IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 14, 2022, accepted May 2, 2022, date of publication May 12, 2022, date of current version June 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174679

Zero Trust Architecture (ZTA): A Comprehensive Survey

NAEEM FIRDOUS SYED[1], (Member, IEEE), SYED W. SHAH[1], (Member, IEEE), ARASH SHAGHAGHI[1,2], (Member, IEEE), ADNAN ANWAR[1], ZUBAIR BAIG[1], (Senior Member, IEEE), AND ROBIN DOSS[1], (Senior Member, IEEE)

[1]Centre for Cyber Security Research and Innovation (CSRI), Deakin University, Geelong, VIC 3217, Australia
[2]Department of Information Systems and Business Analytics, RMIT University, Melbourne, VIC 3000, Australia

Corresponding author: Naeem Firdous Syed (naeem.syed@deakin.edu.au)

This work was supported by the Australian Department of Defence through the Defence Science and Technology Group's Operations Research Network.

ABSTRACT We present a detailed survey of the Zero Trust (ZT) security paradigm which has a growing number of advocates in the critical infrastructure risk management space. The article employs a descriptive approach to present the fundamental tenets of ZT and provides a review of numerous potential options available for successful realization of this paradigm. We describe the role of authentication and access control in Zero Trust Architectures (ZTA) and present an in-depth discussion of state-of-the-art techniques for authentication and access control in different scenarios. Furthermore, we comprehensively discuss the conventional approaches to encryption, micro-segmentation, and security automation available for instantiating a ZTA. The article also details various challenges associated with contemporary authentication mechanisms, access control schemes, trust and risk computation techniques, micro-segmentation approaches, and Software-Defined Perimeter, that can impact the implementation of ZT in its true sense. Based upon our analysis, we finally pinpoint the potential future research directions for successful realization of ZT in critical infrastructures.

INDEX TERMS Zero trust architecture (ZTA), access control, authentication, micro-segmentation, software-defined parameter (SDP).

I. INTRODUCTION

The rapid growth and adoption of the Internet of Things (IoT) and edge computing platforms has challenged the ability of traditional perimeter-based security architectures to effectively protect both enterprise assets and critical infrastructures. The notion of a *Zero-Trust Architecture (ZTA)* has been gaining momentum and is increasingly seen as the security architecture of choice for such infrastructures. As the name implies, ZTA is built on the notions of least privilege, granular access control and dynamic and strict policy enforcement wherein no user or device is implicitly trusted–irrespective of stature or location. In this paper, we undertake a horizon scan to identify the current state-of-the-art as relevant for effective implementation of ZTA in a critical infrastructure context. Although there are some existing works such as [1] and [2] which review the working principles of ZT, our focus in this article is on the basic tenets of ZT and how state-of-the-art approaches can be used to accomplish these. We critically analyse the individual components of ZTA to see whether existing techniques suffice for realization of ZTA in critical infrastructures. Based on this comprehensive survey of high-quality research publications relevant to the ZTA tenets, we then present recommendations that can be referenced as a guiding framework for the crafting of future cyber security strategies for the protection of critical infrastructures and their operations.

As per the National Institute of Standards and Technology (NIST) report on Zero Trust Architecture [3], ZTA is not a single network architecture which can be achieved using just one technology. Rather, ZTA comprises various guiding principles that need to be strategically implemented to secure enterprise assets such as data, devices, users and other components of infrastructure. The key principles for achieving ZTA are authentication and access control, as these are the means by which the user's identity is established and

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Casoni.

privileges ascertained for the conduct of different operations involving protected resource(s). For implementing ZTA in a critical infrastructure context, a strong authentication scheme which identifies both users and devices is required. Beyond this, rather than simply relying on entry-point authentication, the use of a context-aware and continuous authentication scheme which takes into account both the user and device contexts on an ongoing basis to ‘‘actively’’ authenticate is also recommended. In terms of access control, ZTA strategies should include a risk-aware access control scheme which determines the risk associated with the granting of an access request. Further to these two primary principles, ZTA realisation necessitates the adoption of lightweight encryption schemes to account for resource-constrained devices in cyber physical systems. Micro-segmentation and software defined perimeters are also recommended by NIST as core ZTA implementation strategies. However, these technologies require customisation to secure the edge network of IoT devices. Threat intelligence is also critical, as it can serve as a key feedback mechanism to drive automated security technologies within the defence environment. A tailored, reliably responsive system is mandatory for continuous trust evaluation and access control. Given the large volume of data coming from heterogeneous sources within a complex system, a Machine Learning (ML) supported approach is recommended for effectively deducing trust. This article offers a comprehensive survey of state-of-the-art approaches for achieving a zero trust architecture (See Figure 1 for an overview). Our goal is to support informed decision-making by practitioners when drafting security strategies appropriate for their own peculiar operational contexts.

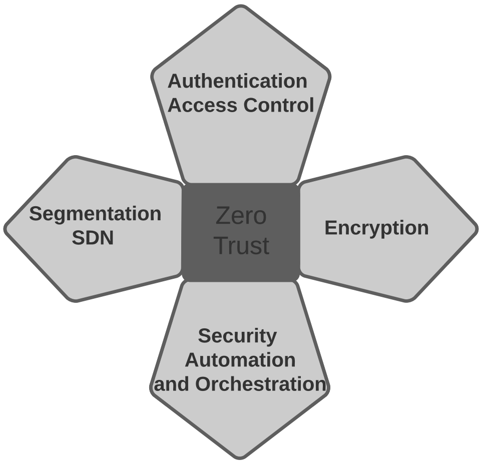


FIGURE 1. **Fundamental requirements for achieving a zero trust environment.**

After thorough investigation of each requirement depicted in Figure 1, we have arrived at the following conclusions:

- Conventional user authentication mechanisms (static, active, or context-aware) are either vulnerable or have limitations. Therefore, lightweight and scalable continuous authentication mechanisms are essential for enabling trust for all tangible and non-tangible organisational resources such as users, devices, systems and processes.
- Different IoT-enabled environments have entirely different access control requirements and therefore demand customised arrangements. For most critical infrastructures, risk-aware access control which incorporates capabilities congruent with a fine-grained access control scheme (such as FBAC, or Function-Based Access Control) seems the optimal choice, as this can actually be used to evaluate the risk inherent to a particular access request. This evaluation is performed by leveraging information collected from diverse sources, including reported threat-intelligence and outputs from cyber-physical systems (CPS) designed to support situation awareness. The result is context-specific access at a high-level of granularity.
- In essence, the goal of ZTA is to protect data, whether it is in rest, in transit, or being processed. Therefore, encryption is an important requirement for achieving a zero trust environment. Conventional encryption techniques are now at a greater risk of compromise due to recent developments in quantum computing. If encryption is to remain a viable means of securing data, post-quantum cryptosystems which can withstand quantum computing attacks are necessary. Fortunately, standardization of post-quantum cryptosystems is currently underway at NIST. However, successfully incorporating such systems across a host of different applications and devices–many of them resource-constrained–is a major challenge that will require dedicated and incremental research efforts.
- **A**nother important element of ZTA is micro-segmentation, which enforces policies closer to the protected resources, thereby breaking the network into smaller logical segments and precluding lateral movement by attackers. Micro-segmentation in the IoT context can be achieved using software-defined networking (SDN) along with the use of network function virtualisation (NFV) and a software-defined perimeter (SDP) that acts as an overlay network to protect resources. While this approach has shown some promise, it necessitates significant changes to networks, clients, and servers, however. Another problem is that the central controller then becomes a single point of failure which can be targeted by malicious entities to severely impair SDP functionality. Thus, a federated, network-and-application-layer-aware segmentation and SDP technique that is resilient to contemporary cyber attacks is required for diverse defence network architectures.
- Threat intelligence (TI) and Security Situation Awareness (SSA) are important elements of ZTA because they provide necessary feedback to the policy engine for making informed decisions. However, defining indicators of compromise (IoC) is generally difficult. The

heterogeneity of sources and the sheer volume of data involved also makes it difficult to accurately identify potential threats within a complex system. Thus, it is essential to have an effective feedback system in place which can incorporate input from heterogeneous data sources and device state monitoring logs to effectively recognise and react to threats.

- The dynamic enforcement of access policies within a ZTA requires a reliable trust evaluation capability. However, trust schemes and risk evaluation frameworks face challenges in cyber-physical systems (CPS) and critical infrastructures due to the numerous sources of input and high volumes of contextual data collected. To meet these challenges, the services of Machine Learning (ML) algorithms are needed to automate the detection and prevention processes (with minimal false positives) for effective implementation of a ZTA. Deep learning has already shown great utility toward cyber attack detection and it will continue to improve as computing capabilities advance and more data is gathered to support the growing body of knowledge. ML is an indispensable component of the strong security automation framework that is required for a functional ZTA.
- Likewise, the heterogeneity of sources and data containing numerical and imprecise information makes it challenging to have a trust mechanism that can leverage the varying input data (e.g. contextual information, behavioral data, device-related information, and location information). The implementation details of risk computation remain obscure in current literature: more work must be done to elaborate the procedures for implementing access control policies at the network and application levels for critical applications.

A. REVIEW APPROACH

The presentation of our literature review is structured as follows. We begin by listing the basic tenets or logical components of ZTA as outlined by NIST. Elaborating on these, in subsequent sections we critically analyse each identified tenet with reference to current practices intended to instantiate or embody it. This is illustrated by means of different use-case scenarios. In each section pertaining to a particular ZTA tenet, the weaknesses of the current practices are identified and explained. In total, we reviewed more than 180 articles and reports relating to the ZTA tenets. The keywords we used for retrieving these documents were zero trust architecture, authentication, access control, encryption, micro-segmentation, and security automation (i.e. the requirements depicted in Figure 1).

While the topic of zero trust has been explored by other authors in previous published work, our approach for the current literature review differs significantly (see Table 1 for a comparison). Our approach is both new and more comprehensive because we examine prevailing practices in context to identify their shortcomings vis-à-vis NIST's full list of identified ZTA tenets.

TABLE 1. Comparison with existing works.

Ref	Approach	Includes all Tenets
[1]	Presents key-technologies of ZT and their applications in different scenarios	No
[2]	Discusses top three factors that drive the need of ZT	No
[4]	Presents ZT architecture, its adoption in firms, impact upon users and technology interaction, and related limitations	No
[5]	Presents challenges and steps to consider when migrating to ZTA	No
Ours	Identifies all tenets that form the basis of ZTA, comprehensively reviews the state-of-the-art, and highlights the related issues that impede realisation of secure/genuine ZTA	Yes.

B. TARGET AUDIENCE

This article is intended for developing cyber-security researchers with aims to 1) provide basic information on ZTA; 2) present the state-of-the-art for practices indeed to meet ZTA requirements; 3) identify the problems with these current approaches; and 4) present a broad overview of future research directions across the ZTA tenets.

C. ORGANISATION

The rest of this Article is structured as follows. Section II presents a generic discussion on ZTA and its essential tenets. State-of-the-art authentication and access control mechanisms are presented in Sections III and IV, respectively. Encryption mechanisms are discussed in Section V, and segmentation techniques are detailed in Section VI. Section VII expands on security automation and orchestration. Finally, the discussion and conclusion is presented in Sections VIII and IX, respectively.

II. ZERO TRUST ARCHITECTURE

The operative definition of zero trust and zero trust architecture (ZTA) in accordance with NIST [3] is as follows: zero trust refers to the collection of ideas that may help to lessen (or ideally eliminate) the ambiguity involved in in enforcing the precise access decisions for each and every request made by viewing the network as compromised, and ZTA refers to the actual overall system design intended to support this.

An abstraction of zero trust access is shown in Figure 2, which illustrates the roles of authentication and authorization via Policy Decision/Enforcement Points (PDP/PEP) to enforce access control for every connection request. The access control relies on device security posture and potential consideration of other contextual factors (e.g. time and location, prior access behaviour) that may impact confidence level before access to a resource is granted as per the defined policies.

NIST defines the seven basic tenets for a ZTA [3] which are aimed at achieving the optimal goal of implementing ZTA (with the option of selectively implementing some tenets and not others, in accordance with perceived need).

TABLE 2. Mapping of forrester's tenets on NIST baseline.

Forrester	NIST Baseline						
	Resource	Commun. Security	Session Security	Access Control	Minimum-Security Posture	Continuous Authentication	Information logging
ZT Network	x			x			
ZT People	x	x	x	x		x	
ZT Devices	x	x	x	x	x	x	
ZT Data		x					
ZT Workload	x			x		x	
Visibility& Analysis							x
Automation				x		x	

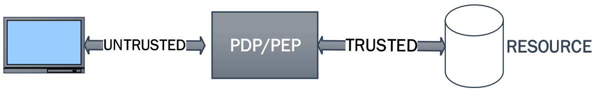


FIGURE 2. Abstract zero trust access control [3].

- *Resource* – any data source or computing service
- *Communication Security* – communication is secured irrespective of location.
- *Session Security* – access to resources is granted on a per-session basis, and authentication and authorization for one resource may not extend privileges to others.
- *Access Control* – access to resources is determined by dynamic policy, including the observable state of client identity, application, and requesting asset.
- *Minimum-Security Posture* – enterprise ensures that all the owned and associated devices are in the most secure state and monitors assets to ensure this.
- *Continuous Authentication* – all resource authentication and authorization is dynamic and strictly enforced. An enterprise that intends to implement ZTA may have an identity, credential, and access management (ICAM) system and multi factor authentication (MFA) for added security. A continuous inspection during the interaction of the user with a possibility of friction free re-authentication/authorization may help.
- *Information Logging* – the enterprise collects as much information as possible about the current state of the network infrastructure and communications, and uses this information to improve its security posture.

Similar zero trust principles are described in Forrester's extended zero trust model [6]. The principles in Forrester's model are data protection centred, with all entities such as user, device, network and workload protected through the analysis and automation of all network operations. Table 2 shows a mapping between tenet terminologies, revealing their interchangeability and suggesting that either scheme is useful toward guiding zero trust security practice. Since these two schemes effectively sort the same core concepts under different headings,, Table 2 situates these concepts within a matrix to illustrate relationship between them.

