14	M8
[47]	T2	M9
[48]	T2	M9
[49]	T1	M9

TABLE 7. *(Continued.)* **Techniques and metrics of primary studies.**

[50]	T2, T9	M9
[51]	T2	M9
[52]	T5	M9
[53]	T10	M9

TABLE 8. **The most used techniques for safety and cybersecurity assessment.**

Technique Id	Number of references	References
T2	10	[1], [2], [20], [22], [23], [46], [47], [48], [50], [51]
T14	9	[1], [3], [4], [5], [7], [8], [9], [17], [37], [46]
T10	7	[29], [32], [33], [35], [36], [38], [53]
T1	4	[2], [17], [44], [49]
T13	4	[20], [21], [22], [23]
T9	3	[22], [38], [50]
T8	3	[20], [23], [26]

TABLE 9. **Types of case studies.**

Case study type Id	Case study type
C0	No case study provided.
C1	The provided case study is only theoretical (formulas are provided, but no calculations are performed).
C2	The provided case study is demonstrated using a simulated environment and artificial input values.
C3	The provided case study is demonstrated using a simulated environment, but real input values are used.
C4	The provided case study demonstrates application on a real system with real values used.

- Generalization (i.e. utilization of techniques initially designed for safety assessment to assess cybersecurity with minor modifications of the technique itself) of assessment techniques;
- Availability of case study and its type according to Table 9 below.

The list of possible types of case studies was prepared after a preliminary analysis of primary studies. Types and corresponding identifiers are provided in Table 9.

For safety assessment (RQ3), modifications of well-known reliability assessment techniques like FMEA/FMECA, FTA, and Bayesian networks are mostly used.

As for cybersecurity (RQ4), either specific modifications are utilized (like IMECA), or in most cases cybersecurity assessment is integrated into the overall safety assessment process. In most cases, the assessment process is risk-based, including risk identification, risk analysis, risk evaluation, and documentation.

The main limitations identified (RQ5) include dimension issues (the approach is not applicable due to a huge number of components to be analyzed) and too strict assumptions (like independent failures or attacks). To overcome such limitations, modifications to methodologies used are being introduced, for example, focusing only on elements that are part of the safety function for complex safety systems, etc.

V. KEY FINDINGS

The discussion on key findings focuses primarily on the most interesting results regarding the adopted assessment techniques, namely their combined usage, proposed modifications, and attempts toward generalization. A few other general findings are also highlighted.

A. USE OF SEVERAL ASSESSMENT TECHNIQUES

As shown in Table 7, altogether 28 studies were focusing on several assessment techniques utilization. The main motivation to use several techniques derives from the fact that the results of one technique usually either don't cover all the non-functional aspects of interest (i.e. the technique is focused on safety and doesn't consider cybersecurity issues) or need to be verified through a different technique (i.e. different techniques are used in parallel and then the obtained results are being compared and processed).

Though cybersecurity analysis is implemented in the overall I&C design procedure, it is generally not combined with the safety analysis development. In several analysed studies, the introduced approaches comprehended the significance of integrated safety and cybersecurity analysis and intended to incorporate both into a joint methodological process. For instance, two applicable techniques, which describe the integration of cybersecurity into safety analysis (cybersecurity-informed safety, or security-informed safety), recommend a merging of fault tree analysis (FTA) with attack tree analysis (ATA) or Boolean-driven Markov processes (BDMP). Other introduced approaches either combine safety and cybersecurity methods, e.g., ATA and bowtie analysis, or integrate both fields (i.e. implement strategies devoted to ''unintentional'' (safety) events as well as to ''intentional'' (cybersecurity) chains).

In [42], the scenario-based approach utilizing HARA and TARA techniques is pursued. In particular, correlation of damage scenario and hazard scenario is performed, so as to show the connection of safety with cybersecurity.

The authors of [37] present a framework for performing safety analyses, risk assessment, and safety requirements management using semi-formal and formal techniques like FMEA, FMECA, and FTA. The framework implements a compositional V-cycle methodology, covering all phases of the system development lifecycle. Future integration of other assessment techniques into the framework is planned by the authors.

TABLE 10. The focus of the primary studies.

Ref.	Safety	Cybersecurity	Several assessment techniques	Modification of assessment techniques	Assessment technique generalization	Availability of case study
[1]	√		√	√		C1
[2], [46]	√		√			C0
[3], [8], [9], [33], [48], [49]	√			√		C1
[4],[22]	√			√		C3
[5], [7]	√			√		C2
[6], [38]	√		√			C2
[10]	√		√	√	√	C2
[11], [17]	√	√	√		√	C1
[12], [51]		√	√	√		C1
[13], [14], [26]		√	√	√		C2
[15], [18], [21], [25]		√		√		C2
[16], [36], [53]		√	√			C1
[20]		√	√			C2
[23], [40], [45]		√		√		C1
[27]	√	√		√		C2
[28], [29], [32]	√		√			C1
[31]	√	√	√		√	C2
[34]	√	√	√	√	√	C1
[35]	√	√	√			C1
[37], [47]	√		√	√		C1
[39]	√	√		√	√	C1
[42]	√	√	√			C2
[43]	√	√	√	√	√	C2
[44], [50]	√		√	√		C2
[52]	√	√	√	√		C1

B. MODIFICATION OF ASSESSMENT TECHNIQUES

In 33 studies, listed in Table 10, modifications of assessment techniques are considered. Among the reasons of modification, the following are mentioned: dimension problem of the technique, reduction of resources required to perform the analysis, and application of well-known approaches to different domains.

In [5] FMEA is modified by the introduction of the risk evaluation methodology for controlling multi-uncertainties in the assessment process. It is shown that the proposed methodology can significantly improve the risk assessment results and the risk discrimination of failure modes, but at the current stage controlling only a single uncertainty is implemented.

Authors of [4] propose a novel approach to calculate risk priority numbers based on factors like severity, occurrence, and detection during the application of FMEA, and outline that classical FMEA only considers risk factors regarding safety, ignoring other factors (i.e. cybersecurity or economic impacts).

In [27] initial events for FTA include not only safety-related issues like failures in components or subsystems but also cybersecurity ones like attacks.

It cannot be too highly stressed that several reviewed studies provide evidence that methods originally intended for reliability assessment could be successfully utilized for safety and/or cybersecurity assessment with minor modifications. For example, the probabilistic risk assessment method which is the most general method to get the risk information could be applied to cybersecurity, safety block diagrams, and cybersecurity block diagrams, etc.

Finally, on the aspect of safety and cybersecurity protection mechanisms, it is suggested that they could be based on recent technologies successfully used in other sectors, such as blockchain technology [41], [52].

C. ASSESSMENT TECHNIQUES GENERALIZATION

Generalization of assessment techniques is addressed only in 7 studies but looks as a promising direction for research. The main idea is to develop generic approaches that could be parametrized, so as to be 'tuned' to a required domain or set of metrics. The relatively limited number of studies could be explained by the complexity of such task and the amount of resources needed to provide representative case studies.

In [43], a hybrid ontology is presented that could be utilized for safety and cybersecurity assessment. The authors claim that a true combined approach also needs to include dependability engineering to harmonize the basic concepts between all three disciplines: safety, cybersecurity and dependability. It is also highlighted that focusing on cybersecurity risks requires more effort compared to safety risk analyses due to risk nature: safety risks are based on systematic faults or quite well-known random faults and allow implementation of a systematic assessment approach, while cybersecurity risks are mainly caused by malicious acts which originate a huge number of possible threat scenarios.

The authors of [31] propose an ontological metamodel that considers safety, cybersecurity, and resiliency. Co-engineering of safety and cybersecurity is based on a system losses approach, i.e. system losses caused either by safety or cybersecurity violations are prioritized so as to provide a structured approach for their mitigation. It is claimed that such an approach allows achieving an overall increase in scalability, usability, and unification of already existing models.

In [17] a generic XMECA (FMECA + IMECA = XMECA) technique is presented, intended to cover different domains – safety and cybersecurity – using a unified approach. Verification of XMECA results is performed using EUMECA (E – error, U – uncertainty) with a focus on decisions and judgments made by experts during the XMECA process.

D. METRICS AND CASE STUDIES

Techniques used in a majority of the analysed studies are tailored to risk assessment (risk-based approach), covering only failures, only vulnerabilities, or covering both of them.

Some cybersecurity risk assessment methods with application on real I&C systems are based on national standards. An example is the Chinese national standard GB/T 36466-2018: Information security technology-Implementation guide [58]. According to this document, four risk elements including asset, threat, vulnerability, and protection capability would be first identified and assessed adopting a combination of qualitative methods of expert evaluation and quantitative methods of numerical calculation. Possibility of, and loss from, security incidents then would be calculated through the above four elements and, finally, the risk value is obtained.

Aiming at providing an internationally valid reference methodology, a common international method for combined safety and security modeling, design and assessment is an open and active research topic.

The major part of the case studies presented in the reviewed publications are theoretical ones or taken from realistic contexts but adopting artificial inputs (case studies classified as C1 and C2 in Table 9). Although application to real systems would be highly desirable, this is not expected to be possible in the foreseeable time due to the limitations stated. Indeed, the assumptions adopted to make the technique manageable (e.g., with reference to scalability) are sources of inaccuracy in the obtained results when analyzing realistic systems that do not fully adhere to such assumptions. Devising assessment techniques suitable to deal with real system contexts is an active, challenging research direction.

E. GENERAL FINDINGS

The performed review shows that the focus of recent publications is more on cybersecurity and less on safety as a whole. This could be explained by the modernization of control systems in critical industries, especially towards more flexibility, but a drawback is that new potential cybersecurity issues are introduced.

With the integration of information systems and physical systems, the cybersecurity of information systems and functional safety of physical systems interact with each other, resulting in a type of new comprehensive problem and introducing serious risks. New approaches addressing this issue are needed.

Existing technologies of the I&C system, including programmable logic controllers (PLCs) and FPGA-based platforms, are vulnerable as they are attractive targets for the cyberattack threats. Appropriate risk assessment that includes not only failure analysis and reliability issues but possible intrusions can strongly contribute to enhancing cybersecurity and safety, by providing support to the development of preventive measures in avoiding/mitigating potential cyberattacks.

VI. THREATS TO THE VALIDITY OF THIS STUDY

In this section, we discuss major threats to the validity of this mapping study.

The possibility exists that some relevant studies have not been chosen due to the expertise of the authors. We mitigated this threat, as much as possible, by examining the titles, abstract, and keywords at the first stage and going deeper into

the checks at the second stage, following the steps shown in Figure 1. Moreover, several meetings have been carried out during the selection process, to discuss possible doubts.

Another potential threat relates to the defined search string, since a different set of primary studies may be derived with even slight variation of the search string. This threat characterizes all systematic surveys. To mitigate it, we discussed in depth the goal of the planned study, for which clear and relevant research questions were then identified and used to build the search question.

Regarding the quality of reviewed studies, we did not adopt any specific quality criteria, as usually recommended when performing systematic literature reviews and mapping studies. However, we excluded studies that had not undergone a peer-review process, thus assuring the scientific quality of the selected papers.

Potential issues on generalization of the obtained results constitute another threat that is common to all the mapping studies. While it is not feasible to generalize the drawn conclusions to the whole universe of primary studies on a specific topic, to mitigate this threat we considered only primary studies published during the last 5 years, thus focusing mainly on current trends in the field.

VII. CONCLUSION

This mapping study analysed 49 papers dealing with cybersecurity and safety assessment. Major concluding points include:

- It is observed that out of the 49 included studies, 16 focus on cybersecurity only, 23 focus on safety only, and the remaining 10 are based on a joint approach to safety and cybersecurity. This distribution trend testifies that needs in the different application domains are rather wide in terms of metrics of primary interest.
- It should be particularly emphasized that the majority of techniques used in studies were either based on simulation analysis or theoretical concepts.
- A great majority of the studies (33 out of 49) propose modifications/extensions of classical assessment techniques, either to address joint safety and cybersecurity analysis, or to accommodate new needs of the application context. This trend shows that classical assessment techniques, well consolidated by long-lasting practice, are still very popular and constitute a basis for enhancements to satisfy more sophisticated analysis needs.

The results of the performed survey indicate the lack of a systematic process of unified safety and cybersecurity assessment.

Among future research directions for safety and cybersecurity integration:

- There is a clear need in putting efforts into developing a generic technique (method or standard) supported by tool to combine cybersecurity and safety, which can be helpful for different applications in critical industries, since the significance of integrating both measures was demonstrated in this mapping study, and a generic approach may offer benefits such as feasibility and flexibility.
- It is observed that there are various approaches for evaluating the indicators of interest, including the usage of different assessment techniques and comparison of their outputs for validation purposes. A more extended investigation is necessary to estimate the accuracy and efficiency of assessment mechanisms, in order to find the optimal option to employ in a specific context, guided by criteria of accuracy and cost.

REFERENCES

[1] X. Zhang, Y. Li, Y. Ran, and G. Zhang, "A hybrid multilevel FTA-FMEA method for a flexible manufacturing cell based on meta-action and TOPSIS," *IEEE Access*, vol. 7, pp. 110306–110315, 2019, doi: 10.1109/ACCESS.2019.2934189.

[2] C. Kymal and O. G. Gruska, "Integrating FMEAs, FMEDAs, and fault trees for functional safety," in *Proc. Annu. Rel. Maintainability Symp. (RAMS)*, May 2021, pp. 1–6, doi: 10.1109/RAMS48097.2021.9605786.

[3] P. Liu, Y. Xu, and Y. Li, "An improved failure mode and effect analysis model for automatic transmission risk assessment considering the risk interaction," *IEEE Trans. Rel.*, early access, Oct. 27, 2022, doi: 10.1109/TR.2022.3215110.

[4] S. K. Akula and H. Salehfar, "Risk-based classical failure mode and effect analysis (FMEA) of microgrid cyber-physical energy systems," in *Proc. North Amer. Power Symp. (NAPS)*, College Station, TX, USA, Nov. 2021, pp. 1–6, doi: 10.1109/NAPS52732.2021.9654717.

[5] Y. Liu, B. Chen, Q. Dong, W. Liu, W. Nie, and C. Yang, "Failure mode risk assessment methodology for controlling multi-uncertainties in the evaluation process," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105470, doi: 10.1016/j.engappai.2022.105470.

[6] A. R. Patel and P. Liggesmeyer, "Machine learning based dynamic risk assessment for autonomous vehicles," in *Proc. Int. Symp. Comput. Sci. Intell. Controls (ISCSIC)*, Rome, Italy, Nov. 2021, pp. 73–77, doi: 10.1109/ISCSIC54682.2021.00024.

[7] L. Pokoradi, S. Kocak, and E. Toth-Laufer, "Fuzzy hierarchical failure mode and effect analysis," in *Proc. IEEE 19th Int. Symp. Intell. Syst. Informat. (SISY)*, Sep. 2021, pp. 71–76, doi: 10.1109/SISY52375.2021.9582523.

[8] J. Li and M. Chignell, "FMEA-AI: AI fairness impact assessment using failure mode and effects analysis," *AI Ethics*, vol. 2, no. 4, pp. 837–850, Nov. 2022, doi: 10.1007/s43681-022-00145-9.

[9] S. E. Fatollah, R. Dabbagh, and A. S. Jalavat, "An extended approach using failure modes and effects analysis (FMEA) and weighting method for assessment of risk factors in the petrochemical industry," *Environ., Develop. Sustainability*, pp. 1–26, Oct. 2022, doi: 10.1007/s10668-022-02609-8.

[10] J. L. de la Vara, A. S. García, J. Valero, and C. Ayora, "Model-based assurance evidence management for safety–critical systems," *Softw. Syst. Model.*, vol. 21, no. 6, pp. 2329–2365, Dec. 2022, doi: 10.1007/s10270-021-00957-z.

[11] C. Schwarzl, N. Marko, H. Martin, V. E. Jiménez, J. C. Triginer, B. Winkler, and R. Bramberger, "Safety and security co-engineering for highly automated vehicles," *Elektrotechnik Informationstechnik*, vol. 138, no. 7, pp. 469–479, Nov. 2021, doi: 10.1007/s00502-021-00934-w.

[12] K. Kandasamy, S. Srinivas, K. Achuthan, and V. P. Rangan, "IoT cyber risk: A holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process," *EURASIP J. Inf. Secur.*, vol. 2020, no. 1, pp. 1–18, Dec. 2020, doi: 10.1186/s13635-020-00111-0.

[13] H. Guo, L. Ding, and W. Xu, "Cybersecurity risk assessment of industrial control systems based on order—A divergence measures under an interval-valued intuitionistic fuzzy environment," *IEEE Access*, vol. 10, pp. 43751–43765, 2022, doi: 10.1109/ACCESS.2022.3169133.

[14] A. Abakumov and V. Kharchenko, "Combining IMECA analysis and penetration testing to assess the cybersecurity of industrial robotic systems," in *Proc. 12th Int. Conf. Dependable Syst., Services Technol. (DESSERT)*, Athens, Greece, Dec. 2022, pp. 1–7, doi: 10.1109/DESSERT58054.2022.10018823.

[15] Y. Wang, B. Yu, H. Yu, L. Xiao, H. Ji, and Y. Zhao, "Automotive cybersecurity vulnerability assessment using the common vulnerability scoring system and Bayesian network model," *IEEE Syst. J.*, vol. 17, no. 2, pp. 2880–2891, Jun. 2023, doi: 10.1109/JSYST.2022.3230097.

[16] S.-G. Tân, I.-H. Liu, and J.-S. Li, "Threat analysis of cyber security exercise for reservoir testbed based on attack tree," in *Proc. 10th Int. Symp. Comput. Netw. Workshops (CANDARW)*, Himeji, Japan, Nov. 2022, pp. 375–379, doi: 10.1109/CANDARW57323.2022.00023.

[17] I. Babeshko, O. Illiashenko, V. Kharchenko, and K. Leontiev, "Towards trustworthy safety assessment by providing expert and tool-based XMECA techniques," *Mathematics*, vol. 10, no. 13, p. 2297, Jun. 2022, doi: 10.3390/math10132297.

[18] W. Wang, A. Cammi, F. D. Maio, S. Lorenzi, and E. Zio, "A Monte Carlo-based exploration framework for identifying components vulnerable to cyber threats in nuclear power plants," *Rel. Eng. Syst. Saf.*, vol. 175, pp. 24–37, Jul. 2018.

[19] J. Peterson, M. Haney, and R. A. Borrelli, "An overview of methodologies for cybersecurity vulnerability assessments conducted in nuclear power plants," *Nucl. Eng. Des.*, vol. 346, pp. 75–84, May 2019.

[20] J. W. Park and S. J. Lee, "A quantitative assessment framework for cyber-attack scenarios on nuclear power plants using relative difficulty and consequence," *Ann. Nucl. Energy*, vol. 142, Jul. 2020, Art. no. 107432.

[21] C. Lee, H. B. Yim, and P. H. Seong, "Development of a quantitative method for evaluating the efficacy of cyber security controls in NPPs based on intrusion tolerant concept," *Ann. Nucl. Energy*, vol. 112, pp. 646–654, Feb. 2018.

[22] Y. Zhao, L. Huang, C. Smidts, and Q. Zhu, "Finite-horizon semi-Markov game for time-sensitive attack response and probabilistic risk assessment in nuclear power plants," *Rel. Eng. Syst. Saf.*, vol. 201, Sep. 2020, Art. no. 106878.

[23] J. W. Park and S. J. Lee, "Probabilistic safety assessment-based importance analysis of cyber-attacks on nuclear power plants," *Nucl. Eng. Technol.*, vol. 51, no. 1, pp. 138–145, Feb. 2019.

[24] N. Chowdhury, "CS measures for nuclear power plant protection: A systematic literature review," *Signals*, vol. 2, no. 4, pp. 803–819, Nov. 2021, doi: 10.3390/signals2040046.

[25] A. Boudermine, R. Khatoun, and J.-H. Choyer, "Attack graph-based solution for vulnerabilities impact assessment in dynamic environment," in *Proc. 5th Conf. Cloud Internet Things (CIoT)*, Marrakesh, Morocco, Mar. 2022, pp. 24–31, doi: 10.1109/CIoT53061.2022.9766588.

[26] D. Liu, Y. Chen, J. Shi, and D. Chen, "Study on cyber security risk assessment of digital instrumentation & control system of nuclear power plant," in *Proc. Int. Conf. Power Syst. Technol. (POWERCON)*, Guangzhou, China, Nov. 2018, pp. 4742–4750.

[27] R. B. Ferreira, D. M. Baum, E. C. P. Neto, M. R. Martins, J. R. Almeida, P. S. Cugnasca, and J. B. Camargo, "A risk analysis of unmanned aircraft systems (UAS) integration into non-segregate airspace," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Dallas, TX, USA, Jun. 2018, pp. 42–51, doi: 10.1109/ICUAS.2018.8453455.

[28] B. H. Davatgar, N. Paltrinieri, and R. Bubbico, "Safety barrier management: Risk-based approach for the oil and gas sector," *J. Mar. Sci. Eng.*, vol. 9, no. 7, p. 722, Jun. 2021, doi: 10.3390/jmse9070722.

[29] M. Bucelli, N. Paltrinieri, and G. Landucci, "Integrated risk assessment for oil and gas installations in sensitive areas," *Ocean Eng.*, vol. 150, pp. 377–390, Feb. 2018.

[30] S. Pirbhulal, V. Gkioulos, and S. Katsikas, "Towards integration of security and safety measures for critical infrastructures based on Bayesian networks and graph theory: A systematic literature review," *Signals*, vol. 2, no. 4, pp. 771–802, 2021, doi: 10.3390/signals2040045.

[31] G. Bakirtzis, T. Sherburne, S. Adams, B. M. Horowitz, P. A. Beling, and C. H. Fleming, "An ontological metamodel for cyber-physical system safety, security, and resilience coengineering," *Softw. Syst. Model.*, vol. 21, no. 1, pp. 113–137, Feb. 2022, doi: 10.1007/s10270-021-00892-z.

[32] Y. Zhou, C. Li, C. Zhou, and H. Luo, "Using Bayesian network for safety risk analysis of diaphragm wall deflection based on field data," *Rel. Eng. Syst. Saf.*, vol. 180, pp. 152–167, Dec. 2018.

[33] H. Xu, Y. Zhang, H. Li, M. Skitmore, J. Yang, and F. Yu, "Safety risks in rail stations: An interactive approach," *J. Rail Transp. Planning Manage.*, vol. 11, Oct. 2019, Art. no. 100148.

[34] C. Chen, G. Reniers, and N. Khakzad, "Integrating safety and security resources to protect chemical industrial parks from man-made domino effects: A dynamic graph approach," *Rel. Eng. Syst. Saf.*, vol. 191, Nov. 2019, Art. no. 106470.

[35] N. U. I. Hossain, R. Jaradat, S. Hosseini, M. Marufuzzaman, and R. K. Buchanan, "A framework for modeling and assessing system resilience using a Bayesian network: A case study of an interdependent electrical infrastructure system," *Int. J. Crit. Infrastruct. Protection*, vol. 25, pp. 62–83, Jun. 2019.

[36] R. Arief, N. Khakzad, and W. Pieters, "Mitigating cyberattack related domino effects in process plants via ICS segmentation," *J. Inf. Secur. Appl.*, vol. 51, Apr. 2020, Art. no. 102450.

[37] M. Adedjouma and N. Yakymets, "A framework for model-based dependability analysis of cyber-physical systems," in *Proc. IEEE 19th Int. Symp. High Assurance Syst. Eng. (HASE)*, Hangzhou, China, Jan. 2019, pp. 82–89, doi: 10.1109/HASE.2019.00022.

[38] M. Galagedarage Don and F. Khan, "Process fault prognosis using hidden Markov model–Bayesian networks hybrid model," *Ind. Eng. Chem. Res.*, vol. 58, no. 27, pp. 12041–12053, Jul. 2019.

[39] H. Abdo, M. Kaouk, J.-M. Flaus, and F. Masse, "A safety/security risk analysis approach of industrial control systems: A cyber bowtie—Combining new version of attack tree with bowtie analysis," *Comput. Secur.*, vol. 72, pp. 175–195, Jan. 2018.

[40] A. Asllani, A. Lari, and N. Lari, "Strengthening information technology security through the failure modes and effects analysis approach," *Int. J. Quality Innov.*, vol. 4, no. 1, pp. 1–14, Dec. 2018, doi: 10.1186/s40887-018-0025-1.

[41] R. Gupta, S. Tanwar, N. Kumar, and S. Tyagi, "Blockchain-based security attack resilience schemes for autonomous vehicles in industry 4.0: A systematic review," *Comput. Electr. Eng.*, vol. 86, Sep. 2020, Art. no. 106717.

[42] M. Khatun, M. Glaß, and R. Jung, "An approach of scenario-based threat analysis and risk assessment over-the-air updates for an autonomous vehicle," in *Proc. 7th Int. Conf. Autom., Robot. Appl. (ICARA)*, Prague, Czech Republic, Feb. 2021, pp. 122–127, doi: 10.1109/ICARA51699.2021.9376542.

[43] J. Alanen, J. Linnosmaa, T. Malm, N. Papakonstantinou, T. Ahonen, E. Heikkilä, and R. Tiusanen, "Hybrid ontology for safety, security, and dependability risk assessments and security threat analysis (STA) method for industrial control systems," *Rel. Eng. Syst. Saf.*, vol. 220, Apr. 2022, Art. no. 108270, doi: 10.1016/j.ress.2021.108270.

[44] K.-L. Lu and Y.-Y. Chen, "Safety-oriented system hardware architecture exploration in compliance with ISO 26262," *Appl. Sci.*, vol. 12, no. 11, p. 5456, May 2022, doi: 10.3390/app12115456.

[45] A. P. H. D. Gusmão, M. M. Silva, T. Poleto, L. C. E. Silva, and A. P. C. S. Costa, "Cybersecurity risk analysis model using fault tree analysis and fuzzy decision theory," *Int. J. Inf. Manage.*, vol. 43, pp. 248–260, Dec. 2018.

[46] G. Xie, Y. Li, Y. Han, Y. Xie, G. Zeng, and R. Li, "Recent advances and future trends for automotive functional safety design methodologies," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5629–5642, Sep. 2020.

[47] M. Ghadhab, S. Junges, J.-P. Katoen, M. Kuntz, and M. Volk, "Safety analysis for vehicle guidance systems with dynamic fault trees," *Rel. Eng. Syst. Saf.*, vol. 186, pp. 37–50, Jun. 2019.

[48] S. Atsushi, "A framework for performing quantitative fault tree analyses for subsystems with periodic repairs," in *Proc. Annu. Rel. Maintainability Symp. (RAMS)*, Orlando, FL, USA, May 2021, pp. 1–6.

[49] J. Famfulik, M. Richtar, R. Rehak, J. Smiraus, P. Dresler, M. Fusek, and J. Mikova, "Application of hardware reliability calculation procedures according to ISO 26262 standard," *Qual. Rel. Eng. Int.*, vol. 36, no. 6, pp. 1822–1836, Oct. 2020.

[50] T. Wang, X. Chen, Z. Cai, J. Mi, and X. Lian, "A mixed model to evaluate random hardware failures of whole-redundancy system in ISO 26262 based on fault tree analysis and Markov chain," *Proc. Inst. Mech. Eng. D, J. Automobile Eng.*, vol. 233, no. 4, pp. 890–904, Mar. 2019.

[51] C.-S. Cho, W.-H. Chung, and S.-Y. Kuo, "Using tree-based approaches to analyze dependability and security on I&C systems in safety-critical systems," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1118–1128, Jun. 2018, doi: 10.1109/JSYST.2016.2635681.

[52] A. Gu, Z. Yin, C. Cui, and Y. Li, "Integrated functional safety and security diagnosis mechanism of CPS based on blockchain," *IEEE Access*, vol. 8, pp. 15241–15255, 2020, doi: 10.1109/ACCESS.2020.2967453.

[53] Y. Tian, J. Li, and X. Huang, "A cybersecurity risk assessment method and its application for instrumentation and control systems in nuclear power plants," *IFAC-PapersOnLine*, vol. 55, no. 9, pp. 238–243, 2022, doi: 10.1016/j.ifacol.2022.07.042.

[54] *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems (E/E/PE, or E/E/PES)*, Standard IEC 61508, 2010.

[55] *Risk-Based Inspection Methodology*, Standard API RP 581, 3rd ed., Oct. 2020.
[56] *Fault Tree Analysis (FTA)*, IEC 61025, 2006.
[57] C. E. Budde, C. Kolb, and M. Stoelinga, "Attack trees vs. fault trees: Two sides of the same coin from different currencies," in *Quantitative Evaluation of Systems* (Lecture Notes in Computer Science), vol. 12846. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-85172-9_24.
[58] *Information Security Technology—Implementation Guide to Risk Assessment of Industrial Control Systems*, Standard GB/T 36466-2018, 2018.
[59] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, Jan. 2022, Art. no. 101895, doi: 10.1016/j.mex.2022.101895.
[60] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, Aug. 2015, doi: 10.1016/j.infsof.2015.03.007.

IEVGEN BABESHKO is currently a Graduate Fellow with the Software Engineering & Dependable Computing Laboratory, Institute of Information Science and Technologies "Alessandro Faedo," and an Associate Professor with the Computer Systems, Networks and Cybersecurity Department, National Aerospace University "Kharkiv Aviation Institute." He is also the Head of the Functional Safety Division, Ukrainian Technical Committee TC185 "Industrial Automation." He covered a contributor roles in several European projects, including TEMPUS/ERASMUS+ (MASTAC, SAFEGUARD, SEREIN, CABRIOLET, CERES, and ALIOT) and Horizon 2020 (ECHO). He is involved as a regular member of Program Committee of IEEE DESSERT Conference. He is the coauthor of more than 50 scientific papers and reports, including ten monographs. His professional and research interests include reliability, safety, cybersecurity assessment, assurance and certification of industrial control systems, the dependability and resilience of IIoT systems, and academia-industry cooperation.

FELICITA DI GIANDOMENICO is currently the Research Director of ISTI-CNR, Pisa, Italy, where she is also leading the Software Engineering and Dependable Computing Research Laboratory. Her research interests include the design of dependable computing systems, software implemented fault/intrusion tolerance, and the modeling and evaluation of dependability attributes, with a focus on critical infrastructures. She covered the role of a principal investigator of CNR and/or the Work-Package Leader in several European projects (including Caution++, CRUTIAL, SAFEDMI, CHESS, and SmartC2Net) and national projects (more recently, TENACE). She has been the Chair of the IEEE Technical Committee on Dependable Computing and Fault Tolerance, from January 2017 to December 2018, and the Chair of the IEEE/IFIP DSN Steering Committee, from January 2017 to December 2018. She is routinely involved in program committee of the most relevant conferences in the dependability area. She was the Program Co-Chair of SRDS 2008, DSN 2009, SAFECOMP 2014, and SERENE 2019. She is a member of the IFIP WG10.4 on Dependable Computing and Fault Tolerance and a member of the Steering Committee of the Conferences IEEE/IFIP DSN and EDCC.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 November 2023, accepted 28 January 2024, date of publication 1 February 2024, date of current version 7 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361039

TOPICAL REVIEW

A Comprehensive Review on Cyber-Attacks in Power Systems: Impact Analysis, Detection, and Cyber Security

NAVEEN TATIPATRI AND S. L. ARUN
School of Electrical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
Corresponding author: S. L. Arun (arun.sl@vit.ac.in)

This work was supported by the Office of Dean, Academic Research Velllore Institute of Technology (VIT), Vellore.

ABSTRACT Continuous communication and information technology advancements facilitate the modernization of the conventional energy grid into an integrated platform. Internet-of-Things (IoT) incorporates power systems, particularly smart grid features and the delivery of new services from the utility side to the end user over a two-way communication channel. However, severe security vulnerabilities have been created due to over-dependency on IoT based communication systems. In addition, critical information exchange between any two entities or devices is always an appealing target for cyber-attackers, especially with financial interest motive by damaging integrity, confidentiality and authenticity in a communication channel. Maintaining data security and preserving privacy in between two entities during the transmission or any data distribution are essential. The potential attacks and impacts of those attacks need to be investigated to develop an effective cyber security infrastructure. Thus, considerable researchers focused on detection and mitigation of these vulnerable cyber-attacks using advanced computation tools. This review article thoroughly investigated possible ways to address cyber security challenges such as smart meter security, end-users privacy, electricity theft cyber-attacks using blockchain and cryptography against communication attacks in smart grid. The operational impacts of cyber-attacks on power system security, as well as the economic impact on deregulated energy markets, have been extensively explored. In addition, the robustness of security features and cryptographic methods against various cyber-attacks is investigated to suggest unexplored cyber-attacks for future scope. Specially, the study of real-world cyber security events, case studies, new findings and new scopes in diverse power industries are carried out. More than 135 research articles has been examined for this review article. This paper mainly concentrates on distribution-side cyber-attacks with impact analysis, detection and protection techniques.