A. ZTA LOGICAL COMPONENTS

ZTA comprises various services consisting of numerous logical components, which are operated either onsite or offsite via a cloud. Of these components, NIST describes three as core: policy enforcement point (PEP), policy administrator (PA), and policy engine (PE), as shown below in Figure 3.

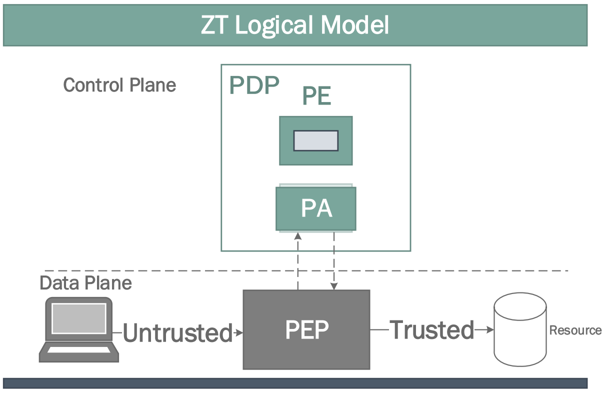


FIGURE 3. Core logical components of ZT [3].

The functionalities of these three core components are as follows:

- Policy engine (PE) - makes the access decision in accordance with enterprise policies by feeding the external inputs to a trust algorithm which functions as the "brain" of the entire system.
- Policy administrator (PA) - works closely with the PE and either allows or denies access as per the PE's decision. It may be incorporated into the PE and it talks to the PEP for policy enforcement.
- Policy enforcement point (PEP) - enables, monitors, and finally terminates the connection between the subject and the resource. It can be further divided into subcomponents, namely, client (e.g. agent on a device) and resource (e.g. a gateway). The area beyond the PEP is usually a *trust-zone*.

In addition to the aforementioned core components, a number of external components are mentioned in [3] that facilitate the realization of zero trust security (e.g. continuous diagnostics and mitigation, data access policies, identity management, security information and event management (SIEM), activity logs, etc.). As described above, the actual access decision is made by the PE by leveraging a *Trust Algorithm* (TA).

1) TRUST ALGORITHM (TA)

The trust algorithm (TA) is the process employed by PE to make a decision by considering inputs such as entries in the policy database, user role, attributes, behavioral information, threat-related information, etc. as per the need of a particular deployment, as shown in Figure 4.

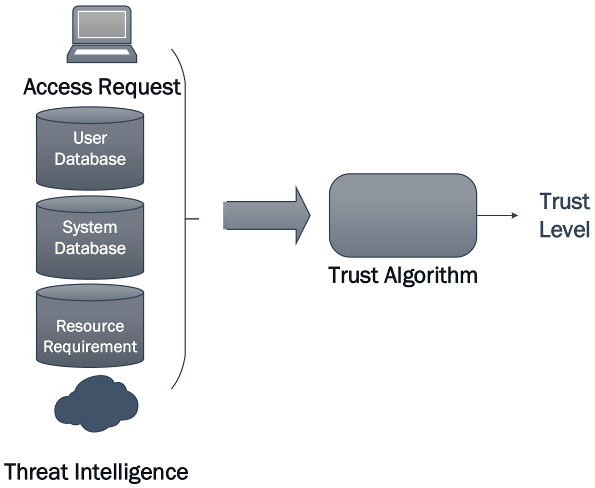


FIGURE 4. **Trust algorithm.**

- *Access Request* - Whenever a user makes an access request, the basic information regarding the resource and the requester is used by the TA (e.g. operating system, patch level, and application used).
- *User Identification, attributes, and Privileges* - User related information, authentication performed by the PE, and other attributes such as time and location may also be used by the TA to compute the confidence level. The collection of privileges given to different users as encoded in the ID management and policy database may also be utilized by the TA.
- *Asset Database and Observable Status* - This database comprises the status of all the resources, which are compared against the observable status of the requester (i.e. operating system, patch level, location) by the TA to make an access decision.
- *Resource Access Requirement* - This includes policies that are based on minimal requirements to access a resource as set by the custodian, e.g. the requirement of multi-factor authentication at a new location.
- *Threat Intelligence* - This includes information such as attack signatures or malware operating on the Internet, as provided by either internal or external sources and used by the TA.

In TA, all aforementioned data sources are assigned different weights as per their importance and the needs of enterprise. Once the TA has made the decision, the PE passes this to the PA which configures all the corresponding PEPs to either enable or disable the communication; for example, it may send the configurations to gateways and agents, thereby requiring re-authentication, re-authorisation, or termination of the connection, as per the defined policies. In accordance with the NIST framework [3], the TA can be implemented in various ways as described below.

- *Criteria vs. Score–based TA* - Criteria-based TA necessitates a combination of attributes be fulfilled before allowing an action. Whereas in score-based TA, the weighted values of input data are used to compute a confidence level which is compared against the threshold value. If the confidence level is greater than the set threshold, access is granted; otherwise it is denied.
- *Singular Vs Contextual TA* - Decisions in singular TA do not take the user's historical information into account while making a decision, which may lead to a faster decision making process but can have repercussions in the sense that some threats can go undetected. On the contrary, contextual trust makes use of a user's historical behavioural patterns while making a decision, which is helpful, as deviation from a normal usage pattern may indicate a threat–one that a contextual TA has the ability to recognise and flag.

As indicated in the NIST guidelines [3], contextual score-based TA may be an optimal choice, as it can compute a confidence score by leveraging contextual information, thereby increasing the possibility of detecting a potentially malicious request. Different methods used for trust computation in different conventional and cyber-physical systems are detailed in Section VII-F.

B. ZTA IMPLEMENTATION TECHNIQUES

ZTA can be implemented in an organization by adopting various technologies and techniques. As explained above, PE consists of policy rules that are derived from a number of data sources [3]. The data sources can be categorised into:

- Identity management (validity of user accounts, cryptographic certificates etc.)
- Access control (access levels of various users, past access history and contextual attributes)
- Network and Application logs collected through SIEM
- Threat Intelligence (internal and external threat knowledge base)
- Continuous Diagnostics and Mitigation (CDM)
- Compliance to various industry defined standards

These data sources are integrated in the TA of the PE, which makes policy decisions based on various factors in compliance with the ZTA principles. The zero trust architectures can also vary based on components deployed and data sources that drive the organisation's PE [3]. Zero trust architectures can be based on:

- Identity Governance: In this approach, the identities of users and devices are the primary factors incorporated in

TABLE 3. Techniques to implement ZTA tenets.

ZTA tenets	Objective	Techniques to implement the tenet
Resources	Resource identification and classification (a precondition for all other tenets)	Identity management – users, devices
All communication is secured irrespective of network location	Enforce same security policy for both internal and external access requests	Segmentation (network and application segmentation to enforce policies closer to the data or resource), authenticate and authorise all requests both internal and external, encrypt all communication both internal and external
Access to resources is granted per connection basis	previous connection or session access rights do not influence subsequent session rights; access is strictly limited to the requested resource	Session authentication, granular access control at the resource level
Access to resource is granted by dynamic policies which consider the state of user/device identity and may include other behavioral attributes	Dynamic access policies need to be implemented that consider state of user identity, state of user device, and their behavioral attributes	Context based authentication and authorization, risk based (calculate risk score based on context and history), adaptive and dynamic access control techniques
All owned resources/devices in most secure state	Continuous diagnostics and mitigation to assess the state of the device; compromised devices should be prevented access	Device state monitoring, Communication monitoring, Behaviour based authentication to deny access to compromised devices
All resource authentication and authorization is dynamic and strictly enforced	Authentication and authorisation are continuous and automatic processes, reevaluating and adapting trust in new and ongoing communications	Continuous authentication and authorization, adaptive access policies, re-authentication and re-authorisation, automation of authentication and authorisation
Collect information to adjust and improve security posture	Continuous data collection to detect threats in the system, with automatic application of required security measures	Activity logging, network monitoring, threat detection and analysis; reactive reconfiguration

the PE. The policy decisions consider the validity of user identities, contextual information and asset sensitivity to measure access risks.

- Micro-Segmentation: In this approach, the PEP is deployed closer to the resource or the data to protect it from unauthorised access. This prevents lateral movement by an intruder. Micro-segmentation can be achieved by placing virtual firewalls to eliminate unauthorised access.
- Network Infrastructure and Software Defined Perimeters: In this type of ZTA implementation, an overlay network is created to control access to resources which are otherwise not accessible directly, thereby creating a software-defined perimeter (SDP). The advantages of software-defined networks are leveraged in this approach to enforce access control and secure communication with back-end applications.

An implementation of zero trust requires identifying the main zero trust tenets applicable to the situation, as well as what techniques are required to enforce these tenets. Although ZTA principles and their logical components have been defined by various organisations, the implementation strategy required for a critical Infrastructure (CI) is still unclear. The reasons behind this is that CIs use a wide range of technologies (new and legacy systems) and endpoints (e.g. IoT devices, CPS devices, traditional network endpoints), which presents various challenges to the effective implementation of ZTA. Hence, it is essential to identify the most suitable techniques for realising ZTA tenets in a CI. Table 3 lists the techniques associated with the implementation of ZT tenets, which will be discussed in detail in subsequent sections. We maintain that to successfully realize ZT in CI, the underlying techniques requiring thorough investigation include authentication, access control, encryption, network segmentation, SDP, and security automation. A detailed discussion of these is provided in subsequent sections (see Sections 3 through 7). We believe this will provide the reader with a succinct overview of the current state-of-the-art for each technique while also pinpointing the weaknesses and knowledge gaps which future research needs to address.

III. AUTHENTICATION

Authentication is the process of verifying the identities of users (or devices, in machine-to-machine scenarios) when they attempt to access resources. This is important for determining whether a subject requesting access is legitimate, which is a fundamental consideration when attempting to establish zero trust. In this section, we present an overview of state-of-the-art techniques for user and device authentication.

A. CONVENTIONAL USER AUTHENTICATION MECHANISMS AND RELATED ISSUES

User authentication is critical to both personal devices and online services. However, it has been repeatedly demonstrated that the traditional methods of authentication are vulnerable to subversion. For example, the use of passwords–the most popular method of authentication–has numerous vulnerabilities. Users often use easy-to-guess passwords (e.g. asd123). It is also not uncommon for users to replicate passwords across multiple accounts; consequently, if any one of these accounts is compromised, then all others are susceptible as well. Moreover, well-crafted, ‘‘strong’’ passwords are also prone to hacking and can be inferred by conducting sophisticated side channel attacks such as those mentioned in [7]–[9]. Authentication mechanisms based on physical biometrics such as fingerprints, face recognition and, iris scans are also not difficult to bypass. For example, fingerprints are easy to capture from a surface to make a dummy fingerprint. It has been shown that face recognition based mechanisms can be compromised by using the victim's photograph or a 3D-printed reproduction of the user's head [10]. Similarly, iris-based mechanisms can be bypassed using a photograph of the user's iris superimposed onto a contact lens [11]. Vein-based approaches [12] and typing/tapping characteristics captured thorough sound, as proposed in [13], are promising but may be impractical if they only work on limited devices or in controlled environments.

In view of the above considerations, the popularity of multi factor authentication (MFA) is increasing. A widely used instantiation of MFA is referred to as two-factor authentication (2FA). 2FA generally combines any two of the following (traditional) authentication factors - i.e., *knowledge* (e.g., password), *inherence* (e.g., fingerprint), and *possession* (e.g., hardware or software token). Through its requirement for two authentication factors, 2FA provides an additional layer of security because if one of the factors (e.g. password) is compromised, the other is still in place (e.g. token) to thwart illicit access. The second factor of authentication can be divided into two broad categories, namely, hardware and software tokens. For example, [14], [15] provides an overview of the hardware token based solutions, which require the user to have a specific hardware which generates a unique one time code (*OTC*) to be used as the second factor of authentication. However, the problem with this approach is that it requires users to carry a dedicated piece of hardware with them at all times, which users commonly find inconvenient [16]. Another disincentive for using this approach is that it incurs an extra cost to the service provider. An alternative to such hardware-based solutions, is the software-token. A typical example of this is the one time code sent to a user's pre-registered mobile phone number via SMS. A significant vulnerability of this approach is that it is susceptible to interception [17]. This approach can also have privacy implications because providing a personal mobile number to multiple service providers can lead to spam. These kinds of shortcomings have been redressed by application-based solutions in which the OTC is generated by an app (and not transmitted). Either approach is less than optimal because it requires significant interaction: the user must wait for the code and then manually enter it (or accept a push notification).

Numerous prior studies have indicated this extensive user interaction as a reason for the low adoption rate of 2FA. Apart from this, the dependence of these approaches on a secondary device (e.g. a mobile phone) is also problematic [18], [19]. If the 2FA device is lost, stolen, without power or otherwise inoperable, then the dependent service may not be accessible. The aforementioned shortcomings of most 2FA solutions necessitates the creation of a mechanism that is easy to use (to improve adoption rate) and which is not dependent upon any secondary device (such as mobile phone). A potential alternative could be the use of behavioural biometrics (e.g. to recognise some type of gesture) as a second factor. Such a solution would mean that the user did not have to carry any additional device for the purposes of user authentication.

B. CONTEXT-AWARE USER AUTHENTICATION

Given the aforementioned problems, the need for context-aware and continuous/active authentication is gaining greater recognition. The word 'context' can be defined as any information that can establish the situation of an entity [20]. Context-aware security uses this situational information (e.g. identity, geolocation, time) to decide whether to provide access to a particular resource.