INDEX TERMS Cyber attacks, cyber security, cryptography, Internet of Things, power systems, smart grid.

I. INTRODUCTION

In recent accomplishments, integration based on Machine-to-Machine (M2M) communication and widespread application of IoT communication technology played a vital role in smart grid. Incorporating IoT into a smart grid enables seamless interactions throughout all energy sectors such as generation, transmission and distribution, [1]. A traditional grid has a mechanized one-way communication infrastructure with fewer sensors. In contrast, a smart grid has digital two-way communication with more sensors. While adopting a future power system incorporating IoT provides effective billing, improved corrective capabilities during failures and enhanced operational efficiency [2].

Using a two-way communication channel, consumers and service providers communicate via smart meters, sensors, Advanced Metering Infrastructure (AMI), meter data management systems, and utility servers [4]. A smart grid includes an intelligent monitoring system that monitors all electricity flowing through the system which can more

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaosong Hu.

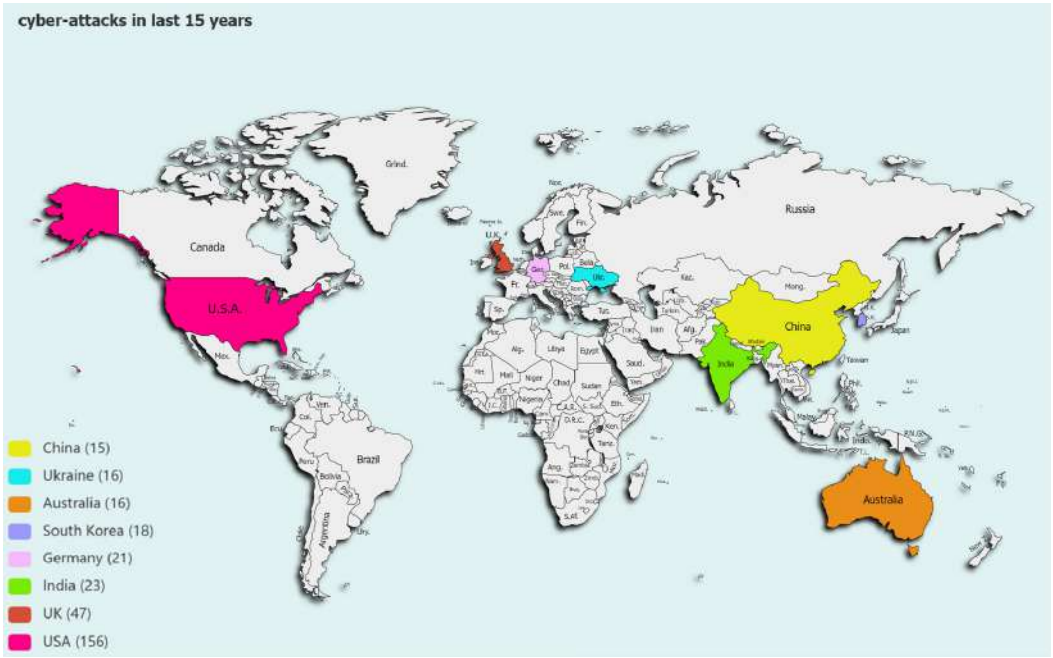


FIGURE 1. Reported cyber-attack incidence in last 15 years around the world [3].

efficiently balance the power flow, detect surges, outages and technical energy losses. In addition, smart grid technology also reduces operational costs, saves energy by using demand side management, demand response and Transactive Energy Management (TEM) technologies.

Through internet-based communications, public solutions on control and monitor the smart grid and high dependency could cause disaster due to vulnerabilities. Further, attackers could find infrastructure desirable [5]. Hence, increased connectivity and digitalisation pose new security challenges. For instance, an attacker can attack electronic devices by corrupting state estimation readings to maintain im-balance in between demand, supply in real time due to device data falsification [6]. Then, the smart grid's sensitivity could make it a cyber-terrorism target [7]. As a result, it is crucial to examine smart grid components and identify past flaws and cyber security problems. It can even cause plant failure and subsequent physical damage. Virtual networks of the power sector are essential, and attacks on them can impact a country's prosperity, public safety, and national defense. According to a survey by United Nations, most of the world's population already lives in cities (55 % in 2018), and by 2050, that figure will be closer to 68% [8]. These people rely heavily on reliable electricity distribution. Brownouts or blackouts can significantly impact safety and security in such urban settings. Since the last few decades, cyber security attacks have been the most serious concern. According to specops sources [3], the USA has witnessed the most cyber-attacks in the recent decade, followed by the United Kingdom, India, Germany, and South Korea as shown in Fig. 1. The simple fact is that most urban electric infrastructures are ageing and pushed to their breaking points. As mentioned earlier, the urban population data highlights the critical need to secure the utility operations. Deploying an Intrusion Detection System (IDS) and firewalls to secure power grid data, account management, non-segregated networks are necessary.

In recent years, cryptographic primitives are becoming essential solution to provide security for critical information transfer in communication channel by using message authentication codes, hash functions for authentication and Authenticated Key Agreement (AKA) schemes to encrypt messages while maintaining privacy and confidentiality in smart meters to the divisional network [9]. Therefore, this review paper aims to analyze the cyber-attack vulnerabilities and suggests research aspects to meet smart grid security requirements and fulfil security objectives by using detection and mitigation techniques such as cryptography, artificial intelligence, and blockchain. Meanwhile, study how security criteria affect data security, privacy, and cyber threats during data transmission. In [10], researchers have discussed deep learning and machine learning with different network operations, algorithms, and datasets to create a functional IDS which provides cyber security to the system. Arezoo Hasankhani et al., identified the following primary areas for blockchain technology applications in smart grids: demand response, EVs, IoT technology, decentralized energy balance, energy marketing [11]. In addition, a realistic aspect of the main advantages and disadvantages of using blockchain technology in smart grids have been discussed. In [12] authors reviewed about different cyber-attacks, strategies, and approaches for providing cyber security in energy systems. In [13], authors discussed cryptographic approaches as well

as key management techniques. In addition, discussed the security and integrity verification tools for communication protocols.

IoT incorporated power systems, particularly smart grid features posing cyber security vulnerabilities due to over dependency on communication systems. Therefore, the ideal approach for protecting smart grids and energy systems from cyber attacks is to provide accurate, up-to-date, and efficient overviews and details regarding identifying and dealing to cyber-attacks. As of now, researchers have put together several review articles in the literature on block chain, machine learning and deep learning based techniques for cyber security in power systems. However, prior work has not been done in power systems on communication attacks such as Denial of Service (DoS), Man-In-The-Middle (MITM), replay attacks, and so on. This review article assesses the feasibility of identifying the primary fields for cryptographic technology applications in smart grid sectors such as energy marketing systems, M2M and substation communications.

A. ABBREVIATIONS AND ACRONYMS

AMI -	Advanced Metering Infrastructure
AKA -	Authenticated Key Agreement
AVISPA -	Automated Validation of Internet Security Protocols and Application
BAN logic -	Burrows-Abadi-Needham logic
CK -	Canetti and Krawczyk
CPS -	Cyber Physical Security
DoS -	Denial of Service
DER -	Distributed Energy Resources
ECC -	Elliptic Curve Cryptography
ECQV -	Elliptic Curve Qu-Vanstone
FDIA -	False Data Injection Attack
GNY -	Gong, Needham and Yahalom logic
GOOSE -	Generic Object-Oriented Substation Event
IoT -	Internet of Things
IDS -	Intrusion Detection System
MITM -	Man-In-The-Middle
M2M -	Machine-To-Machine
NPP -	Nuclear Power Plant
PMU -	Phasor Measurement Unit
PLC -	Programmable Logic Controller
PMAKE -	Privacy-preserving Multi-factor Authenticated Key Establishment
PF-DA -	Pairing Free-Data Aggregation
PUF -	Physical Unclonable Function
PIDMS -	Proactive Intrusion Detection and Mitigation System
RES -	Renewable Energy Sources
ROM -	Random Oracle Model
SCADA -	Supervisory Control And Data Acquisition
SVM -	Support Vector Machine
TEM -	Transactive Energy Management
TES -	Transactive Energy System
TESP -	Transactive Energy Simulation Platform

II. TAXONOMY OF CYBER ATTACKS IN POWER SYSTEMS

Taxonomy is the structured classification of things or concepts. Our proposed taxonomies aim to classify various types of vulnerabilities or cyber attacks across the generation, transmission, and distribution sectors. The damages incurred through cyber attacks and the vulnerabilities of attacks on power grids will vary based on the field and strategies employed by the attackers. The majority of cyber-attack exploitation is directly or inversely associated with grid instability. While cyber attacks on the generation sector have primarily relied on False Data Injection Attacks (FDIA) [20], the transmission sector has become a victim of physical access-based attack vectors such as time delay attacks [21], load redistribution attacks, time synchronisation attacks [22], load altering attacks [23], false command injection attacks,and cyber-physical attacks [24]. Most cyber attack vulnerabilities in the distribution sector are network access-based, including MITM attacks [25], DoS attacks [26], Replay attacks [27],and malware attacks. In addition, taxonomy of cyber attacks to power grid with impacts on power systems is presented in fig. 2.

III. RESEARCH MOTIVATION AND CONTRIBUTIONS

The motivation for this survey arises from the quote, ''wherever IoTs are present, cyber-attack vulnerabilities are also present''. IoT applications include intelligent information transfer, monitoring of pollution, green infrastructure, smart homes, and connected healthcare. Smart grid is the major IoT application, which provides the structure sensing, communication and processing methods necessary for a smart energy systems. The rapid improvements in IoT technology give the new potential for the seamless operation of the smart grid systems. On the other hand, IoTs are becoming cyber-attack vulnerabilities like critical information leakage, and infrastructure damage.

Furthermore, IoT vulnerabilities may leads to grid blackouts like Ukraine's electricity grid attacks in 2015 and 2016 where attackers try to open circuit breakers to stop the electricity supply by using malicious bad firmware injection. In addition to that attackers implemented DoS attack on telecommunication system to block the communication in between consumers and grid. This study describes a public network-based smart grid defensive mechanism, opened possibilities of cyber attacks on smart grids and assists potential researchers and participants in this field in understanding the structure of an IoT-enabled smart grid system, as well as security breaches, prevention, and detection of those security breaches in smart energy systems. The significant contributions of the article are as follows:

1) A thorough examination of random cyber risks across various power sectors such as generation, transmission, distribution, and consumption have been conducted.
2) Detection techniques for various cyber-attacks such as impersonation, replay, privileged insider, man-in-the-middle, denial of service, ephemeral secret leakage,

TABLE 1. Comparison of proposed work with Existing Literature.

Ref.no	Year	A	B	C	D	E	F	G	H	I	J	K	L
[14]	2023	✔	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘	✘
[15]	2023	✔	✘	✔	✔	✘	✘	✘	✘	✘	✘	✘	✘
[13]	2023	✔	✘	✔	✔	✘	✘	✔	✔	✘	✘	✘	✘
[16]	2023	✔	✔	✔	✔	✘	✘	✔	✘	✘	✘	✔	✘
[17]	2023	✔	✔	✔	✘	✔	✘	✘	✘	✘	✘	✔	✘
[18]	2023	✔	✘	✔	✘	✔	✔	✔	✘	✘	✘	✘	✔
[19]	2023	✔	✘	✔	✘	✘	✘	✔	✘	✘	✘	✔	✘
Proposed Work	-	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

✔ - Assessment existing literature provided ;
✘ - Assessment existing literature didn't provided
A - Cyber-attacks in Power Grid ; **B** - Energy Marketing System ; **C** - AI based Detection and Protection scheme; **D** - Smart Meters Security ; **E** - Case study ; **F** - Real-world Cyber-attacks ; **G** - Cryptography Algorithms in Power Grid ; **H** - Verification Tool ; **I** - Cyber-attack Robustness Assessment ; **J** - Security features Assessment ; **K** - Block chain ; **L** - Cyber-attack impact Assessment

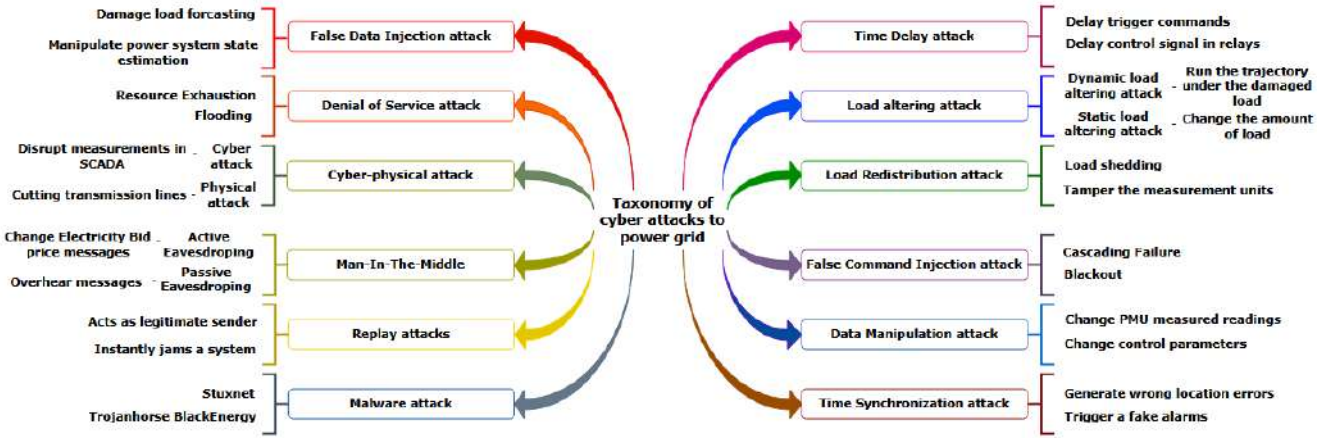


FIGURE 2. Possible Cyber attacks Impact on Power systems.

resending, masquerade and device stolen have been investigated. Furthermore, cyber security strategies against these threats have been examined.

3) The significance of different cryptographic algorithms in data privacy and protection while sharing data between two entities has been analyzed.
4) The study of real-world cyber security events and case studies in diverse power industries are carried out. Further, research gaps in smart grid cyber security are identified and highlighted.

The organization of this paper is as follows: section IV provides a review of cyber-attacks in power generation systems. Section V addresses a review of cyber-attacks in power transmission systems. Section VI presents a review of cyber-attacks in power distribution systems consists impact analysis, detection and security using various advanced methods. Section VII provides case studies and addresses real-world cyber-attack incidences on power systems. Finally, section VIII concludes, highlighting the findings, future directions and suggesting possible future research perspectives.

IV. REVIEW ON CYBER-ATTACKS IN POWER GENERATION SYSTEM

Electricity generation mainly depends upon Nuclear Power Plant's (NPP), hydroelectric, and thermal power plants. Based on the present literature survey, power sectors are changing their approaches to sustainable energy by increasing the integration of renewables even though they have meteorological origins of variability in energy generation [28]. In addition, Renewable Energy Sources (RES) locations are decentralized, which may require fewer employees to report updates in person, which can become a time-consuming process for grid operators and leads to the installation of remote operation tools and IoT devices in power generation systems. In [29], researchers have reviewed wind farm threats, security, unauthorized wind turbine control, disruption and mitigation techniques used to improve confidentiality in the system. In addition, the researchers highlighted future work focused on understanding and combating persistent threats which will increasingly target wind farm assets. The authors of [30] provided a comprehensive review of Cyber Physical Security (CPS), particularly FDIA, in power generation

systems based on the national institute of standards and technology security framework. In [31], the authors have discussed the CPS of photo-voltaic systems and vulnerabilities under various cyber-attacks, such as replay attacks, FDIA, and infrastructure tampering attacks. Furthermore, challenges and opportunities in creating cyber-secure power electronics systems for the next generation have been addressed to assist readers with future research paths. In [32], using real network data and energy generation measurements collected by a wind turbine at Lancaster University, the researchers investigated the amount of malicious scans carried out by Mirai-infected bots in order to penetrate the wind turbine. Furthermore, ping to death with big ICMP packets was investigated.

A. CYBER-ATTACKS DETECTION IN POWER GENERATION SYSTEM

Cyber-attacks, such as FDIA and MITM, are becoming significant threats in power systems, intending to modify the power system condition, which may lead to improper control actions. Fayha almutairy et al., have developed deep learning models such as Wavelet and Temporal convolutional network to detect FDIA in power systems with high RES penetration. In addition, the performance of the developed models has been evaluated on IEEE 14 bus system with an detection rate of more than 99% and 118-bus system with an detection rate of 97% [33]. In [34], researchers have developed a hybrid deep convolution–recurrent neural network to detect electricity theft in renewable energy-based distributed generation-units with detection rate of 99.3% and low false alarm rate of 0.22%. In [35], the authors have implemented a novel distribution algorithm for detecting cyber-attacks such as adversary manipulation of wind farms turbine-specific control logic parameters. In addition, the implemented algorithm has been tested at the Horns Rev wind farm in Denmark. The presented work shows that the implemented algorithm can also provide cyber security for wind farms. In [36], Huang et al. have developed an online platform to detect cyber-attacks in automated generation control using dynamic watermarking techniques without hardware upgrades on generation units. In addition, the developed technique can also be used for large-scale power systems.

B. CYBER-SECURITY IN POWER GENERATION SYSTEM

Cyber security is an essential countermeasure to mitigate cyber-attacks and protect critical infrastructure in power generation systems. In [37], researchers have presented a protection approach using a digital frequency relay to protect equipment from large power fluctuations for longer duration in wind energy systems. In [38], authors have implemented an operating reliability evaluation mechanism for multi-state power systems to achieve dynamic system reliability by considering cyber malfunctions. In [39], researchers have implemented a comprehensive algorithm with the help of the proportional fairness index to coordinate defence countermeasures of microgrids during any cyber-attack. Furthermore, it analyzed cyber defence based on coalitional game theory.

In [40], Lee and Huh have presented the system information and event management analysis method to prevent the leakage of peak information and hackings through insecure web services in NPPs. In [41], researchers have developed a framework using knowledge-based hidden makrove modelling to analyze the integrative cyber-attack reaction in NPPs. In addition, researchers have developed a security state estimation method utilizing online updated hidden Markov models to analyze the functional impact. Poong Hyun Seong et al., have designed a cyber-attack reaction planning approach based on Markov decision process model and the Monte-Carlo tree search algorithm to develop optimal reaction plans that improve response margin time and conserve time essential to secure NPPs safety [42].

FDIAs are becoming a primary threat to the generation system, causing disruptions in control logic parameters and state estimation readings to damage the electricity generation quantity, power market by maintaining imbalance in power generation. It is necessary for power plants to provide security for equipment and critical information with improved marginal time and a low false alarm rate against FDIA.

V. REVIEW ON CYBER-ATTACKS IN POWER TRANSMISSION SYSTEM

Because of its size and the need for high system availability, the energy sector has adapted to digital technology, leading to cyber-attack or cyber-physical attack vulnerabilities to the transmission system. Attackers can use various attack vectors, such as malicious activities, malware injections, and viruses, to compromise the networks, measurements and also changes power flow of the transmission system which can cause a blackout or significant disruption in the power grid [43]. In addition, several sensors have been deployed to analyse the real-time operation of a power system by monitoring bus injection powers, bus voltages, and line currents. The control center assesses the stability of the grid based on redundant measures transmitted through the Supervisory Control And Data Acquisition (SCADA) system. Transferring measured data can also lead to cyber-attacks vulnerabilities. Power systems security threats have been classified into three types:

1) Physical attacks on networks can be considered as terrorist attacks, which may cause disrupting substation operations, cutting transmission lines, or generator units to fail.
2) Cyber-attacks that disrupt measurements or data transmission in SCADA systems.
3) Cyber-physical or coordinated attacks, such as the tripping of transmission lines are the consequence of FDIA's capabilities [24].

In [44], Hossein Rahimpour et al., presented a potential cyber attack vulnerabilities and their risks pertaining to power transformers in power networks. In [45], researchers have considered timing attacks, replay attacks and FDIA to analyze the impact of cyber-attacks on High Voltage Direct Current transmission-based oscillation damping control. In addition, the implementation of cyber-attack preventive measures for Alternative Current-High Voltage Direct Current systems, which have strong, robust control schemes and accurate detection algorithms considered for future scope. Habib Rajabi Mashhadi et al., have proposed an analytical method to analyse the influence of renewable energy power plants on transmission network congestion [46]. In [24], researchers have developed mixed integer linear program model to implement load transmission attacks via FDIAs, which may cause many transmission lines to overflow. The developed model established a standard to analyze realistic cyber-attacks that may disrupt transmissions and cause a blackout. In addition, developing a detection strategy for cyber-attacks aimed at transmission line congestions in Direct Current state estimation is considered for future studies. In [47], Yury Dvorkin et al., have implemented a bi-level optimization model to analyze the impact of distributed cyber-attacks on the distribution and transmission electrical grid. In addition, future research is projected into how attackers can use publicly available grid sources to create more harmful attack strategies.

A. CYBER-ATTACKS DETECTION IN POWER TRANSMISSION SYSTEM

Mohsen ghafouri et al., have implemented a new detection scheme based on thevenin equivalent system parameters, which performs fast and accurate detection of possible cyber-physical attacks on the voltage stability monitoring of transmission system [48]. In addition, the implemented scheme has been utilized to calculate an indicator that detects Phasor Measurement Unit (PMU) data attacks. In [49], Wilson et al. proposed a deep-learning based stacked autoencoder framework for developing machine-learning features against transmission SCADA attacks. Also, presented unsupervised learning framework to detect automatic and adaptive attacks in the transmission SCADA system. Furthermore, the framework can also be improved so that it not only detects but also locates the event on each line planned. In [50], researchers have presented a cyber-physical data analysis using a deep-autoencoder to monitor transmission protection systems. A ridge regression-based classifier has been deployed to identify cyber anomalies. In addition, the outcomes of the presented models can be investigated as the underlying cause of reported incidents with the aid of cyber log data from protection equipment.

Transformer taps have mostly been used in transmission networks to manage bus voltages. Therefore, tap change commands carried across the SCADA network are always appealing targets for attackers to disrupt system operation. To address the issue, the authors have developed an algorithm that detects the presence of a concealed misleading tap change command in the on-load tap changer [51]. In [22], the authors have proposed a detection technique for cyber-attacks against line current differential relay by using a learning-based framework which employs a multi-layer perceptron model to detect FDIAs, and time synchronization attacks and to divide them from faults. In [52], Pal et al. proposed a mechanism for detecting PMU data manipulation attacks by using that mechanism which continuously monitors the equivalent impedance of transmission lines and divides observed anomalies to detect the presence and location of attacks.

B. CYBER-SECURITY IN POWER TRANSMISSION SYSTEM

Security systems are one of the most crucial components for transmission system. With ongoing automation, they are becoming more digital and more efficient at delivering electricity which exposing them to cyber-attack vulnerabilities and generating many challenges. In [53], researchers have proposed an algorithm to detect the additional placement of PMUs for maximum security against FDIAs. The algorithm has been evaluated for a range of IEEE-30, 57 and 118 bus-based electric transmission network models. Future research analyze the impact of cyber-attacks on PMU placement strategies in realistic transmission networks. In [21], Lou et al. designed a time delay attack, which delays the delivery of system control commands and a recurrent neural network is used to predict delay values from input traces. The results demonstrated that long short-term memory-based deep learning approach could work well in power plant control systems based on data traces from three sensor measurements such as pressure, temperature, and power generation. In [54], Dehghani et al. launched an FDIA on the information exchange between independent system operator and under-operating agents in the power transmission system to evaluate system security levels. Blockchain has developed to increase the data confidentiality between independent system operator and under-operating agents.

The transmission system is more vulnerable to time delay and PMU data manipulation attacks. Due to a time delay attack, a delay in trigger commands can cause critical infrastructure damage or cascade failure, and PMU data manipulation may lead to load shedding or power overflow in a transmission system. Late detection of FDIA can cause power transmission lines tripping, change in power flow, and large-scale cascade failure [55], some defences against power transmission line attacks include maintaining PMU placement strategies, implementing fast key agreement protocols for secure communication and using blockchain to maintain confidentiality while sending commands.

VI. REVIEW ON CYBER-ATTACKS IN POWER DISTRIBUTION SYSTEM

Distribution networks are more susceptible to cyberattacks due to their vast size and decentralized nature. In addition,

IoT applications become integral part in distribution system components such as electricity marketing, substations, smart meters as shown in the Fig. 2. As a result, they are exposed to major cyber-security risks, such as attacks, vulnerabilities, and consequences. Due to their control and communication requirements, even Distributed Energy Resources (DER) and battery storage installation may pose negative impact on the grid. In [5], the authors have reviewed the threats and potential cyber-security vulnerabilities, attack countermeasures, and security requirements in IoT-based smart grids. The below mentioned literature evaluated impact analysis of cyber attacks in smart distribution system.

In [56], Ma explained the cyber security challenges in smart cities, such as critical information leakage and intentional cyber-attacks by considering four essential components in smart cities such as smart grid, the smart homes, the smart transmission system, and the smart healthcare system. Furthermore, future research focus on cyber security challenges, threats to user privacy, and relevant authorities and policymakers. Researchers utilized the observer-based method and the decomposition form of the system matrices to analyze the impact of DoS attacks on state estimation in [26]. In addition, the detection and estimation of distributed attacks have been considered for future study. Researchers have implemented a game theory model for power plants, transmission lines, and distribution networks to analyze cyber-attacks and defence probability [57]. In addition, it allows decentralized defence strategies to make defenders separate decision-makers planned for future studies.

In [58], researchers have analyzed the security and privacy of emerging peer-to-peer electricity trading markets. Further, designing privacy-preserving protocols for the defined scenarios using the specified requirements as a guideline is proposed for future studies. In [59], Marufu et al. proposed a strategy for determining how successful cheating attacks on power marketing schemes can be executed in resource-constrained smart microgrids. In addition, mitigation techniques are implemented to prevent cheating attacks. The authors have presented a systematic detection of possible cyber-attacks and examined the influence of attacks on power market operation in association with TEM–based power systems [60]. Furthermore, intend to examine and analyze the impact of additional attacks, such as DoS and replay attacks, on the microgrid's peer-to-peer markets, as well as deploy detection schemes in the microgrid considered for future work. In [61], researchers have presented an ensemble decision tree approach based on the bagging technique to find possible anomalies in the electricity market and physical measurements within the Transactive Energy System (TES), which can reduce the impact of outliers. In addition, the presented approach may be tested on advanced use scenarios to depict a few realistic TES behaviors. In [62], Zhang et al. implemented a deep-stacked autoencoder algorithm to identify possible anomalies in the electricity market and physical measurements with an accuracy rate of 96.9%. The proposed algorithm analyzed the main cause and provide appropriate control actions to protect utilities and prosumers from cyber intrusion. In [63], Wang et al. detected possible anomalies in TES by considering zero-mean FDIA and analyzed cyber-attacks impact on price, quantity and market clearing price. In [64], researchers build a electrical trading protocol in the java programming language to maintain privacy preserve in between prosumers and utility by using blockchain and Elliptic Curve Cryptography (ECC). Pal et al. analyzed the influence of data integrity attacks on electricity pricing exchange in between distribution system operator and price-responsive loads in TES [65]. The system performance is evaluated using four metrics such as operational, financial, comfort, and reliability metrics under data integrity attack by using a 240-bus western electric co-ordinating council transmission model with modeling of the distribution system at specific buses. In [66], researchers have analyzed the impact of manipulated malicious bid prices and quantity cyber-attacks in TEM. Moreover, the TEM operation has been studied through proxy attacks simulated on the TE30 test system. In [67], the authors have created a comprehensive simulation-based transactive energy valuation method to systematically assess the system value, process, object, and design. In addition, a co-simulation-based Transactive Energy Simulation Platform (TESP) has been developed based on the valuation method to perform marketing operations. The work remarks that can deploy agents and market mechanisms without reprogramming any simulators.

A. CYBER-ATTACKS THROUGH DEVICES

AMI is a catch-all word for whole infrastructure, from smart meters to control center equipments which establish communication between two-entities or devices. The purposes of AMI can include remote meter reading for error-free data, identifying network problems, load profiling, and energy audits by sending energy usage data in near real-time. Unfortunately, sophisticated cyber-attacks on AMI are an open and apparent vulnerability. Attackers typically target less secure system elements such as AMIs to manipulate energy and consumption reports based on financial motives. When the system is heavily loaded, the failure of a single critical component can cause a chain reaction of component failures, eventually leading to blackout. To safeguard the necessary infrastructure from cyber-attacks several researchers analyzed attack detection and protection techniques.

1) DETECTION OF DEVICES THROUGH CYBER-ATTACKS

Energy theft has become one of the most concerning attacks in electricity distribution system. In [68], researchers utilized support vector regression and impact difference to detect possible anomaly pricing cyber-attacks that influence the guidelines of smart meter electricity rates in smart home systems. Bhattacharjee and Das implemented a two-tier approach to detect the FDIA in data consumption without increasing the false alarms in smart meters by using

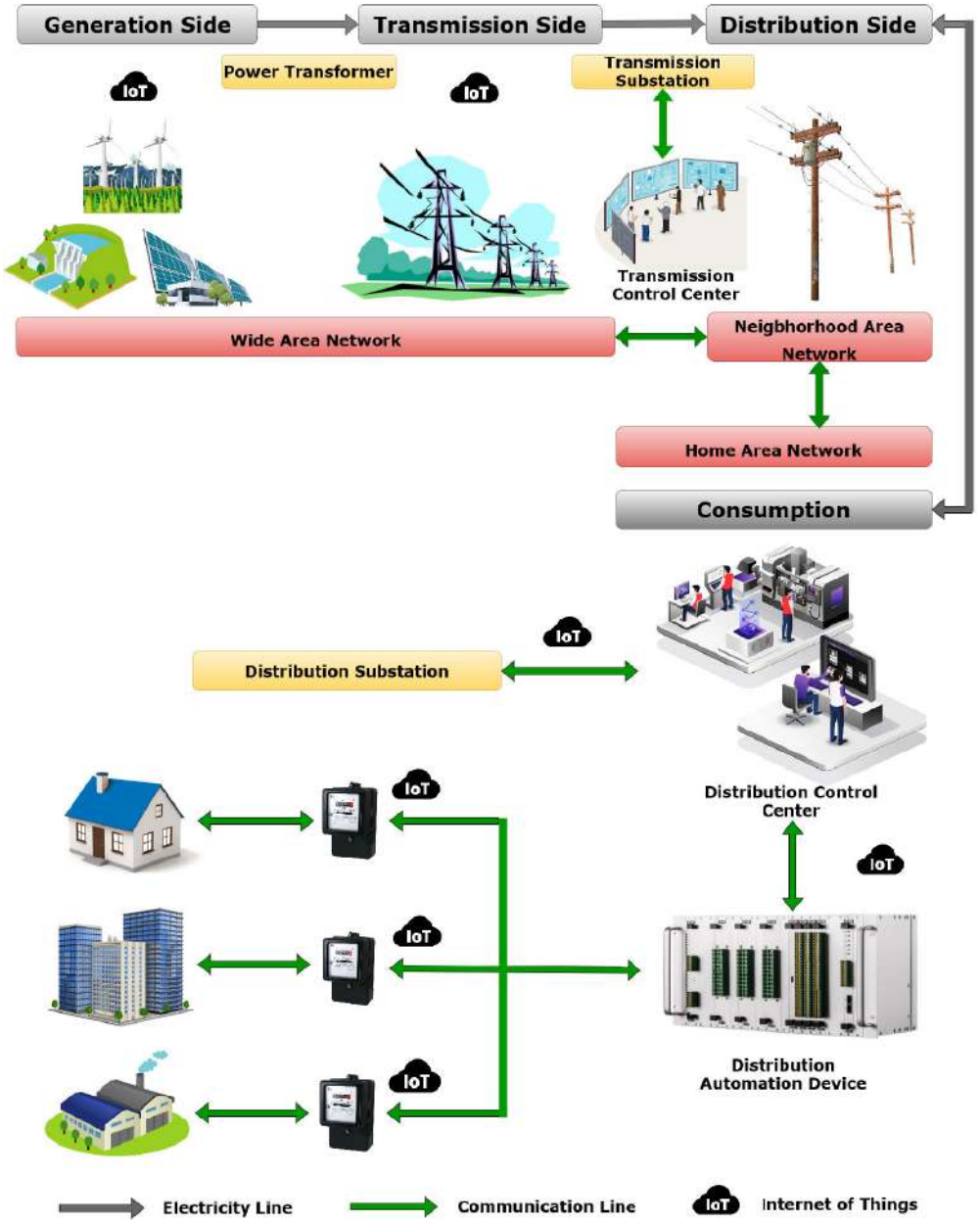


FIGURE 3. Advanced Communication Infrastructure for Power network.

harmonic - arithmetic mean ratio as tier-1 and residential under the curve as tier-2 [20]. The work highlights that the analysis on ON-OFF and data omission attacks with minor modifications to the tier-2 detection level approach can also be considered for future studies. In [69], researchers have developed an isolation forest-based detection method to detect FDIA without a pre-training procedure for detection labels in a power system with an fast detection accuracy rate 96.3% at time 1.944 sec. In [70], the authors have developed a highly randomized tree algorithm to detect FDIA, which jeopardizes power system state estimation by applying FDIA into smart meter measurements. The developed algorithm has achieved detection accuracy of 99.76% with IEEE-118 bus, 99.39% with IEEE-57 bus and 97.8% with IEEE-14 bus systems respectively. Furthermore, developed algorithm showed more accurate results than Support Vector Machine (SVM), k-nearest neighbor and random forest. Furthermore, a stacked autoencoder has been used in co-ordination with a highly randomized tree classifier to deal with dimensionality. In [71], researchers have developed multiple-stage IDS techniques, such as temporal failure propagation graph, SVM, for intrusion detection and generating attack pathways for recognizing attack events in smart meters.

2) CYBER-SECURITY FOR DEVICES THROUGH CYBER-ATTACKS

Yankson and Ghamkhari developed an attack-thwarting technique for preventing load-altering attacks, which can rectify

frequency disturbances in power grids [23]. Furthermore, the developed technique has given effective results when tested on IEEE 33-bus power distribution system. In [72], the authors presented a bi-level optimization strategy for determining the most compromised and effective attacks as well as independent system operators effective response. Besides that, a defense strategy has been developed to reduce network losses and maintain rated voltage and current values.

B. CYBER-ATTACKS THROUGH ADVERSAL USERS

Adversary users may act maliciously by tampering their meters to decrease the electricity consumption, resulting in financial losses to utility companies or service providers and grid instability. Researchers proposed various detection and mitigation techniques in the existing literature.

In [73], Ahmadian et al. incorporated FDIA into the measurement system, in which the attacker acts as a virtual bidder in the day-ahead and real-time markets to maximize its profit by trading and proposed the mathematical programming equilibrium constraint-based single-level optimization problem to determine the optimal cyber-attacks against state estimation. In [74], researchers designed a easy-to-implement detection algorithm based on a co-variance estimator to detect and identify coordinated electricity theft incidence by evaluating both dependent & independent smart meter data generation process.

In [75], Yang et al. described a resilience technique for defending Programmable Logic Controller's (PLC) from critical information tampering attacks. In addition, generated a data authentication mechanism with an accuracy of 97.4% for the message digest in PLC-to-PLC communication. In [76], researchers have developed a privacy-aware AKA scheme to provide secure communication in between smart meters and service providers. Moreover, the developed scheme ensures the physical security of smart meters by utilizing light-weight cryptographic primitives such as one-way hash functions and PUF. Gope, implemented an efficient Privacy-preserving Multi-factor Authenticated Key Establishment (PMAKE) scheme based on reverse fuzzy extractor, one-way hash function and PUF to achieve secure smart grid communication [77]. Furthermore, the implemented scheme can guarantee the physical security for smart meters. The Table 2 and 3, presented the simulation platforms used to evaluate the performance of proposed techniques or algorithms against cyber-attacks.

TABLE 2. Proposed schemes with analyzed IEEE-bus Networks.

Ref.no	IEEE Buses
[78]	IEEE 37 - Bus
[34]	IEEE 123 - Bus Test System
[38]	IEEE RTS (Reliability Test System)
[45]	IEEE New England 39 - Bus AC-HVDC
[24]	IEEE 118 - Bus
[47]	IEEE RTS & IEEE 13 - Bus Distribution feeder
[48]	IEEE New England 39 - Bus & IEEE 118 – Bus
[51]	IEEE 118 - Bus
[22]	IEEE 39 - Bus
[53]	IEEE 14 - Bus for impact analysis & IEEE - 14, 30, 57, 118
[54]	IEEE 14 - Bus Network
[79]	IEEE 14 - Bus
[69]	IEEE 118 - Bus
[70]	IEEE 14, 30, 57, 118 - Bus
[23]	IEEE 33 - Bus Power Distribution System
[72]	IEEE 94 - Bus
[80]	IEEE 57 & 118 - Bus
[81]	IEEE 9 - Bus
[82]	IEEE 57 - Bus
[83]	IEEE 33 - Bus
[33]	IEEE 14 & 118 - Bus

TABLE 3. Proposed schemes with analyzed Simulation Networks.

Ref.no	Different Simulation Scenarios
[65]	Western Electric Co-ordinating Council 240 - Bus model
[35]	High Fidelity Simulation Test - Bed
[36]	NPCC 140 - Bus System
[62]	TESP & IEEE 9 - Bus System
[66]	TESP & IEEE 9 - Bus System
[67]	TESP & IEEE 9 - Bus System
[61]	TESP & IEEE 9 - Bus System
[84]	OPNET Simulator
[85]	SUMO & OMNET ++
[86]	Speed Goat Real-Time Digital Simulator
[87]	Power World Transient Simulation Tool
[73]	5 - Bus PJM(Pennsylvania-Jersey-Maryland) System
[88]	Xilinx ISE 14.7

C. ATTACKS THROUGH COMMUNICATION CHANNELS

A smart grid is an IoT-based application that allows energy providers to exchange electricity information with their customers or devices. However, the distribution systems reliance on communication networks makes it highly vulnerable to cyber-attacks. Attackers exploit this by attempting to steal information transferring through communication lines via DoS attacks and MITM, which can result in service interruption, energy theft or critical data theft. In addition, an attacker may try to eavesdrop on crucial messages transmit to the market operator. As a result, the attacker could know the identities of users, smart meter readings, bidding-offer information, electricity supply and demand information from these critical messages. Any drawback happen while providing security may leads to grid instability, AMI damage, and blackouts. Many researchers published various analysis methods, detection, and protection techniques in the literature to control the communication attacks. In [89], researchers reviewed about cyber systems and cyber physical systems, as well as the communication standards and protocols utilized in smart grids. In [90], authors presented a trust-based multi-path routing protocol for secure communication in the Mobile ad hoc network by minimising packet losses and detecting malicious nodes. In addition, cryptography and block chain approaches for providing high security to Mobile ad hoc network are being considered for future scope.

1) DETECTION OF COMMUNICATION CHANNEL CYBER-ATTACKS

In [80], the researchers proposed a cyber-attack detection scheme based on kernel principal component analysis

and randomized trees algorithm for dimensional reduction between the sensor and gathered measurements in smart grid networks. The performance of proposed scheme has been evaluated using standard IEEE 57 and 118 bus systems. In [91], researchers have proposed an artificial feed-forward network using a true data integrity agent-based model to detect false data cyber-attacks in smart grid systems for security assessment. The proposed model has detection accuracy of 98.91% through replay cyber-attacks. The proposed model can also be used in intelligent transportation systems for cyber-security.

In [81], the authors have considered the cosine similarity matching and chi square detector approach for use to detect cyber-attack in smart grid. In addition, the Kalman filter estimation method has been utilized to measure the divergence between actual and estimated data in order to detect attacks. In [84], researchers have developed supervised machine learning algorithms such as tree classification, naive bayes, multilayered perceptron, and multinomial logistic regression algorithms for classification tasks between network abnormality effects such as cyber-attacks and faults on energy-aware smart home systems. In [92], the authors have presented an IoT micro-security add-on that leverages a convolution neural network model to identify phishing attacks on IoT devices. In addition, the recurrent neural network-long short-term memory model has been hosted on back-end services to identify botnet attacks on IoT devices. In [93], researchers have developed an attack detection technique based on a deep belief network and interval state estimator to detect malicious attacks and electrical load forecasting. Moreover, the proposed mechanisms been evaluated on IEEE 14 and 118-bus systems. In [94], the authors have presented an adaptive and resilient N-IDS model using deep learning architectures to monitor network traffic, detect and classify network attacks such as jamming attacks, DoS attacks, and MITM attacks.

In [95], the authors have proposed a data integrity-based effective IDS with two phases: data sampling and selecting features to protect the network with accurate detection rate of 0.936 sec and false alarm rates of 0.33%. Even though the proposed system performs better in unstable conditions, it only detects data integrity-based attacks. In [79], researchers have developed an IDS architecture to monitor and detect lethal attacks such as price manipulation attacks, DoS attacks with detection rate of more than 95% and false positive rate is below 5% using a cumulative sum algorithm in smart grid. In [96], the authors have proposed an SVM algorithm to detect active eavesdropping attacks with detection probability of 95% using artificial training data in the wireless communication channel. From the presented work, adding more hidden features to the proposed algorithms can improve detection performance.

In [97], Sahoo et al. presented a cooperative mechanism based on the cooperative vulnerability factor to detect potential deceptive cyber-attacks in cyber-physical Direct Current microgrids. Furthermore, the proposed mechanism performance has been examined in MATLAB environment. In [85], authors developed a cross-layer IDS based on random forest and k-nearest neighbor to detect spoofing attacks in inter-vehicle communications. Based on the results of the IDS, attackers have been barred from using the wireless charging mechanism.

2) COMMUNICATION CHANNEL CYBER-ATTACKS DETECTION AND CYBER-SECURITY

The use of the internet for data communication between building controllers, such as smart meters and the electric grid, renders the system susceptible to cyber-attacks. A skilled adversary may be able to manipulate the exchanged data which will harm the system. In [98], researchers designed, implemented, and evaluated a monitoring system for open-flow networks and injected proxy attack in between open-flow controller and open-flow switches to capture messages and monitor traffic data. From the presented work, it is to be noted that they can deploy multiple monitoring systems for load balancing and upgrade open-flow versions from 1.0 to 1.1. The authors of [99] created a singular value decomposition technique and private pilot to identify active attacks by authenticating the sender based on the wireless channel. Furthermore, passive eavesdropping and active attacks have been defended using the concept of one-time pad by encrypting wireless channels with a private plot. In [100], researchers have proposed an authentication method to overcome false data flow and improve false data detection with less detection time 4.67 sec without increasing end-user overload in smart grid communication. In [82], the authors have proposed a deep learning-based dual denoising auto-encoder and unified scheme to protect the cyber physical system from eavesdropping attacks and to detect typical cyber-attacks, such as FDIA, DoS, and relay attacks. The proposed scheme performance has been evaluated on IEEE-57 bus system. In [86], researchers have presented a dynamic state estimation technique based on an unknown input observer to estimate the presence of unknown inputs in the microgrid communication channel for stable operation. In addition, a residual function has been generated that detects the presence of FDIA and triggers a detection alarm for attack isolation and mitigation.

3) CYBER-SECURITY FOR COMMUNICATION CHANNEL ATTACKS

A smart meter is an essential component of the smart grid and transmits real-time data to a utility centre. According to the united nation-national institute of standards and technology, bi-directional communication between the two parties opens doors for cyber-attack vulnerabilities. Implementing security for that kind of attack is one of the challenging tasks. Cryptography is one of the efficient technique to provide security against communication channel attacks such as DoS, MITM, jamming attacks, replay attacks, impersonation attacks not only in smart grid but also in health care purpose. In [101], authors proposed a logistic map based

key generation for secure communication by maintaining confidentiality and authenticity in Mobile ad hoc based health care network. Fig. 4. express the structure of cryptography primitive.

In [102], researchers presented an AKA scheme with privacy preservation for smart grid communication. When an adversary compromises a smart meter device, this scheme considers reducing the possibility of a critical leakage attack. In addition, the work highlights that blockchain technology can be used to better authentication schemes for privacy protection in smart grid.

In [103] and [104], Abbasinezhad-Mood and Nikooghadam proposed an ECC-based self-certified key distribution mechanism to address the issue of public key infrastructure maintenance in between smart meters and service providers in smart grid. However, Khan et al. discovered security problems such as the inability to provide security against DoS attacks, insider attacks, anonymity and failure to update the identity and keys from Mood and Nikooohadam's work. Therefore, the design of an authentication scheme to mitigate the flaws mentioned above has been considered for future work. In [105], Ashok Kumar Das et al., proposed a new anonymous signature-based authenticated key exchange scheme for IoT-enabled smart grid, called AAS-IoTSG which allows a smart meter to establish a session key for encrypted communication by mutually authenticating with a service provider. From the presented work, the proposed method can be tested on a Raspberry Pi to demonstrate its viability for IoT-enabled devices with limited resources such as smart meters. In [106], researchers proposed identity-based signature to demonstrate an anonymous key agreement methodology for smart grid system. Moreover, the proposed protocol not only provides authentication but also provides smart meter anonymity. In [107], the authors have presented a novel symmetric homo-morphic scheme to achieve lightweight aggregation for encryption which can provide secure and efficient authentication in smart grids. In [108], Braeken et al. presented a Elliptic Curve Qu-Vanstone (ECQV) certificate-based key agreement paradigm for smart metering communications, which does not require a secure network during entity registration and is resistant to key escrow. Furthermore, the proposed scheme can also be secure under Random Oracle Model (ROM). In [109], researchers have developed a key management scheme based on ECC to mitigate MITM and re-transmission attacks between smart meters and power companies (outside the HAN). The results revealed that the false factor is directly proportional to detection time. For example, the identification time increases by 1.4 seconds as the false factor rises from 0.1 to 0.3.

Cryptographic algorithms have been used to mitigate communication attacks between devices like smart meters and service providers in the power distribution system, which typically uses cryptographic keys to maintain perfect secrecy. Whenever cryptographic algorithms need to be strengthened, it is often possible to use larger keys or hybrid with two algorithms. In [110], researchers presented a defence strategy using event-based cryptography to keep attackers away from obtaining critical information in the sensor communication channel between the plant and the cyber physical system supervisor. In [111], the authors have developed a secure communication methodology using recursive inter-networking architecture, which addresses almost all communication attacks in a closed environment like LAN. Besides, recursive inter-networking architecture capabilities and features can replace existing communication technology while providing increased security. Furthermore, the work highlights that the developed method can be extended to open environments such as wide area networks and neighbourhood area networks. In [112], researchers have designed an Advanced Encryption Standard-512 bit algorithm for faster processing speed with a stable surface environment in web-based applications with more secure communication.

Elakrat and Jung developed a field-programmable gate array security mechanism to minimize information-gathering attacks based on a cryptographic approach to secure data confidentiality and prevent the injection of malware into the vital digital assets data communication system of NPP [113]. In [114], researchers have developed a secure access control scheme based on certificate-less signcryption with a proxy re-encryption scheme which can secure in ROM. The work remarks that the presented scheme can also be extended to merge attribute-based signcryption with proxy re-encryption schemes. In [115], the authors have implemented a lightweight privacy-preserving Q-learning (LiPSG) framework for smart grid energy monitoring. Moreover, four additive secret-sharing-based sub-protocols such as secure action selection, SMAX, SEle and SGry were developed to perform the atomic operations efficiently and securely. Kumar et al. developed a hardware chip integrated S-box advanced encryption standard algorithm to secure the smart grid SCADA system and chip performance is evaluated using field programmable gate array with different key sizes and grid sizes [88]. In [116], the authors present a lightweight fault-tolerant privacy-preserving data aggregation strategy using modified Paillier cryptosystem, ECC, Chinese-remainder theorem, and hash function technique. Furthermore, the proposed scheme robust against all security features. In [117], researchers have developed a novel Pairing Free-Data Aggregation (PF-DA) scheme based on certificate-based cryptography to reduce the impact of certificate pre-checking problems in the energy internet-based smart grid communication networks. Furthermore, designing the decentralised data aggregation scheme can be provided more security for smart grid communication.

In [25], the authors have presented single and multi-antenna models by applying the Stackelberg game with renewed intelligent simulated annealing algorithm and the stochastic algorithm with feedback to provide security

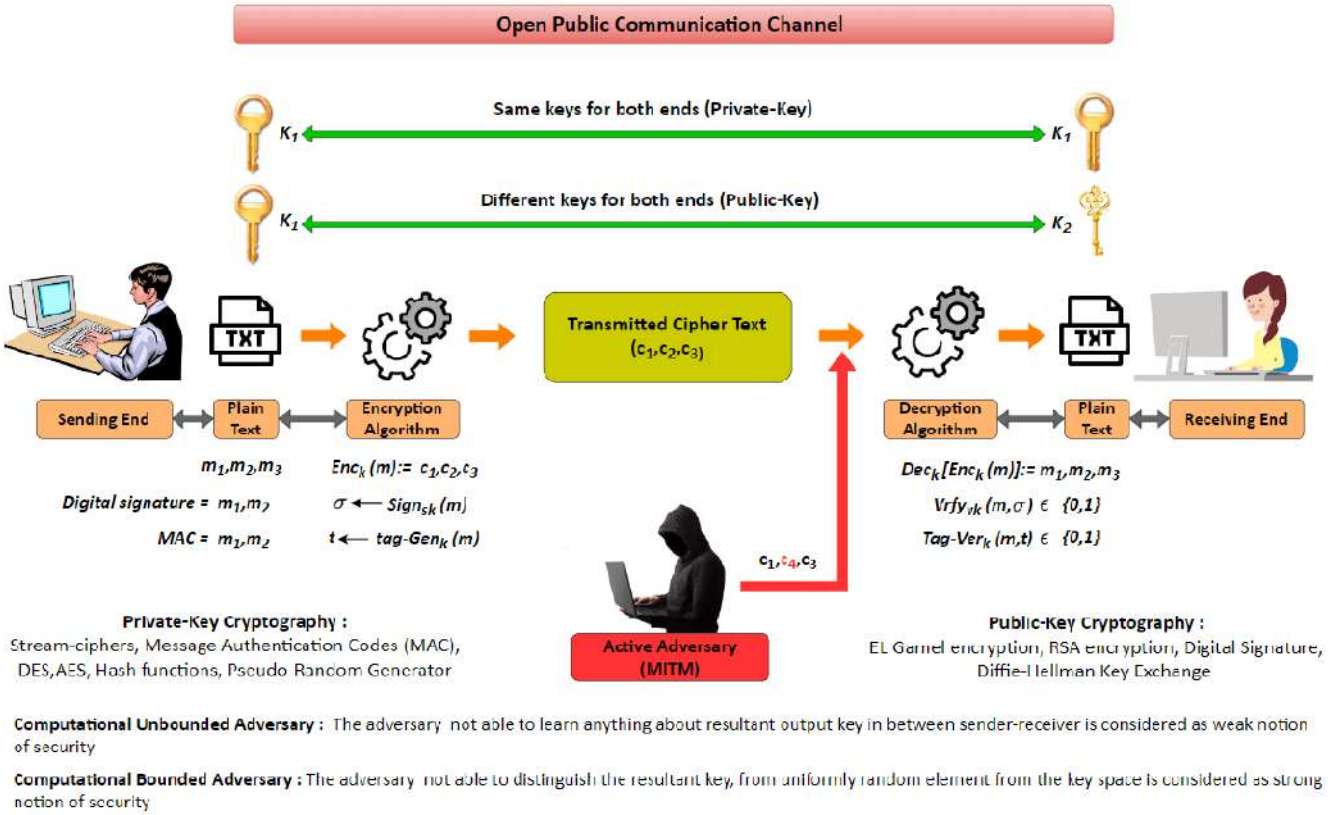


FIGURE 4. A Detailed Overview of Cryptographic Primitives in Communication Channel.

against jamming and MITM attacks in green cyber-physical communication systems. In [118], researchers have proposed a cyber-security architecture that integrates identity-based security mechanism and intelligent security system for energy management to provide appropriate security and privacy for components, data, and actions in the energy internet. The evaluated results of the proposed architecture expressed safety and efficiency for energy internet. Marcos Vicente Moreira et al., presented a security module to prevent MITM attacks between controller and sensor communication channel in cyber-physical systems. Furthermore, the extension of the security module offered NA-safe controllability [119]. In [120], researchers have developed a super-lightweight security protocol using a logical XOR and one-way hash function to secure the smart grid neighbourhood area network communications. The work highlights the implementation of a lightweight scalable blockchain-based multi-party computational protocol that can be employed for resource constraint networks. In [121], the authors have proposed a resilient scheduling strategy that uses additional metrics based on the difference between forecasted and actual bills to detect FDIA in interconnected multiple smart buildings. Besides, the support vector regression method has been used to calculate predicted bills.

Moghadam et al. developed a lightweight protocol based on hash and private key to mitigate IEC62351 security flaws while facilitating key agreement in smart grid [27]. The developed protocol can agree the session key within 0.057ms. Furthermore, it explored privacy, authentication, and private data transfer security between two entities and tested several sorts of cyber-attacks such as impersonation, replay, and MITM attacks. In [122], researchers have presented the timing performance of the RSASSA-probabilistic signature scheme digital signature algorithm to secure the Generic Object-Oriented Substation Event (GOOSE) messages in power system control operations. The work highlights the requirement of cyber-security and time domains that an authentication scheme can achieve.

In [123], [124], and [125], researchers have developed multiple techniques such as PUF- based AKA scheme and reconfigurable authenticated key exchange scheme to secure communication channels by mitigating energy theft attacks, such as ephemeral leakage attacks in between service providers and smart meters. In [126], researchers have developed a lightweight mutual authentication scheme based on PUF to encrypt communications in between smart meters and neighbourhood gateways. In [127], the authors have presented a novel authentication key exchange approach based on low-cost memristor-PUF to investigate security between the head-end system and smart meters. Furthermore, the work highlights that the presented scheme for analysing various other attack scenarios, such as replay attack, MITM,

and impersonation attack, has been considered as future scope.

In [128], researchers developed an anonymous authentication approach based on a group signature scheme with configurable linkability with tokens to reduce double spending and billing scam in the smart grid. In [129], the authors developed a hash function, ECC, and symmetric encryption-based anonymous and reliable authentication scheme for the smart grid to ensure the integrity of information transmitted between the smart meter and central service provider. Limbasiya and Arya have discussed various attacks, authentication schemes, and security parameters for secure communication in the smart grid system [130]. Researchers have presented a Diffie-hellman-based message authentication protocol for smart grid communications between the HAN-gateways and BAN-gate ways [131]. In [132], researchers have developed a lightweight ECC-based mutual authentication scheme with trifling operations to secure communication between consumers and substations for smart-grid environments. The presented work shows that the developed scheme can also analyze real-time data communications in smart grid. In [133], researchers have introduced a new AKA protocol using an ECQV implicit certificate to access data securely by providing mutual authentication in smart meters for smart grid environments. Aziz et al. implemented a lightweight authentication protocol based on the hash function with masked identity to secure information exchange between the control centre and smart breakers in a smart grid [83]. The work highlights that the proposed scheme injecting into the REF542plus controller using manufacturing software such as CAN open digital field bus can provide low computation and communication costs for real-time smart grid applications. Gope, proposed a lightweight authentication scheme, while ensuring strong user anonymity support to satisfy all the security features of M2M based home network services [134]. In [135], the authors have introduced mutually authenticated key establishment scheme to provide secure communication between the multiple smart meters and service providers in a cloud-enabled smart grid system. The work remarks the necessity to design two protocols to mitigate: one to store the gathered data in the cloud server and the other to obtain the processed data from the cloud server. In [136], [137], and [138], researchers have developed an ECC-based authentication protocol to mitigate considered communication attacks between smart grid devices and utility centres. Besides, another researcher proposed a privacy-preserving lightweight authentication scheme based on pseudo-identity and secret parameters to address the shortcomings of the ECC authentication protocol. Such shortcomings are the protocol insecurity against masquerade, smart grid device theft, and failure to ensure robust mutual authentication.

In [139], the authors have proposed a blockchain and homo-morphic encryption-based privacy-preserving data aggregate model to prevent internal and external attacks such as MITM, privileged-insider attacks, and impersonation attacks with low computational cost in a cloud computing-based smart grid system. In [140], the researchers have developed a blockchain-based secure and lightweight authentication protocol with centralized register authority to mitigate the majority of common attacks, such as replay attacks, jamming attacks, and DoS attacks in practical smart grid environments. In addition, the work highlights that the developed protocol can be used for batch verification and to evaluate dynamic issues. The Table 4. gives detailed view about the vital characteristics for cyber security against cyber attacks.

As discussed earlier, the proposed schemes can withstand a wide range of attacks, which is critical for communication networks. The security of proposed schemes is evaluated both formally and informally depending on their robustness against major cyber-attacks. From the mentioned literature, the effectiveness of proposed schemes against all possible cyber-attacks are examined in Table 5. Furthermore, the proposed schemes are primarily focused on providing security against well-known attacks such as the replay attack, MITM, impersonation attack, and failing to provide security against insider attacks, which is very difficult to detect, as shown in Fig. 5.

The proposed schemes needs to meet common security requirements such as data integrity, privacy, confidentiality and availability in order to develop good security. Table 6 presented how well the proposed schemes are defended against all potential security vulnerabilities. From the discussed literature, un-tracebility is one of the important security feature, schemes are failing to provide strong security which will become a biggest concern as shown in Fig. 6.

VII. CYBER-ATTACK INCIDENCE IN POWER SYSTEMS

A. CASE-STUDIES

Based on 2015 Ukraine cyber-attack, the authors have implemented the cascading outage analysis to analyze the impact of various cyber-attacks by opening all devices, generators, and loads connected to the lines of every transmission and distribution system provider in the North American regional interconnection system [87]. In [141], the authors have implemented electrical power system analysis software in a petrochemical plant to analyze the influence of electrical parameters on modified remote data transmitted cyber-attacks in SCADA systems. Furthermore, the designed cyber-attacks can be mitigated by using cryptography. The authors looked into cyber terrorism in NPPs after the 2014 cyber-attack on the South Korean NPP [142]. In addition, GEN-4, radiation control, and secure information management have been explored as potential solutions to the problem of cyber terrorism in NPPs. In [143], the authors developed a multi-state markov model to analyse the impact of integrity attacks such as command messages for circuit breakers and modifying IED parameter. In [144], researchers analysed IoT-related vulnerabilities, potential mitigation, and prevention techniques for real-world cyber security incidents

TABLE 4. Vital characteristics to provide cyber-security against cyber-attacks.

Ref.no	Year	Platform	Cryptographic Library	Verification Tool	Cryptographic Algorithm
[128]	2017	gcc Apple LLVM Version 8.0.0	TEPLA 2.0	-	Group Signature
[134]	2017	-	Crypto++ library	-	Light weight anonymous authentication & key agreement protocol
[131]	2017	-	-	proverif	Diffie-hellman based message authentication
[103]	2018	STM32 F4 DISCOVERY & Nano pi M3 board for to ends	Stm32 & open SSL	Proverif, ROM	ECC- based self-certified key distribution scheme
[83]	2018	MATLAB R2014a	Java class cryptosystem	-	Crypto hash function, SKA, SGMA
[106]	2018	-	-	Proverif, ROM	Identity based AKE protocol
[108]	2018	-	-	AVISPA	Secure key agreement model based on ECQV certificates
[27]	2019	LAN employed the switched Ethernet network	-	AVISPA	ECC based authentication
[76]	2019	Ubuntu 12.04 virtual machine	JPBC library Pbc 05.14 & JCE library	-	Privacy - preserving authentication protocol using PUF
[122]	2019	python	-	-	RSASSA-Probabilistic Signature Scheme
[138]	2019	Grid smart home hardware testbed, Pentium IV, Hiper smart card	-	AVISPA	ECC based authentication
[126]	2020	AT91SAM3X8E micro controller board	Arduino Libs as a cryptographic library	Mao, Boyd's logic	Light-weight mutual authentication protocol based on PUF
[137]	2020	Pentium IV, Hiper smart card	-	AVISPA, BAN logic, ROR	Light weight authentication using pseudo-identity and secret parameters
[125]	2020	-	-	Scythe, AVISPA	End-to-end PUF based AKE
[77]	2020	Ubuntu 12.04 virtual machine	JPBC library Pbc-05.14	-	PMAKE scheme based on PUF
[102]	2021	NS-3 version 3.28	C/C++ open SSL library	-	Authenticated Key Agreement
[109]	2021	Communication inside the network on IEEE 802-15-4 & network outside the building on IEEE 802-16 WiMAX	-	-	ECC based authentication
[132]	2021	NS – 2.35	PBC library version 05.12	AVISPA	ECC based authentication
[105]	2021	Philips Hiper smart card	-	AVISPA, ROR	ECC-based schnorr's signature based AKE
[136]	2021	Pentium IV, Hiper smart card	-	Proverif, BAN logic	Light weight authentication using pseudo-identity and secret parameters
[129]	2021	Pycrypto, Rasberry pi-3	-	ROM, scythe based security	ECC-based AKE protocol (ARAP-SG)
[120]	2021	AT91SAM3X8E for smart meter & Intel® core™ i7- 3612QM CPU @ 2.10 GHz and 6GB-RAM for NG	Arduino Libs as a cryptographic library	Mao, Boyd's logic	Super light-weight secure protocol based on one-way hash function and logical XOR
[135]	2021	-	-	GNY logic, Proverif	Mutually authenticated key establishment protocol
[133]	2021	-	-	CK security model	Authenticated key agreement protocol based on ECQV implicit certificate
[117]	2022	Ubuntu 12.04 virtual machine	JPBC library Pbc-05.14 & JCE library	ROM	PF-DA designed by using certificate-based cryptography
[127]	2022	MATLAB (Mathworks) using okamoto protocol	-	NIST 800-22 statistical tests	Memristor based - PUF

TABLE 5. Robustness of proposed schemes against Cyber-attacks.

Ref.no	A	B	C	D	E	F	G	H	I
[105]	✔	✔	✔	✔	✘	✔	✘	✘	✔
[102]	✔	✔	✔	✔	✘	✘	✘	✘	✘
[131]	✔	✔	✘	✔	✘	✘	✘	✘	✘
[27]	✔	✔	✘	✔	✘	✘	✘	✘	✘
[109]	✘	✘	✘	✔	✘	✘	✔	✘	✘
[132]	✔	✔	✔	✔	✘	✘	✘	✘	✔
[76]	✔	✔	✔	✔	✔	✘	✘	✘	✔
[136]	✔	✔	✔	✘	✔	✘	✘	✘	✔
[137]	✘	✔	✔	✔	✘	✘	✘	✔	✔
[138]	✔	✔	✘	✔	✘	✘	✘	✘	✔
[129]	✔	✔	✔	✔	✔	✔	✘	✘	✔
[125]	✔	✔	✔	✔	✘	✔	✘	✘	✔
[103]	✔	✔	✔	✘	✘	✔	✘	✘	✘
[106]	✔	✔	✘	✔	✘	✘	✘	✘	✘
[83]	✔	✔	✘	✘	✘	✘	✘	✘	✘
[77]	✘	✘	✔	✔	✘	✘	✘	✘	✔
[117]	✔	✔	✔	✔	✘	✘	✘	✘	✘
[134]	✘	✔	✘	✘	✔	✘	✘	✔	✘
[120]	✔	✔	✘	✔	✔	✘	✘	✘	✘
[126]	✔	✔	✔	✔	✘	✘	✘	✘	✔
[135]	✔	✔	✔	✔	✘	✔	✘	✘	✘
[133]	✔	✔	✘	✔	✔	✘	✘	✘	✘
[108]	✔	✔	✘	✔	✔	✘	✘	✘	✘

✔ - proposed scheme is strong against specified attack ;
✘ - proposed scheme is not strong against specified attack
A - Impersonation Attack ; **B** - Replay attack ; **C** - Privileged insider attack ; **D** - MITM ; **E** - DoS attack ; **F** - Ephemeral secret leakage attack ; **G** - Resending attack ; **H** - Masquerade attack ; **I** - Device stolen attack

TABLE 6. Robustness of proposed schemes against security features.