Modern mobile computing devices are increasingly integrating sensors (e.g. GPS, accelerometer, gyroscope) which generally have enormous computing capabilities that could be used to support context-aware authentication. For instance, many online services such as online banking and email ask users additional questions when they attempt to login from a new IP address. In this case, the IP address is the contextual information being used to flag that the user may still be an adversary despite having provided the correct credentials. However, many commonly asked security questions have answers (e.g., place of birth, pet's name etc.) that are often readily available online (e.g. in social media posts and profiles) and are therefore no deterrent at all for an attacker willing to do some research. An interesting extension of this approach is presented in [21], where a user's mobile location provides the contextual passive information (using WiFi and cell tower data) and, based on this information, an active factor of authentication (PIN, Password, or None) is modulated to ascertain the identity of the user. The paper proposes a probabilistic framework that leverages the location information and decides (through risk-assessment) which active form of user authentication should be deployed. This provides a particular usability advantage in that the user may not be authenticated through an active mode (i.e. PIN or password) when the location in which the transaction occurs is illogical. Similarly, the authors of [22] showed that the mobility pattern of the user can be modelled by leveraging an n-gram model to determine anomalous instances where a mobile device may be stolen (and access to private data should be therefore restricted). The authors of [23] also used behavioural features (i.e. GPS location, time since email checked) to calculate a score which can be compared against a threshold to decide whether to perform implicit authentication.

The authors of [24] devised a context-aware user authentication mechanism for a scenario in which the user attempts to access an application or service hosted on a cloud. When a user attempts to access a particular service hosted on a cloud, an agent installed on the user's mobile device (e.g., phone) gathers the contextual information and sends it to the cloud-hosted context aware authentication system, which in turn compares this information against the saved information of the user and makes an authentication decision.

TABLE 4. Summary of context-aware authentication mechanisms.

Sr.	Contextual Information	Purpose	Ref
1	Location using WiFi and cell tower data	Modulate the explicit and implicit authentication	[21]
2	Location & emails data	Risk computation	[23]
3	Location, time, phone status	Risk computation	[27]
4	Location, time zone, OS details, phone details	Risk computation	[24]
5	Location, phone placement (hand, table, pocket)	Combine the implicit and explicit modes of authentication	[25]
6	Location, time, behavioral data	Access the smart home devices	[26]

Specifically, the mobile agent collects the time zone and GPS location, which are compared to proceed further. Then, the system checks the OS details like OS type, phone manufacturer and model. Finally, the mechanism computes a Cosine similarity between the applications installed and process running on the mobile device. If the similarity of applications and processes is greater than the defined threshold, then the access is granted (or denied otherwise).

The authors of [25] present an interesting contextual authentication mechanism (context-aware multimodal FIDO authentication, or CAMFA) for mobile phones used to access any service hosted remotely. CAMFA is compliant with FIDO (Fast IDentity Online-a technical standard for authentication systems). Here, the FIDO server defines the relying party's level of authentication (RP LoA) required for accessing a particular service. The FIDO client (on the relying mobile device) then utilizes this to request that CAMFA meet the level of authentication required for that service by means of explicit (PIN, Face) or implicit (keystroke, location, placement) methods. The CAMFA mechanism monitors the risk level associated with the user's current situation through utilization of sensors embedded in the device. For example, the user's risk level changes in accordance with location and where the mobile phone is placed (e.g. hand, table, pocket). When the user's sensed behavioural information corresponds with the situational information, the risk level is computed to be low. With reference to both the computed risk level and the LoA required for the service the user is attempting to access, CAMFA combines different implicit and explicit methods (e.g. PIN, keystroke, location, face) to authenticate users. The authors of [26] propose a context-aware authentication mechanism for smart homes that utilizes the user's location, time, and other behavioural data for accessing the devices integrated into these homes. For making an access decision, this mechanism assigns different weights to different pieces of contextual information (e.g., location $= 0.2$, time $= 0.1$, calendar $= 0.2$, password $= 0.3$, preference $= 0.2$). The confidence level is computed utilizing the defined weights and then compared against a threshold value to make the access decision. The evaluation demonstrates the flexibility of the approach for assigning security levels to different users, as well as the appropriateness of the aforementioned contextual information (which can be obtained reasonably quickly) for making access decisions.

Issues With Context-Aware Authentication: Table 4 presents a summary of contextual information used in different mechanisms. It is conspicuous that, in almost all of the mechanisms, location is used as the primary source of contextual information to be combined with some other form of information, such as time, phone details, or behavioral data. All of the approaches involve comparing different types of contextual information to compute a risk score which determines further requirements for authentication.

Although the context-aware authentication mechanisms described above have shown success in specific scenarios, they do have some limitations. For example, modern users access the same online services using many different types of devices (mobile phones, laptops, desktops, etc.) and the sensors required for establishing the necessary contextual information may not be present in many devices. For example, context–aware mechanisms that need an accelerometer to determine the position of a device for risk assessment (e.g., in [25]) will not work for laptops or desktops that have no accelerometer. Therefore, having a mechanism that can leverage rich contextual information–but which is also widely usable across all sorts of devices–is challenging. A suitable alternative might be to pair the user's location (i.e. contextual information) with the user's daily mobile phone activity to generate questions. Numerous works such as [28], [29] propose such authentication mechanisms, which reference call, SMS and web logs to as questions such as "Who did you call first today?" However, a problem with this approach is that such questions are easy to answer for close relations (e.g. partners, family or friends) who may in some cases be the adversary. To counter this issue, the user's historical location information may be referenced when presenting such questions. For example, when a person is at home, the authentication mechanism may ask questions related to activities performed at work and vice versa. Another anticipated problem with this approach is that it requires asking a series of questions, which may result in usability issues (and may only be suitable for fall-back or second-factor authentication). The selection of questions that are easy to answer for an actual user but difficult for an attacker to guess is also an open problem.

C. CONTINUOUS AUTHENTICATION

Traditional modes of authentication (e.g. passwords, biometrics) only provide entry-point security, which is to say that they only establish the identity of the subject (user, device, process) when the subject is attempting to access a secure service (or device). Once the subject passes this stage, there is generally no procedure to ensure that the authenticated subject is in on-going control of the session (or device). As mentioned earlier, passwords are frequently leaked or hacked. If any critical service is password protected and is somehow compromised, there is no way to confirm the subject's identity beyond the login stage. To address this issue, continuous authentication (also referred to as active, transparent, or implicit authentication) is widely advocated. For example, authors of [30] use the colours of a user's clothes and facial skin to continuously authenticate the user during a login session. However, continuously capturing the user's photograph and sending it (to a remote service) for authentication may not only be computationally expensive, but can also have serious privacy implications. In [31], authors proposed a continuous authentication mechanism that leverages the pattern of user's hand movement while typing on the keyboard. The webcam (as on the laptop) is pointed towards the keyboard and a continuous video stream is fed to the algorithm, which in turn attempts to repeatedly authenticate the user. There are two significant issues with this approach. First, continuously streaming the video will be onerous and may degrade the performance of the system if any other intensive computation is also being carried out. Second, requiring that a webcam remain pointed toward a keyboard means it cannot be used for anything else, which necessitates the use of and additional external device that can be oriented in this way without interfering with use of the keyboard or monitor. Likewise, the authors of [32] also use the user's typing behaviour for continuous user authentication. They use two features for accomplishing continuous authentication: key hold-time and inter-key time. However, this procedure may only work in situations where the user is actively engaged in typing activity. In many scenarios, the user may not be performing any task that involves typing, thus rendering the approach useless. A number of works have proposed continuous authentication mechanisms for mobile devices (e.g., smartphones). As the smartphone designs incorporate numerous sensors (e.g. touch sensors, accelerometer and gyroscope), they offer opportunity to profile the user's behavioural characteristics (e.g. how the screen is tapped), and reference this profile for continuous authentication. For example, [33], [34] use the touch dynamics on the screen and extract features such as orientation of the finger, pressure exerted on the screen, area occluded and time instances. These features are converted to vectors and fed to machine learning classifiers which then model the user's behavioural profile using the extracted features for future reference.

Many works have utilized gait patterns (i.e. the way person walks) for enabling continuous user authentication. Gait related data is captured by leveraging accelerometer and gyroscope sensor data. These methods either use the raw data corresponding to a gait pattern (and then use correlation or machine learning for performing continuous authentication, e.g. [35] [36]), or they extract features (e.g. fast Fourier transform or Wavelet coefficients) from the captured data to train the classifier and perform continuous authentication (e.g. [37], [38]). A class of continuous authentication techniques tap passive biometrics such as medical signals which do not require the user's active cooperation, unlike methods such as facial recognition, fingerprint recognition and speaker voice recognition, which can distract the user. Many medical signals such as brain activity tracked by Electroencephalogram (EEG) [39], heart rate monitoring using Electrocardiogram (ECG) [40], electrical activity of muscles monitored by Electromyogram (EMG) [41] and an ensemble of various other medical signals [42] have also been explored.

Continuous authentication mechanisms for devices which do not maintain physical contact with humans cannot leverage the aforementioned authentication schemes that monitor human user behavioural or rely on biometric-based parameters. In such scenarios, ubiquitous parameters, such as radio frequency (RF) signals, and ambient parameters, such as light, temperature, and sound that can be acquired by the devices without human interaction, have been proposed [43], [44]. Wireless channel parameters such as channel state information (CSI) and received signal strength (RSS), which change in the presence of users and their movements, can be used to continuously authenticate devices without human interaction. Ambient parameters can be similarly leveraged to verify the device location, as such parameters do not change drastically within short periods of time.

Issues With Continuous Authentication: Although the aforementioned approaches have demonstrated success in enabling continuous authentication in some particular scenarios, a few limitations are rather conspicuous. Almost all of these approaches are device-specific, i.e. they only work on the devices they are designed for. For example, an approach that uses the touch dynamics (e.g., [33]) for continuous authentication will not work on some other devices such as laptops or desktops. Therefore, the extension of such approaches for enabling continuous authentication for online services will be difficult, as such services can generally be accessed from a variety of devices (e.g. smartphones, laptops, desktops). The aforementioned approaches for continuous authentication are also scenario-specific. For example, approaches based upon gait pattern will only work while user is walking. Likewise, most of the continuous authentication mechanisms are based on behavioural biometrics (e.g. the way person types, taps, or walks) that tend to change under different circumstances (e.g. the tapping behaviour of a user may be different while sitting or walking). Therefore, the development of a continuous user authentication mechanism that can work across all sorts of devices and situations

TABLE 5. Summary of continuous authentication mechanisms.

Sr.	Approach	Hardware Used	Purpose	Ref
1	Typing behaviour	Webcam to record hand movement	Continuous authentication (on computers)	[31]
2	Typing behaviours	Keyboard	Continuous authentication (on Computers)	[32]
3	Touch dynamics	Touch screen	Continuous authentication (on smartphones)	[33], [34]
4	Gait patterns	Accelerometer, gyroscope	Continuous authentication (on smartphones, wearables, and in smart spaces)	[37], [38]
5	Physiological signals (EEG, ECG, EMG)	Corresponding signal capturing devices	Continuous authentication (in smart spaces)	[39], [40], [41], [42].
6	Radio signals	Corresponding RF hardware	Continuous authentication (in smart spaces)	[43], [44]

remains an open research problem. The aforementioned discussion reveals that most of the approaches for continuous authentication leverage the user's physical or behavioral biometrics. This presents obvious challenges for device design, as it requires the integration of new and improved components that have been optimised for authentication purposes.

D. DEVICE AUTHENTICATION

The Forrester extended ZTA ecosystem [6] considers devices such as IoT devices as potential threats to enterprise networks and have suggested a ZTA principle that allows enterprises to segment, protect and restrict devices connecting to the network. This principle is also referred to as Zero Trust Device (ZTD). According to the NIST ZTA, all resources that generate data are considered a resource. Resources could include several categories of devices such as servers, workstations, mobile devices, and IoT and operational technology (OT) devices. Devices can also be categorised into enterprise-owned or personal devices. In order to implement zero trust for devices, it is essential to identify all the devices that connect to the network and what they access. The traditional methods to authenticate people may not to be appropriate for identifying and authenticating IoT and OT devices. Some of the main challenges include:

- Most IoT and OT devices operate without human assistance, hence human-associated authentication factors are irrelevant.
- Machine-to-Machine (M2M) communications exist in IoT and OT networks which require new authentication mechanisms (e.g. mutual authentication) to authenticate devices [45].
- However, due to the typical computational limitations of IoT devices, existing identity verification methods might be impractical for authentication between them [46].

In an IoT or OT-based system, more devices than people will connect to the network and all must be authenticated. In the context of a ZTA, devices must be authenticated before messages can be exchanged between them for M2M communication. Popular authentication methods include symmetric key authentication, lightweight public key infrastructure (PKI), and Open Authorization 2.0 (OAuth2.0). In symmetric key authentication, a shared key is used between the sender and the receiver. Even though this method is easier to deploy, it is less secure than asymmetric key techniques. In asymmetric key authentication, digital certificates can be utilised to prove the identity of a device before communication is established. OAuth2.0 is a token-based authentication and authorisation scheme appropriate for authenticating IoT devices. In order to enhance the authentication process, hardware based methods such as Trusted Platform Module (TPM) and Trusted Execution Environment (TEE) are being increasingly used to secure and process authentication. As most IoT devices are resource constrained and most authentication schemes discussed thus far are incorporated during device enrolment, authors in [47] proposed a authentication and access control scheme for the entire IoT device life-cycle. The IoT device life-cycle consists of pre-deployment, ordering, deployment, functioning and retirement. The main advantage of this scheme is that an IoT device might exist in multiple domains during its life-cycle and the authentication scheme can work across multiple domains. The proposed scheme is based on attribute based access control which is certificateless and can reduce computation costs on constrained devices. In addition to the above mentioned authentication schemes for IoT devices, constrained devices require unique identities which are tamper-proof and cannot be easily imitated. This is discussed in detail in the following section.