Ref.no	A	B	C	D
[103]	✔	✔	✔	✘
[106]	✔	✔	✘	✔
[83]	✔	✘	✔	✘
[125]	✔	✔	✔	✔
[105]	✔	✘	✔	✔
[102]	✘	✔	✔	✔
[131]	✘	✔	✔	✘
[27]	✔	✔	✔	✘
[109]	✘	✘	✘	✘
[132]	✔	✔	✔	✘
[76]	✔	✔	✘	✘
[136]	✔	✘	✔	✔
[137]	✔	✘	✔	✘
[138]	✔	✔	✘	✔
[129]	✔	✘	✘	✔
[77]	✔	✔	✔	✔
[117]	✘	✘	✘	✘
[134]	✔	✔	✔	✔
[120]	✘	✔	✔	✘
[126]	✘	✘	✔	✘
[135]	✔	✔	✔	✔
[133]	✔	✔	✔	✔
[108]	✔	✘	✔	✘

✔ - proposed scheme is strong against security feature ;
✘ - proposed scheme is not strong against security feature
A - Anonymity; **B** - Perfect Forward Secrecy; **C** - Mutual Authentication; **D** - Untracebility

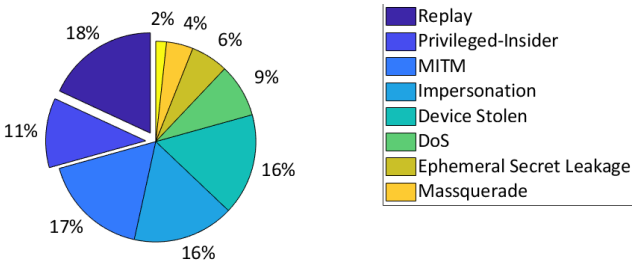


FIGURE 5. Percentage of cyber-attacks strong against proposed schemes.

affecting electricity consumers. In [145], researchers have proposed a generalized stochastic petri net to investigate the impact of integrating multi-level preventive, responsible measures on security indicators like mean-time-to-disrupt and system availability against cyber-attacks in NPPs. Furthermore, the proposed approach can also be extended to examine system dependability. In [146], the authors have provided a case study demonstration on the Proactive Intrusion Detection and Mitigation System (PIDMS) to examine the packet replay attack scenario in photo voltaic inverter communication. Moreover, PIDMS is extensible to other smart grid devices to send and receive cyber-physical data streams. In [78], researchers have presented tandem stability and machine learning-based classifiers to analyze the influence of time delay attacks on automatic generation control in the power grid. The presented approach has been evaluated and verified on the IEEE 37-bus system model.

B. REAL-WORLD CYBER-ATTACK INCIDENCES ON POWER SYSTEMS

1) ZERO-DAY ATTACK ON DAVIS-BESSE NPP, 2003

By injecting a zero-day attack into a micro-soft SQL server, attackers gained access to the secret and control

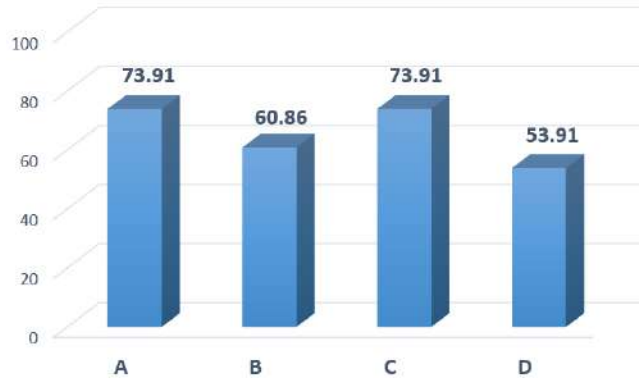


FIGURE 6. Percentage of proposed schemes strong against specified security features.

networks of Davis-Besse NPP in Ohio. The injected malware generated a massive amount of traffic in order to disrupt communication networks between corporate and control networks. In addition, the worm had left a safety monitoring system inoperable for more than five hours. Employees were unable to monitor the core temperature sensors at the plant.

2) STUXNET ATTACK ON NUCLEAR PROGRAM OF IRAN, 2010

On November 29, 2010, the Iranian president stated that the stuxnet virus had destroyed hundreds of centrifuges used to enrich uranium at Natanz nuclear enrichment. According to estimates, the stuxnet worm destroyed Nine Eighty-four uranium enrichment centrifuges and destroyed enrichment efficiency by thirty percent. Another virus corrupted government computers at nuclear plants and stolen data in 2012.

3) CYBER-ATTACK ON KOREA HYDRO AND NUCLEAR POWER, 2014

On December 23, 2014, Korea Hydro and Nuclear Power announced that their computer systems had been hacked. ''Unless you stop operating the nuclear power plants until Christmas and give us $1 billion, we will continue to release the facility's secret data'', the hackers posted on their twitter page. Furthermore, two nuclear reactor manuals from Korea Hydro and Nuclear Power were posted online, exposed ten thousand employee's personal data.

4) SANDWORM ATTACK ON UKRAINE ELECTRICITY COMPANY, 2015

On December 23, 2015, remote cyber intrusions at three electric power distribution companies caused a blackout that left over 2,25,000 customers without power for 16 hours in Prykarpattyaoblenergo, Ukraine. The attackers injected malware through spear phishing emails with malicious attachments, gaining access to the SCADA control and then opened breakers at over 30 substations. Furthermore, serial-to-ethernet servers, backup power was disabled with bad firmware, and a DoS attack on the utility telephone system was also carried out.

5) INDUSTROYER ATTACK ON UKRAINE ELECTRICITY COMPANY, 2016

On December 17, 2016, a remote cyber intrusion occurred at a local substation that supplies power to the capital city of Kyiv. Attackers opened breakers at a substation again. However, this time they attempted to compromise the relays.

6) DTRACK ATTACK ON KUDANKULAM NPP, 2019

On September 4, 2019, malware was discovered on a personal computer belonging to a user who was connected to an administrative internet network. The nuclear power corporation of India limited issued an official statement confirming the incident. ''This PC has been disconnected from the critical internal network and networks are constantly monitored.''

7) REVIL RANDSOMWARE ATTACK ON UK ELECTRICITY MARKET, 2020

On May 12, 2020, hackers attacked internal IT systems at Elexon, which is center of balancing and settlement system, works for the energy system operators of Great Britain's national grid. Elexon's official response to this incident was as follows: ''the attack is to our internal IT systems and ELEXON's laptops only. Electricity supply is not affected.''

8) CYBER-ATTACK ON LADAKH ELECTRICITY DISTRIBUTION CENTER, 2022

On March 2022, unknown hackers attempted but failed to hack into an electricity distribution center. ''two attempts by the hackers to target electricity distribution centers near Ladakh were unsuccessful. We have already strengthened our defenses to counter such attacks,'' said India's minister of power and renewable energy.

VIII. CONCLUSION

This paper reviewed various approaches for cyber-attacks detection, protection and impact analysis in multiple areas such as wind farms, PV systems, transmission systems, smart meters and communication channels. A need of cyber security for IoT-based smart grid systems has been examined. This review article analyzed the literature to provide an overview of the need and potential methods for detecting and mitigating cyber attacks, particularly communication attacks, using artificial intelligence, block chain and cryptographic primitives. When it comes to the analysis of proposed literature, it suggested vital characteristics, simulation platform, libraries to do practical design, simulation and verification of cryptographic primitives for secure communication between two endpoints in a smart grid system. and Furthermore, the robustness of security properties, cryptographic algorithms against various cyber attacks was analyzed to suggest an unexplored attack.

A. FINDINGS

Based to the literature, FDIA is the most serious concern in the power system. The authors presented unique detection strategies for FDIA by employing thevenins equivalent parameters [48], an extremely randomised tree algorithm [70], and auto regressive models such as wavelet and TCN instead of the recurrent family model [33]. Not only FDIA, eavesdropping attack also one of the cyber security vulnerable attacks in smart communication system. Researchers developed a deep learning architecture [94], SVM [96], decomposition form of the system matrices [26], dual denoising auto-encoder based encryptor [82] and a certificate-less signcryption [114] to analyse impact and detect DoS, eavesdropping attacks. Furthermore, researchers utilised a zero-knowledge proofs & the pailiers crypto system [102], as well as a blockchain & homomorphic encryption based aggregation architecture [139], to minimise smart meter data manipulation attacks.

Based on presented literature, the following new findings are highlighted in the field of power systems cyber-attacks:

- FDIA's is one of the concerned attacks in power systems. Machine learning-based techniques such as extremely randomized tree and isolation forest could deliver accurate and fast detection of FDIAs with an accuracy of more than 99.75% and a detection time of less than 1.944 seconds, respectively. Because, false factor increases then detection time also increases.
- Cryptographic algorithms such as elliptic curve-based encryption incorporated with block chain, will facilitate electricity trading without the mediator as well as provide low computation cost for data aggregation. In addition, the kind of approach will provide authenticated security for deregulation energy markets such as TEMS, Demand Response.
- To satisfy the IEC 61850 protocols in the standard of IEC 62351, the control commands must transmit within 4ms between substation to circuit breakers. Hash function and private key based protocol can agree on the session key within 0.057ms, which satisfies the time restrictions of GOOSE and sampled value protocols and will provide private key privacy and session key security against communication attacks such as MITM, replay attack and DoS attack. Furthermore, RSASSA-PKCS-V1_5 fails to meet the timing standards of GOOSE messages, which may leads to revisit the IEC 62351-6 standard with new considerations for better cyber security.

B. FUTURE DIRECTIONS

The comprehensive review has opened up new scopes in power systems cyber securities.

- Establishment of a detection approach based on dynamic watermarking to detect sophisticated adversaries that can be scaled up to large-scale power systems [36].
- A deep stacking auto-encoder technique can also be used to identify the root cause and implement appropriate control measures to prevent cyber intrusion between utilities and consumers in smart grid [62].
- Despite FDIA detection in smart grid, artificial feed-forward networks based on a true-data integrity agent model can be employed for cyber security in intelligent transportation systems [91].
- Merging of attributed-based signcryption with proxy re-encryption scheme to secure data from communication attacks in smart grid [114].
- A lightweight multiparty computation protocol based on blockchain that can also be suitable for resource-constrained networks like the smart grid [120].
- Blockchain technology can also provide better authentication protocols for the smart grid privacy protection [102].

The following research areas are suggested in the field of cyber-attacks in power systems based on existing research:

1) Development of decentralized defense system to identify and mitigate threats in renewable energies such as wind and photo-voltaic systems, based control networks using artificial intelligent control techniques that can be extend to large-scale power systems. Furthermore, evaluation of the impact of renewable energy power plants temporal characteristics and participation in power markets.
2) Development of effective strategies for analyzing the impact, detecting and protecting against cyber-attacks on state estimation (PMU, Direct Current, Alternative Current -High Voltage Direct Current) in transmission lines, as well as improving a framework for locating the events at each line.
3) Implementation of a framework to analyze a few realistic behaviours of agents, operators and electricity market mechanisms and the development of cyber security actions for data manipulation attacks and energy theft attacks using intelligent transportation in TES.
4) There is a need to analyze the impact of cyber-attacks such as DoS and MITM in various environments like peer-to-peer energy trading, M2M communication, TES, demand side management, and distribution side, as well as design the detection schemes and monitoring systems to identify meters that inject false power consumption data and to handle zero-day sort of attacks.
5) The scope of IoT- based smart grid projects is limited to closed environments (LAN, recursive internet working architecture). There is a need to extend the real-world smart grid infrastructures, such as the implementation and evaluation of a prototype in collaboration with smart grid operators or service providers, to analyze real-world data communication in smart grids.
6) Establishment of a lightweight authentication scheme to address specific cyber security challenges such as privacy of users and policy makers, reducing the protocol message size, lowering the computational cost

and shortening the time domains for a distributed secret-key management scheme. To enhance the privacy protection in smart grid communication channels blockchain technology shall be adopted.

IX. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "A survey on smart grid potential applications and communication requirements," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 28–42, Feb. 2013.

[2] S. Garg, K. Kaur, G. Kaddoum, J. J. P. C. Rodrigues, and M. Guizani, "Secure and lightweight authentication scheme for smart metering infrastructure in smart grid," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3548–3557, May 2020.

[3] C. Ang. (2021). *The Most Cyber Attsignificant Acks From 2006-2020, by Country*. [Online]. Available: https://www.visualcapitalist.com/cyber-attacks-worldwide-2006-2020/

[4] M. Benmalek and Y. Challal, "MK-AMI: Efficient multi-group key management scheme for secure communications in AMI systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

[5] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Comput. Netw.*, vol. 169, Mar. 2020, Art. no. 107094.

[6] R. E. Pérez-Guzmán, Y. Salgueiro-Sicilia, and M. Rivera, "Communication systems and security issues in smart microgrids," in *Proc. IEEE Southern Power Electron. Conf. (SPEC)*, Dec. 2017, pp. 1–6.

[7] M. Z. Gunduz and R. Das, "Analysis of cyber-attacks on smart grid applications," in *Proc. Int. Conf. Artif. Intell. Data Process. (IDAP)*, Sep. 2018, pp. 1–5.

[8] H. Ritchie and M. Roser. (2018). *Two-Thirds of Global Population Will Live in Cities By 2050, Our World in Data*. [Online]. Available: https://ourworldindata.org/urbanization

[9] M. Benmalek, Y. Challal, and A. Derhab, "Authentication for smart grid AMI systems: Threat models, solutions, and challenges," in *Proc. IEEE 28th Int. Conf. Enabling Technologies: Infrastructure Collaborative Enterprises (WETICE)*, Jun. 2019, pp. 208–213.

[10] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecuriy: A review," *IEEE Access*, vol. 10, pp. 19572–19585, 2022.

[11] A. Hasankhani, S. M. Hakimi, M. Bisheh-Niasar, M. Shafie-khah, and H. Asadolahi, "Blockchain technology in the future smart grids: A comprehensive review and frameworks," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106811.

[12] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, and N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future," *Electr. Power Syst. Res.*, vol. 215, Feb. 2023, Art. no. 108975.

[13] M. Abdalzaher, M. Fouda, A. Emran, Z. Fadlullah, and M. Ibrahem, "A survey on key management and authentication approaches in smart metering systems," *Energies*, vol. 16, no. 5, p. 2355, Mar. 2023.

[14] M. Sewak, S. K. Sahay, and H. Rathore, "Deep reinforcement learning in the advanced cybersecurity threat detection and protection," *Inf. Syst. Frontiers*, vol. 25, pp. 589–611, Aug. 2022.

[15] S. Banik, S. K. Saha, T. Banik, and S. M. M. Hossain, "Anomaly detection techniques in smart grid systems: A review," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, Jun. 2023, pp. 331–337.

[16] J. Kua, M. B. Hossain, I. Natgunanathan, and Y. Xiang, "Privacy preservation in smart meters: Current status, challenges and future directions," *Sensors*, vol. 23, no. 7, p. 3697, Apr. 2023.

[17] K. Y. Yap, H. H. Chin, and J. J. Klemeš, "Blockchain technology for distributed generation: A review of current development, challenges and future prospect," *Renew. Sustain. Energy Rev.*, vol. 175, Apr. 2023, Art. no. 113170.

[18] S. S. Koduru, V. S. P. Machina, and S. Madichetty, "Cyber attacks in cyber-physical microgrid systems: A comprehensive review," *Energies*, vol. 16, no. 12, p. 4573, Jun. 2023.

[19] M. Lydia, G. E. P. Kumar, and A. I. Selvakumar, "Securing the cyber-physical system: A review," *Cyber-Phys. Syst.*, vol. 9, no. 3, pp. 193–223, Jul. 2023.

[20] S. Bhattacharjee and S. K. Das, "Detection and forensics against stealthy data falsification in smart metering infrastructure," *IEEE Trans. Depend. Secure Comput.*, vol. 18, no. 1, pp. 356–371, Jan. 2021.

[21] X. Lou, "Learning-based time delay attack characterization for cyber-physical systems," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2019, pp. 1–6.

[22] A. Ameli, A. Ayad, E. F. El-Saadany, M. M. A. Salama, and A. Youssef, "A learning-based framework for detecting cyber-attacks against line current differential relays," *IEEE Trans. Power Del.*, vol. 36, no. 4, pp. 2274–2286, Aug. 2021.

[23] S. Yankson and M. Ghamkhari, "Transactive energy to thwart load altering attacks on power distribution systems," *Future Internet*, vol. 12, no. 1, p. 4, Dec. 2019.

[24] J. Khazaei, "Cyberattacks with limited network information leading to transmission line overflow in cyber–physical power systems," *Sustain. Energy, Grids Netw.*, vol. 27, Sep. 2021, Art. no. 100505.

[25] K. Wang, L. Yuan, T. Miyazaki, Y. Chen, and Y. Zhang, "Jamming and eavesdropping defense in green cyber–physical transportation systems using a Stackelberg game," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4232–4242, Sep. 2018.

[26] R. Gao and G.-H. Yang, "Sampled-data distributed state estimation with multiple transmission channels under denial-of-service attacks," *Appl. Math. Comput.*, vol. 429, Sep. 2022, Art. no. 127229.

[27] M. F. Moghadam, M. Nikooghadam, A. H. Mohajerzadeh, and B. Movali, "A lightweight key management protocol for secure communication in smart grids," *Electr. Power Syst. Res.*, vol. 178, Jan. 2020, Art. no. 106024.

[28] I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," *Energy*, vol. 145, pp. 65–78, Feb. 2018.

[29] J. Staggs, D. Ferlemann, and S. Shenoi, "Wind farm security: Attack surface, targets, scenarios and mitigation," *Int. J. Crit. Infrastructure Protection*, vol. 17, pp. 3–14, Jun. 2017.

[30] J. Y. Siu and S. K. Panda, "A review of cyber-physical security in the generation system of the grid," in *Proc. IECON 46th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2020, pp. 1520–1525.

[31] J. Ye, A. Giani, A. Elasser, S. K. Mazumder, C. Farnell, H. A. Mantooth, T. Kim, J. Liu, B. Chen, G.-S. Seo, W. Song, M. D. R. Greidanus, S. Sahoo, F. Blaabjerg, J. Zhang, L. Guo, B. Ahn, M. B. Shadmand, N. R. Gajanur, and M. A. Abbaszada, "A review of cyber–physical security for photovoltaic systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 4, pp. 4879–4901, Aug. 2022.

[32] A. Jindal, A. K. Marnerides, A. Scott, and D. Hutchison, "Identifying security challenges in renewable energy systems: A wind turbine case study," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, Jun. 2019, pp. 370–372.

[33] F. Almutairy, L. Scekic, R. Elmoudi, and S. Wshah, "Accurate detection of false data injection attacks in renewable power systems using deep learning," *IEEE Access*, vol. 9, pp. 135774–135789, 2021.

[34] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.

[35] N. Trantham and A. Garcia, "Reputation dynamics in networks: Application to cyber security of wind farms," *Syst. Eng.*, vol. 18, no. 4, pp. 339–348, Jul. 2015.

[36] T. Huang, B. Satchidanandan, P. R. Kumar, and L. Xie, "An online detection framework for cyber attacks on automatic generation control," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6816–6827, Nov. 2018.

[37] I. Gandhi, L. Ravi, V. Vijayakumar, and V. Subramaniyaswamy, "Improving security for wind energy systems in smart grid applications using digital protection technique," *Sustain. Cities Soc.*, vol. 60, Sep. 2020, Art. no. 102265.

[38] H. Jia, C. Shao, D. Liu, C. Singh, Y. Ding, and Y. Li, "Operating reliability evaluation of power systems with demand-side resources considering cyber malfunctions," *IEEE Access*, vol. 8, pp. 87354–87366, 2020.

[39] N. Fardad, S. Soleymani, and F. Faghihi, "Cyber defense analysis of smart grid including renewable energy resources based on coalitional game theory," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 2063–2077, Aug. 2018.

[40] S. Lee and J.-H. Huh, "An effective security measures for nuclear power plant using big data analysis approach," *J. Supercomput.*, vol. 75, no. 8, pp. 4267–4294, 2019.

[41] C. Lee, Y. Ho Chae, and P. H. Seong, "Development of a method for estimating security state: Supporting integrated response to cyber-attacks in NPPs," *Ann. Nucl. Energy*, vol. 158, Aug. 2021, Art. no. 108287.

[42] C. Lee, S. M. Han, Y. H. Chae, and P. H. Seong, "Development of a cyberattack response planning method for nuclear power plants by using the Markov decision process model," *Ann. Nucl. Energy*, vol. 166, Feb. 2022, Art. no. 108725.

[43] T. Krause, R. Ernst, B. Klaer, I. Hacker, and M. Henze, "Cybersecurity in power grids: Challenges and opportunities," *Sensors*, vol. 21, no. 18, p. 6225, Sep. 2021.

[44] H. Rahimpour, J. Tusek, A. Abuadbba, A. Seneviratne, T. Phung, A. Musleh, and B. Liu, "Cybersecurity challenges of power transformers," 2023, *arXiv:2302.13161*.

[45] R. Fan, J. Lian, K. Kalsi, and M. Elizondo, "Impact of cyber attacks on high voltage DC transmission damping control," *Energies*, vol. 11, no. 5, p. 1046, Apr. 2018.

[46] M. J. P. Jaghargh and H. R. Mashhadi, "Structural and behavioural evaluation of renewable energy power plants' impacts on transmission network congestion using an analytical approach," *IET Renew. Power Gener.*, vol. 14, no. 7, pp. 1164–1173, May 2020.

[47] Y. Dvorkin and S. Garg, "IoT-enabled distributed cyber-attacks on transmission and distribution grids," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2017, pp. 1–6.

[48] M. Ghafouri, M. Au, M. Kassouf, M. Debbabi, C. Assi, and J. Yan, "Detection and mitigation of cyber attacks on voltage stability monitoring of smart grids," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5227–5238, Nov. 2020.

[49] D. Wilson, Y. Tang, J. Yan, and Z. Lu, "Deep learning-aided cyber-attack detection in power transmission systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[50] A. Ahmed, V. V. G. Krishnan, S. A. Foroutan, M. Touhiduzzaman, C. Rublein, A. Srivastava, Y. Wu, A. Hahn, and S. Suresh, "Cyber physical security analytics for anomalies in transmission protection systems," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 6313–6323, Nov. 2019.

[51] S. Chakrabarty and B. Sikdar, "Detection of hidden transformer tap change command attacks in transmission networks," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5161–5173, Nov. 2020.

[52] S. Pal, B. Sikdar, and J. H. Chow, "Classification and detection of PMU data manipulation attacks using transmission line parameters," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5057–5066, Sep. 2018.

[53] Q. Yang, L. Jiang, W. Hao, B. Zhou, P. Yang, and Z. Lv, "PMU placement in electric transmission networks for reliable state estimation against false data injection attacks," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1978–1986, Dec. 2017.

[54] M. Dehghani, M. Ghiasi, T. Niknam, A. Kavousi-Fard, M. Shasadeghi, N. Ghadimi, and F. Taghizadeh-Hesary, "Blockchain-based securing of data exchange in a power transmission system considering congestion management and social welfare," *Sustainability*, vol. 13, no. 1, p. 90, Dec. 2020.

[55] F. Mohammadi and R. Rashidzadeh, "Impact of stealthy false data injection attacks on power flow of power transmission lines—A mathematical verification," *Int. J. Electr. Power Energy Syst.*, vol. 142, Nov. 2022, Art. no. 108293.

[56] C. Ma, "Smart city and cyber-security; technologies used, leading challenges and future recommendations," *Energy Rep.*, vol. 7, pp. 7999–8012, Nov. 2021.

[57] X. G. Shan and J. Zhuang, "A game-theoretic approach to modeling attacks and defenses of smart grids at three levels," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106683.

[58] M. Montakhabi, A. Madhusudan, S. van der Graaf, A. Abidin, P. Ballon, and M. A. Mustafa, "Sharing economy in future peer-to-peer electricity trading markets: Security and privacy analysis," in *Proc. Workshop Decentralized IoT Syst. Secur. (DISS)*, San Diego, CA, USA, 2020, pp. 1–6.

[59] A. Marufu, A. V. Kayem, and S. D. Wolthusen, "The design and classification of cheating attacks on power marketing schemes in resource constrained smart micro-grids," in *Smart Micro-Grid Systems Security and Privacy*. Cham, Switzerland: Springer, 2018, pp. 103–144.

[60] R. Dasgupta, A. Sakzad, and C. Rudolph, "Cyber attacks in transactive energy market-based microgrid systems," *Energies*, vol. 14, no. 4, p. 1137, Feb. 2021.

[61] A. Arman, V. V. G. Krishnan, A. Srivastava, Y. Wu, and S. Sindhu, "Cyber physical security analytics for transactive energy systems using ensemble machine learning," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2018, pp. 1–6.

[62] Y. Zhang, V. V. G. Krishnan, J. Pi, K. Kaur, A. Srivastava, A. Hahn, and S. Suresh, "Cyber physical security analytics for transactive energy systems," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 931–941, Mar. 2020.

[63] P. Wang, K. Ma, J. Lian, and D. J. Hammerstrom, "On anomaly detection for transactive energy systems with competitive market," *Int. J. Electr. Power Energy Syst.*, vol. 128, Jun. 2021, Art. no. 106662.

[64] S. Fkaier, M. Khalgui, G. Frey, Z. Li, and J. Yu, "Secure distributed power trading protocol for networked microgrids based on blockchain and elliptic curve cryptography," *IET Smart Grid*, vol. 6, no. 2, pp. 175–189, Apr. 2023.

[65] S. Pal, S. Biswas, S. Sridhar, A. Ashok, J. Hansen, and V. Amatya, "Understanding impacts of data integrity attacks on transactive control systems," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2020, pp. 1–5.

[66] V. V. G. Krishnan, Y. Zhang, K. Kaur, A. Hahn, A. Srivastava, and S. Sindhu, "Cyber-security analysis of transactive energy systems," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo. (T&D)*, Apr. 2018, pp. 1–9.

[67] Q. Huang, T. E. McDermott, Y. Tang, A. Makhmalbaf, D. J. Hammerstrom, A. R. Fisher, L. D. Marinovici, and T. Hardy, "Simulation-based valuation of transactive energy systems," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 4138–4147, Sep. 2019.

[68] Y. Liu, S. Hu, and T.-Y. Ho, "Leveraging strategic detection techniques for smart home pricing cyberattacks," *IEEE Trans. Depend. Secure Comput.*, vol. 13, no. 2, pp. 220–235, Mar. 2016.

[69] Y. Song, Z. Yu, X. Liu, J. Tian, and M. Chen, "Isolation forest based detection for false data attacks in power systems," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia)*, May 2019, pp. 4170–4174.

[70] S. H. Majidi, S. Hadayeghparast, and H. Karimipour, "FDI attack detection using extra trees algorithm and deep learning algorithm-autoencoder in smart grid," *Int. J. Crit. Infrastruct. Protection*, vol. 37, Jul. 2022, Art. no. 100508.

[71] C.-C. Sun, D. J. Sebastian Cardenas, A. Hahn, and C.-C. Liu, "Intrusion detection for cybersecurity of smart meters," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 612–622, Jan. 2021.

[72] P. A. Giglou and S. N. Ravadanegh, "Defending against false data injection attack on demand response program: A bi-level strategy," *Sustain. Energy, Grids Netw.*, vol. 27, Sep. 2021, Art. no. 100506.

[73] S. Ahmadian, X. Tang, H. A. Malki, and Z. Han, "Modelling cyber attacks on electricity market using mathematical programming with equilibrium constraints," *IEEE Access*, vol. 7, pp. 27376–27388, 2019.

[74] J. Tao and G. Michailidis, "A statistical framework for detecting electricity theft activities in smart grid distribution networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 205–216, Jan. 2020.

[75] K. Yang, H. Wang, H. Wang, and L. Sun, "An effective intrusion-resilient mechanism for programmable logic controllers against data tampering attacks," *Comput. Ind.*, vol. 138, Jun. 2022, Art. no. 103613.

[76] P. Gope and B. Sikdar, "Privacy-aware authenticated key agreement scheme for secure smart grid communication," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3953–3962, Jun. 2018.

[77] P. Gope, "PMAKE: Privacy-aware multi-factor authenticated key establishment scheme for advance metering infrastructure in smart grid," *Comput. Commun.*, vol. 152, pp. 338–344, Feb. 2020.

[78] X. Lou, C. Tran, R. Tan, D. K. Y. Yau, Z. T. Kalbarczyk, A. K. Banerjee, and P. Ganesh, "Assessing and mitigating impact of time delay attack: Case studies for power grid controls," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 141–155, Jan. 2020.

[79] M. Attia, S. M. Senouci, H. Sedjelmaci, E.-H. Aglzim, and D. Chrenko, "An efficient intrusion detection system against cyber-physical attacks in the smart grid," *Comput. Electr. Eng.*, vol. 68, pp. 499–512, May 2018.

[80] M. R. C. Acosta, S. Ahmed, C. E. Garcia, and I. Koo, "Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks," *IEEE Access*, vol. 8, pp. 19921–19933, 2020.

[81] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1652–1656, Oct. 2015.

[82] S. Wu, Y. Jiang, H. Luo, and X. Li, "Deep learning-based defense and detection scheme against eavesdropping and typical cyber-physical attacks," in *Proc. CAA Symp. Fault Detection, Supervision, Saf. Tech. Processes (SAFEPROCESS)*, Dec. 2021, pp. 1–6.

[83] I. Aziz, H. Jin, I. Abdulqadder, Z. Hussien, Z. Abduljabbar, and F. Flaih, "A lightweight scheme to authenticate and secure the communication in smart grids," *Appl. Sci.*, vol. 8, no. 9, p. 1508, Sep. 2018.

[84] G. Tertytchny, N. Nicolaou, and M. K. Michael, "Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine learning," *Microprocessors Microsyst.*, vol. 77, Sep. 2020, Art. no. 103121.

[85] D. Kosmanos, A. Pappas, L. Maglaras, S. Moschoyiannis, F. J. Aparicio-Navarro, A. Argyriou, and H. Janicke, "A novel intrusion detection system against spoofing attacks in connected electric vehicles," *Array*, vol. 5, Mar. 2020, Art. no. 100013.

[86] A. O. Aluko, R. P. Carpanen, D. G. Dorrell, and E. E. Ojo, "Real-time cyber attack detection scheme for standalone microgrids," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21481–21492, Nov. 2022.

[87] B. Huang, M. Majidi, and R. Baldick, "Case study of power system cyber attack using cascading outage analysis model," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[88] N. Kumar, V. M. Mishra, and A. Kumar, "Smart grid security by embedding S-Box advanced encryption standard," *Intell. Autom. Soft Comput.*, vol. 34, no. 1, pp. 623–638, 2022.

[89] M. K. Hasan, A. A. Habib, Z. Shukur, F. Ibrahim, S. Islam, and M. A. Razzaque, "Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations," *J. Netw. Comput. Appl.*, vol. 209, Jan. 2023, Art. no. 103540.

[90] M. Sirajuddin, C. Rupa, C. Iwendi, and C. Biamba, "TBSMR: A trust-based secure multipath routing protocol for enhancing the QoS of the mobile ad hoc network," *Secur. Commun. Netw.*, vol. 2021, pp. 1–9, Apr. 2021.

[91] S. Sengan, V. Subramaniyaswamy, V. Indragandhi, P. Velayutham, and L. Ravi, "Detection of false data cyber-attacks for the assessment of security in smart grid using deep learning," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107211.

[92] G. D. L. T. Parra, P. Rad, K.-K.-R. Choo, and N. Beebe, "Detecting Internet of Things attacks using distributed deep learning," *J. Netw. Comput. Appl.*, vol. 163, Aug. 2020, Art. no. 102662.

[93] H. Wang, J. Ruan, Z. Ma, B. Zhou, X. Fu, and G. Cao, "Deep learning aided interval state prediction for improving cyber security in energy internet," *Energy*, vol. 174, pp. 1292–1304, May 2019.

[94] L. Ashiku and C. Dagli, "Network intrusion detection system using deep learning," *Proc. Comput. Sci.*, vol. 185, no. 1, pp. 239–247, 2021.

[95] R. B. Benisha and S. R. Ratna, "Detection of data integrity attacks by constructing an effective intrusion detection system," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 11, pp. 5233–5244, Nov. 2020.

[96] T. M. Hoang, T. Q. Duong, H. D. Tuan, S. Lambotharan, and L. Hanzo, "Physical layer security: Detection of active eavesdropping attacks by support vector machines," *IEEE Access*, vol. 9, pp. 31595–31607, 2021.

[97] S. Sahoo, S. Mishra, J. C. Peng, and T. Dragicevic, "A stealth cyber-attack detection strategy for DC microgrids," *IEEE Trans. Power Electron.*, vol. 34, no. 8, pp. 8162–8174, Aug. 2019.

[98] Y. Taniguchi, H. Tsutsumi, N. Iguchi, and K. Watanabe, "Design and evaluation of a proxy-based monitoring system for OpenFlow networks," *Sci. World J.*, vol. 2016, pp. 1–10, Jan. 2016.

[99] Y. Huang, L. Jin, Z. Zhong, Y. Lou, and S. Zhang, "Detection and defense of active attacks for generating secret key from wireless channels in static environment," *ISA Trans.*, vol. 99, pp. 231–239, Apr. 2020.

[100] A. Tolba and Z. Al-Makhadmeh, "A cybersecurity user authentication approach for securing smart grid communications," *Sustain. Energy Technol. Assessments*, vol. 46, Aug. 2021, Art. no. 101284.

[101] M. Sirajuddin, C. Rupa, S. Bhatia, R. N. Thakur, and A. Mashat, "Hybrid cryptographic scheme for secure communication in mobile ad hoc network-based E-healthcare system," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–8, Jun. 2022.

[102] X. Xiang and J. Cao, "An efficient authenticated key agreement scheme supporting privacy-preservation for smart grid communication," *Electric Power Syst. Res.*, vol. 203, Feb. 2022, Art. no. 107630.

[103] D. Abbasinezhad-Mood and M. Nikooghadam, "An anonymous ECC-based self-certified key distribution scheme for the smart grid," *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 7996–8004, Oct. 2018.

[104] A. A. Khan, S. Itoo, V. Kumar, M. Ahmad, and S. Jangirala, "Cryptanalysis and design flaws of anonymous ECC based self-certified key distribution scheme for smart grid," *Mater. Today, Proc.*, vol. 57, pp. 2185–2189, Jan. 2022.

[105] J. Srinivas, A. K. Das, X. Li, M. K. Khan, and M. Jo, "Designing anonymous signature-based authenticated key exchange scheme for Internet of Things-enabled smart grid systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4425–4436, Jul. 2021.

[106] K. Mahmood, "Pairing based anonymous and secure key agreement protocol for smart grid edge computing infrastructure," *Future Gener. Comput. Syst.*, vol. 88, pp. 491–500, Nov. 2018.

[107] C. Guo, X. Jiang, K.-K. R. Choo, X. Tang, and J. Zhang, "Lightweight privacy preserving data aggregation with batch verification for smart grid," *Future Gener. Comput. Syst.*, vol. 112, pp. 512–523, Nov. 2020.

[108] A. Braeken, P. Kumar, and A. Martin, "Efficient and provably secure key agreement for modern smart metering communications," *Energies*, vol. 11, no. 10, p. 2662, 2018.

[109] T. Chen, X. Yin, and G. Wang, "Securing communications between smart grids and real users; providing a methodology based on user authentication," *Energy Rep.*, vol. 7, pp. 8042–8050, Nov. 2021.

[110] P. M. Lima, L. K. Carvalho, and M. V. Moreira, "Confidentiality of cyber-physical systems using event-based cryptography," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1735–1740, 2020.

[111] N. B. Samyuel and B. A. Shimray, "Securing IoT device communication against network flow attacks with recursive internetworking architecture (RINA)," *ICT Exp.*, vol. 7, no. 1, pp. 110–114, Mar. 2021.

[112] A. N. Nazarov and A. N. A. Koupaei, "An architecture model for active cyber attacks on intelligence info-communication systems: Application based on advance system encryption (AES-512) using pre-encrypted search table and pseudo-random Functions(PRFs)," in *Proc. Int. Conf. Eng. Telecommun. (EnT)*, Nov. 2019, pp. 1–5.

[113] M. A. Elakrat and J. C. Jung, "Development of field programmable gate array–based encryption module to mitigate man-in-the-middle attack for nuclear power plant data communication network," *Nucl. Eng. Technol.*, vol. 50, no. 5, pp. 780–787, Jun. 2018.

[114] E. Ahene, Z. Qin, A. K. Adusei, and F. Li, "Efficient signcryption with proxy re-encryption and its application in smart grid," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9722–9737, Dec. 2019.

[115] Z. Wang, Y. Liu, Z. Ma, X. Liu, and J. Ma, "LiPSG: Lightweight privacy-preserving Q-learning-based energy management for the IoT-enabled smart grid," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3935–3947, May 2020.

[116] I. A. Kamil and S. O. Ogundoyin, "EPDAS: Efficient privacy-preserving data analysis scheme for smart grid network," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 33, no. 2, pp. 208–217, 2021.

[117] G. K. Verma, P. Gope, and N. Kumar, "PF-DA: Pairing free and secure data aggregation for energy internet-based smart meter-to-grid communication," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2294–2304, May 2022.

[118] A. S. Sani, D. Yuan, J. Jin, L. Gao, S. Yu, and Z. Y. Dong, "Cyber security framework for Internet of Things-based energy Internet," *Future Gener. Comput. Syst.*, vol. 93, pp. 849–859, Apr. 2019.

[119] P. M. Lima, M. V. S. Alves, L. K. Carvalho, and M. V. Moreira, "Security against communication network attacks of cyber-physical systems," *J. Control, Autom. Electr. Syst.*, vol. 30, no. 1, pp. 125–135, Feb. 2019.

[120] S. Aghapour, M. Kaveh, M. R. Mosavi, and D. Martín, "An ultra-lightweight mutual authentication scheme for smart grid two-way communications," *IEEE Access*, vol. 9, pp. 74562–74573, 2021.

[121] B. K. Sethi, A. Singh, D. Singh, and R. Misra, "Optimal energy m anagement of s mart b uildings under cyber attack," *Int. J. Energy Res.*, vol. 45, no. 14, pp. 19895–19908, Nov. 2021.

[122] S. M. Farooq, S. M. S. Hussain, and T. S. Ustun, "Performance evaluation and analysis of IEC 62351-6 probabilistic signature scheme for securing GOOSE messages," *IEEE Access*, vol. 7, pp. 32343–32351, 2019.

[123] P. Gope and B. Sikdar, "Lightweight and privacy-preserving two-factor authentication scheme for IoT devices," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 580–589, Feb. 2019.

[124] P. Gope and B. Sikdar, "A privacy-aware reconfigurable authenticated key exchange scheme for secure communication in smart grids," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5335–5348, Nov. 2021.

[125] M. Tahavori and F. Moazami, "Lightweight and secure PUF-based authenticated key agreement scheme for smart grid," *Peer Peer Netw. Appl.*, vol. 13, no. 5, pp. 1616–1628, Sep. 2020.
[126] M. Kaveh and M. R. Mosavi, "A lightweight mutual authentication for smart grid neighborhood area network communications based on physically unclonable function," *IEEE Syst. J.*, vol. 14, no. 3, pp. 4535–4544, Sep. 2020.
[127] H. M. Ibrahim, H. Abunahla, B. Mohammad, and H. AlKhzaimi, "Memristor-based PUF for lightweight cryptographic randomness," *Sci. Rep.*, vol. 12, no. 1, pp. 1–18, May 2022.
[128] H. Kishimoto, N. Yanai, and S. Okamura, "An anonymous authentication protocol for the smart grid," in *Smart Micro-Grid Systems Security and Privacy*. Cham, Switzerland: Springer, 2018, pp. 29–52.
[129] M. Tanveer, A. U. Khan, H. Shah, A. Alkhayyat, S. A. Chaudhry, and M. Ahmad, "ARAP-SG: Anonymous and reliable authentication protocol for smart grids," *IEEE Access*, vol. 9, pp. 143366–143377, 2021.
[130] T. Limbasiya and A. Arya, "Attacks on authentication and authorization models in smart grid," in *Smart Micro-Grid Systems Security and Privacy*. Cham, Switzerland: Springer, 2018, pp. 53–70.
[131] X. Li, F. Wu, S. Kumari, L. Xu, A. K. Sangaiah, and K.-K.-R. Choo, "A provably secure and anonymous message authentication scheme for smart grids," *J. Parallel Distrib. Comput.*, vol. 132, pp. 242–249, Oct. 2019.
[132] D. Sadhukhan, S. Ray, M. S. Obaidat, and M. Dasgupta, "A secure and privacy preserving lightweight authentication scheme for smart-grid communication using elliptic curve cryptography," *J. Syst. Archit.*, vol. 114, Mar. 2021, Art. no. 101938.
[133] M. Qi and J. Chen, "Two-pass privacy preserving authenticated key agreement scheme for smart grid," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3201–3207, Sep. 2021.
[134] P. Gope, "Anonymous mutual authentication with location privacy support for secure communication in M2M home network services," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 1, pp. 153–161, Jan. 2019.
[135] V. Sureshkumar, S. Anandhi, R. Amin, N. Selvarajan, and R. Madhumathi, "Design of robust mutual authentication and key establishment security protocol for cloud-enabled smart grid communication," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3565–3572, Sep. 2021.
[136] A. Irshad, S. A. Chaudhry, M. Alazab, A. Kanwal, M. S. Zia, and Y. B. Zikria, "A secure demand response management authentication scheme for smart grid," *Sustain. Energy Technol. Assessments*, vol. 48, Dec. 2021, Art. no. 101571.
[137] S. Yu, K. Park, J. Lee, Y. Park, Y. Park, S. Lee, and B. Chung, "Privacy-preserving lightweight authentication protocol for demand response management in smart grid environment," *Appl. Sci.*, vol. 10, no. 5, p. 1758, Mar. 2020.
[138] N. Kumar, G. S. Aujla, A. K. Das, and M. Conti, "ECCAuth: A secure authentication protocol for demand response management in a smart grid system," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6572–6582, Dec. 2019.
[139] P. Singh, M. Masud, M. S. Hossain, and A. Kaur, "Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107209.
[140] W. Wang, H. Huang, L. Zhang, Z. Han, C. Qiu, and C. Su, "BlockSLAP: Blockchain-based secure and lightweight authentication protocol for smart grid," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 1332–1338.
[141] M. Stănculescu, S. Deleanu, P. C. Andrei, and H. Andrei, "A case study of an industrial power plant under cyberattack: Simulation and analysis," *Energies*, vol. 14, no. 9, p. 2568, Apr. 2021.
[142] H. S. Cho and T. H. Woo, "Cyber security in nuclear industry—Analytic study from the terror incident in nuclear power plants (NPPs)," *Ann. Nucl. Energy*, vol. 99, pp. 47–53, Jan. 2017.
[143] M. Bahrami, M. Fotuhi-Firuzabad, and H. Farzin, "Reliability evaluation of power grids considering integrity attacks against substation protective IEDs," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1035–1044, Feb. 2020.
[144] C. Xenofontos, I. Zografopoulos, C. Konstantinou, A. Jolfaei, M. K. Khan, and K.-K. R. Choo, "Consumer, commercial, and industrial IoT (in) security: Attack taxonomy and case studies," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 199–221, Jan. 2021.
[145] D. Tripathi, A. K. Tripathi, L. K. Singh, and A. Chaturvedi, "Towards analyzing the impact of intrusion prevention and response on cyber-physical system availability: A case study of NPP," *Ann. Nucl. Energy*, vol. 168, Apr. 2022, Art. no. 108863.
[146] S. Hossain-McKenzie, A. Chavez, N. Jacobs, C. B. Jones, A. Summers, and B. Wright, "Proactive intrusion detection and mitigation system: Case study on packet replay attacks in distributed energy resource systems," in *Proc. IEEE Power Energy Conf. Illinois (PECI)*, Apr. 2021, pp. 1–6.

NAVEEN TATIPATRI received the B.Tech. degree in electrical and electronics engineering and the M.Tech. degree in power systems from Jawaharlal Nehru Technological University (JNTU), Anantapur, Andhra Pradesh, India, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering, VIT, Vellore. His research interests include transactive energy management systems and cyber security for communication channel attacks in power systems.

S. L. ARUN received the B.E. degree in electrical and electronics engineering from the Institute of Road and Transport Technology, Erode, India, in 2010, the M.Tech. degree in power systems from NIT Calicut, Kerala, India, in 2013, and the Ph.D. degree in electrical engineering from NIT Tiruchirappalli, India, in 2019. He is currently an Assistant Professor with the School of Electrical Engineering, VIT, Vellore, India. He has published many research papers in reputed international journals and international and national conferences. His research and teaching interests include smart grid technology, demand response, P2P energy transactions, cyber security for smart grid, power system analysis, operation and control, distributed generation, and micro-grid.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 August 2022, accepted 15 November 2022, date of publication 15 December 2022, date of current version 19 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3229766

SURVEY

Cyber Security Maturity Assessment Framework for Technology Startups: A Systematic Literature Review

MOHAMED NOORDIN YUSUFF MARICAN[1], SHUKOR ABD RAZAK[2], (Senior Member, IEEE), ALI SELAMAT[1,3,4,5], (Member, IEEE), AND SITI HAJAR OTHMAN[1], (Member, IEEE)

[1]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[2]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu 21300, Malaysia
[3]Malaysia-Japan Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia
[4]MaGICX–Media and Game Innovation Centre of Excellence, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[5]Faculty of Informatics and Management, University of Hradec Kralove, Hradec Kralove 50003, Czech Republic

Corresponding author: Ali Selamat (aselamat@utm.my)

The author would like to acknowledge the financial support from the Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2022/ICT08/UTM/01/1).

ABSTRACT Cybersecurity has gained increasing importance among firms of different sizes and industries due to the significant rise of cyber-attacks over time. Technology startups are particularly vulnerable to cyber-attacks due to the lack of cyber security measures. This is because of limited human capital and financial resources to quantify cyber risks and allocate appropriate investments to cyber security. Technology startups are suppliers and vendors to large organisations such as MNCs, government and financial institutions. They could possibly have a network connection back to the large organisations and might even store confidential information of these large organisations such as financial records, personal data and other proprietary information. As such, with the lack of appropriate cyber security measures, technology startups may be an attack vector for malicious hackers to gain entry to the large organisations. Focusing on technology startups, this study conducted a systematic literature review on cyber security maturity assessment frameworks. This study addressed five research questions on the existing cyber security maturity assessment frameworks in various industries, the target for implementation, cyber security maturity level, shared control domains of these frameworks, and the quantification of the return of cyber security investments. Referring to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist, a detailed analysis was performed on 24 published research articles (out of 650) from reputable journals and conference proceedings from January 2011 to June 2022. The results revealed the lack of an end-to-end cyber security maturity assessment framework for technology startups. Despite the similarities in the cyber security maturity level for certain frameworks, the results revealed no singular framework that can evaluate the cyber security maturity level of technology startups. The results further revealed the lack of studies on the quantification of the return of cyber security investments in an end-to-end cyber security maturity assessment framework for technology startups. This put the startup in a vulnerable position since management is not able to obtain relevant data on the startup's cyber maturity posture and without such information, they are not able to appropriately justify their security investments to mitigate the evolving cyber risks.

INDEX TERMS Cyber security risk, cyber security maturity, cyber security framework, cyber risk quantification, return of security investment, technology startup.

I. INTRODUCTION

Following the growing connectivity in this digital era, the occurrence of cyber-attacks has continued to increase tremendously. There are different cyber-attacks, such as ransomware attacks, distributed denial of service, phishing, and exploiting vulnerable web and mobile applications. Taking the case of the Southeast Asian region, Singapore encountered a significant increase in cyber-attacks on a weekly basis, with an annual increase of 145% in 2021 [1]. The number of

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Merlino.

cyber-attacks has increased inevitably; it is only a matter of time before these attacks occur since anyone with the knowledge of hacking can execute malicious intentions. Being a victim of a cyber-attack is financially taxing which may cost businesses thousands of dollars in recent times [2]. Therefore, it is crucial for the cyber security functions in organisations to have the capability in addressing the potential cyber security threats on a timely basis.

Cyber risks critically affect businesses following widespread cyber-attack cases [3]. Organisations of different sizes, small and medium enterprises (SMEs) or multinational companies (MNCs) are susceptible to these attacks. The size of an SME is no different from a startup [4]. The substantial effects of cyber-attacks in terms of revenues and clients' trust have positioned cyber risks as the top agenda during board meetings. An SME in Singapore reports annual revenue of $100 million or has less than 200 employees [5]. The effects of cyber-attacks are more critical for startups. Startups have limited financial resources to invest in cybersecurity, which makes them more vulnerable to cyber-attacks [6]. Poor security measures put startups at higher risk against these attacks, which have made it particularly challenging for startup founders to gain clients' trust, especially with the rising cases of cyber-attacks [7].

Cyber security issues are no longer an information technology (IT) problem. It has now become a business risk which should be handled with due care at the highest level in the organisation. Most malicious perpetrators have shifted their focus to smaller organisations since they are easier targets than larger organisations [6]. The smaller organisations do not have adequate financial resources to strengthen their information security capabilities in order to protect the business [8]. Smaller organisations like technology startups need to allocate the appropriate investments to implement the required security measures to combat against these cyber threats. The significance of cyber security has propelled the need to establish a specific framework that can help businesses to recognise, prevent, respond, and recover from cyber-attacks [9]. Implementing a cyber security maturity assessment (CSMA) framework equips businesses to deal with cyber threats. Startups demonstrate low cyber security maturity levels due to their lack of cyber security measures, making them susceptible to cyber-attacks [10]. Thus, it is imperative to determine the cyber security maturity level in order to comprehend the current and target maturity level so that startups are able to implement the appropriate cyber security measures to deal with cyber-attacks based on the identified gaps uncovered during the cyber security maturity assessment.

Focusing on technology startups, the current study aims to review the existing CSMA frameworks that are commonly used by cyber security practitioners in the industry. This study specifically examined the comprehensiveness of these frameworks from an end-to-end perspective to assess cyber risks, determining cyber security maturity levels and quantifying the returns of cyber investments for technology startups. Moreover, this study compared the existing and commonly-used cyber security frameworks, underlined their common features and extrapolated the key control domains which can be streamlined to conduct a cyber security maturity assessment for technology startups in a more effective manner. The objective is to provide management with the information to make a more informed decision so that the right amount of investment can be allocated to implement cyber security solutions for the technology startup in order to mitigate the cyber security risks. With the appropriate security measures in place, this will give added protection for the startup to mitigate against cyber-attacks by malicious threat actors.

II. BACKGROUND AND RELATED WORKS

Cybersecurity is one of the most effective methods to counter business risk [3] and is a key determinant in the decision-making process at the organisational level [11]. As of January 2022, there were more than 3,800 startups in Singapore [12]. The Ponemon Institute conducted a survey and revealed that the majority of SMEs experienced cyber-attacks (66%) and data breaches (63%) in the past 12 months [13]. These attacks affected the financial standing, operations, and reputation of organisations. The increasing connectivity and the upsurge of digital transformation initiatives have created a thriving environment for malicious perpetrators, increasing the rate of cyber-attacks. This has called for the need to establish CSMA frameworks and standards in the industry [11].

Various CSMA frameworks are available for cyber security practitioners in the industry to evaluate the cyber security maturity of organisations. Through these existing frameworks, organisations' current cyber security maturity level can be determined to establish a roadmap towards attaining the desired maturity level. Despite the importance of a cyber security framework against cyber-attacks for organisations [8], startups experience difficulties developing an appropriate framework for building up their cyber security maturity [11]. Without a clear framework, technology startups cannot invest properly in the suitable security measures. Poorly executed security measures result in poor cyber security, which reflects a low cyber security maturity level. Organisations can defend themselves from cyber-attacks that cause data breaches and financial losses by investing in the latest security measures [14].

A. CYBER SECURITY FRAMEWORKS

There are existing cyber security frameworks used by industry practitioners to assess cyber risks and determine the cyber security maturity posture of their organisations. Some of the commonly-used cyber security frameworks include the National Institute of Standards and Technology (NIST), International Organisation for Standardization (ISO) 27001, Control Objectives for Information and Related Technologies (COBIT 5), Cyber Security Capability Maturity Model (C2M2), Capability Maturity Model Integration (CMMI). However, these frameworks lack the end-to-end structure on

assessing cyber risks, determining the cyber security maturity levels and quantifying the returns of security investments based on the mitigation measures. Technology startups do not have the budget to invest in cyber security [7]. As such, the ability to obtain an end-to-end viewpoint on the cyber maturity posture will allow management to make proper decisions on the investment that they make to implement cyber security measures. In order to do this, the ability to assess cyber risks, determining the cyber security maturity level and quantifying the returns of security investments are necessary to be included in the end-to-end framework.

The existing cyber security frameworks are also generally used in traditional setups. The control objectives in the frameworks are broad and aplenty which take a significant amount of time (e.g., 3 to 6 months) to complete. Technology startups are known to be lean and agile, and build products with speed through innovation [41]. Thus, they do not have the luxury of time to complete a cyber security assessment which takes 3 to 6 months. As such, the control objectives in the cyber security frameworks need to be more streamlined and focused for technology startup. With a leaner framework for technology startup, this will assist in shortening the time frame to complete the cyber security assessment.

B. CYBER SECURITY MATURITY LEVELS

Cyber Security Maturity Levels help technology startups to determine their current and target maturity level [28]. It provides a good understanding for the startup to determine their existing cyber security posture and the gaps which need to be remediated in order to achieve their target maturity level. Knowing the cyber security maturity levels help cyber security practitioners to better manage the security of their organisations. According to [30], 12 cyber security maturity models have been identified between 3 to 5 maturity levels. From a maturity scale of 1 to 5, a startup with level 1 in the maturity scale has the lowest cyber security posture with very weak cyber defences which make the company susceptible to cyber-attacks. On the other hand, a startup with a 4 in the maturity scale have an above average cyber security posture with strong defences against malicious perpetrators.

The cyber security maturity level is determined by the number of effective cyber security and data protection controls implemented in the organisation. The number of effective cyber security and data protection controls is in turn determined by the amount of cyber security investments that have been allocated to mitigate cyber risks identified in the organisation. Knowing the cyber security maturity levels is important especially for technology startups as it helps management to appropriately cater cyber security investments so that they can right-size their cyber security measures depending on the current and target maturity level of the startup.

Cyber security frameworks have been extensively explored and discussed in the literature. However, end-to-end cyber security maturity assessment frameworks for SMEs, especially technology startups, have not been systematically reviewed, which is addressed in the current study. Focusing on technology startups, this study presented a comprehensive overview of the cyber security maturity assessment framework and a quantification approach to determine the return of cyber security investments. The end-to end framework will help technology startups to effectively review risks, appropriately identify the cyber security maturity level and provide management with sufficient data to make decisions in justifying investments related to cyber security. With such a framework, this will help technology startups to secure their enterprise against cyber-attacks and reduce the risk of becoming an attack vector to their clients which can be organisations such as MNCs, governments and financial institutions.

III. SYSTEMATIC LITERATURE REVIEW

The systematic literature review (SLR) can determine future research in a particular field. SLR helps researchers to obtain a firm grasp of the field of study and recognize the current research trends and gaps [16]. SLR must be comprehensively methodologically performed with rigor to eliminate bias. It should be beyond a collection of research articles; these research articles should be analytically and objectively reviewed and summarised [17]. With that, SLR was performed in this study to identify, evaluate, and summarise the findings of prior studies within a particular field of study [15].

For this study's SLR, several research questions were clearly established:

1) What cyber security maturity assessment (CSMA) frameworks are available for use in various industries?
2) Are these existing CSMA frameworks targeted for implementation in technology startups?
3) Do these existing CSMA frameworks determine the cyber security maturity level?
4) What are the shared control domains among these existing CSMA frameworks?
5) Do the existing CSMA frameworks incorporate the quantification of the return of cyber security investments?

This study gathered research articles from the following digital databases: IEEE explore (ieeexplore.ieee.org); Scopus (www.scopus.com); Springer (www.springer.com); Web of Science (http://apps.webofknowledge.com). This study targeted research articles from January 2011 to June 2022 using the following keywords: ''Cyber Security Maturity Assessment Model''; ''Cyber Security Maturity Assessment Framework''; ''Cyber Security Maturity Assessment''; ''Cyber Security Maturity Assessment'' AND ''Technology Startup''; ''Cyber Security Maturity Assessment Framework'' AND ''Technology Startup''; Cyber Security Maturity Assessment Model'' AND ''Technology Startup''; ''Cyber Security Maturity Assessment'' AND ''SME''; ''Cyber Security Maturity Assessment Framework'' AND ''SME''; ''Cyber Security Maturity Assessment Model'' AND ''SME''; ''Cyber Security Maturity Assessment'' AND ''Startup''; ''Cyber Security Maturity

TABLE 1. Inclusion and exclusion criteria.

S/N	Inclusion Criteria	Exclusion Criteria
1	Cyber security maturity assessment framework/model in technology startups	Research articles on cyber security maturity assessment framework/model with no reference to assessing cyber risks
2	Cyber security maturity assessment framework/model in startups	Research articles on cyber risk assessment framework/model with no reference to assessing cyber security maturity
3	Cyber maturity assessment framework or model for SMEs	Research articles with no reference to cyber security maturity assessment or cyber risk assessment
4	Only research articles written in English	Research articles written in languages other than English
5	Cyber security maturity assessment in various industry sectors and different organisational types	Unpublished articles, theses, references, or textbooks
6	Related research articles published between January 2011 to June 2022	Related research articles published before January 2011

Assessment Framework'' AND ''Startup''; ''Cyber Security Maturity Assessment Model'' AND ''Startup''. As SMEs and startups share similar size [5], the search included ''SME'' to cover all related small businesses.

Table 1 summarises the inclusion and exclusion criteria of this study to ensure a targeted search with respect to the research questions. All selected research articles were saved in Mendeley (www.mendeley.com), which is a reference management software that manages scholarly publications.

The PRISMA methodology, which incorporates an evidence-based minimum set of items, was employed in this study for efficient reporting of systematic reviews and meta-analyses. Figure 1 presents this study's PRISMA flowchart [18].

Based on the keywords used in the systematic literature review, this study identified 1,772 (including duplicates) research articles published in IEEE explore, Scopus, Springer, and Web of Science using the described search strings in the identification stage. There were 51 articles extracted from IEEE Explore, 175 from Scopus, 74 from Web of Science and 1472 from Springer as shown in Figure 1.

The initial screening retained a total of 620 research articles after all duplicates were removed. The screening process further excluded a total of 550 research articles according to the inclusion and exclusion criteria that have been identified as per Table 1.

The abstracts of the remaining 70 research articles were then reviewed which excluded 27 research articles. Although the excluded research articles consisted of relevant keywords in the title, abstract, and content, these research articles

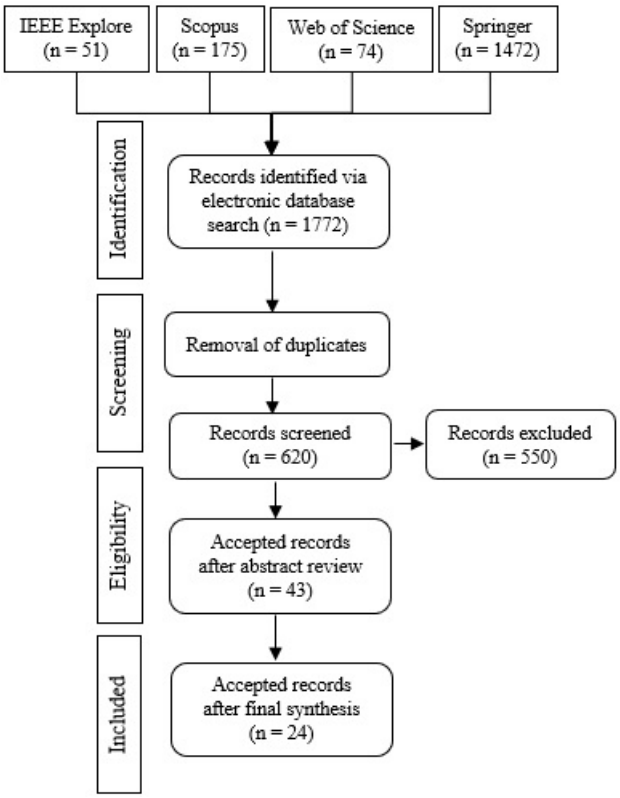


FIGURE 1. PRISMA flowchart.

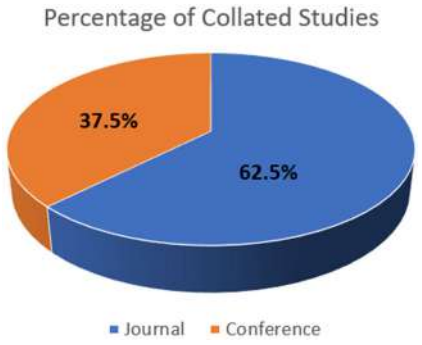


FIGURE 2. Percentage of collated studies.

focused on cyber risk with no context of cyber security maturity and vice versa. These research articles also did not address this study's research questions. After the final synthesis, 24 research articles have been accepted for an in-depth analysis.

A. OVERVIEW OF SELECTED STUDIES

24 articles were selected for this research. Among them, 9 papers appeared in conference proceedings while 13 papers were published in journals. The numbers in percentages are represented in Figure 2.

Figure 3 shows the number of papers by year of publication based on the 24 papers that have been selected in this systematic literature review. The graph indicates that there

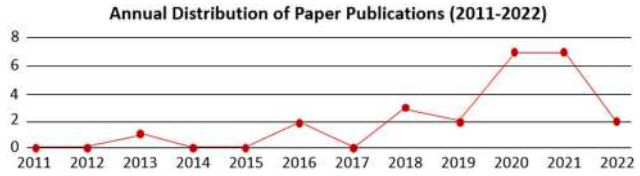


FIGURE 3. Number of papers by year of publication.

is an increase of publication from 2019 onwards. The low distribution of papers in 2022 was as of 31 Aug.

IV. THREATS TO VALIDITY

The potential biasness and the data extraction in an imprecise manner could constrain our findings and pose a major threat to how this SLR is conducted. The four common threats to validity have been taken into account: constructing validity, internal validity, external validity and conclusion validity [40]. Initially, the search terms used may not be able to extract all relevant papers in the identified databases, but manual scrutiny was conducted in the reference section of each paper to further drill down and extract the papers that fall under the research area's realm. An independent evaluation of each of the 43 papers was conducted to ensure relevance to the research area and questions. The selection of the 24 journal papers was conducted as per the PRISMA guidelines [18] to reduce the risk of missing relevant papers and ensure the selected papers can address the research questions and consider the inclusion and exclusion criteria. Several combinations of the search terms were used to avoid the accidental exclusion of relevant papers. Following the PRISMA guidelines provide reasonable assurance, without bias and using the objective criteria, the selected and reviewed papers are among the most relevant studies related to the research area and relevant to the research questions that have been determined.

V. FINDINGS

Data extraction is conducted based on the analysis of the keywords in the 24 selected papers and depicted in Figure 2 below using the VOSviewer software. The VOSviewer helped to identify the keywords which appeared most often in the articles and the links between the authors of the articles. The bigger bubbles showed the keywords which appeared most often.

This analysis is required to gather the results of the research in order to address the research questions (RQs) which have been determined for this systematic literature review. Data extraction was performed on the selected research articles (n=24), and the results are discussed with respect to this study's research questions (RQs).

RQ1: What are the cyber security maturity assessment (CSMA) frameworks available for use in various industries?