1) IDENTITY OF DEVICES

Research efforts are being made to specify an identity of things (IDoT) scheme that can be used to enforce strong access policies within a ZTA. Simple device attributes such as International Mobile Equipment Identity (IMEI) number, manufacturer details, and model and firmware version would not be sufficient, as these can be sniffed from the network and used for imitating genuine devices [46]. A viable IDoT scheme should support the following properties:

- Unique device identification
- Tamper resistance and unclonability
- Adaptive authentication and access control

- End-to-end encryption
- Scalability

Similar to user authentication factors, four categories of information can be leveraged to verify the identity of devices during authentication [46]. These are shown in Figure 5. Inherited information is that which a device inherits from its hardware components. Information of this type is fetched by recognising attributes such as physically unclonable functions (PUFs). PUFs are unique design characteristics that are used only in one piece of hardware. For example, circuits can be integrated in specific way that is verifiable via challenge-response behaviour between components [46]. By applying an electrical stimulus to the PUF, a response based on the interaction of the stimulus and the physical micro-structure of the device is produced. Due to the unclonable and unpredictable nature of the PUF response, this technique has emerged as a way to create lightweight, identity-based cryptosystems [48]. To date, PUF-enabled signature detection has been proposed for identity verification [49], [50] and securing IoT communications [51], [52]. However, this approach has also been shown to be vulnerable to modelling attacks that can be used to infer characteristics well enough to support cloning and redistribution [53]. In addition, multiple challenge-response pairs need to be collected to enroll the devices. Adversaries can use machine learning based techniques to map detected responses against various challenges, which can in turn be used for device masquerading [54]. These issues have been addressed by using cryptographic schemes such as Elliptic curve cryptography [54]. Similarly, the authors of [55] proposed a device authentication protocol for wireless devices which leverages the frequency response of the speaker-to-microphone (S2M). Such hardware components have unique characteristics due to factors such as minute differences in manufacturing processes and other uncontrollable variables which result in unique design instances which behave in unique ways that partner devices can be made to recognise. Whenever device D_1 must authenticate itself to another device, D_2, D_1 sends a sound to D_2, which computes the received frequency and compares it against the known fingerprint of that device (i.e. D_1) to make an authentication decision. While experiments demonstrated that this approach is resilient against replay attacks, however, it is also only applicable to devices with embedded speakers and microphones. Numerous IoT devices do not feature these components, thereby rendering it an unusable approach in such instances. Authors in [56] proposed a continuous device-to-device (d2d) authentication protocol that leverages Wi-Fi channel state information and uses a dynamic function for frequent updating of keys to accomplish authentication continuously. However, this mechanism may only be employed on Wi-Fi-enabled devices.

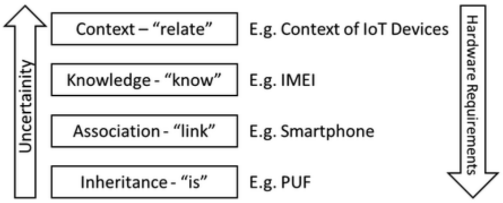


FIGURE 5. Various components of the identity of IoT (IDoT) devices, adapted from [46].

The second category of IoT identity is the association it has with other devices such as IoT gateways or smartphones. As IoT devices do not posses any identity generating information such as hardware tokens, their association with IoT gateways can be used instead as associations that are not expected to change regularly [46]. A common example of this is that a personal wearable device will transmit data to the cloud via the user owned smartphone that this device has been paired with. Though this is practicable in the context of personal devices owned by specific individuals, however, it is not a feasible approach to identification in the context of an integrated industrial environment. The third category of IoT device identity is knowledge about the device. Information such as manufacturer, IMEI number, firmware version number, and serial number can all be used as ways to establish the identity of a device. The fourth category is contextual information about the operational environment within which IoT or OT devices are being used. Such information can be collected by monitoring the behaviour of a device in relation to its neighbouring devices to establish a baseline of historical patterns. Current behaviour can then be compared against this baseline to detect agreement or deviation.

Commercially, Public Key Infrastructure (PKI) has been widely adopted to provide unique identities to IoT devices. However, with PKI, certificates need to be managed (distributed, revoked, stored and provisioned) with regard to their use by IoT devices. Intrinsic ID, on the other hand [57] provides a leaner solution by using a static random-access memory (SRAM) PUF to generate a unique identity. IoT Identity management tools such as Ericsson's IoT identity access management (IAM) platform and Vouch's decentralised IoT IAM are two commercially available solutions currently being used to take IAM beyond traditional human-centric frameworks [58].

Issues With Device Authentication: The broader categories of device authentication are summarized in Table 6. However, these approaches have some associated problems: they are prone to vulnerability via known attack vectors and also cannot be extended to many of the device types ubiquitous in critical infrastructures. Various open research directions for device authentication are presented in [59].

TABLE 6. Summary of device authentication mechanism limitations.

Approach	Limitations
PUF-based Solutions	Vulnerable to modelling attacks through ML techniques
Cryptography-based solutions	ECC is vulnerable to quantum attacks
Device fingerprinting	Can only be used on specific devices.

TABLE 7. NIST guidelines for authentication at different levels (adapted from [60]).

Requirement	AAL1	AAL2	AAL3
Permitted authenticator types	Memorized secret; Look-up secret; out-of-band; SF OTP device; MF OTP device; SF crypto software; SF crypto device; MF crypto software; MF crypto device MF OTP device; MF crypto software;	MF crypto device; or memorized secret plus: • Look-up secret • Out-of-band • SF OTP device • SF crypto software • SF crypto device	MF crypto device; SF crypto device plus memorized secret; SF OTP device plus MF crypto device or software; SF OTP device plus SF crypto software plus memorized secret
FIPS 140 validation	Level 1 (government agency verifiers)	Level 1 (government agency authenticators and verifiers)	Level 2 overall (MF authenticators) Level 1 overall (verifiers and SF crypto devices) Level 3 physical security (all authenticators)
Reauthentication	30 days 12 hours or 30 minutes inactivity;	MAY use one authentication factor	12 hours or 15 minutes inactivity; SHALL use both authentication factors
Security controls	SP 800-53 Low Baseline (or equivalent)	SP 800-53 Moderate Baseline (or equivalent)	SP 800-53 High Baseline (or equivalent)
MitM resistance	Required	Required	Required
Verifier-impersonation resistance	Not required	Not required	Required
Verifier-compromise resistance	Not required	Not required	Required
Replay resistance	Not required	Required	Required
Authentication intent	Not required	Recommended	Required
Records retention policy	Required	Required	Required
Privacy controls	Required	Required	Required

a: NIST's REQUIREMENTS FOR DIGITAL IDENTITY ESTABLISHMENT

Verifying digital identity over open networks presents unique technical challenges associated with impersonation and other forms of attack. NIST provides recommendations for authentication processes and authenticators (called tokens in some specifications) which can be used to achieve various Authenticator Assurance Levels (AALs) [60]. The three AALs are defined as follows:

AAL1 - AAL1 provides some confidence that the claimant has control over an authenticator associated with the subscriber's account. AAL1 requires single-factor or multi-factor authentication, which can be accomplished using a variety of secure authentication mechanisms that the claimant can use to establish custody and control of the authenticator.

AAL2 - AAL2 offers higher confidence that the claimant has control over an authenticator. Through use of secure authentication protocols, possession and control of two separate authentication factors is established. At AAL2 and higher, approved cryptographic algorithms are necessary.

AAl3 - AAL3 gives a high level of assurance that the claimant is in control of the authenticator(s) associated with the subscriber's account. Authentication at AAL3 is based on a cryptographic protocol that verifies the possession of a key. AAL3 authentication requires both a hardware-based authenticator and an impersonation-resistant authenticator; the same device can meet both requirements. Claimants must verify possession and control of two separate authentication elements using a secure authentication protocol to authenticate at AAL3. Approved cryptographic techniques are required.

Table 7 shows the requirements for different AALs. Interested readers are referred to [60] for more details on requirements and elaboration on different means of authentication (i.e. authenticator types) such as look-up secrets, out-of-band devices, and single factor and multi-factor one-time password (OTP) devices. Various standards exist that can be used for authentication processes within ZTA. An overview of these standards is presented in Table 8. One or more of these standards can be used by an organization to achieve a particular authentication level. Although some of these standards do offer certification (e.g. FIDO functional certification

TABLE 8. Various authentication standards.

Standard	Description
FIDO [61]	The FIDO Alliance is an open industry association that develops and promotes authentication standards with a focus on reducing the over-reliance on passwords
FIDO2 [62]	A joint effort between the FIDO Alliance and the World Wide Web Consortium, with goal is to create strong authentication for the web. At its core, FIDO2 consists of the W3C Web Authentication standard and the FIDO Client to Authenticator Protocol 2.
U2F [63]	The Universal 2nd Factor is an open standard that strengthens and simplifies two-factor authentication using specialised Universal Serial Bus or near-field communication devices based on similar security technology found in smart cards.
WebAuthn [64]	Web Authentication is a web standard published by the World Wide Web Consortium (W3C). WebAuthn is a core component of the FIDO2 Project under the guidance of the FIDO Alliance. The project aims to standardise an interface for authenticating users to web-based applications and services using public-key cryptography.
FIPS(1-4) [65]	The United States Federal Information Processing Standards are publicly announced standards developed by NIST for use in computer systems by non-military government agencies and contractors.
OATH [66]	Open Authentication is an industry-wide collaboration to develop an open reference architecture using open standards to promote the adoption of strong authentication.
OAuth [67]	Open Authorisation is an open standard for access delegation, commonly used as a way for Internet users to grant websites or applications access to their information on other websites but without giving them the passwords.

to measure compliance and ensure interoperability among products and services that support FIDO specifications), it is advisable that certification be offered which covers all that is required at each AAL. As it stands, the higher the AAL, the closer it is to fulfilling the theoretical requirements for a ZTA.

IV. ACCESS CONTROL

The foundational requirement for ZTA is access control–the ability to ascertain the privileges of a subject (an authenticated user or a process executed on that user's behalf) and restrict access accordingly. In essence, the purpose of logical access control is to protect resources such as devices, data, and applications (referred to as objects) with respect to operations available to a subject (e.g. read, write, execute). To accomplish a particular operation on a specific object, the subject must satisfy the access control policies, i.e. if the policy is satisfied, access is granted to an object. These policies are part of the organisation's access control mechanism (ACM), and are derived from the business and security requirements of the enterprise. As described by NIST, the ACM is a logical component that assesses access requests and decides whether the subject is authorized to execute the requested operation [68]. ACMs can deploy numerous methods to define and enforce access control policies.

A. IDENTITY-BASED ACCESS CONTROL

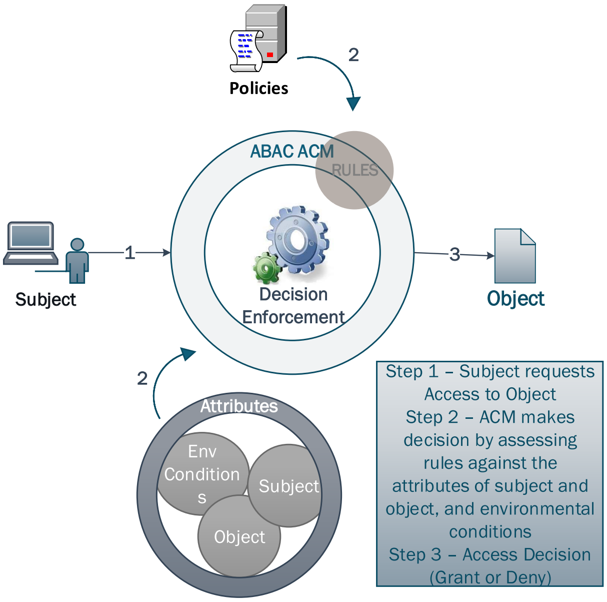
Identity-based access control (IBAC) is a simple and coarse-grained approach to access control, where access authorization is directly mapped to the subject's identifier. One approach to adopting IBAC is through the use of an Access Control List (ACL), which requires the system administrator to define access rights for objects with regard to different identities which can be recognized as subjects. The problem with this approach is that it is not scalable for efficient use within settings that involve dynamically changing groups of subjects and objects because it would require constantly revising access authorisations every time a change is made. Moreover, the access decisions are not made with respect to contextual considerations such as business functions or characteristics but solely the identifiers themselves.

B. ROLE-BASED ACCESS CONTROL (RBAC)

In contrast to IBAC, RBAC utilizes the roles of the different subjects within the organization to implement the access control mechanism. For example, a subject with the designated role of "research-scientist" will be allowed to access "R&D" related documents. In contrast, a "procurement officer" may not be allowed to operate on the aforementioned documents. In RBAC, the access is predefined at the time of defining the roles (which in turn are explicitly related to the privileges). Whenever a subject attempts to access a resource, it is the subject's role which the ACM compares against rules to allow or deny an access request. In theory, RBAC enables the central management of access control without need for unwieldy ACLs. However, in practice, RBAC can easily result in "role-explosion," or the accumulation of roles and privileges that endure beyond the times they are justified, which can also require significant amounts of work to prevent or correct. Variations of RBAC have been proposed in the access control literature which improve upon the original RBAC scheme. However, adoption of these solutions has been rather limited due to factors such as projected deployment costs, the infeasibility of certain assumptions in real world settings, and inherent limitations on achieving fine-grained access control when to more recent schemes (such as attribute-based access control).

C. ATTRIBUTE-BASED ACCESS CONTROL (ABAC)

As the name suggests, the ABAC positions its access control mechanism on many attributes. In a sense, IBAC and RBAC can be seen as special cases of ABAC. IBAC uses the subject's "identity," while RBAC leverages the subject's "role" to enforce access control. The difference is that the policies define a complex Boolean rule set through which multiple attributes can be checked. As described in [68], the ABAC can be defined as an ACM in which subjects' access requests to perform operations on objects are evaluated by taking into account the attributes of the subject, object and environment, along with the policies defined around the aforementioned attributes and conditions. Environmental conditions refer to the context in which the access request is sent through, e.g. time, week, location, and risk level. An overview of ABAC is presented in Figure 6. Extensible Access Control Markup Language (XACML) and Next Generation Access Control (NGAC) support and implement the ABAC model. However, they differ significantly from each other in their approaches to defining and managing attributes, as well as in how access decisions are made and enforced. A comparison of XACML and NGAC is conducted by NIST and discussed



Step 1 – Subject requests Access to Object
Step 2 – ACM makes decision by assessing rules against the attributes of subject and object, and environmental conditions
Step 3 – Access Decision (Grant or Deny)

FIGURE 6. Overview of attributes-based access control.

in [69]. In a nutshell, XACML defines the polices by using logical formulae involving attributes, while NGAC utilizes enumerations that involve configurations of relations. Below, an overview of the advantages NGAC has over XACML is offered:

- NGAC utilizes a single linear decision-making algorithm that is applied over non-conflicting policies. By contrast, XACML involves multiple complex processes around the collection of attributes, condition matching, computation of rules, and resolution of any conflicts.
- NGAC allows complete separation of access control logic from the operational environment, unlike XACML, which only allows for partial separation.
- NGAC allows an easy inculcation of Discretionary Access Control (DAC), which is difficult to accomplish in XACML.
- Unlike XACML, NGAC allows per-subject and per-object reviews of combined policies.