Table 2 presents the identified CSMA frameworks from all 24 research articles, which were identified as available for use in various industries across different countries. A few research articles highlighted the same CSMA framework. For instance, seven research articles [25], [28], [30], [32], [33], [35], [37] focused on the Cyber Security Capability Maturity Model (C2M2), whereas four research articles [27], [9], [33], [37] utilized the Control Objectives for Information and Related Technologies (COBIT) Framework. Several research articles also repeated and described the same framework in their literature review.

RQ2: Are these CSMA frameworks targeted for implementation in technology startups?

The analysis further revealed only one CSMA framework [28] was targeted for implementation in technology startups, which proved the lack of a CSMA framework for technology startups. In the research article entitled "Adoption of COBIT 5 Framework in Risk Management for Startup Company", a risk management model was described concerning the processes of the COBIT 5 Framework. Considering that SMEs and startups are similar in terms of size [5], seven other research articles that focused on CSMA frameworks for SMEs were also identified:

1) Cybersecurity Risk Management in Small and Medium-Sized Enterprises: A Systematic Review of Recent Evidence [9]
2) The framework of Effective Risk Management in Small and Medium Enterprises (SMEs): A Literature Review [19]
3) A Dynamic Simulation Approach to Support the Evaluation of Cyber Risks and Security Investments in SMEs [22]
4) A Novel Cybersecurity Framework for Countermeasure of SMEs in Saudi Arabia [26]
5) Calculated Risk? A Cybersecurity Evaluation Tool for SMEs [29]
6) Cyber Resilience Self-Assessment Tool (CR-SAT) for SMEs [31]
7) Reference Framework "HOGO" for Cybersecurity in SMEs based on ISO27002 and 27032 [38]

Overall, this study identified 37 CSMA frameworks from 24 research articles. Adding to that, only seven frameworks were reported to be specifically targeted for SMEs, whereas only one framework for startups was identified. These results reaffirmed the need to emphasize the CSMA framework for technology startups.

RQ3: Do the existing CSMA frameworks assess the cyber security maturity level?

Table 3 presents CSMA frameworks that determine the cyber security maturity level. Assessing risk without determining the cyber security maturity level limits the ability of organisations to assess their current cyber security posture and to determine the intended or target cyber security posture. Having insights on the cyber security maturity level enables organisations to allocate the appropriate investments to enhance their cyber security maturity or posture [14].

Referring to Table 3, these frameworks were highlighted in 15 research articles. The remaining eight research articles

TABLE 2. Relevant papers describing CSMA framework.

No.	CSMA Framework	Ref
1	Committee of Sponsoring Organisations of the Treadway Commission (COSO)	[19]
2	ISO 21827	[20], [33], [35]
3	Electricity Subsector Cybersecurity Capability Maturity Model (ES-C2M2)	[21], [28], [32], [33], [35]
4	SMECRA (SME Cyber Risk Assessment) Methodology	[22]
5	Holistic Cybersecurity Maturity Assessment Framework (HCYMAF)	[23]
6	CyberGov (Cybersecurity Governance) Framework	[24]
7	Cyber Security Capability Maturity Model (C2M2)	[25], [28], [30], [32], [33], [35], [37]
8	Information Security Management Maturity Model (ISM3)	[25], [30], [33], [37]
9	The Publisher's Programme Overview for Information Security Management Assistance (PRISMA)	[25]
10	ISO 27002	[25], [38]
11	Holistic Cybersecurity SME's Coordination Model	[26]
12	COBIT Framework	[27], [9], [33], [37]
13	Oil and Natural Gas Subsector Cybersecurity Capability Maturity Model (ONG-C2M2)	[28], [33], [35]
14	National Initiative for Cybersecurity Education Capability Maturity Model (NICE)	[28], [30], [32], [33], [35]
15	Federal Financial Institute of Examination Council Capability Maturity Model (FFIEC-CMM)	[28], [32]
16	African Union Maturity model for Cybersecurity (AUMMCS)	[28], [32]
17	National Institute of Standards and Technology (NIST)	[9], [29], [30], [33]
18	Health Information Trust Alliance (HITRUST CSF)	[9]
19	A Pedagogic Cybersecurity Framework (PSF)	[9]
20	Centre for Internet Security (CIS)	[9]
21	Cloud Security Alliance (CSA)	[9]
22	SME Cybersecurity Evaluation Tool (CET)	[29]
23	Information Security Evaluation Maturity (ISEM) Model	[29]
24	Systems Security Engineering Capability Maturity Model (SSE-CMM)	[30]
25	ISO 27001	[30], [35], [36]
26	Information Security Maturity Model (ISM2)	[30]
27	Gartner's Information Security Awareness Maturity Model (GISMM)	[30]
28	Information Security Framework (ISF)	[30]
29	Resilience Management Model (RMM)	[30]
30	Community Cyber Security Maturity Model (CCSMM)	[30], [33]
31	Cyber Resilience Self-Assessment Tool	[31]
32	Citigroup's Information Security Evaluation Maturity model (ISEM)	[33]
33	IBM Information Security Framework	[33]
34	Saudi Cybersecurity Maturity Assessment Framework (SCMAF)	[34]
35	ISO 15408	[35]
36	ISO 27032	[38]
37	Cyber Security Governance Maturity Model (CSGMM)	[39]

emphasised risk assessment that did not specifically include the assessment of cyber security maturity level. The detailed analysis of all 23 frameworks also revealed the application of different approaches in assessing cyber security maturity levels. However, this study identified similarities in certain frameworks. For instance, the following cyber security maturity models consist of five cyber security maturity levels but the maturity levels have been defined differently [30]:

1) Information Security Evaluation Maturity Model: 1–Complacency; 2–Acknowledgment; 3–Integration; 4–Common Practice; 5–Continuous Improvement
2) Information Security Management Maturity Model: 1–Undefined; 2–Defined; 3–Managed; 4–Controlled; 5–Optimised
3) Information Security Framework: 1–Initial; 2–Basic; 3–Capable; 4–Efficiency; 5–Optimising
4) Community Cyber Security Maturity Model: 1–Initial; 2–Advanced; 3–Self-Assessed; 4–Integrated; 5–Vanguard

On the other hand, the following cyber security maturity models consist of three to four cyber security maturity levels but define cyber security maturity level differently [30]:

1) Gartner's Information Security Awareness Maturity Model: 1–Blissful Ignorance; 2–Awareness; 3–Corrective; 4–Operational Excellence
2) Resilience Management Model: 1–Incomplete; 2–Performed; 3–Managed; 4–Defined
3) Nice Cyber Security Capability Maturity Model: 1–Limited; 2–Progressing; 3–Optimised

Overall, the results demonstrated the absence of a singular CSMA framework to determine organisations' cyber security maturity level, including technology startups.

RQ4: What are the shared control domains between the existing CSMA frameworks?

Fundamentally, control domains are necessary as key controls for risk assessment. Table 4 presents the extracted shared control domains among the CSMA frameworks reported in seven research articles [20], [25], [28], [31], [32], [34], [38].

Based on the obtained results, common control domains that can be streamlined and evaluated in the risk assessment stage were found evident. These common control domains

TABLE 3. Frameworks which include CSMA.

No.	CSMA Framework	Ref
1	Electricity Subsector Cybersecurity Capability Maturity Model (ES-C2M2)	[21], [28], [32], [33], [35]
2	Holistic Cybersecurity Maturity Assessment Framework (HCYMAF)	[23]
3	CyberGov (Cybersecurity Governance) Framework	[24]
4	Cyber Security Capability Maturity Model (C2M2)	[25], [28], [30], [32], [33], [35], [37]
5	Information Security Management Maturity Model (ISM3)	[25], [30], [33], [37]
6	The Publisher's Programme Overview for Information Security Management Assistance (PRISMA)	[25]
7	Oil and Natural Gas Subsector Cybersecurity Capability Maturity Model (ONG-C2M2)	[28], [33], [35]
8	National Initiative for Cybersecurity Education Capability Maturity Model (NICE)	[28], [30], [31], [33], [35]
9	Federal Financial Institute of Examination Council Capability Maturity Model (FFIEC-CMM)	[28], [32]
10	African Union Maturity Model for Cybersecurity (AUMMCS)	[28], [32]
11	National Institute of Standards and Technology (NIST)	[9], [29], [30], [33]
12	SME Cybersecurity Evaluation Tool (CET)	[29]
13	Information Security Evaluation Maturity Model (ISEM)	[30]
14	Systems Security Engineering Capability Maturity Model (SSE-CMM)	[30]
15	Information Security Maturity Model (ISM2)	[30]
16	Gartner's Information Security Awareness Maturity Model (GISMM)	[30]
17	Information Security Framework (ISF)	[30]
18	Resilience Management Model (RMM)	[30]
19	Community Cyber Security Maturity Model (CCSMM)	[30], [33]
20	Cyber Resilience Self-Assessment Tool	[31]
21	Citigroup's Information Security Evaluation Maturity (ISEM) Model	[33]
22	Saudi Cybersecurity Maturity Assessment Framework (SCMAF)	[34]
23	Cyber Security Governance Maturity Model (CSGMM)	[39]

can be classified as the highest priority, which ultimately exhibit substantial risk impact on organisations. The common key control domains can be generalised as follows:

- People: This domain incorporates the organisation's human capital under the management of the Human Resource function. It consists of workforce management and the capabilities and educational qualifications of employees in key positions.
- Process: This domain covers all organisational processes from document maintenance, change and configuration management, asset management, and cybersecurity to programme management. It helps identify and manage all related security development and management processes.
- Technology: This domain focuses on the application, development, implementation, and maintenance of devices and technologies. This implements a data loss prevention tool that prevents data leakage.
- Compliance: This domain involves monitoring the organisation's compliance with information security policies, regulatory standards, and industry certifications. For instance, the organisation must comply with the ISO27001 certification.

TABLE 4. Shared control domains.

No.	Control Domains	Ref
1	Technology, Vulnerability, Risk, Impact, System, Entity, SubSystem, Capability, Threat and Process	[20]
2	Risk Management, Security Policy and Plan Management, Human Resource Management, Physical Security Management, IT Security Management, Communication Security Management, Security Technology Management, Security Event and Incident Management, Security Audit and Compliance Management	[25]
3	Asset, Change and Configuration Management, Cybersecurity Programme Management, Event and Incident Response, Continuity of Operation, Identify and Access Management, Information Sharing and Communications, Risk Management, Situational Awareness, Supply Chain and External Dependencies Management, Threat and Vulnerability Management and Workforce Management	[28], [32]
4	Risk, Assets, Access, Threat, Situation, Sharing, Response, Dependencies, Workforce and Cyber	[28], [32]
5	Asset Management, Threat and Vulnerability Management, Incident Analysis, Awareness and Training, Information Security, Detection Processes and Continuous Monitoring, Business Continuity Management, Information Sharing and Communication	[31]
6	Governance, Asset Management, Cybersecurity Risk Management, Physical Security, Third Party Security and Logical Security	[34]
7	People, Organisational Document, Process and Technology	[37]

Instead of having comprehensive control domains, this study identified five key domains which can be examined during the risk assessment stage.

RQ5: Is quantifying the return of cyber security investments embedded as part of the CSMA framework?

This study identified one research article entitled "A Dynamic Simulation Approach to Support the Evaluation of Cyber Risks and Security Investments in SMEs" [22] that highlighted its framework's capability to evaluate SMEs' cyber security investments. The study examined the targeted investments based on the risks posed and incorporated various scenarios to evaluate the cyber security investments according to several standard parameters. In one of its simulations, an organisation experiences losses due to cyber-attack, suggesting its need to allocate more investments in cyber security. As a result, the organisation's losses reduced

and eventually stabilised with increased investments in cyber security.

The lack of a quantification model embedded in an end-to-end cyber security maturity assessment framework for technology startups is a critical concern, especially when startups are highly vulnerable against the increasing rise of cyber-attacks. Technology startups cannot quantify and allocate the appropriate investments in cyber security without risk quantification.

A. GAP ANALYSIS

The existing cyber security frameworks are used by cyber security practitioners in various industries but there is a lack of a cyber security framework to assess the maturity level specifically for a technology startup from a cyber security standpoint. Out of the 37 frameworks reviewed, only seven were specifically targeted for SMEs, and only one framework was identified for technology startups. Though SMEs and startups are similar in terms of size [5], the fundamental difference is that technology startups are known for agility and thrives on innovation with information technology.

Based on the frameworks reviewed to determine the cyber security maturity levels, there are different approaches towards assessing the cyber security maturity levels. Though there are similarities in the maturity level, they are defined differently and are not suitable for a technology startup. The cyber security maturity levels for technology startups should be aligned with the stages of the startup lifecycle for clear understanding based on the investments the startup received in each stage. Figure 3 shows an appropriate cyber security maturity level based on each stage of the startup lifecycle.

Different cyber security frameworks have a variety of control domains. However, there is no framework which has control domains to assess the key controls specific for a technology startup. After analysing the control domains from the cyber security frameworks included in this study, five key domains have been extrapolated to be analysed as part of the Risk and Controls Assessment phase.

There is also a lack of a Cyber Quantification phase embedded in the cyber security framework. Since technology startups is a lean organisation, it is important to ensure that the security budget is used prudently. In order to do this, there should a cyber quantification model to calculate the return of security investments based on the mitigation costs for the control deficiencies. The return of security investments would allow management to make a proper decision when allocating the budget to invest in cyber security measures.

First and foremost, the analysis of the cyber security frameworks selected in this study have shown that there is a lack of a specific cyber security framework to examine the key control domains in a technology startup. There is no framework which assess the cyber security maturity level specifically for a technology startup. Finally, there isn't an end-to-end framework which is available to assess cyber risk, determine the cyber security maturity level and calculate the returns of cyber security investments. An end-to-end cyber security framework provides an overview to assess cyber security risk and justify mitigating measures in a more effective manner.

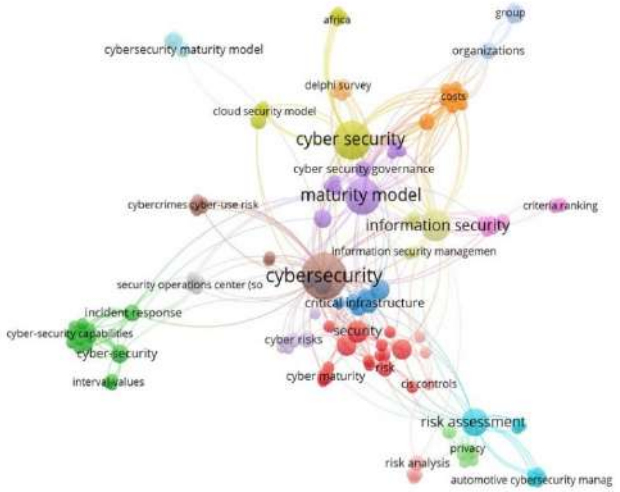


FIGURE 4. VOSviewer network visualization.

B. PROPOSED CSMA FRAMEWORK

There are existing frameworks to assess the cyber security maturity of organisations. However, the frameworks are broad and generic, and thus not specific enough to be applied in technology startups. Since startups tend to be a lean organisation with limited resources, a new framework needs to be developed which is customised and focused in identifying, mitigating and quantifying risks in a technology startup. The new Cyber Security Maturity Assessment (CSMA) framework targets specifically at technology startups and provide a holistic and end-to-end framework. The CSMA framework consists of three phases; Risk and Control Assessment, Cyber Security Maturity Level and Cyber Quantification as shown in Figure 4 below.

After extrapolating the key control domains for technology startups from the existing cyber security frameworks, the People, Process, Product, Platform and Compliance or the 4P1C domains are introduced. The 4P1C domains would allow a more streamlined approach in conducting a cyber security maturity assessment and at a much quicker pace.

In Phase 1, a Risk and Control Assessment is conducted to assess the cyber risks. Each of the 4P1C domains are broken down into several sub-domains, and each of the sub-domains may contain one or more key control objectives which need to be assessed. In the Risk and Control Assessment phase, the risk assessment rating and risk treatment are derived. Phase 2 determines the cyber security maturity level of each of the key controls, sub-domains, the 4P1C domains and the overall cyber security maturity level of the technology startup. Finally, Phase 3 calculates the return of security investment for each of the mitigating measures using an enhanced version of the Return of Security Investment (ROSI) formula [42].

The proposed CSMA framework provide an avenue to effectively assess cyber risks using the 4P1C model,

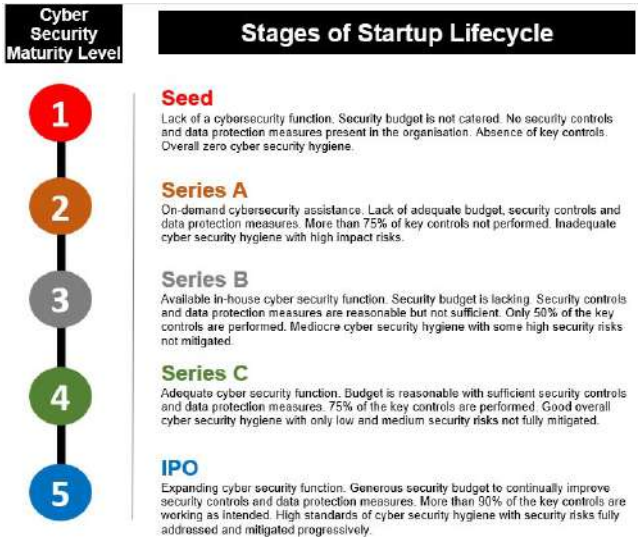


FIGURE 5. Cyber security maturity level for technology startups.

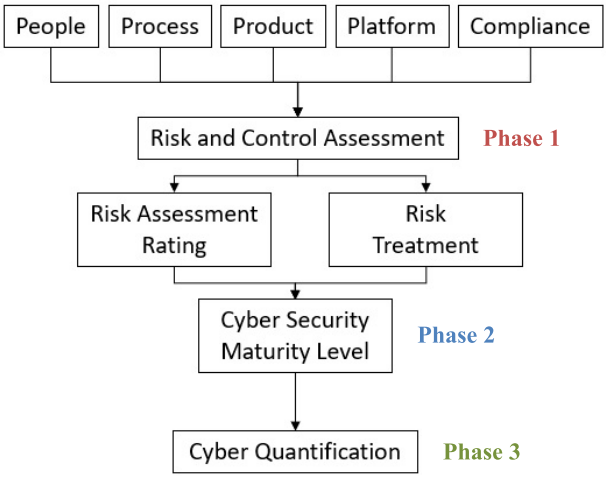


FIGURE 6. Cyber security maturity framework.

determine cyber security maturity level and quantify the returns of security investment in a technology startup. Instead of using a comprehensive framework with significant number of controls which are not applicable for a technology startup, the proposed framework can be used to assess cyber risks in a more objective, focused and streamlined manner. Determining the cyber security maturity level allow the required security controls to be implemented in order to address the identified gaps and finally quantifying the costs of mitigations and the returns of security investments provide management with sufficient data to justify the need to invest in appropriate cyber security solutions.

VI. CONCLUSION AND FUTURE WORK

Technology startups are subjected to cyber-attacks on a frequent basis [2]. The impact of cyber-attacks on smaller organisations like startups is more severe than what larger organisations experience due to their limited financial resources. It may even result in the closure of a startup. Startups with limited financial resources to properly invest in cyber security are more likely to be targeted by malicious perpetrators [6]. Startups must gain their clients' trust and confidence by withstanding against cyber-attacks and building a secure and reliable product for their clients. A cyber security maturity assessment framework can substantially benefit technology startups in evaluating their cyber risks, recognizing their current and future cyber security posture, and quantifying the return of their cyber security investments based on the mitigation costs. Such a framework enables technology startups to allocate appropriate investments in cyber security to implement the required security measures based on the identified cyber risks.

This study performed a systematic literature review on cyber security maturity assessment frameworks for technology startups. Referring to the PRISMA checklist, all five research questions were addressed through the analysis of 24 selected research articles, which revealed several key points. Firstly, there is a lack of CSMA framework specifically for technology startups. This study extracted a total of 37 CSMA frameworks from the 24 research articles. However, only seven frameworks were specifically meant for SMEs, but only one framework was targeted for startups. These results proved the need to implement a streamlined CSMA framework for technology startups. Secondly, despite the shared similarities in the cyber security maturity levels among specific frameworks, the levels were defined differently, which proved the absence of a singular framework that can assess the cyber security maturity level of technology startups. Finally, in the review of 24 selected research articles, only one highlighted the aspect of investments in cyber security for SMEs. No other research articles highlighted the quantification of the return of cyber security investments for technology startups.

From this literature review, it can be highlighted that the existing cyber security frameworks used by industry practitioners are not suitable to be implemented in an agile and lean technology startup. The cyber security maturity model in the existing frameworks is not appropriately defined to suit the different stages in the startup lifecycle. The existing frameworks are also not embedded with a cyber quantification phase which is key to calculate the return of security investments for the startup. Without an end-to-end cyber security maturity assessment framework, management in technology startups is not able to obtain relevant data in order to justify the need to invest in cyber security measures.

As this study only targeted literature from IEEE explore, Scopus, Springer, and Web of Science, other relevant publications may have been excluded from this analysis. Therefore, it is recommended for future research to also explore other repositories. Researchers can also use the proposed model for technology startups in the different industry sectors such as fintech, logistics and e-commerce. Each country has different cyber security and data protection regulations; hence the proposed framework can also be tested on technology startups in the different countries to evaluate the effectiveness

of conducting the assessment. SMEs and MNCs in different industry sectors may also want to adopt this proposed framework instead of using a broad framework with significant number of controls which take a long time and plenty of resources to complete. This framework can thus be utilised as a lightweight approach for the SMEs and MNCs to conduct the assessment.

REFERENCES

[1] Singapore Business Review, Singapore. (2022). *Singapore Cyber Attacks Soar 145% YoY in 2021*. [Online]. Available: https://sbr.com.sg/information-technology/news/singapore-cyber-attacks-soar-145-yoy-in-2021

[2] H. S. Lallie, L. A. Shepherd, J. R. C. Nurse, A. Erola, G. Ephiphaniou, C. Maple, and X. Bellekens, "Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," *Comput. Secur.*, vol. 105, pp. 1–20, Mar. 2021.

[3] B. Cerin, "Cyber security risk is a board-level issue," in *Proc. 43rd Int. Conv. Inf., Commun. Electron. Technol. (MIPRO)*, Sep. 2020, pp. 384–388.

[4] C. Zuszsanna, "Startup: Hype or tendency?" *J. Org. Culture, Commun. Conflict*, vol. 24, no. 3, pp. 1–9, 2020.

[5] *Ministry of Trade and Industry*. Accessed: Jul. 20, 2022. [Online]. Available: https://www.mti.gov.sg

[6] A. L. Mitrofan, E. V. Cruceru, and A. Barbu, "Determining the main causes that lead to cybersecurity risks in SMEs," *Bus. Excellence Manage.*, vol. 10, pp. 38–48, Dec. 2020.

[7] T. Mshvidobadze, "Security issues for digital technology entrepreneurships and startups," *Sci. Practical Cyber Secur. J.*, vol. 4, no. 4, pp. 66–73, 2020.

[8] L. Sanchez, A. S. Olmo, E. F. Medina, and M. Piattini, "Security culture in small and medium-sized enterprise," *Commun. Comput. Inf. Sci.*, vol. 110, pp. 315–324, Oct. 2010.

[9] A. A. Garba and A. M. Bade, "An investigation on recent cyber security frameworks as guidelines for organizations adoption," *Int. J. Innov. Sci. Res. Technol.*, vol. 6, pp. 103–110, Feb. 2021.

[10] A. Alahmari and B. Duncan, "Cybersecurity risk management in small and medium-sized enterprises: A systematic review of recent evidence," in *Proc. Int. Conf. Cyber Situational Awareness, Data Anal. Assessment (CyberSA)*, Jun. 2020, pp. 1–5.

[11] A. Rabii, S. Assoul, K. O. Touhami, and O. Roudies, "Information and cyber security maturity models: A systematic literature review," *Inf. Comput. Secur.*, vol. 28, no. 4, pp. 627–644, Jun. 2020.

[12] Action Community for Entrepreneurship, Singapore. (2022). *Creating a Future-Ready Startup Ecosystem*. [Online]. Available: https://ace.org.sg/wp-content/uploads/2022/01/ACE-Position-Paper-Jan-2022.pdf

[13] Ponemon Institute, Singapore. (2019). *2019 Global State of Cybersecurity in Small and Medium-Sized Businesses*. [Online]. Available: https://www.cisco.com/c/dam/en/us/products/collateral/security/ponemon-report-smb.pdf

[14] T. Ncubukezi, L. Mwansa, and F. Rocaries, "A review of the current cyber hygiene in small and medium-sized businesses," in *Proc. 15th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Dec. 2020, pp. 1–6.

[15] Angraini, R. A. Alias, and Okfalisa, "Information security policy compliance: Systematic literature review," *Proc. Comput. Sci.*, vol. 161, pp. 1216–1224, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919319465 and https://www.researchgate.net/scientific-contributions/Okfalisa-Okfalisa-2212519726

[16] I. Tikito and N. Souissi, "Meta-analysis of systematic literature review methods," *Int. J. Mod. Educ. Comput. Sci.*, vol. 2, pp. 17–25, Feb. 2019.

[17] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Sprouts, Work. Papers Inf. Syst.*, vol. 10, no. 26, pp. 1–51, May 2010.

[18] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Ann. Internal Med.*, vol. 89, no. 9, pp. 873–880, Sep. 2009.

[19] N. Ekwere, "Framework of effective risk management in small and medium enterprises (SMEs): A literature review," *Bina Ekonomi*, vol. 20, no. 1, pp. 23–46, Apr. 2016.

[20] R. Anass, A. Saliha, and R. Ounsa, "A concept & compliance study of security maturity models with ISO 21827," in *Proc. 22nd Int. Conf. Enterprise Inf. Syst.*, 2020, pp. 385–392.

[21] R. M. Adler, "A dynamic capability maturity model for improving cyber security," in *Proc. IEEE Int. Conf. Technol. Homeland Secur. (HST)*, Nov. 2013, pp. 230–235.

[22] S. Armenia, M. Angelini, F. Nonino, G. Palombi, and M. F. Schlitzer, "A dynamic simulation approach to support the evaluation of cyber risks and security investments in SMEs," *Decis. Support Syst.*, vol. 147, Aug. 2021, Art. no. 113580.

[23] A. Aliyu, L. Maglaras, Y. He, I. Yevseyeva, E. Boiten, A. Cook, and H. Janicke, "A holistic cybersecurity maturity assessment framework for higher education institutions in the United Kingdom," *Appl. Sci.*, vol. 10, no. 10, p. 3660, May 2020.

[24] M. Yassine, M. Belaissaoui, and S. Abdelkebir, "A maturity framework for cybersecurity governance in organizations," *EDP Audit, Control, Secur. Newslett.*, vol. 63, no. 6, pp. 1–22, May 2021.

[25] F. Ghaffari and A. Arabsorkhi, "A new adaptive cyber-security capability maturity model," in *Proc. 9th Int. Symp. Telecommun. (IST)*, Dec. 2018, pp. 298–304.

[26] L. Ajmi, Hadeel, N. Alqahtani, A. U. Rahman, and M. Mahmud, "A novel cybersecurity framework for countermeasure of SME's in Saudi Arabia," in *Proc. 2nd Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, May 2019, pp. 1–9.

[27] Y. Kusumaningrum, "Adoption of COBIT 5 framework in risk management for startup company," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 3, pp. 1446–1452, Apr. 2021.

[28] A. A. Garba, M. M. Siraj, and S. H. Othman, "An explanatory review on cybersecurity capability maturity models," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 4, pp. 762–769, 2020.

[29] M. Benz and D. Chatterjee, "Calculated risk? A cybersecurity evaluation tool for SMEs," *Bus. Horizons*, vol. 63, no. 4, pp. 531–540, Jul./Aug. 2020.

[30] N. T. Le and D. B. Hoang, "Can maturity models support cyber security?" in *Proc. IEEE 35th Int. Perform. Comput. Commun.*, Dec. 2016, pp. 1–7.

[31] J. F. Carias, S. Arrizabalaga, L. Labaka, and J. Hernantes, "Cyber resilience self-assessment tool (CR-SAT) for SMEs," *IEEE Access*, vol. 9, pp. 80741–80762, 2021.

[32] A. Garba, A. M. Bade, M. Yahuza, and Y. Nuhu, "Cybersecurity capability maturity models review and application domain," *Int. J. Eng. Technol.*, vol. 9, no. 3, pp. 779–784, Sep. 2020.

[33] R. Kour, R. Karim, and A. Thaduri, "Cybersecurity for railways—A maturity model," *J. Rail Rapid Transit*, vol. 234, no. 10, pp. 1–20, Oct. 2019.

[34] I. Almomani, M. Ahmed, and L. Maglaras, "Cybersecurity maturity assessment framework for higher education institutions in Saudi Arabia," *PeerJ Comput. Sci.*, vol. 7, no. 2, pp. 1–26, Sep. 2021.

[35] H. Imran, M. Salama, C. Turner, and S. Fattah, "Cybersecurity risk management frameworks in the oil and gas sector: A systematic literature review," in *Advances in Information and Communication*, vol. 2. New York, NY, USA: Springer, Mar. 2022, pp. 871–894.

[36] D. Proenca and J. Borbina, "Information security management systems—A maturity model based on ISO/IEC 27001," in *Proc. Int. Conf. Bus. Inf. Syst.*, vol. 320, Jun. 2018, pp. 102–114.

[37] M. Zammani, R. Razali, and D. Singh, "Organisational information security management maturity model," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, pp. 668–678, 2021.

[38] C. F. Cruzado, L. S. Rodriguez-Baca, L. G. Huanca-Lopez, and E. I. Acuna-Salinas, "Reference framework 'HOGO' for cybersecurity in SMEs based on ISO 27002 and 27032," in *Proc. 12th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2022, pp. 35–40.

[39] S. R. Hamidi, A. A. Aziz, S. M. Shuhidan, A. A. Aziz, and M. Mokhsin, "SMEs maturity model assessment of IR4.0 digital transformation," in *Proc. Int. Conf. Kansei Eng. Emotion Res.*, in Advances in Intelligent Systems and Computing, vol. 739, Mar. 2018, pp. 721–732.

[40] X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang, "A map of threats to validity of systematic literature reviews in software engineering," in *Proc. 23rd Asia–Pacific Softw. Eng. Conf. (APSEC)*, 2016, pp. 153–160.

[41] E. S. Rasmussen and S. Tanev, "The emergence of the lean global startup as a new type of firm," *Technol. Innov. Manage. Rev.*, vol. 5, no. 11, pp. 12–19, Nov. 2015.

[42] T. Yaqoob, A. Arshad, H. Abbas, M. F. Amjad, and N. Shafqat, "Framework for calculating return on security investment (ROSI) for security-oriented organizations," *Future Gener. Comput. Syst.*, vol. 95, pp. 754–763, Jun. 2019.

MOHAMED NOORDIN YUSUFF MARICAN received the master's degree (Hons.) in internet security management from the Curtin University of Technology, Australia. He is currently pursuing the Ph.D. degree in computer science with Universiti Teknologi Malaysia. He is also an Adjunct Lecturer in cyber security with universities in Singapore, Australia, and U.K. In August 2022, he was a cyber security professional for more than 20 years working in various sectors of the industry, such as government, banking, oil and gas, consulting, social enterprise, and technology startups. He is also a member of the Information Systems Audit and Control Association (ISACA), International Information System Security Certification Consortium (ISC2), and Association of Certified Fraud Examiners (ACFE). He also holds industry certifications, such as Certified Information Security Manager (CISM), Certified Information Systems Auditor (CISA), Certified Information Systems Security Professional (CISSP), Certified Fraud Examiner (CFE), and PRINCE2 Foundation and Practitioner.

SHUKOR ABD RAZAK (Senior Member, IEEE) is currently a Professor at Universiti Teknologi Malaysia (UTM) and currently seconded as the Deputy Vice Chancellor of Universiti Sultan Zainal Abidin (UNISZA), Terengganu, Malaysia. He also actively conducts several types of research in digital forensic investigation, wireless sensor networks, and cloud computing. He is the author or coauthor for many journals and conference proceedings at national and international levels. His research interests include the security issues for mobile *ad-hoc* networks, mobile IPv6, vehicular *ad-hoc* networks, and network security.

ALI SELAMAT (Member, IEEE) has been the Dean of the Malaysia Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM), Malaysia, since 2018. An academic institution established under the cooperation of the Japanese International Cooperation Agency (JICA) and the Ministry of Education Malaysia (MOE) to provide the Japanese Style of Education in Malaysia. He is currently a Full Professor with UTM, where he is also a Professor with the Software Engineering Department, Faculty of Computing. He has published more than 60 IF research papers. His H-index is 20 and his number of citations in WoS is more than 800. His research interests include software engineering, software process improvement, software agents, web engineering, information retrievals, pattern recognition, genetic algorithms, neural networks, soft computing, computational collective intelligence, strategic management, key performance indicator, and knowledge management. He is on the Editorial Board of the *Journal Knowledge-Based Systems* (Elsevier). He has been serving as the Chair for the IEEE Computer Society Malaysia, since 2018.

SITI HAJAR OTHMAN (Member, IEEE) received the Ph.D. degree from the University of Wollongong, Australia. She is currently a Senior Lecturer with the Department of Computer Science, Universiti Teknologi Malaysia (UTM). Her current research interests include cybersecurity, security management, computer forensic, conceptual modeling, disaster management, disaster recovery, and business continuity planning.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2022, accepted June 3, 2022, date of publication June 13, 2022, date of current version June 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3182383

ICSTASY: An Integrated Cybersecurity Training System for Military Personnel

DONGHWAN LEE[1,2], (Graduate Student Member, IEEE), DONGHWA KIM[1], CHANGWON LEE[1], MYUNG KIL AHN[1], AND WONJUN LEE[2], (Fellow, IEEE)

[1]Cyber/Network Technology Center, Agency for Defense Development, Seoul 05771, Republic of Korea
[2]School of Cybersecurity, Korea University, Seoul 02841, Republic of Korea

Corresponding author: Wonjun Lee (wlee@korea.ac.kr)

This work was supported in part by the Agency of Defense Development, Republic of Korea; and in part by the National Research Foundation (NRF) of Korea Grant by the Korean Government through the Ministry of Science and ICT (MSIT) under Grant 2019R1A2C2088812.

ABSTRACT Cyberwarfare can occur at any moment, anywhere on the planet, and it happens more often than we realize. The new form of warfare is wreaking havoc on not only the military but also on every aspect of our daily lives. Since cybersecurity has only recently established itself as a critical element of the military, the military community relies heavily on the private sector to ensure cyber mission assurance. Given the military's secrecy, such reliance may increase the danger of mission degradation or failure. To address this issue, the military has attempted to build a dedicated cybersecurity training system for the purpose of internalizing cybersecurity training. However, existing cybersecurity training systems frequently lack comprehensive support for effective and efficient cybersecurity training. In this study, we propose ICSTASY, a scenario-based, interactive, and immersive cybersecurity training platform that supports a variety of training features holistically. The primary requirements and design principles required to overcome the challenges inherent in developing a cyber training system were offered based on a review of prior work. Through the demonstration of our prototype, we have proven the feasibility of efficient and truly realistic cyber training, not only for the military environment but also for the private sector.

INDEX TERMS Cybersecurity training, cybersecurity training system, cyber trainer, prototype demonstration.

I. INTRODUCTION

Cybersecurity is indispensable to the attainment of success in military operations nowadays. According to the results of a recent survey of military professionals, over the next five years, cyber attacks will be the greatest concern for the national security enterprise [1]. One of the biggest challenges lying ahead of us is a dearth of the forces capable of repelling enemy attacks. There are highly trained, intellectual criminals behind cyber attacks whereas defense's human resource pool is limited and heavily relies on automated devices for the majority of their defensive activities. Due to these constraints, the defensive operations in cyberwarfare can barely protect critical assets, and other actions such as backtracking and identifying threat actors are practically impossible. Many organizations including the military have attempted to address this issue by introducing a cybersecurity training program that is specifically tailored to train and educate their defense forces.

Cybersecurity training is a critical prerequisite to the military being fully cyberized. The military has attempted to build its own specialized cybersecurity training system, or the cyber range. SIMTEX (Simulator Training Exercise Network) [2] is one of the earliest examples of military-developed cybersecurity training systems. SIMTEX was designed initially for the US Air Force's training purposes and later selected as the operational platform for US military-hosted cybersecurity exercises such as Black Daemon and Cyber Flag. SIMTEX offered virtualized hosts to simulate the information assets targeted by cyber attacks and a VPN tunnel to isolate attack flow across remote sites for the exercises.

CAAJED (Cyber And Air Joint Effects Demonstration) [3] is another USAF effort that combines a commercial wargame simulator called MAP (Modern Air Power) and a cyber simulation model called SECOT (Simulated Enterprise for

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

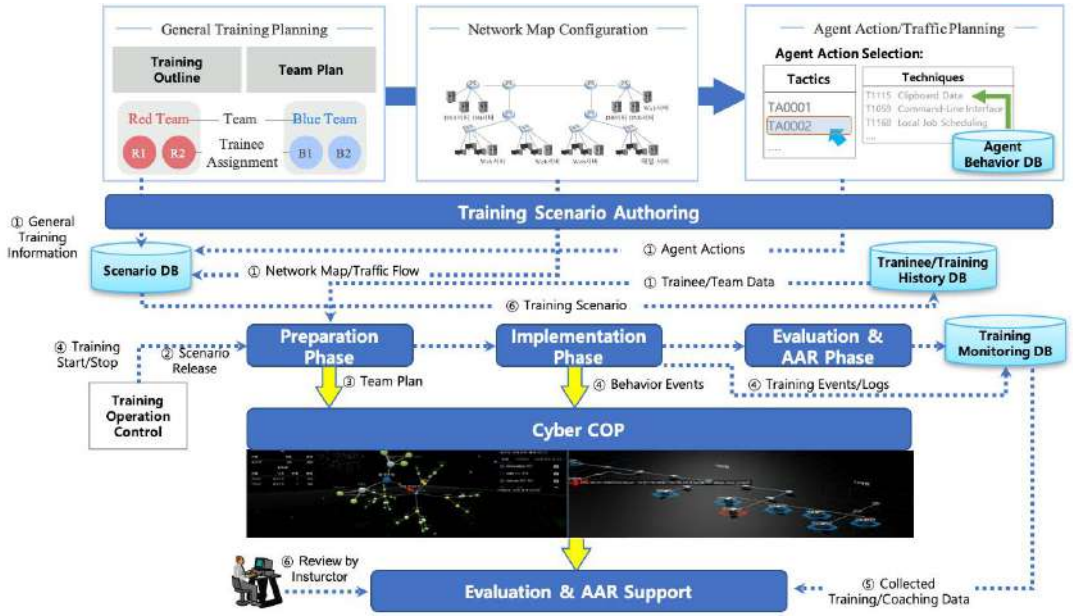


FIGURE 1. Operational concept and procedure of ICSTASY.

Cyber Operations Training). CAAJED was utilized in Cyber Defense Exercise 2007 (CDX 2007), a cyber exercise based on capture-the-flag tactics. During the exercise, red team members conduct simulated cyber attacks, and once the attacks are successful, SECOT calculates the cyber attack's impact on mission performance in the kinetic world.

SAST (Security Assessment Simulation Toolkit) [4] was developed by Pacific Northwest National Laboratory to provide high-level, specialized training to USAF CNO personnel. SAST provides an isolated network that simulates a large network under cyber attack. Additionally, SAST comprises MUTT, a Multi-User Training Tool that generates millions of simulated users to simulate realistic background traffic, and CAT, a Coordinated Attack Tool that incorporates cyber attacks into simulations.

StealthNet [5] is a US Army-funded LVC (Live-Virtual-Constructive) platform for cyber-related testing, evaluation, and training on the Army's tactical networks. StealthNet contains emulation models of cyber attacks such as jamming, DDoS, and worm propagation to determine the impact of cyber threats on tactical networks. StealthNet, in particular, provides simulation models for wireless tactical networks, enabling LVC co-simulation for cybersecurity training in tactical networks.

However, military-developed cybersecurity training programs are constrained in two ways: To begin, they were created for large-scale, short-term exercises such as capture-the-flag drills or cyber wargames. While these events may be beneficial for strengthening capabilities for existing cybersecurity responsibilities, they are detrimental in developing the highly qualified individuals required in the long run. Second, they frequently lack the capabilities necessary for training or are overly focused on tactical applications. These restrictive, overly-specific cybersecurity training systems have hampered the development of a realistic and effective cybersecurity training process.

Meanwhile, in the private sector, many research efforts have been conducted in recent days to build more comprehensive training systems that would address more fundamental, long-term cybersecurity training demands. These modern cybersecurity training systems include graphical user interfaces for configuring the training environment, autonomous red/blue team agents, and automated scoring. However, private-developed cybersecurity trainers have limitations in that such features are not fully integrated, limiting the ability to provide comprehensive, practical cybersecurity training. To overcome these shortcomings, we present a novel cybersecurity training platform, ICSTASY, in this work. ICSTASY provides holistic support for a variety of training capabilities.

To ascertain prerequisites and capability gaps and to develop a blueprint concept for a fully integrated cybersecurity training system, we begin by formulating the desired training system's operational concept. ICSTASY's operational concept and procedure are depicted in Fig. 1. To begin, the initial (preparation) phase defines all of the preliminary information for cybersecurity training, such as a team plan, network map configuration, and agent actions. The following step (implementation) manages a training session, via which trainees interact with the system and associated tasks such as user/agent behavior monitoring and progress/situation visualization. The last (evaluation & AAR) phase of the system facilitates the instructor's assessment and After-Action-Review (AAR) activities by consolidating training logs into trainees' scores and providing reports and replays

of completed sessions. The considerations identified with the operational concept are condensed to the ICSTASY design requirements, which we will use to demonstrate that our prototype is developed in accordance with our initial goals and conceptions throughout the development process. The contribution of our study is as follows:

- The operating concept and procedure were suggested to develop a fully integrated cybersecurity training platform for the military environment that requires more discreet but realistic and comprehensive cybersecurity training.
- We defined requirements and specifications and presented a system architecture that enables the implementation of the operating concept and procedure.
- Finally, we built a prototype of the desired platform, ICSTASY, and demonstrated its capabilities of accommodating the numerous features necessary to deliver effective and realistic cyber training dedicated to (but not limited to) the military.

The rest of this paper is arranged as follows: Section II introduces related studies. Section III provides the design concepts and requirements for ICSTASY, and Section IV elaborates on the overall architecture and system design of ICSTASY by expending on the previously stated design principles and requirements. Section V illustrates the development process through several, detailed screenshots of ICSTASY and compares our cybersecurity training system to others. The concluding section recaps and summarizes this paper.

II. RELATED WORK

As with the military, there is little completed research on integrated cybersecurity training systems in the private sector, including academia; nonetheless, there is some notable work on each technological part of cybersecurity training systems. This section will highlight some of the essential work proposed in the private sector.

CyRIS [6] is a cyber range instantiation system developed by JAIST in which KVM-based virtual hosts are set up and created automatically following a script-based scenario file. Additionally, the scenario script specifies the types of emulated attacks to be executed and the target nodes. Their latest cybersecurity training system, CyTrOne [7], includes these technologies.

Nautilus [8] devised its own script language called SDL (Scenario Description Language) for automating the deployment and configuration of a virtualization-based cybersecurity training environment. As with CyRIS, SDL specifies virtual hosts and network configurations for a training environment, but instead of emulated attacks, it defines vulnerabilities embedded in hosts. A CVE (Common Vulnerability Enumerator) code [9] identifies each vulnerability and, based on a predetermined script, automatically plants it upon the instantiation of the vulnerable host.

ASL (Attack Specification Language) [10] provides an integrated representation of cyber threat scenarios for cybersecurity trainers. Considering the dynamic nature of the cyber threat scenarios, ASL is built with the innate feature to deduce the most advantageous attack technique given the conditions using machine learning based inference. Taking a step forward, GHOSTS [11] introduced the concept of a Non Player Character (NPC) into cyber training systems, which aims to emulate the hostile behaviors of an enemy and the benign activities of regular users.

CybOrg [12] is a cyber gym platform dedicated to the training of autonomous agents. The platform is built on a commercial cloud platform, AWS, and intends to provide a repeating training environment for autonomous agents to practice cyber attack and defense techniques using reinforcement learning. Each repeat uses a YAML-based script to duplicate and diversify the episodes given to the agents. Agents trained in this manner get deployed as red and blue team agents that face off against trainees.

However, the linked work discussed above concentrated on specific technological aspects rather than proposing a comprehensive platform. The Swedish research agency FOI launched CRATE (Cyber Range and Training Environment), a pioneering cybersecurity training platform [13], [14]. In contrast to the other previous effort, CRATE's objective was to create an integrated cybersecurity training platform by combining the fragmented technology elements. For example, CRATE's NodeAgent and Core API services facilitate the configuration and deployment of virtualized hosts and networks. Its CRATE Exercise Control (CEC) platform enables situational monitoring and evaluation of cybersecurity training [15]. Additionally, SVED (Scanning, Vulnerabilities, Exploits, and Detection) identifies vulnerabilities in a training environment and assists automated red-team agents with attack planning. [16]. Although a significant portion of the features rely on commercial off-the-shelf software such as OpenVAS [17], snort [18], or TCPdump and thus provide only partial, limited capabilities, CRATE retains meaning as the initial attempt to integrate the technology elements of a cybersecurity training system.

KYPO [19] is another notable study that takes into account the exhaustive design principles of a cybersecurity training system. KYPO acknowledges the importance of real-time monitoring and evaluation (so-called post-mortem analysis) by suggesting a highly fine-grained log production and collection architecture.

III. DESIGN CONCEPTS AND REQUIREMENTS

A detailed assessment of existing cyber trainers showed that most of their systems could not handle the inclusion of additional elements needed for full-fledged cybersecurity training. Our novel training system addresses these limitations by being developed according to the standard V model, i.e., based on identified capability gaps. We first developed a set of principles and requirements to consider when designing

a novel cybersecurity training system. The system design, implementation, and evaluation follow in due order.

A. EDITABLE AND REUSABLE SCENARIO WITH TEMPLATES

A scenario is a critical component of any cybersecurity training system. A training session is essentially a reproduction of a training scenario, and a robust cyber training system is one with rich scenarios. However, many cyber trainers supply scenarios as a bundled package which usually does not allow instructors to change the scenarios. This precludes the trainer from diversifying training scenarios and providing variance within a single training session. As a result, a novel cyber training system must enable an editable and reusable scenario. To ensure that this design principle is adhered to, we propose the following requirements.

1) A scenario should contain all of the elements necessary to conduct a training session, such as host and network configuration, agent behavior schedule, expected user events, and other relevant information.
2) A scenario should be exportable and re-importable as an editable script or markup language.
3) A scenario should have a layered structure with numerous layers, enabling the training system and scenario editor to access and locate required data.

B. AUTONOMOUS OPPONENT FOR TRAINEE INTERACTION

In a typical cyber training system, interactive experiences are confined to engagement with a human opponent or to unidirectional activities outlined in a script file [20]. This prevents trainees from encountering a variety of situations that can arise in cyberspace. More intelligent *agents* capable of interacting with learners in a training environment is necessary to offer as many situations as possible. Additionally, trainees can benefit from a variety of cybersecurity experiences if an autonomous blue team and an autonomous red team are available, which is the only form of team permitted in the majority of conventional cyber trainers. In summary, we propose the following requirements for this design principle.

1) Autonomous agents capable of varying their behavior in response to changing variables in the training environment should be provided.
2) A user should be able to plan and edit the essential actions of agents throughout the scenario authoring process.
3) Agents from either the red or blue teams should be able to be chosen for an autonomous opponent team.

C. FULL VISIBILITY INTO TRAINING SESSIONS

Because comprehensive situation awareness in cyberspace has been one of the most significant issues in the cybersecurity area, a question-and-answer-based test for trainees was an indirect method used to provide visibility into the training environment. We can promote productive interactions between a trainee and their instructor if we can automatically recognize and notify the instructor about the trainees' behavior. This enables an instructor to adjust their teaching methods while keeping a close eye on the trainee's progress. The following requirements would provide complete visibility into training sessions.

1) The training system should recognize and collect all potential events associated with trainee activity into raw logs, which are the system's most granular logs.
2) The expected training event for a training session should be definable during the scenario creation phase using the logical operations of the raw log events.
3) The training system should notify the instructor immediately upon detecting expected training events, using a visually effective method such as a dashboard and/or a Common Operational Picture (COP) for the cyber training environment.

D. AUTOMATED EVALUATION AND AFTER-ACTION-REVIEW

As indicated previously, a question-and-answer-based test has typically been the primary approach for enabling visibility into the training environment. For instance, if a trainee responds with a string only obtained through successive privilege escalation, it implies the trainee successfully executed the privilege escalation. The evaluation process is identical in the majority of conventional cyber trainers. However, if we can automatically detect and analyze trainees' behaviors, we will be able to evaluate trainee behavior as well. The following requirements are needed to substantiate our cyber training system's automatic evaluation and AAR features.

1) The trainee behavior events that occur during a training session should be recorded in a database to be utilized post-session by the evaluation and AAR features.
2) The training system should provide an interface via which an instructor can assign scores to observed behaviors ahead of time so that the score is automatically attributed when the trainee exhibits that expected behavior.
3) When the system detects important behaviors, the main screens, such as the dashboard/COP screen and the trainee's screens, should be captured as screencasts for the AAR phase debriefing.

IV. OVERALL ARCHITECTURE AND SYSTEM DESIGN

This section presents the proposed training system's overall architecture and system design based on the principles and requirements described in the preceding section. We begin by proposing the system's overall architecture, followed by a description of the system's design in three primary components.

A. OVERALL ARCHITECTURE

To begin, we determine the operational procedures for our training system to create a design for the overall architecture.

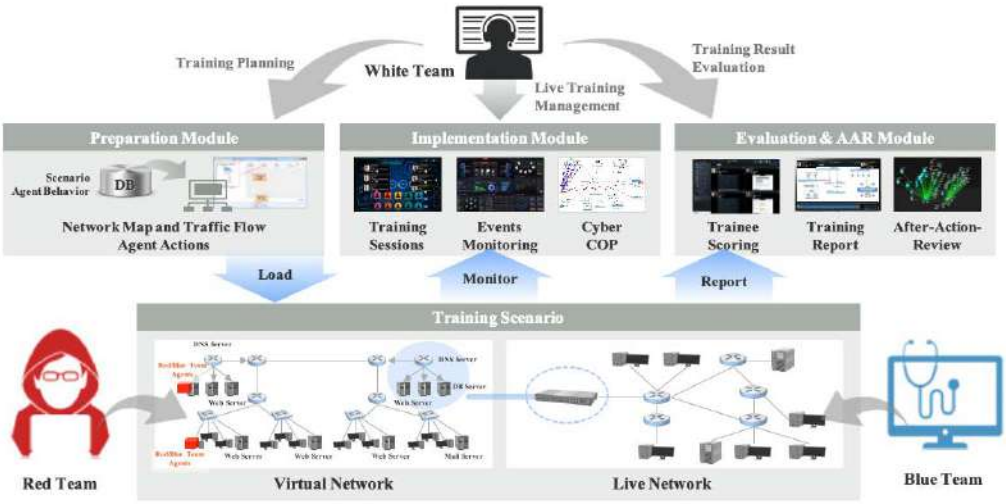


FIGURE 2. System architecture and modules of the ICSTASY prototype.

When considering the cybersecurity training system's use cases, the key user is the instructor. They create the scenario required for a training session, conduct the session, and lead and evaluate trainees. These tasks may be accomplished collaboratively by members of several teams, such as white, yellow, and green. Unless otherwise specified, we refer to a user who can participate in any of these teams as an instructor. As briefly mentioned above, the ICSTASY operational procedure has three phases, mainly from the instructor's perspective:

1) The preparation phase: contains activities such as scenario creation, network map/virtual machine configuration, agent behavior design, and training session management.
2) The implementation phase: includes initiating, managing, and terminating a training session. Automated agents, live monitoring, and coaching activities are performed throughout the session.
3) The evaluation and AAR phase: the final stage of training, during which an instructor can assess trainees' progress and advise them based on the information acquired throughout the assessment.

Given the operational procedure stated above, ICSTASY has three modules that correspond to the three phases: the preparation, implementation, and evaluation & AAR modules. As shown in Fig. 2, each module performs the functionality necessary for each operational phase.

B. MODULE-WISE FEATURES AND SYSTEM DESIGN

This section details the features and specifications of each module focused on meeting the aforementioned significant requirements.[1]

[1]The requirements are referred to by their section and item numbers, for example, III-A-1 refers to the first requirement in Section III.A.

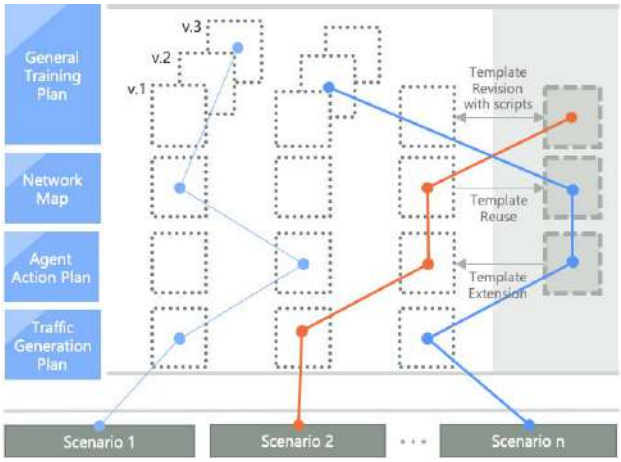


FIGURE 3. Layered structure of the ICSTASY training scenario.

1) PREPARATION MODULE: SCENARIO AUTHORING/MANAGEMENT

First of all, we divided the scenario authoring process into multiple steps to accommodate the multi-layered structure of a training scenario (Requirements III-A-1 and III-A-3): 1) Defining general training concepts and organizing teams 2) Configuring network map/virtual hosts; 3) Scheduling the actions of agents and listing expected events. Each step intends to aid instructor teams in their preliminary work. For instance, in the second step, ICSTASY provides a drag-and-drop UI that enables the instructor to easily and efficiently build virtualized infrastructure, which is distinct green team work. In this manner, based on the objectives of training, the instructor can readily deploy security appliances: from virtualizable IDS/IPS/firewalls like pfSense, snort, Suricata, and Bro to any hardware-type appliances supporting IP networks such as firewalls, IDS/IPSes, and the anti-DDoS and anti-spam devices. The third permits instructors to more easily

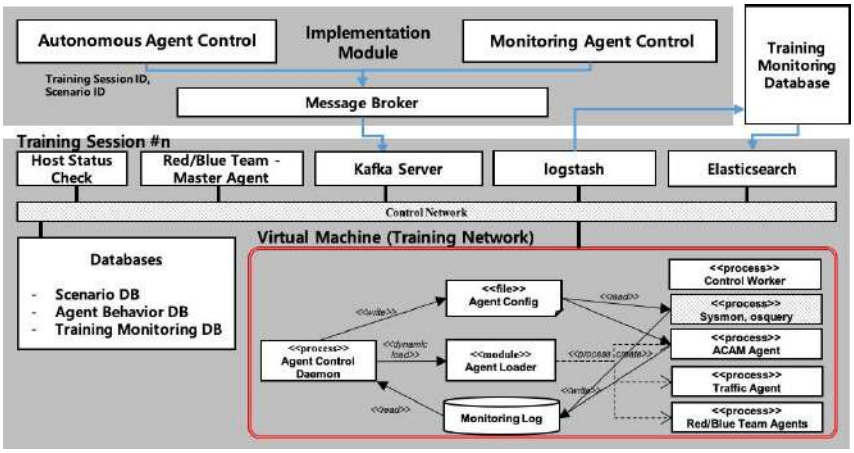


FIGURE 4. Structure and data flow of the ICSTASY log collection.

monitor and evaluate trainees' actions and performance, which may be related to the work of a yellow or white team.

To meet the scenario's requirements (Requirements III-A-2, III-C-2, and III-D-2), we designed a scenario as an editable XML file containing gathered data from each step in each layer. A proficient instructor may quickly locate and edit specific sections of a scenario file, resulting in a more sophisticated scenario than one prepared via the GUI. Fig. 3 illustrates a scenario file reflecting the layered structure of the ICSTASY training scenario. Scenarios are saved as templates in each layer, enabling scenario reuse and editing on a template-by-template basis.

We added a session management step before the implementation phase, in addition to the scenario authoring activity. An instructor must complete this step by creating a session where a selected scenario will be loaded. This enables us to build several sessions from a single scenario and easily manage temporal data such as trainee logs.

2) PREPARATION MODULE: AGENT ACTION PLANNING

The agent action planning feature is one of the most prominent features of the preparation module. The agents' actions are produced and structured automatically by specifying a few parameters, such as the starting and ending points of attack (Requirement III-B-2). Each activity of a red team agent is associated with a Technique Instance (TI), which is defined as an instantiated technique in MITRE's ATT&CK framework [21]. Each TI has pre-and-post conditions that allow us to simulate the attack path before training and pre-determine the agent's availability. Our prior work [22] and [23] have the particular automation strategies upon which our red team and blue team agents were built, respectively. Thus, an instructor can assign agents alternative roles and courses of action according to the training objective, giving a high level of diversity and flexibility for an advanced cyber training experience (Requirements III-B-1 and III-B-3). It eliminates the need for a costly white/green team and allows for the potential of a one-person white team, whereas many existing cyber trainers rely on pre-determined, immutable agent activities. Appendix contains the complete list of TIs included in ICSTASY.

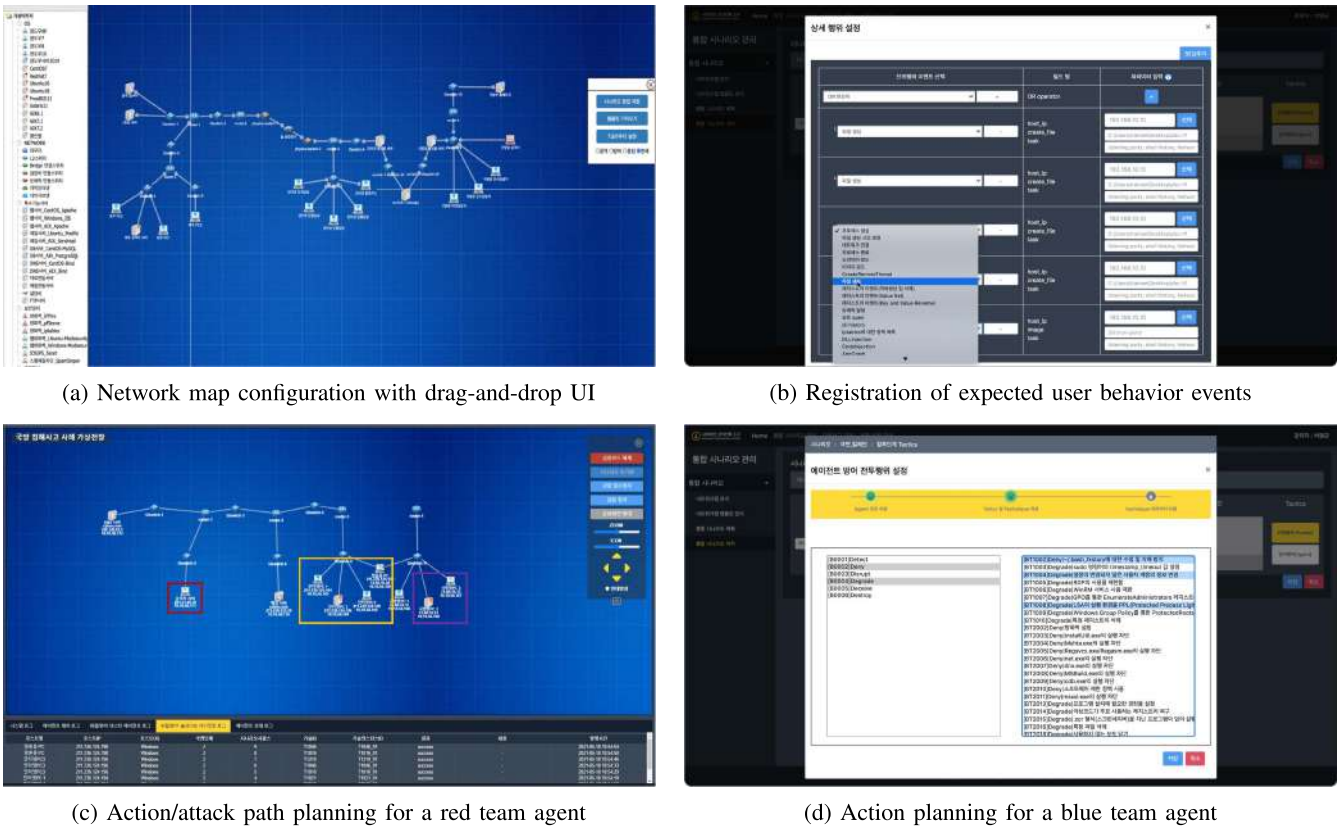
In addition to the autonomous agents, we enabled ICSTASY to imitate background and network traffic using pre-stored pcap files [24]. The preparation phase allows for the configuration of source-and-destination pairs for traffic creation. The source and destination can be hosts in a simulation or a real-world environment and in [25], indicating that our training system is LVC interoperation ready.

3) IMPLEMENTATION MODULE

The implementation module includes a variety of features for managing live training sessions. The implementation module's primary feature is session management, which enables an instructor to control the flow of a training session via an initiation/pause/termination interface. A notable feature of ICSTASY is the ability to pause a session, which is rarely available in other cyber trainers. This capability suspends all system activities, including the log collection for the session and all associated virtual machines. We integrated VMware vShpere API [26] and IBM PowerVC API [27] into ICSTASY connecting the session management feature to the backend that manages all the virtual machines.

Another critical feature included in the implementation module is the ability to visualize training sessions. The visualization function provides a visual representation of the session statuses and enables event-driven monitoring of learner behavior. The implementation module utilizes Logstash [28] to capture all data associated with a training session, accumulating it in the monitoring database (Requirements III-C-1 and III-D-1). Elasticsearch is used to retrieve and analyze the stored logs [29].

The key concept behind the visualization feature is a *behavior event*, which is pre-defined metadata during the preparation process. It provides expected user/agent behaviors during a training session to meet the objectives. In this manner, we may focus our search on a subset of the massive



(a) Network map configuration with drag-and-drop UI

(b) Registration of expected user behavior events

(c) Action/attack path planning for a red team agent

(d) Action planning for a blue team agent

FIGURE 5. Demonstrative screenshots of the preparation phase.

amount of logs. On the other side, we ensured that we collected as many fine-grained logs as possible from hosts and networks, referred to as an emphatomic event. In addition to the usual IDS-based detection method, live forensic/EDR (End-Host Response)-based techniques were incorporated. For instance, Microsoft's Sysinternals Suite [30] and Facebook's OSquery [31], as well as a self-developed mini-filter driver named ACAM (Advanced Cyber Activity Monitoring), collected a large amount of host-related data. These are deployed in each host, enabling highly sensitive detection of kernel level changes such as privilege escalation, driver loading/unloading, process crashes, and DLL/code injections. The atomic logs are stratified, so at least one comprises an *interim event*, and one or more interim events constitute a behavior event at the highest level. For visualization purposes, Fig. 4 illustrates the hierarchical structure of the log collection. The resulting behavioral events appear in the form of a cyber COP (Common Operational Picture) (Requirement III-C-3). ICSTASY can efficiently and precisely detect trainee/agent behaviors using this approach, whereas the majority of existing trainers rely on the instructor's skill.

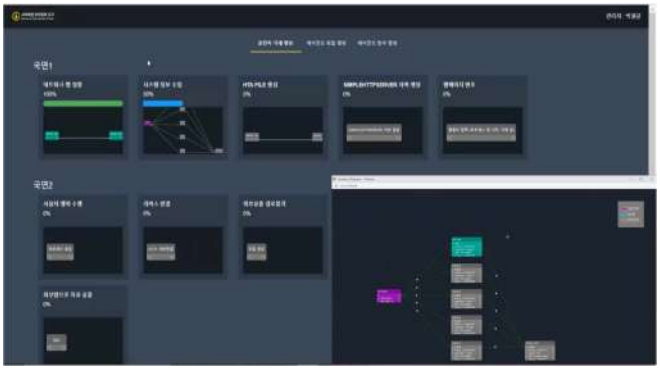
Visualization also requires the log collection of autonomous agents. Once the training session initiates, the implementation module triggers automated agents to begin trace the pre-programmed path planned in the preparation phase. Unlike human behaviors, agents report agent behaviors and hence do not require the module to gather granular logs and detect them. After determining the success or failure of an action, an agent generates a behavior log.