Recently, access control is of particular interest for IoT systems such as those employed in homes. For example, the authors of [70] proposed a lightweight ABAC for IoT networks. Specifically, they considered a scenario in which a more powerful mobile device is accessing a smart device (i.e. a smart camera). This scenario is of particular interest as the powerful devices could potentially take control of any device in a smart environment to violate user security and privacy (e.g. by accessing a smart camera to see inside the home). For enabling ABAC in an IoT context, [70] proposes to collect contextual information, i.e. location, time, and network information such as MAC address and network type. Using this contextual information, the authors first compute the trust locally to ensure that the requesting device meets a predefined trust threshold to be allowed access to any smart device or sensor. Once the local trust value is established, the access request, along with the trust value and context information, is sent to a cloud-based decision point. If the access request is assessed as safe, an access grant decision is sent to the gateway that manages access to the smart device.

The authors of [71] calculate the risk associated with an access request based on the user's contextual information and use this for controlling access to a device by an unauthorized person while also protecting information about the context of the device from an adversary. They presented an access control mechanism that captures and adapts to the user's perception of the context (i.e. feedback through user interaction) and conducts autonomous classification of this contextual information. Specifically, the location context was detected via a Global Positioning System (GPS) and Wi-Fi Access Points, and social context was established by detecting people's proximal devices via Bluetooth. Subsequently, the context was classified by leveraging machine learning fed with scrupulously computed features from the contextual information. The outcome of the analysis was next retrieved as a score, which was then used to make a conclusive decision about access request. Similarly, the authors of [72] also proposed a combined trust and attribute-based access control scheme for use in the IoT environment. They determined the user's trust by leveraging his/her behavioral characteristics. To compute the trust level they leveraged fuzzy sets to determine the current trust level and then combined it with the previous trust levels to arrive at a final trust level. Once the trust was computed, they updated the trust attributes database and combined it with the other static attributes, i.e. subject, object, operation, and environmental condition.

The authors of [73] also investigated the problem of an IoT device disrupting other devices within a home context, either by accident or maliciously. They demonstrated that ABAC is the appropriate way to enforce access control in smart homes as it can incorporate user, device, and environmental conditions into the access decision process. The authors argue that XACML is not an appropriate choice within an IoT context and that NGAC affords an easier and more efficient approach to adaptive policy definition and management.

D. RISK-BASED ACCESS CONTROL (RbAC)

Compared to the previously discussed access control models, Risk-Based Access Control is covered in far fewer proposals, as interest in it has been limited predominantly to the context of military operations. Recently, however, a few proposals have emerged which investigate the adoption of RbAC in the general IoT context. With RbAC, rather than having static and predefined policies, risk analysis is used to determine the risk associated with the particular request. This is then compared against the access policies and acceptable risk to ascertain whether to allow access when receiving an access request.

Authors in [74] used fuzzy inference systems that leverage the security levels of the subject and the object to determine the risk for making access decisions. However, this approach is not scalable for the IoT environment as it requires an excessive amount of time to establish the risk value. Authors in [75] also used fuzzy modeling for measuring risk using action severity, risk history, and data sensitivity. However, instead of using real-time contextual information, the fuzzy rules were based on prior knowledge about the deployed scenario. More recently, authors in [76] used fuzzy inference in conjunction with expert knowledge to estimate risk when granting access in an IoT context. This approach used real-time contextual attributes of the subject making the access request to estimate the risk and make the appropriate access decision. These attributes included the user's context, resource sensitivity, action severity, and risk history to determine the risk.

E. CAPABILITY-BASED ACCESS CONTROL

Capability-based access control leverages the concept of capability to define the privileges of the subject. This concept was initially introduced in [77] as a token that gives its possessor privileges to access an entity within a computer system. The scheme relies on cryptographic signed tokens which determine the privileges of a subject to conduct a particular operation on an object. Authors in [78] proposed a distributed capability-based access control for an IoT environment. This solution has two phases. In the first phase, this approach computes a session key by performing the authenticated key-exchange. In the second phase, the session key is used to establish secure communication and a capability token is used to gain access to the protected resource. However, capabiliy-based access control demands that all devices act as a PDP, which may be deemed precarious for constrained devices (such as those which typically predominate in IoT applications).

F. USAGE CONTROL (UCON)

The Usage Control Model is a more flexible model for authorization which focuses on the granularity of access decisions. In conventional models, the attributes of subject and object can only change either before or after the authorization, but not once the authorization is permitted. The UCON introduces the concept of mutable attributes, obligations, and conditions. The mutable attributes can change their values during the time that the subject is accessing a particular object, thereby enabling policy enforcement before the authorization and also continuously during access of the object. Therefore, the proper remedial actions (e.g. blocking access) can be taken immediately after the attributes are updated, even if the subject is still in the process of accessing the object. The authors of [79] showed the efficacy of UCON toward the enforcement of energy-saving and safety policies in a smart home. For example, they demonstrated the suitability of UCON in implementing policies such as ''smart oven may only turn-on if there is an adult present.'' This is to ensure that there is no gas leakage or safety risk to the children in a home. The authors demonstrated that their access control scheme was capable of registering local and remote attributes in real-time, allowing permissions to be revoked if any of the attributes were updated. However, this approach requires a separate usage control system as well as an attributes manager on every node, which may again be precarious where resource constrained devices are concerned

In line with the UCON approach, a number of proposals have emerged which focus on enhanced granularity in access authorizations. One of the main threats motivating this research has been increased recognition of insider threats within organizations. For instance, the authors of [80] propose a Linux container-based solution for isolating the system administrators from resources irrelevant to their current ticket's task while enabling them to obtain additional permissions when approved by the permission broker. A more granular and generic approach is proposed by Desmedt and Shaghaghi in [81]. Inspired by the concept of Functional Encryption, the authors propose Function-Based Access Control (FBAC). With FBAC, access authorizations are no longer stored as a two-dimensional Access Control Matrix (ACM). Instead, FBAC stores access authorizations as a three-dimensional tensor (called an access control tensor). Hence, applications no longer blindly give execution rights and users can only invoke commands which have been authorized at different levels such as data segments. Simply put, one might be authorized to use a certain command on one object while being forbidden to use the same command on another object. Evidently, this level of granularity and customization can not be efficiently modeled using the classical access control matrix. The authors discuss the theoretical foundations of FBAC and argue that their proposed model results in a new generation of applications capable of enforcing access restrictions at unprecedented granularity. The proposed solution has not to date been studied in the context of the IoT but it may bring enhanced access control capabilities to this domain if so deployed. It may be particularly relevant for such emerging IoT software architectures as the modular framework proposed in [82].

G. ARCHITECTURES FOR ACCESS CONTROL

As indicated in [83], there are three main architectures for access control, namely, *Policy-based*, *Token-based*, and *Hybrid* architectures. In the following, we succinctly review the basics of these architectures:

1) POLICY-BASED ARCHITECTURE

A typical example of this architecture is XACML, which comprises a PEP, PDP, Policy Administration Point (PAP), and Policy Information Point (PIP). Figure 7 illustrates interaction between the different modules involved. The policies are designed by the PAP and are made available to the PDP (1). The subject makes the access request (2), which is received by PEP and in turn forwarded to the PDP module (3). The PDP evaluates the access request against the set of available policies and if any further information is needed, the PDP obtains it by consulting the PIP (4,5). Finally, the PDP sends the access decision to the PEP (6), which implements the restrictions.

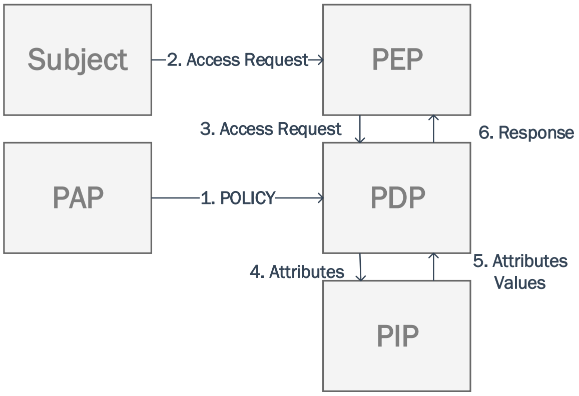


FIGURE 7. Access control - policy based architecture.

2) TOKEN-BASED ARCHITECTURE

The policy-based solution generally has a single, centralized point for policy evaluation and enforcement, which may not be suitable for situations where the resources are distributed across multiples nodes, as in an IoT context [83]. Therefore,

TABLE 9. Summary of different access control methods.

Approach	Pros	Cons
RBAC	minimum privileges	role explosion, scalability, interoperability, Coarse granularity, no contextual info
ABAC	fine-grained, scalable, interoperable	complexity, not user driven
UCON	fine-grained, scalable, mutabililty	complexity, not user driven, no functionality for administration
CapBAC	distributed, user-driven, simplicity, support revocation	coarse grained, no contextual info

the token-based architecture has recently been introduced as a suitable alternative in which permissions to subjects are encoded in tokens by the authorization services. These tokens are then used for granting access to the objects. Amongst many token-based standards, *OAuth* is widely used for allowing client applications to access the resources hosted on the HTTP servers. Figure 8 shows the different components of a typical token-based access control. The subject makes an access request to the resource owner (1), which provides the authorization grant depicting the authorization for the subject (2). Using the authorization grant, the client requests the access token from the authorization server, which upon verifying the request generates a token. Once the token is obtained by the client, it requests a desired resource from the resource server, which validates the token and serves the request only if the token is deemed valid. The entire process of this approach is shown in Figure 8.

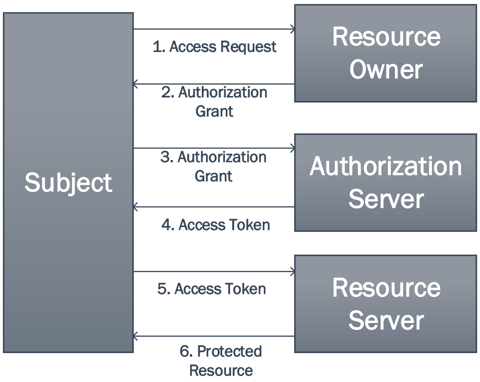


FIGURE 8. Access control - token based architecture.

3) HYBRID ARCHITECTURE

As indicated in [83], the token-based approach requires a level of user interaction which may be considered burdensome by the user. In view of this, the authors of [84] proposed a new approach referred to as User Managed Access (UMA), which extends OAuth, thereby allowing the configuration of policies in the authorization server so as to automate the generation of tokens without user interaction. In this way, UMA combines features of the Policy and Token-based approaches.

a: ISSUES WITH CONVENTIONAL ACCESS CONTROL METHODS

Table 9 presents a cross-comparison of the aforementioned approaches to access control [85]. As indicated in [83], there is no one-size-fits-all approach which will work across all different scenarios, from smart homes, to smart buildings, to IoT-context health applications. The reason for this is that different use-cases impose different requirements for policy management and evaluation. Irrespective of the specific scenario, an authorization framework should be fine-grained and context-aware. Both of these conditions are satisfied when either ABAC or UCON is used as an access control model. The most important requirement for an access control framework for smart homes is *usability*, meaning that the framework should demand minimal effort from the home owner to define policies. The optimal solution for this scenario will be a centralized and policy-based architecture in which policies are autonomously generated by leveraging contextual information. In addition, the PAP should aid the home owner in configuring and modifying policies. Latency can be tolerated to a certain extent in a smart home, thereby allowing for a run-time policy evaluation strategy. The PDP can be deployed on an edge device, such as IoT gateway or a local cloud. As indicated in [83], most of the frameworks designed for smart homes do not fulfil the aforementioned requirements. They tend to define policies which are either coarse-grained or involve overly simplistic consideration of environmental conditions. Similarly, the authorization framework for IoT applications in healthcare should require a *minimal effort for administering the policies for multiple devices*. In addition, it should have low *latency* (as latency in this context can potentially lead to a life-threatening condition), and should also take into account the *constrained capabilities of medical devices*. Therefore, the PDP should be deployed on edge-devices, while the PEP might reasonably be located in the device itself. Most of the authorization frameworks proposed for healthcare either do not reasonably consider usability requirements or adopt run-time evaluations of policies that can lead to latency issues. By contrast, contexts such as smart buildings and interconnected vehicles mostly involve direct d2d communication with little or no requirement for user interaction. This means that usability is not a significant requirement in a framework designed for this scenario. On the contrary, *interoperability, latency, and automated* *decision making capability* are the most important requirements for fully automated scenarios. A similar framework as described for health IoT applications can possibly work for this use-case with an added functionality of a PDP that allows the correct interpretation of policies from different administrative domains. As indicated in [83], however, most of the frameworks proposed for this scenario do suffer from latency issues.

H. BLOCKCHAIN BASED ACCESS CONTROL

The traditional centralized access control models have been well documented and analyzed. For example, the authors of [86] indicated that the existence of a trusted third party for access control can lead to a single point of failure. Recently, blockchain has shown some promise as a suitable alternative for access control. As the blockchain is inherently distributed, it helps to alleviate problems associated with the centralized approach (i.e. single point of failure and risk of privacy leakage). Furthermore, it also aids in maintaining a trusted log, and a smart contract can help with enforcing complex access privileges. For instance, the authors of [87], leverage blockchain to implement a solution that continuously inspects access authorisations with the goal of providing auditability of data access for personal health record system users and facilitating detection of anomalies. Authors in [88] proposed leveraging blockchain for publishing the access polices corresponding to a resource and for distributed transfer of access rights amongst the users. This allows the users to monitor the policies related to a particular resource and determine who has rights to access that resource. This also helps to prevent a fraudulent denial of access rights granted by the enforceable policy. The authors demonstrated the efficacy of this approach by deploying the policies defined in XACML on a Bitcoin blockchain. The authors of [89] used blockchain-enabled smart contracts to enforce the access control policies, thereby enabling subjects to verify that the policies are correctly enforced while also minimizing the chances of fraudulent denial of access by a malicious third party. They store the smart contract representing the access control policy on blockchain with a proper transaction whenever it is created by the resource owner. When a subject makes an access request, a transaction is generated with a reference to evaluate the policy and make the access decision. They demonstrated the efficacy of this approach by codifying the XACML policies into a smart contract (using Solidify language) and deployed it on Ethereum. This approach provides a benefit against malicious denial of access (e.g. a policy enforcement party can fraudulently enforce the system to deny the access), as the subject can see how the policies are being enforced. Likewise, the authors of [90] proposed attribute based access control using consortium blockchain for the IoT. This scheme has two main components: attribute authorities and IoT devices. The attribute authorities simultaneously act as consortium nodes and as the key generation center. They act as the managers of the blockchain, and use a consensus mechanism to jointly manage the distributed ledger. Simultaneously, for every IoT device that registers with the system, they generate a pair of public and secret keys based upon its identity (i.e. by using identity-based cryptography), using which the devices can mutually authenticate one another and agree on a session key. IoT devices use the attributes assigned by the attribute authorities to prove permission before they can exchange the data.

There are a few *vulnerabilities* when using blockchain for access control. For example, the blockchain needs all transactions to be recorded on all peers, for which a consensus mechanism is used. Recently, a lightweight consensus approach had been adopted to improve performance. However, this performance is still not comparable to that of centralized solutions [91]. In addition, the transactions in blockchain are inherently transparent. However, this is not desirable from a privacy perspective. Due to this reason the permissioned blockchain emerged, which provides privacy at the cost of decentralization. Likewise, as pointed out in [91], maintaining and improving the security of smart contracts and blockchain is also challenging.

The evolution of quantum-computers poses a serious threat to public-key cryptosytems and digital signatures, thereby warranting modifications of blockchain for the post-quantum era. In view of this, many efforts are currently being undertaken to standardize post-quantum cryptosystems (PQCs). For example, the authors of [92], [93] proposed some modifications to blockchains for the post-quantum era. However, since the standardization of post-quantum cryptosystems is still in process, the proposed modifications to blockchains will only be validated once the standardization process is concluded. Similarly, there are numerous issues with post-quantum blockchains, as pointed out in [94]. These include large key and signature sizes with correspondent impacts on performance. We refer readers to [94] for more detailed discussion on post-quantum blockchains and related challenges.

V. ENCRYPTION

As indicated in Section 1, zero trust implies tight control over the data. Given this, encryption is important to protect data at rest, in transit, or during processing. Encryption must be used to protect important enterprise data stored (i.e., at rest) in computing devices and portable storage devices (e.g. USB Flash drives). However, modern attackers have crafted numerous methods to retrieve encrypted data at rest. Numerous methods involving insiders and cryptographic or data integrity attacks have proven successful. Different methodologies such as data fragmentation and active defence attempt to remedy these problems. For example, [95] showed an effective way (referred to as ''Horus'') for data encryption in high-performance computing systems. Data fragmentation techniques such as Tahoe Least-Authority File Store and Storj [96], and active defence technologies such as CryptoMove [97] helps with protecting data by distributing, transferring and mutating the encrypted data in such a way that it is difficult to identify, retrieve or damage the data. In addition to the aforementioned simplistic situation in which data at rest is encrypted, it is also important to protect the data while in the processing stage. Data is increasingly being managed and processed in public (or private) clouds due to their ubiquitousness and other associated advantages. This creates a problem, as the cloud server would need access to encryption keys for processing the data, leading to security concerns. The processing of data locally (i.e. by downloading the data and decrypting it using the secret key) is challenging and computationally expensive. Given this, two emerging techniques, homomorphic encryption and Secure Multi-Party Computation (SMPC), are of particular interest where computation on encrypted data is concerned. Unlike conventional encryption methodologies, homomorphic encryption allows for the performance of computations on encrypted data without needing the secret key, resulting in encrypted output (of computation), allowing the owner of the data to retrieve the plaintext using the key. Prior research shows that homomorphic encryption–being based upon Ring-Learning with Errors (RLWE) and its relation to the hard mathematical problem of high-dimensional lattices–s secure against the current quantum computer. This makes them more secure than RSA and other cryptography approaches that are based upon elliptic curves. Given that more and more enterprises are switching towards cloud environments for saving and computing data, the importance of homomorphic encryption is becoming apparent. However, a lack of standardization of homomorphic encryption is making it difficult to enable its widespread use. Also, this makes it difficult to have a uniformed and simplified Application Programming Interface (API), making it difficult for application developers to understand and use APIs in this area. In addition to outsourcing data to clouds, homomorphic encryption can also help critical infrastructures such as healthcare to enable privacy-sensitive computation which is otherwise not possible. For example, due to data privacy issues in healthcare, predictive analytics is difficult to conduct. However, homomorphic encryption can realize such analysis by performing computation on encrypted data, thereby reducing the privacy concerns. Numerous implementations of homomorphic encryption are enlisted in [98]. The recent advancements in pervasive computing necessitate the cooperative computation on data shared by many parties while maintaining the data confidentiality of individual parties. This joint computation can be accomplished through cryptographic primitive SMPC in a privacy-preserving way. Existing solutions of SMPC including the overview of cloud-assisted cooperative methods, their architecture, and SMPC protocol for different scenarios such as privacy-preserving machine learning that is needed in many applications are discussed in [99].

A. LIGHTWEIGHT ENCRYPTION

The abundance of constrained sensing and computing devices such as IoT and sensor networks (and M2M communication in some situations) necessitates lightweight cryptographic methods which can conveniently work on devices with limited processing, storage, and power resources. Although conventional methods like AES (encryption), SHA-256 (hashing), and RSA/Elliptic Curve (signing) work desirably on most computing devices such as laptops, desktops, and smartphones, they fail to perform optimally in IoT device and embedded systems (e.g. RFID devices and sensor networks) contexts. For such constrained devices, lightweight cryptography is highly desirable as indicated in NIST's report [100]. The purpose of lightweight cryptography is to consume fewer resources (i.e. processing power, memory usage, and energy consumption) so that it can be accomplished on constrained devices. To accomplish this, we often see the smaller block, key and simple rounds of calculation in lightweight cryptography. However, this simplification comes at the cost of security (e.g. [101] demonstrated that 128-bit AES implemented on Arduino needed only 30 minutes to break by leveraging the differential and correlation power analysis). In hardware and software implementations of lightweight cryptography, RAM, energy, implementation size, and throughput are the important metrics to be considered.

The NIST-recommended methods for hashing (as a part of its early initiative on lightweight cryptography [100]) are SPONGENT, Quark, PHOTON, and Lesamnta-LW, as they all have a small memory footprint and input (of just 256 characters). SPONGENT makes use of sponge function and is based upon the finite state machine and cycles through the states as the input data is added. In sponge construction, a fixed-length permutation and padding is used to transform an input X^* of any length to X^o, where 'o' is defined as a part of the process. Precisely, the sponge construction makes use of a function (*f*) and has two phases, namely, "absorption" and "squeezing." In the absorption phase, r bits of input and state are XORed and interleaved with *f*. In the squeeze phase, *r* bits of state are blocked as output and interleaved with *f*. The length of the output (in bits) is defined as part of the hashing process. Lesamnta-LW is based on AES (with S-box structure similar to that of AES), and is 5-times faster than SHA-256 and requires a RAM of only 50 bytes on an 8-bit processor. Quark also uses sponge function and can be used for both hashing and stream encryption. Three different variants of Quark (i.e., u-Quark, d-Quark, s-Quark) can accomplish the 64-112 bit security. PHOTON, on the other hand, is also based on AES and creates an 80 - 256 bits hash. PHOTON can accept the input of any length and produce an output of variable-length. The detailed method of PHOTON is described in [102]. A comparison of different lightweight methods for hashing is presented in [103].

One of the promising alternatives of AES for lightweight encryption is PRESENT [104], which uses either an 80- or 128-bit encryption key. It operates on a block of 64 bits and employs SPN (substitution-permutation network) for encrypted output. PRESENT generally has 32 rounds with key-operation, S-box and P-box layers in each round. In operation, the key round performs the 'xor' operation between the key and input data, followed by an S-box (i.e., substitution) of 4 x 4 bits which helps in reducing the processing power in comparison with AES. Another alternate is XETA [105] which also operates on the 64-bit block and uses a key of 64-bits. XTEA is fast and has a small code size. Other options are SIMON (for optimized hardware implementation) with a key-size of 64-256 bits and a block-size of 32-128 bits, and SPECK for optimized software implementation. Mickey V2, Trivium, Grain and Enocor are the lightweight stream ciphers with low resource requirements as indicated in [103]. Similarly, CLEFIA is a lightweight block cipher that requires only 6k gates for implementation and has a block-size of 128 bits and variable key-size ranging from 128-256 bits. CLEFIA [106] is also included in ISO/IEC 29192 as a standard for lightweight encryption. RC5 is a flexible method that can support key-sizes of up to 2048 bits and block sizes of 32-128 bits. These parameters can be matched with the resources available on the device and its security requirements. For the lightweight signing of messages, Chaskey is a suitable option that uses a 128-bit key and requires around three thousand gates in contrast with SHA-256 that needs approximately fifteen thousand gates.

It is noteworthy that, NIST is currently standardizing the lightweight cryptography methods for constrained devices through an open competition-like process (the second round of the process completed in 2020) [107]. Therefore, the aforementioned methods may only be seen as a reference which were included in NIST's initial report on lightweight cryptography and until the standardization process is concluded. We refer readers to [107] for further details on the standardization process and recent entries.

1) ISSUES WITH LIGHTWEIGHT CRYPTOSYSTEMS

Table 10 presents a summary of lightweight hashing and encryption methods. However, the so-called Cryptographically Relevant Quantum Computer (CRQC) [108] is envisaged to be particularly precarious for asymmetric encryption (and even symmetric encryption and hashing in general). Though the symmetric encryption mechanism (e.g. AES) and hashing are secure against quantum attacks [109], they still need larger key size and hash length to maintain the necessary level of security [110]. This is particularly problematic for resource constrained devices with limited memory and weak computational power. In view of this, recent research works [111], [112] have attempted to design quantum safe lightweight cryptosystems using quantum permutation pads. However, significant research efforts are needed to analyze these approaches against known attack vectors and to evaluate their suitability for legacy hardware.

TABLE 10. Summary of lightweight cryptosystems.

Approach	Purpose
SPONGENT, Quark, PHOTON, Lesamnta-LW	Hashing
PRESENT [104]	Encryption (80 or 256 bit key)
XETA [105]	Encryption (64 bit keys)
SIMON	Encryption (64-256 bit key)
CLEFIA [106]	Encryption (128-256)

B. LIGHTWEIGHT MUTUAL AUTHENTICATION

Mutual authentication is of particular interest in the IoT environment. For example, an attacker can take over a vulnerable device (e.g. a sensor) and feed falsified data to the server to intentionally induce a bad decision, which in critical infrastructure can prove to be fatal (consider a traffic control system). This suggests that both sensor and server should mutually authenticate one another prior to data exchange. Again, as the devices in a particular IoT environment are generally constrained, this necessitates that the protocol should be lightweight. In this pursuit, a number of lightweight mutual authentication approaches have been introduced. For example, [113] presents a lightweight key agreement protocol (Algebric Eraser, -AE) which makes use of one way E-multiplications. However, their complexity increases linearly with the needed security level. Besides this, the vulnerabilities of [113] are indicated in [114]. Likewise, another approach that is referred to as NTRU is proposed in [115]. NTRU is based upon probability theory and polynomial algebra, and has received much attention recently due to its speed. Unfortunately, it has an associated problem with large key-size. Similarly, authors in [116] proposed a public-key encryption approach and demonstrated the efficacy of using it for a mutual authentication mechanism in constrained devices. The proposed encryption scheme is not computationally expensive, so it may be suitable for constrained devices. This mutual authentication protocol is shown to have benefits over other methods such as AE, NTRU, and Elliptic Curve Cryptography (ECC). Furthermore, the authors demonstrated that this scheme takes only around 125ms for mutual authentication on constrained devices. However, the evaluation has only been conducted on a Texas development kit. Thus, its actual performance on constrained devices within an operational environment remains unclear. Likewise, the authors of [117] proposed a lightweight encryption, key management, and authentication suite for IoT devices. They use one-key-for-one-file encryption, where the encryption key is generated using a random number and keystroke seed which is hard coded in the hardware security module of the device, thereby requiring no key to be either maintained by the devices or transported between them. Therefore, for encryption, the device picks a random number and uses it along with the keystroke seed to generate the key. Once encryption is done, the key is deleted and a random number is sent to the receiving party. The authors assume that all IoT devices within a network have obtained the unique identification of every other device during the configuration process. For mutual authentication, the sender sends the time stamp and its identification is encrypted with a random key (i.e. generated as described above) and a random number sent (n_1) to the receiver. The receiver upon receipt, uses the random number (n_1) and keystroke seed to generate the key and decrypts the message to ascertain the device identity and time stamp. The receiver then sends a random number (n_2), and modulo-2 addition of his unique identification and sender identification, along with a time stamp, all encrypted using a key generated by a random number, back to the sender. The sender can decrypt this message using the key generated from n_2 to confirm its identity and complete the mutual authentication. Although this approach is lightweight and can potentially be deployed in IoT replacing the IPsec, however, the assumption that all the devices within the network possess the unique identity of each other may not always hold.