The module's final feature is live coaching, which provides an engaging experience for trainees and increases training efficiency. To avoid possible intrusion into the training world while conducting coaching activities, we isolated the training network from the 10 GbE network. All data not used for training, including atomic logs for visualization, flows up through this network. The coach can monitor each trainee's shared screen guide any trainees using the live coaching feature.

4) EVALUATION AND AAR MODULE

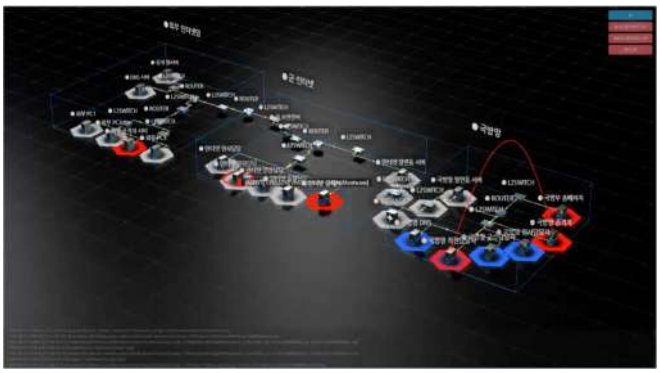
The module for evaluation and AAR relies heavily on the implementation module. From a design standpoint, the evaluation and AAR module can be defined as an implementation module that uses archived data rather than real-time data. After a training session, the instructor can playback recorded COP and trainee screens and examine saved behavior events and other records (Requirements III-D-2 and III-D-3). To facilitate evaluation and AAR, we first built a central time server and timestamped all visualization and coaching data collected during a training session. This simplifies data synchronization and enables instructors to

(a) Progress diagram of the red team agent action execution



(b) Dashboard with the status of expected user behavior events

(c) Cyber COP in the force-directed graph expression



(d) Cyber COP in the network map expression

FIGURE 6. **Demonstrative screenshots of the implementation phase.**

navigate trainees' training records by moving around a single timeline. Second, we assured that the module visualized COP using the latest web standards, including CSS3, and that the visualization data was stored after time stamping. As a result, a more informative and lightweight COP-centric replay feature emerged, especially when compared to video recording techniques that consume considerable system resources. Thus, an instructor can review the training situation collectively and conveniently for the chosen time period without losing any knowledge.

V. DEVELOPMENT RESULTS

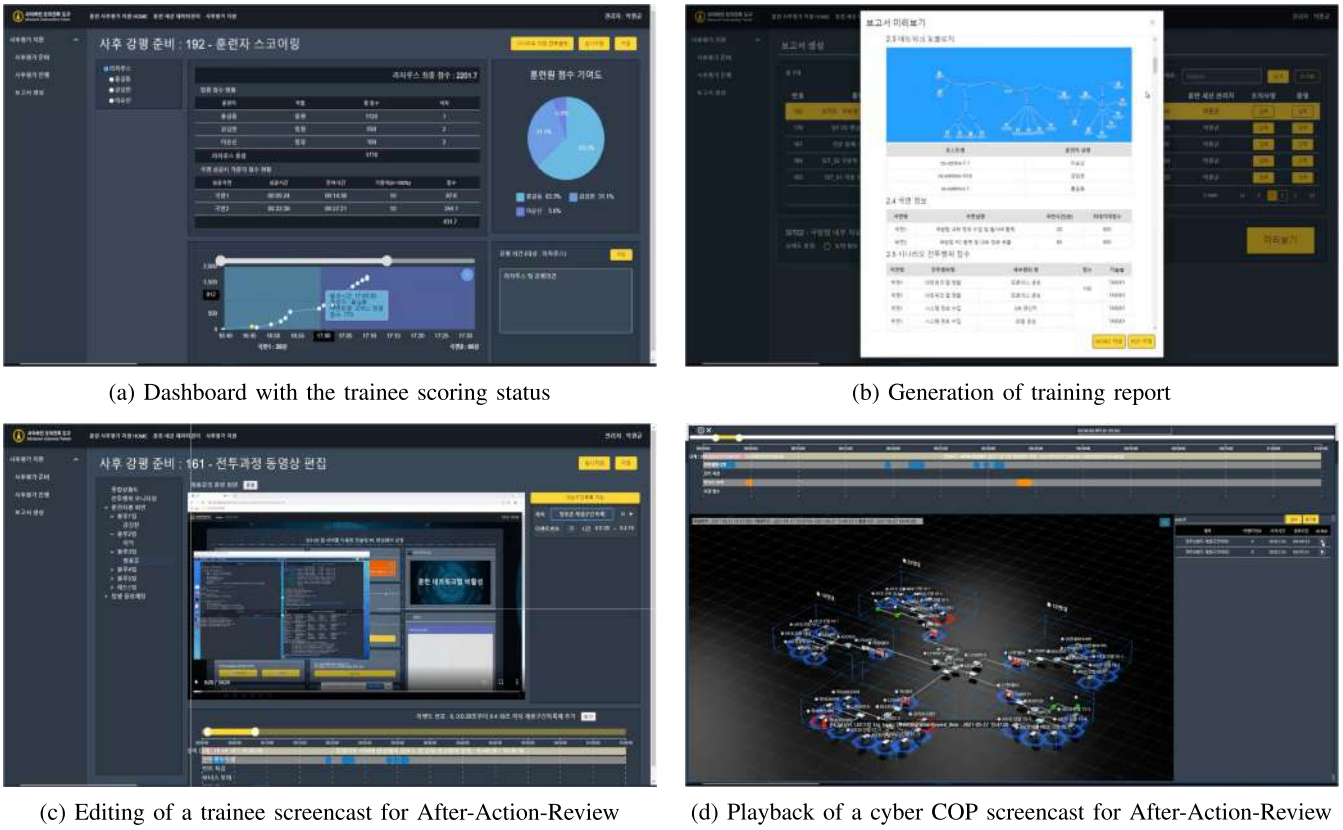
We developed a prototype of ICSTASY based on the module design described previously to illustrate the feasibility and usability of an advanced, immersive cyber training experience. We cannot give detailed training situations due to the possibility of disclosing confidential information; nonetheless, we have attempted to provide as many different screenshots as possible to understand our training system.

A. PREPARATION PHASE

Given that the preparation phase is the most labor-intensive, we placed focus on the instructor interface during the development process. As specified in the module design, drag-and-drop-based UI for configuring the network map/virtual hosts was implemented. Refer to Fig. 5a for a screenshot of the network map configuration tool. An instructor can drag and drop the desired host template from the tool's left side panel to the tool's main panel. The host is then instantiated, allowing the instructor to update the host's different metadata, including the network configuration and user account/credential information. The metadata is initially recorded in the scenario database and is then simultaneously sent to VMware vCenter and IBM PowerVM via the vSphere API and the PowerVC API.

Fig. 5b shows the expected trainee behavior events listed in the preparation phase. As of the prototyping stage, 79 different types of atomic logs are available for an instructor to select an interim log. Given that only the AND operation is permitted for combining atomic logs, combinations can generate $(79-1)(79-2)/2 = 3003$ interim logs. As a two-step logical operation using AND or OR is permitted while composing a behavior log, $(3002 \times 3001 - 1) \times (3002 \times 3001 - 2) \approxeq 8.11 \times 1013$, i.e., a nearly infinite number of behavior logs, can be constructed in our training system. However, providing all of the behavior logs expected for a training session might be challenging for an instructor. Therefore, we enabled the prototype to reuse behavior logs utilized in prior sessions to address this issue.

Fig. 5c and Fig. 5d illustrate the action planning processes of the red and blue team agents, respectively. As for a red team agent, the network map built in the previous process



(a) Dashboard with the trainee scoring status

(b) Generation of training report

(c) Editing of a trainee screencast for After-Action-Review

(d) Playback of a cyber COP screencast for After-Action-Review

FIGURE 7. Demonstrative screenshots of the evaluation and AAR phase.

allows an instructor to automatically configure the attack path and assign offensive TIs utilized in the attack by selecting the attack's start and end points. As indicated previously, the walk-through feature is used to rehearse with the autonomous red team agents. At the bottom of Fig. 5c, we can see the logs generated by the red team agents as they walk through a penetration scenario in which an external host (red square) infiltrates victim hosts (violet square) via intermediary nodes (amber square). In the case of a blue team agent, we supplied a detailed configuration UI that allowed an instructor to fine-tune the blue team's behaviors using a variety of defensive TIs in addition to the fundamental defense measures that may be done automatically with a few inputs. On the right side of the panel in Fig. 5d, we can confirm the many defensive TIs provided under the 6D categories of defense course-of-actions: detection, deny, disrupt, degrade, deceive, and destroy [32].

B. IMPLEMENTATION PHASE

The implementation phase focused on the visualization of COP, which enables instructors to assess the training situation and trainee progress quickly. Because the achievement of expected behavior events is the primary indicator of the flow of the training process and the trainee progress, we developed and organized the forms of COPs expected to be the most effective at representing the state of behavior events.

The graphic in Fig. 6a depicts the progress of agent behavior events, specifically the status of TI executions. TIs are activated per the execution route determined by the pre-and post-conditions specified. With the diagram, an instructor can verify that each TI was successfully run and quickly determine which TI caused the flow to fail to complete as expected. The progress diagram for the blue team agent action execution is constructed similarly to the red team agent's but is displayed in parallel, as the blue team agent's activities are not serialized as the red team agent's actions are.

The dashboard seen in Fig. 6b monitors the status of expected behavior occurrences. Once a trainee's actions identify interim events, the progress bar for the corresponding behavior event reflects the percentage of completed interim events. By clicking on any behavior event, an instructor can view the detailed state of event detection and the logical breakdown of that behavior event.

Fig. 6c and 6d illustrate the two primary Cyber COP displays produced for ICSTASY. The first is a cyber COP with a force-directed graph, which serves as the primary COP for assisting an instructor's situational awareness via a conceptual data model. Specifically, when an agent or trainee node has a new behavior event as a child node, it is added to the center node. The child nodes are added up whenever the agent or trainee experiences a new behavior event. If the conditions between events are dependent, the event nodes have a

TABLE 1. Comparison between cybersecurity training system/platforms.[2]

Phase	Features	Cybersecurity Training System/Platforms					
		CyRIS [6]	Nautilus [8]	CybOrg [12]	CRATE [15]	KYPO [19]	ICSTASY
Preparation	Script/markup language-based training environment configuration (III-A-1, III-A-2)	Yes	Yes	Yes	Yes	No	Yes
	GUI-based training environment configuration (III-A-1, III-A-3)	No	Yes	No	No	No	Yes
	Automated agent action planning (III-B-1, III-B-2)	P/S	No	Yes	Yes	No	Yes
Implementation	Automated training environment provisioning (III-1-1, III-1-3)	Yes	Yes	Yes	Yes	Yes	Yes
	IDS-based basic event monitoring (III-C-1)	No	No	No	Yes	Yes	Yes
	Dedicated agent-based fine-grained event monitoring (III-C-1, III-C-2)	No	No	No	No	N/A	Yes
	Visualization via cyber COP (III-C-3)	No	No	No	P/S	P/S	Yes
	Autonomous red/blue team agents (III-B-3)	P/S	No	Yes	Yes	No	Yes
	Background traffic generation/traffic injection (III-B-1)	No	No	No	P/S	No	Yes
Evaluation & AAR	Automated trainee scoring (III-D-1, III-D-2)	No	No	No	Yes	Yes	Yes
	Screen recording & replay (III-D-3)	No	No	No	No	No	Yes
	Training report generation (III-D-1, III-D-2)	No	No	No	Yes	N/A	Yes

subordinate connection. On the right side of the cyber COP, trainees' achieved behavior events are also listed. By selecting an event from the list, a video with the trainee's screencast at the time of the event will play. ICSTASY accomplishes this by maintaining video recordings of trainees' screencasts with a 30-second window size.

The second is a cyber COP with a conventional network map diagram, which serves as a secondary COP to aid intuitive network plane knowledge. The red and blue hexagonal loops surrounding the nodes in Fig. 6d denote the region of the red and blue teams, respectively. The white team designates the color-coded information prior to the train session and can swap to another color when an instructor confirms a trainee's occupation report. The red parabolic line in the figure represents the network flow associated with a cyber attack, connecting the attack's origin and destination. These visual elements provide instructors and observers of cyber training with an instantaneous perception of a training situation and create a highly immersive, competition-like (i.e., gamified) environment for trainees when combined with the varied coaching experience supported by various media.

C. EVALUATION AND AAR PHASE

In terms of user experience, the assessment and AAR phase is divided into two distinct components: evaluation and AAR. Fig. 7a and 7b illustrate the trainee scoring and training report generation features, respectively, which are mostly used for the instructor's evaluation work. The trainee scoring dashboard summarizes and displays the current training session's point-scoring and learning progress. For instance, an instructor can use the dashboard to determine how trainees earned scores and which trainee contributed the most to their team's point total. The training history saved in the training monitoring database, along with relevant data contained in other databases, is assembled into a single training report, including the scoring data. Additionally, the training report includes additional statistics about the training that are not displayed in the UI, such as the status of file/network/process access and privilege escalation, as well as the CPU/RAM consumption on each host.

Regarding the AAR part shown in Fig. 7c and 7d, we attempted to maximize the use of screencasts of trainees' screens and cyber COPs captured during a training session. However, because retaining complete screencasts of all the screens displayed during a session could result in an enormous strain on the ICSTASY system's storage, the editing process for the screencasts to be used in AAR was added immediately upon the session's conclusion. As illustrated in Fig. 7c, only the partial, selected segments of the trainees' screencasts required for AAR remain after a training session. Although the screencasts of cyber COPs are editable in the same way as those of trainees, they are substantially more lightweight to process since only the visualization data for each COP is recorded and replayed.

D. COMPARISONS WITH OTHER TRAINING SYSTEMS

Table 1 outlines and contrasts the primary features of each work. We can certify that ICSTASY offers the most comprehensive features over any other training solution. While certain training systems, such as CRATE, may contain a number

[2]P/S and N/A denote *Partially Supported* and *Not Available* (unknown), respectively.

TABLE 2. List of Technique Instances for Red Team Agents.

No.	Tactic	Technique	TI ID	Target Platforms	Description
1	TA0001 (Initial Access)	Valid Accounts	T1078	Linux, Windows	Using ordinary user accounts
2		Spearphishing Link	T1192	Linux, Windows	Spearphishing using URL links
3		Spearphishing Attachment	T1193	Linux, Windows	Spearphishing using file attachments
4	TA0002 (Execution)	Service Execution	T1035	Windows	Windows service execution
5		Scheduled Task	T1053	Windows	Execution of a program via scheduled task
6		Command-Line Interface	T1059	Linux, Windows	Execution of an executable file with CLI
7		PowerShell	T1086	Windows	Execution with Windows PowerShell
8		Clipboard Data	T1115	Linux, Windows	Collecting data stored in clipboard
9		Space after Filename	T1151	Linux	Adding a space after a filename
10		Source	T1153	Linux	Execution of a function or a file using the source command
11		Local Job Scheduling	T1168	Linux	Registering a task on the cron daemon
12		User Execution	T1204	Linux, Windows	Execution of an executable file with the ordinary user privilege
13	TA0003 (Persistence)	Winlogon Helper DLL	T1004	Windows	Registering DLL on Windows Registry (e.g., HKLM\Software [Wow6432Node]Microsoft\Windows NT\CurrentVersion\Winlogon\ and HKCU\Software\Microsoft\Windows NT\CurrentVersion\Winlogon\)
14		Port Monitors	T1013	Windows	Manipulating a Windows Registry key to modify the DLL path called by spoolsv.exe
15		Modify Existing Service	T1031	Windows	Modifying sc.exe or a Windows Registry key to change binPath of a running service
16		New Service	T1050	Windows	Registering a new service with sc.exe
17		Service Registry Permissions Weakness	T1058	Windows	Modifying binPath or imagePath of services registered on Windows Registry (HKLM\SYSTEM\ CurrentControlSet\Services)
18		Registry Run Keys / Startup Folder	T1060	Windows	Adding/Modifying Windows Registry key related to starting programs
19		AppInit DLLs	T1103	Windows	Modifying a Windows Registry key (HKEY_LOCAL_MACHINE\ Software\Microsoft\Windows NT\CurrentVersion\Windows) used by AppInit
20		Netsh Helper DLL	T1128	Windows	Modifying a Windows Registry key (HKLM\SOFTWARE\Microsoft\Nets) used by NetSH
21		Authentication Package	T1131	Windows	Modifying a Windows Registry key (HKEY_LOCAL_MACHINE\ Software\Microsoft\Windows NT\CurrentVersion\Windows) used by Local Security Authority (LSA)
22		Create Account	T1136	Linux, Windows	Creation of an account

TABLE 2. List of Technique Instances for Red Team Agents.

23		.bash_profile and .bashrc	T1156	Linux	Adding a script code in .bash_profile or .bashrc
24		AppCert DLLs	T1182	Windows	Modifying a Windows Registry key (HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Session Manager) used by AppCert DLL
25		Port Knocking	T1205	Linux	Conduct of network port scanning
26		Time Providers	T1209	Windows	Modifying a Windows Registry key (HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\W32Time\TimeProviders\) used by the W32Time service
27	TA0004 (Privilege Escalation)	Sudo	T1169	Linux	Execution of a program with the administrator privilege
28		Sudo Caching	T1206	Linux	Using the privilege of root-privileged executable before its returning to the normal user privilege
29	TA0005 (Defense Evasion)	Masquerading	T1036	Linux, Windows	Disguise the names of malicious programs/processes as normal ones
30		File Deletion	T1107	Linux, Windows	Deletion of files using a normal delete operation
31		Modify Registry	T1112	Windows	Creation/modification/deletion/hiding of a Windows Registry entry/key
32		Clear Command History	T1146	Linux	Deletion of CLI command history
33		File Permissions Modification	T1222	Linux, Windows	Modifying the file permissions
34	TA0006 (Credential Access)	Credential Dumping	T1003	Linux, Windows	Dumping of the account credentials
35		Credentials in Files	T1081	Linux, Windows	Using the contents in stored files for account access management (e.g., xml files used for FileZilla to manage sessions)
36		Bash History	T1139	Linux	Accessing the .bash_history file
37		Exploitation of Remote Services	T1210	Linux, Windows	Penetrate into a host using vulnerability of the SMB protocol (e.g., Eternal Blue)
38	TA0007 (Discovery)	System Service Discovery	T1007	Windows	Collecting the system service information
39		Application Window Discovery	T1010	Windows	Accessing the application list
40		System Network Configuration Discovery	T1016	Linux, Windows	Accessing the system network configurations
41		System Owner/User Discovery	T1033	Linux, Windows	Accessing the system administrators/users informations
42		Network Service Scanning	T1046	Linux, Windows	Scanning the network services
43		System Network Connections Discovery	T1049	Linux, Windows	Accessing informations on the active network connections
44		Process Discovery	T1057	Linux, Windows	Accessing the running processes list

TABLE 2. List of Technique Instances for Red Team Agents.

45		System Information Discovery	T1082	Linux, Windows	Accessing the system informations
46		File and Directory Discovery	T1083	Linux, Windows	Exploring files and directories
47		Account Discovery	T1087	Linux, Windows	Accessing the user accounts list
48		System Time Discovery	T1124	Windows	Accessing the system time information
49		Network Share Discovery	T1135	Windows	Scanning shared networks
50		Remote Service	T1021	Windows	Making a connection using a remote service
51		Taint Shared Content	T1080	Windows	Uploading a malware on a shared file server
52	TA0009 (Collection)	Data Staged	T1074	Linux, Windows	Storing data in a temporary space
53		Automated Collection	T1119	Linux, Windows	Collecting files in an automatic manner
54		Data from Information Repositories	T1213	Linux, Windows	Acquiring data via information repositories such as databases and file servers
55	TA0010 (Command & Control)	Data Compressed	T1002	Linux, Windows	Conducting the compression of data
56		Data Encrypted	T1022	Linux, Windows	Conducting the encryption of data
57		Exfiltration Over Command and Control Channel	T1041	Linux, Windows	Exfiltration of data to a C2 server with the commonly used communication protocols
58		Exfiltration Over Alternative Protocol	T1048	Linux, Windows	Exfiltration of data to a C2 server with a special communication protocol
59	TA0011 (Exfiltration)	Standard Cryptographic Protocol	T1032	Linux, Windows	Conducting C&C communication for data exfiltration with the standard encryption techniques
60		Commonly Used Port	T1043	Linux, Windows	Conducting C&C communication for data exfiltration with the commonly used ports
61		Uncommonly Used Port	T1065	Linux, Windows	Conducting C&C communication for data exfiltration with non-commonly used ports
62		Standard Application Layer Protocol	T1071	Linux, Windows	Conducting C&C communication for data exfiltration with the standard application layer protocols
63		Data Encoding	T1132	Linux, Windows	Conduct of data encoding for data exfiltration
64	TA0040 (Impact)	Data Destruction	T1485	Linux, Windows	Deleting all the data in a storage
65		Data Encrypted for Impact	T1486	Linux, Windows	Conducting data encryption for sabotage (e.g., ransomware)
66		Service Stop	T1489	Windows	Termination of a running service
67		Stored Data Manipulation	T1492	Linux, Windows	Modifying stored data such as documents, email files and databases

of features comparable to ICSTASY, given the publicly available data, the technological maturity of each feature appears to be less than that of ICSTASY.

VI. CONCLUSION

This paper introduces ICSTASY, a novel cybersecurity training system for military personnel. It outlines the essential

requirements and design architectures that must be met for trainees to have an immersive training experience and to facilitate instructors in their capacity to coach and manage cybersecurity training effectively. The development outcome of ICSTASY as a prototype proved that design concepts and requirements were concretely represented and incorporated into the system, demonstrating the feasibility of integrated, comprehensive cybersecurity training. Our next effort will include integrating LVC interoperability with ICSTASY.

APPENDIX
LIST OF TECHNIQUE INSTANCES FOR RED TEAM AGENTS

See Table 2.

REFERENCES

[1] (Oct. 2021). A. Mehta. *Cyber Concerns, Classification Disagreements Lead Space Survey Results*. Breaking Defense. [Online]. Available: https://breakingdefense.com/2021/10/cyber-concerns-classification-disagreements-lead-space-survey-results/

[2] M. G. Wabiszewski, T. R. Andel, B. E. Mullins, and R. W. Thomas, "Enhancing realistic hands-on network training in a virtual environment," in *Proc. Spring Simul. Multiconf. (SpringSim)*, San Diego, CA, USA, Mar. 2009, pp. 1–8.

[3] R. S. Mudge and S. Lingley, "Cyber and air joint effects demonstration (CAAJED)," Inf. Directorate, Air Force Res. Lab, Rome, NY, USA, Tech. Rep. AFRL-RI-RS-TM-2008-12, Mar. 2008.

[4] W. D. Meitzler, S. J. Ouderkirk, and C. O. Hughes, "Security assessment simulation toolkit (SAST) final report," Pacific Northwest Nat. Lab. (PNNL), Richland, WA, USA, Tech. Rep. PNNL-18964, Nov. 2009.

[5] G. Torres, K. Smith, J. Buscemi, S. Doshi, H. Duong, D. Xu, and H. K. Pickett, "Distributed stealthnet (D-SN): Creating a live, virtual, constructive (LVC) environment for simulating cyber-attacks for test and evaluation (T&E)," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2015, pp. 1284–1291.

[6] C. Pham, D. Tang, K.-I. Chinen, and R. Beuran, "CyRIS: A cyber range instantiation system for facilitating security training," in *Proc. 7th Symp. Inf. Commun. Technol.*, Ho Chi Minh, Vietnam, Dec. 2016, pp. 251–258.

[7] R. Beuran, D. Tang, C. Pham, K.-I. Chinen, Y. Tan, and Y. Shinoda, "Integrated framework for hands-on cybersecurity training: CyTrONE," *Comput. Secur.*, vol. 78, pp. 43–59, Sep. 2018.

[8] G. Bernardinetti, S. Iafrate, and G. Bianchi, "Nautilus: A tool for automated deployment and sharing of cyber range scenarios," in *Proc. 16th Int. Conf. Availability, Rel. Secur.*, Vienna, Austria, Aug. 2021, pp. 1–7.

[9] S. Christey and R. A. Martin, "Vulnerability type distributions in CVE," MITRE, McLean, VA, USA, Tech. Rep., May 2007. [Online]. Available: https://cwe.mitre.org/documents/vuln-trends/vuln-trends.pdf

[10] S. Arshad, M. Alam, S. Al-Kuwari, and M. H. A. Khan, "Attack specification language: Domain specific language for dynamic training in cyber range," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2021, pp. 873–879,

[11] D. D. Updyke, G. B. Dobson, T. G. Podnar, L. J. Osterritter, B. L. Earl, and A. D. Cerini, "Ghosts in the machine: A framework for cyber-warfare exercise npc simulation," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU/SEI-2018-TR-005, Dec. 2018.

[12] M. Standen, M. Lucas, D. Bowman, T. J. Richer, J. Kim, and D. Marriott, "CybORG: A gym for the development of autonomous cyber agents," in *Proc. 1st Int. Workshop Adapt. Cyber Defense*, Aug. 2021, pp. 1–7. [Online]. Available: https://arxiv.org/html/2108.08476v1

[13] T. Sommestad, "Experimentation on operational cyber security in CRATE," in *Proc. NATO STO-MP-IST Spec. Meeting*, Copenhagen, Denmark, 2015, pp. 7:1–7:12. [Online]. Available: http://www.sommestad.com/teodor/

[14] T. Gustafsson and J. Almroth, "Cyber range automation overview with a case study of CRATE," in *Proc. 25th Nordic Conf. Secure IT Syst. (NordSec)*, Nov. 2020.

[15] J. Almroth and T. Gustafsson, "CRATE exercise control—A cyber defense exercise management and support tool," in *Proc. 5th IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Sep. 2020, pp. 37–45.

[16] H. Holm and T. Sommestad, "SVED: Scanning, vulnerabilities, exploits and detection," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Baltimore, MD, USA, Nov. 2016, pp. 976–981.

[17] Greenbone Networks GmbH. *OpenVAS–Open Vulnerability Assessment Scanner*. Accessed: Nov. 13, 2021. [Online]. Available: https://www.openvas.org

[18] M. Roesch, "Snort–lightweight intrusion detection for networks," in *Proc. 13th USENIX Large Installation Syst. Admin. Conf. (LISA)*, Seattle, WA, USA, Nov. 1999, pp. 1–11.

[19] P. Čeleda, J. Čegan, J Vykopal, and D Tovarňák, "KYPO—A platform for cyber defence exercises," in *Proc. Modelling Simulation Support Oper. Tasks Including War Gaming, Logistics, Cyber Defence (NATO STO-MP-MSG)*, Munich, Germany, Oct. 2015. [Online]. Available: https://www.sto.nato.int/publications/STOMeetingProceedings/STO-MP-MSG-133/MP-MSG-133-COVER.pdf

[20] N. Chouliaras, G. Kittes, I. Kantzavelou, L. Maglaras, G. Pantziou, and M. A. Ferrag, "Cyber ranges and TestBeds for education, training, and research," *Appl. Sci.*, vol. 11, no. 4, p. 1809, Feb. 2021.

[21] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, "MITRE ATT&CK: Design and philosophy," MITRE, McLean, VA, USA, Tech. Rep. MP180360R1, Jul. 2018.

[22] S. Y. Enoch, Z. Huang, C. Y. Moon, D. Lee, M. K. Ahn, and D. S. Kim, "HARMer: Cyber-attacks automation and evaluation," *IEEE Access*, vol. 8, pp. 129397–129414, 2020.

[23] S. Y. Enoch, C. Y. Moon, D. Lee, M. K. Ahn, and D. S. Kim, "A practical framework for cyber defense generation, enforcement and evaluation," *Comput. Netw.*, vol. 208, May 2022, Art. no. 108878.

[24] C. Lee, "Method for providing background traffic using IP random assigning in cyber range," *Electron. Lett.*, vol. 57, no. 6, pp. 261–263, Feb. 2021.

[25] D. Lee, D. Kim, M. K. Ahn, W. Jang, and W. Lee, "Cy-through: Toward a cybersecurity simulation for supporting live, virtual, and constructive interoperability," *IEEE Access*, vol. 9, pp. 10041–10053, 2021.

[26] VMware. *vSphere Automation API Reference*. Accessed: Nov. 13, 2021. [Online]. Available: https://developer.vmware.com/apis/vsphere-automation/latest

[27] *IBM Power Virtualization Center APIs*. Accessed: Nov. 13, 2021. [Online]. Available: https://www.ibm.com/docs/en/powervc/1.4.3?topic=power-virtualization-center-apis

[28] J. Turnbull, *The Logstash Book*. Research Triangle, NC, USA: Lulu Press, 2013.

[29] C. Gormley and Z. Tong, *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics engine*. Sebastopol, CA, USA: O'Reilly Media, 2015.

[30] *Sysinternals Suite*. Accessed: Nov. 13, 2021. [Online]. Available: https://docs.microsoft.com/en-us/sysinternals/downloads/sysinternals-suite

[31] *Osquery*. Accessed: Nov. 13, 2021. [Online]. Available: https://github.com/osquery/osquery

[32] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *Leading Issues Inf. Warfare Secur. Res.*, vol. 1, no. 1, pp. 80–106, Apr. 2011.

DONGHWAN LEE (Graduate Student Member, IEEE) received the B.E. degree in industrial engineering and the M.S. degree in computer science and engineering from Korea University, Seoul, Republic of Korea, in 2006 and 2008, respectively, where he is currently pursuing the Ph.D. degree in cybersecurity. He is a Senior Researcher at the Cyber/Network Technology Center, Agency for Defense Development, Seoul. His research interests include wireless communication, parallel and distributed computing, wireless security, and virtualization technologies for cybersecurity.

DONGHWA KIM received the B.S. and M.S. degrees from the School of Electrical Engineering, Korea University, Seoul, Republic of Korea, in 2004 and 2007, respectively. He is currently a Senior Researcher at the Cyber/Network Technology Center, Agency for Defense Development, Seoul. His research interests include cybersecurity training systems and red team automation.

CHANGWON LEE received the B.S., M.S., and Ph.D. degrees in electronics and computer engineering from Hanyang University, in 1999, 2001, and 2019, respectively. He is currently a Principal Researcher at the Cyber/Network Technology Center, Agency for Defense Development, Seoul, Republic of Korea. His current research interests include cyber security and hardware security.

MYUNG KIL AHN received the B.S. degree in information and communication engineering from Chungnam National University, Daejeon, Republic of Korea, in 1997, the M.S. degree in computer engineering from Sogang University, Seoul, Republic of Korea, in 2003, and the Ph.D. degree in electrical and electronics engineering from Chung-Ang University, Seoul, in 2021. She is currently a Principal Researcher at the Cyber/Network Technology Center, Agency for Defense Development, Seoul. Her research interests include computer security and cyberwarfare modeling and simulation.

WONJUN LEE (Fellow, IEEE) received the B.S. and M.S. degrees in computer engineering from Seoul National University, Seoul, Republic of Korea, in 1989 and 1991, respectively, the M.S. degree in computer science from the University of Maryland, College Park, MD, USA, in 1996, and the Ph.D. degree in computer science and engineering from the University of Minnesota, Minneapolis, MN, USA, in 1999. In 2002, he joined the Faculty of Korea University, Seoul, where he is currently a Professor with the School of Cybersecurity. He has authored or coauthored over 220 papers in refereed international journals and conferences. His research interests include communication and network protocols, optimization techniques in wireless communication and networking, security and privacy in mobile computing, and RF-powered computing and networking. He has served as the TPC and/or an Organizing Committee Member for IEEE INFOCOM, from 2008 to 2023, the PC Vice Chair for IEEE ICDCS 2019 and the ACM MobiHoc, from 2008 to 2009, and over 130 international conferences.