Issues With Lightweight Mutual Authentication: In addition to the specific problems identified above, the generic problem for PKI-based mutual authentication is the threat of Cryptographically Relevant Quantum Computer (CRQC) attacks. As discussed above, CRQC are envisaged to be capable of breaking the conventionally difficult mathematical problem(s) which form the basis of PKI. Therefore, research efforts are need to have quantum safe lightweight mutual authentication mechanisms. NIST is currently undergoing a standardization process for PKI. However, it is not clear that current proposed methods in Round 3 of the process (code-based, isogeny-based, hash-based, lattice-based, and multivariate system-based solutions) will be sufficiently lightweight to work on the kinds of resource constrained devices ubiquitous in critical infrastructures.

VI. SEGMENTATION AND SOFTWARE DEFINED PERIMETER

According to NIST, ZTA can be applied in an enterprise using various approaches [3]. These approaches inform how a PEP is implemented and what the driving policies are. The two most common approaches, which we will discuss here, are micro-segmentation and software defined perimeters.

A. MICRO-SEGMENTATION

Micro-segmentation defines security measures and where those measures are implemented in the network. The core principle of micro-segmentation is the implementation of security policies closer to the resource being protected, effectively breaking a network infrastructure into smaller logical "segments" to efficiently protect a single resource (or logical group of them). Micro-segmentation enables only authorised entities within the data centre to access the application or data on protected resources, thereby preventing lateral movement by an attacker. Devices such as Next Generation Firewalls (NGFW) or security gateways, which will act as a PEP and enforce policies defined in the PE, have been proposed by NIST in its ZTA proposal [3].

Traditional network segmentation techniques like Virtual LANs (VLAN), routers and firewalls prove to be ineffective in providing granular security to workflows. In order to protect the east-west traffic flowing in the data-centres, granular security controls are required to enforce strict security policies between individual resources. Figures 9 and 10 show the traditional and micro-segmented network architectures. The micro-segmented network architecture has a micro-perimeter (firewall or security gateway) enforcing access policies to applications and the data in resources.

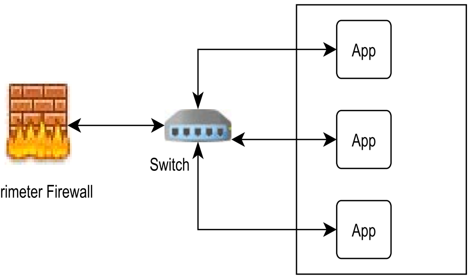


FIGURE 9. **Traditional network architecture.**

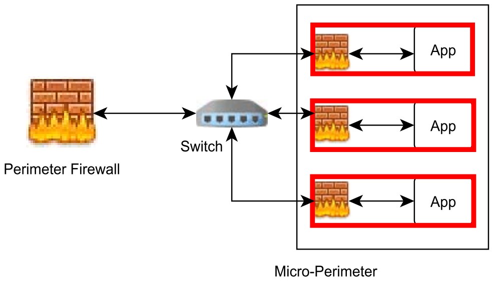


FIGURE 10. **Network architecture with micro-segmentation.**

Micro-segmentation techniques can be applied using various deployment models such as [118]:

- **Native Micro-segmentation**: In this model, micro-segmentation is achieved "natively" using the underlying infrastructure, such as a hypervisor or Operating System (OS) used to deploy the application servers. This allows the access policies to be deployed on the infrastructure or OS directly without the use of external hardware or software solutions. In addition, the security policies can be implemented to all applications instead of just few high priority applications. The advantage of this approach lies in applying fine-grained security controls closer to the applications without additional hardware. However, this model can only be implemented in the virtual environments where workloads are operating.
- **Third-party model**: In this model virtual firewalls are deployed by third-party firewall vendors. Virtual firewalls are used to configure and deploy access policies between virtualised servers. This allows controlling of a large number of distributed virtual firewalls from a single location and applying global access policies for workload needs. However, this approach requires enabling virtual firewall visibility to the workload traffic.
- **Overlay model**: This model uses agent software running on servers and a central controller or orchestration device to gain visibility into workflow communications and enforce dynamic access policies. The advantage of this approach is that the agents have greater visibility over an individual workload's communication patterns, allowing for dynamic deployment of access policies using a central controller.

TABLE 11. Commercial micro-segmentation products and their deployment and access control models.

Commercial Product	Deployment Model	Micro-Perimeter Approach
Cisco Application Centric Infrastructure (ACI) [123]	Native, Hybrid	Network-dependent and higher layer visibility using other Cisco products
VMware NSX [124]	Native, Hybrid	Network-dependent and higher layer visibility using other third-party products
vArmour [125]	Overlay, Hybrid	Works with Cisco ACI to provide security
Illumio [121]	Overlay	Label-based workflow identification
Unisys Stealth [126]	Overlay	User Identity based access policy
Microsoft	Native	Can work with third-party applications to provide higher level access control
Firewall Vendors (Checkpoint, Juniper, Cisco, PaloAlto etc.)	Third-party	DPI based
Cilium [122]	Overlay	API-aware
Consul [127]	Overlay	API-aware
istio [128]	Overlay	API-aware

- **Hybrid model**: In this model, a combination of native and other micro-segmentation approaches is deployed to effectively protect different layers of communication.

Most micro-segmentation implementations are network-dependent and require programmable software-based network equipment such as firewalls and switches where access policies are managed using centralised controllers [119]. Whereas implementations that use virtual firewalls or overlay networks are network-independent and can operate independent of the underlying network technologies.

The policies that need to be configured on these micro-perimeters require understanding the complete life-cycles of workflows and the complex interactions these entail within the enterprise network. In a network-dependent approach, the identified network flows related to a particular workflow need to be translated into network-based access rules using the network identities of the applications that require access. An example of such an access rule would be "application-A can access database-A," which can be translated into network based policies as "IP-address-App-A can access IP-address-DB-A:port2." The drawback of this approach is that workflows are identified and granted access solely using their network identities, which can be spoofed/forged. Network-independent approaches, on the contrary, use workload identities to create fine-grained policies. A few network-independent micro-segmentation approaches are given in [119].

1) TRANSPORT-LEVEL ACCESS CONTROL [120]

In this approach access control is achieved first by packet authentication in a TCP/IP communication. A stenographic overlay technique is used to embed identity tokens in a TCP connection initiation packet (TCP-SYN). The identity is first verified and the remainder of the TCP handshake process is carried out only if access is granted for the requesting identity. The drawback of this approach is that it can only be used with the TCP protocol and is not compatible other (connectionless) protocols such as UDP. In addition, the heavy cryptographic load before TCP connection establishment can be exploited by denial of service attacks to overwhelm the server resources.

2) LABEL-BASED ACCESS CONTROL [121]

This approach assigns labels to various workflows which are used to group and apply access policies based on said labels. This approach makes the access control policies independent of the protocols used and applicable to various types of workflows. However, as is the case with network based identities, these labels can be susceptible to spoofing attacks.

3) DPI-BASED ACCESS CONTROL

In this approach, Deep Packet Inspection (DPI) engines are used to inspect packet contents at various layers to either allow or reject connections [119].

4) API-AWARE ACCESS CONTROLS

This approach depends on breaking up workflows into smaller container-based (e.g. Dockers and Kubernets) services that communicate with each other using API [122].

Table 11 lists the various commercially available micro-segmentation products and their adopted deployment approaches.

a: GENERIC ISSUES WITH GRANULAR POLICY ENFORCEMENT

Allowing only specific authorised hosts from within the LAN reduces lateral movement in malicious activities; nevertheless such translation of workflow access rules into network-level access rules can lead to misconfiguration [119]. In addition, in a complex data center with a large number of workflows, identifying all the possible interactions between workflows and translating them into accurate access policies can be challenging and lead to disruptions in existing workflows. Furthermore, maintaining and updating such access policies due to workflow reconfiguration, new business policies and/or introduction of new workflows can be a challenge to these micro-perimeters. In addition to these challenges, with constantly evolving cyber security threats, network level access policies may not provide granular perimeter security to prevent sophisticated cyber attacks against workflows. Hence workflow access control policies must be context-aware and must be adaptable to changes in workflows as highlighted in [119] and [3].

b: NETWORK-BASED PERIMETERISATION

An improved approach to network-based perimeterisation in which security efforts are concentrated on workflows rather than network endpoints is proposed in [119]. This network-independent perimeterisation approach (eZTrust) can be realised using microservices which run in lightweight containers and which can be monitored. In the eZTrust model, the packets generated by the microservices are stamped with a tag which contains the detailed set of identities of the microservice. Identities such as name, version, kernel version, library version, user details and deployment specific identities can be fetched from the microservice or the service orchestrator. These contextual attributes are used to build context-driven access policies that verify the identity and current state of access-requesting workflows before access is granted. This approach has been found to result in 2–5 times lower packet latency and 1.5–2.5 times lower CPU overhead when compared to other network-based perimeterisation approaches or network-independent techniques such as Transport-level, Label-based, Deep Packet Inspection (DPI) based or proxy based perimeterisation.

The micro-segmentation techniques mentioned thus far only bring perimeters closer to the applications hosted in data centers. It is important to note that Lateral movement by attackers occurs not only in data centers but also in edge networks where M2M communications take place, however. The IoT and OT devices deployed at the edge lack advanced defence capability due to their resource constraints. With the increased deployment of IoT devices and with the boundaries between IT & OT technologies progressively blurring, the need for IoT and ICS device security must be addressed urgently.

c: MICRO-SEGMENTATION IN IoT

In IoT devices, micro-segmentation can be achieved using Software Defined Networking (SDN) along with Network Function Virtualisation (NFV). The authors of [129] proposed an SDN-based IoT framework which leverages the NFV technology to implement fine-grained network functions close to the IoT device, such as routing and access control, to secure IoT communication and enable Quality of Service (QoS) for critical network traffic. An SDN OpenFlow-based controller and IoT gateway were configured to deploy IoT devices and implement fine-grained network functions. In addition, the IoT gateway creates a secure IPsec tunnel with the application servers to secure the communication between IoT devices and the application. Authors in [130], coin the term Policy Enforcement as a Service (PEPS), which enables the provision of a network-level enforcement point, which access control systems (both application-layer and network-layer) can subscribe to, whether they share the same network domain or are external. The resulting inter-layer and inter-domain access Control makes it possible to limit threats closer to their originating source (e.g. a compromised IoT device).

In order to provide fine-grained access control for users to various types of IoT or ICS devices, [131] propose an SDN-based micro-segmentation approach. The end-users are considered the regular users of the IoT/ICS device for their daily tasks; administrators manage access to various end-users and devices; and maintenance personal provide specialised services such as maintenance and repairs. These user types require different levels of access to the devices. Fine-grained access rights proposed in [131] are controlled by deploying separate device proxies per user profile, which are loaded into containers in the IoT gateway. Thus, a single gateway will contain multiple device proxy containers for each user, so different access rights can be provided upon request. The SDN-based network equipment that implements SDN micro-segmentation rules manages the automatic routing of user requests to the appropriate device proxies in the IoT gateways and prevents any unauthorised access to unrelated device proxies.

The authors of [132] have proposed a novel security architecture to implement fine-grained security policies closer to the edge of an IoT network, based on micro-segmentation principles. The authors argue that traditional centralised access controls do not provide sufficient security to the IoT network edge, where most computing is expected to occur as the use of IoT devices increases. The security architecture proposed in [132] consists of a traffic policer, an asset policy database and a network discovery module. The traffic policer is a transparent bridge device (can be implemented using a low-cost hardware device such as Raspberry Pi) that is deployed closer to the edge and controls the traffic flowing through it. The policy database contains fine-grained policies that are used by the traffic policer to make control decisions on the packets originating from or destined for the edge. The network discovery module updates the information about various devices discovered on the edge network. The policy database contains device specific policies as well as generic device policies which are applied when devices are discovered. The administrators can add fine-grained policies as required. The aforementioned approaches are summarized in Table 12.

5) ISSUES WITH REALISATION OF EFFECTIVE MICRO-SEGMENTATION

The main challenge associated with using micro-segmentation for ZTA is that the complex workload interactions which often exist in large networks make it difficult to achieve effective segmentation of applications. In addition, effectively translating workload access requirements into network or application-level access control policies is a challenging task; network-level access restrictions might not prevent malicious activities at the application layer. Furthermore, managing

TABLE 12. Summary of approaches to IoT Micro-segmentation.

Ref	Approach	Function Implemented
[129]	SDN-based IoT framework (with NFV)	Routing, Access Control, Secure IoT communication
[131]	SDN-based (access rights are managed and enforced by deploying separate device proxies per user profile and loading them into containers in the IoT gateway)	Automatic routing of user requests and prevention of unauthorized access request
[132]	Adopts traffic policer, asset policy database and net-work discovery module to enforce granular security policies	Fine-grained policy enforcement

and maintaining various access control policies becomes difficult with constantly changing application requirements and the introduction of new applications which can also result in misconfiguration and errors. For these reasons, dynamic workflow access detection techniques and access control policies which can dynamically identify workflow interactions and update access policies accordingly are very much needed.

B. SOFTWARE DEFINED PERIMETERS

Another approach proposed by NIST for implementing ZTA is the use of a software defined perimeter (SDP) which acts as an overlay network to secure resource access [3]. The main principle of SDP is to verify and authenticate the client's identity before communication is established with the client [133]. This is in contrast to traditional networks which allow the client to establish a connection (e.g. TCP/IP) before authenticating. The SDP implementations consist of an SDP controller (which authenticates and authorises the clients) and an SDP gateway (which connects to the applications). The client has no visibility over the application servers and the client's communication to application servers is authorised and facilitated by the controller and the gateway (as shown in Figure 11).

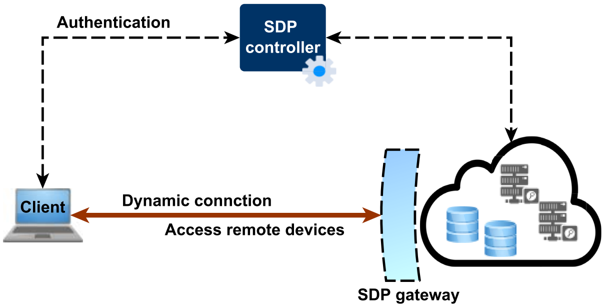


FIGURE 11. A typical SDP architecture, adopted from [133].

The SDP architecture relies on five layers of security to protect client-to-application-server communication [134]. The five layers are:

- **Single Packet Authentication (SPA)**: This is a passive authentication technique which allows legitimate clients to connect securely to the application servers. With the implementation of SPA, the SDP controller listens for connections arriving on closed ports and does not respond to any connection requests. It thus remains obscure to the kinds of port scan tools typically used in reconnaissance attacks. Only for clients that send a valid SPA packet, the controller authenticates and verifies the client before access is allowed to the service. One drawback of the SPA process is the lack of a link between the authentication phase and the following TCP connection establishment process [135], which can result in session hijacking attacks.
- **Mutual Transport Layer Security (mTLS)**: A strict mutual authentication mechanism is enforced in communication between the device/client to the server. In this scheme both server and client need to provide valid identities to authenticate, as opposed to only the server identifying itself. This ensures only valid clients can connect to the back-end resources upon being authenticated and authorised. In resource constrained systems, such encryption-based authentication schemes are challenging to implement due to the computation overhead involved. Several light-weight mutual authentication schemes are discussed in Section V-A.
- **Device Validation**: Beyond merely matching the cryptographic identities presented by users or devices to corresponding rights under policy, the legitimacy of the entity behind the identity must be confirmed. As certain device identities can mimicked or stolen, spoof-resistant device identities need to be used for validating IoT and Industrial IoT (IIoT) devices. Authors in [136], have proposed a unified definition of device identities that can be validated and are spoof-resistant, which are based on existing protocols and device resources currently available. The authors argue that device information from the physical and data link layers of the Open Systems Interconnection (OSI) model are more resistant to spoofing and can be leveraged to create unique device signatures. Information extracted from RF waveforms, device hardware, frame inter-arrival times

from the Medium Access Control (MAC) layer, details from Bluetooth Low Energy (BLE) protocol stack can be effectively extracted to create device fingerprints [136]. Section III-D contains further elaborations on various device identities that can be leveraged and implemented in the SDP.

- **Dynamic Firewalls**: SDP implements dynamic firewalls that contain explicit deny rules, instead of multiple static access rules, to deny all incoming connections. Dynamic firewall rules are created once the previous authentication steps are completed.
- **Binding secure Tunnels to Applications**: This step ensures the communication between the device to application back-end is encrypted, ensuring that the communication between the client and back-end server is protected from various communication channel attacks.

Issues With SDP: These comprehensive authentication and authorisation steps along with encrypted tunnels makes SDP a potential candidate for implementing ZTA in IoT or IIoT networks [137] without heavily depending on perimeter firewalls and network segmentation. SDP will also enable securing existing communications based on IoT protocols such as MQTT which lack support for end-to-end encryption [134]. Even though SDP provides increased security to networks, there exist challenges which need to be overcome for successful implementation. One of the challenges of SDP is that it requires comprehensive changes to the network because it differs significantly from traditional networking practices. Clients as well as servers require modifications to make them work with SDP requirements. Furthermore, the central SDP controller can become a target for malicious cyber attacks which can adversely impact the SDP-based network [134].

VII. SECURITY AUTOMATION AND ORCHESTRATION

Security automation is one of the important principles that can help with the successful realization of ZT security. Security automation can abstractly be defined as a process that aims to curtail frequent mediation by security professionals through the automated detection and prevention of threats. Orthodox security logging approaches typically record a large volume of information that is subsequently used for generating alerts to the security operation teams, who effectively investigate the perceived threats. Often, these alerts are repetitive and include a large number of false positives, resulting in a waste of time and resources on inconsequential analysis. ML techniques can be adopted in such scenarios to support security technology's automatic detection of anomalies and determination of appropriate courses of action in light of the threats and vulnerabilities. This results in quick and seamless actions which help the security teams to focus on the threats at hand, as most of the repetitive threats and false positives are automatically taken care of. In the context of ZTA implementation, security automation focuses on the process of automating access decisions, re-evaluating trust in existing connections, and refining policy generation and enforcement using threat intelligence feeds, situation awareness, network activity logs, and system activity logs. The state-of-the-art approaches to accomplishing these individual components of security automation are presented below before making the final comment on challenges in automating and orchestrating this process for realization of ZTA.

A. THREAT INTELLIGENCE

To prevent catastrophic consequences in a highly connected world comprising a network of IoT devices, organisations have the responsibility of ensuring quick and effective reactions to cyber threats. In order to achieve this, it is necessary to collect and analyse information from various internal and external sources. Such information will refer to threats, vulnerabilities and cyber attacks, and provides the requisite data to enable *Threat Intelligence* (TI). Thus, TI information will provide organisations the latest information regarding cyber threats and will also govern the techniques for countering the threats. With numerous sources of such information, an effective feedback system that automates security control is required to enable the necessary actions.

Threat intelligence involves information gathering on existing and future threats, with the ultimate goal of preventing and mitigating such threats. However, it also involves the dissemination of threat-related information to support decision-making about what proactive security measures are appropriate for the effective management of existent and emerging threats. The authors in [138], proposed a conceptual design to integrate disparate cyber security domains of threat detection and policy controlled systems to automate the task of threat response. The authors argue that effective security countermeasures require risk-based security policies that use contextual information to calculate risk or threat level. However, these risk-based policies pose challenges due to the nature of the contextual information collected, as information local to the system does not provide sufficient information about threats emerging from outside the system. In a Cyber-Physical System (CPS) or an IoT network based critical infrastructure (CI), there exist many challenges for gathering threat intelligence, which include use of heterogeneous devices, standards, and protocols; various layers of data sources (physical, Fog, and Cloud); and the generation of large volumes of data. In order to overcome these challenges, the authors in [139], proposed a novel threat intelligence scheme for Industry 4.0 systems based on Beta Mixture and Hidden Markov Models (HMM). They proposed a threat intelligence architecture comprising smart data management, data pre-processing, and threat discovery modules. The smart data management module collects data from heterogeneous sources such as logs obtained from middleware platforms that connect sensors and actuators, as well as network activity obtained from intrusion detection systems. The evaluation is carried out on publicly available CPS-based power system logs. The pre-processing involves independent component analysis for reducing the dimensionality of features, and the threat discovery uses a Beta Mixture Model that fits the multivariate time series obtained from CPS and network traffic. The output is then fed to the HMM which learns the legitimate and suspicious states of CPS and network activity. Instead of just relying on physical layer parameters of CPS, authors in [140] have proposed a cyber threat intelligence framework which combines information from monitoring at the cyber and physical layers feed actionable countermeasures to the CPS. The proposed approach uses analysis of active malware samples and network intrusions on a CPS honeypot to gather cyber-layer threats. In addition, the physical layer measurements observe the state of physical processes, operation states and physical plant parameters. The CPS threat detector combines the malware attack signatures from the cyber-layer with the CPS data flows from the physical layer using semantic behavioural graphs. As these graphs can become complex with complex network activity, a subgraph approach is utilised to extract a compact representation of the threat activity. In [141], the authors proposed an actionable threat intelligence framework to implement a security response which incorporates feeds received from various sources. The proposed model uses Structured Threat Information eXpression (STIX) documentation, which structures the threat information received from various TI sources to feed a threat response system. The TI document contains the details of the cyber incident, indicators of compromise, and suggested action points. The TRM then communicates the possible actions to various security enforcement tools including host security endpoints, SDN controllers and DNS sinkholes. The authors used extensible Messaging and Presence Protocol (XMPP) as the communication protocol between the TI sources and TRM so as to leverage the scalability and security features of the protocol, with the ultimate goal of securing the communication channel.

B. DEVICE STATE MONITORING

Internally, threat intelligence data can be gathered from various resources that form the network, which can also be referred to as Technical Threat Intelligence (TTI) [142]. For effective policy enforcement and preventing compromised devices from accessing the existing resources and to mitigate the lateral movement of attacks, the TTI system needs to be acquainted with various factors that indicate a compromise, also referred to as Indicators of Compromise (IOC) [142]. The two main categories of resources from which technical information can be obtained are:

- *Network*: these indicators can be obtained from the network activities, including features such as IP addresses, domain names, or URLs. The network-based indicators have short lifetimes and may contain spoofed identities, as adversaries constantly change such parameters.
- *Host-Based*: these indicators can be obtained from the Operating System and from software artifacts such as the hashes of malware binaries, Dynamic Link Libraries (DLL), registry keys, etc.

IOCs are essential artifacts that help in early detection of attacks as well as the implementation of mitigation policies. However, most IOCs are defined by cybersecurity experts and, given the growing sizes of networks and the numbers of connected devices they comprise, tasks at this level quickly become time-consuming and labour-intensive. To overcome these challenges, the authors in [143] have proposed automatic identification of IOCs using neural-based sequence modelling of data obtained from various cyber-security reports. The bi-directional long short-term memory model with conditional random fields was applied in accruing IOCs from cyber-security reports obtained from news articles and patient notes. This technique can be implemented in a threat intelligence network to gather IOC, so that security policies can subsequently be implemented with great effect. However, in IoT networks, obtaining device states requires collecting information such as resource usage, device activation state, firmware versions, hardware and application states [144]. Recently, Manufacturer Usage Description (MUD) data has been proposed to define the intended behaviour of IoT devices to enforce behaviour-based access control [145]. A typical MUD profile contains inbound and outbound access control entries that specify the intended behaviour of the IoT device in question. The authors in [146] have proposed an approach to automatically generate the MUD profiles of IoT devices based on its network traces. MUD profiles can be effectively used to enforce security compliance and also to identify anomalous device behaviour by comparing current behaviour data with pre-existing default MUD profiles. An example network access profile of an IoT device will consist of Domain Name Service (DNS) resolution of Fully Qualified Domain Names (FQDN) of the application servers to which the device is pre-configured to connect, as well as data about the actual connection request made, based on TCP or UDP on specific ports to the application. There are however challenges in using MUD-based access policies, as many commercial IoT devices such as Amazon Echo may connect to a wide range of IP addresses, as validated in [146]. Furthermore, in constrained IoT devices as well as in industrial control system devices such as embedded devices and programmable logic controller systems, there can be limited on-device hardware resources to store logs and events that can be necessary for state monitoring. The heterogeneity of protocols and applications implemented by such devices increases the challenges in acquiring such security-related information.

C. SECURITY SITUATION AWARENESS

As the concept of ZTA gains acceptance in various domains, effective implementation of the same requires integrating network and system state into the policy decision framework. This would enable a policy engine to make situation-aware access decisions. Situation awareness with regard to cyber security threats can be gathered from both internal and external networks to effectively calculate the potential risks to the existing resources. In addition, the resource access policies in ZTA depend on the level of trust of the users and devices rather than the location of the requester in the network which requires greater contextual information such as authentication logs, device states and network activity [147]. Access policies also need to be dynamically enforced and constantly re-evaluated [3], which facilitates effective decision making based on various data sources [147].

In [147], the authors present a policy management framework for ZTA referred to as FURZE (Fuzzy Risk Framework for ZT Networks), which incorporates the enterprise's security situation awareness (SSA) into the Risk-Adaptable Access Control (RAdAC) scheme to facilitate dynamic access policy application reflecting changes in network security settings. The SSA provides information regarding the current state of threats to assets and the criticality of those assets, which collectively indicate the potential impact on enterprise missions. One advantage of the proposed framework is that the RAdAC scheme adapts a dynamic and probabilistic calculation of risk score to make access control decisions instead of a static comparison of context-based system attributes.

ICS have received considerable attention in recent times with regard to cyber situational awareness, owing to repercussions of the Stuxnet malware, according to the review conducted in [148]. Simulated environments have been deployed to show the effectiveness of the different kinds of information sources in CI such as those that render measurements of voltage and current waveforms in power grids to field sensors measurements, to help identify the potential impact of a cyber threat on the operations of the CI.

In [149], the authors proposed a Big Data analysis based situation awareness architecture for smart grids, which incorporates information obtained from electrical devices, substation buses, network devices, station controllers, and control centers, to calculate the awareness scores of the various threats in the smart grids. The awareness score is calculated based on a neural network and game theory-based big data analysis. Such awareness scores can be incorporated into a RAdAC scheme proposed in [147] to dynamically deploy access policies in CIs such as smart grids.

However, challenges such as heterogeneous devices and systems, responding quickly to emerging threats, and interoperability issues in the deployed security tools, impedes the implementation/adoption of a comprehensive automated security system [150]. In order to alleviate this challenge, authors in [150] have proposed a SIEM-based security automation scheme which relies on the near real-time data collected from various existing resources of a network. SIEM provides various advantages which can be leveraged to automate the operations of a security control. The centralised logging and analysis feature of various network and device activities can provide greater visibility over the security state of the system being monitored. Correlated events collected by the SIEM system can be used to identify the root causes of security issues and for automating and applying effective mitigation controls at the most apt locations in the network. Furthermore, Machine Learning (ML) algorithms can be implemented to reduce false alarms and also to increase the effectiveness of the various security controls. Integration of various security tools is another advantage of SIEM solutions, as these can be plugged in to and adapted to various security platforms.

D. MACHINE LEARNING FOR SECURITY AUTOMATION

With a large volume of security logs being collected, ML can act as an enabler for security automation by understanding the behavior of security threats and recognizing the patterns to automatically take defensive actions as highlighted in [151]. AI has not been explored to its fullest potential for realizing an effective security automation procedure [151]. ML has shown promise in a number of critical scenarios for threat detection and prevention. For example, SDNs are being increasingly adopted for automated security monitoring in cloud computing platforms. As described in [152], SDN platforms are more susceptible to threats because of the separation that exists between their control and data planes. Conventional approaches for preventing DoS attacks in SDNs are not helpful, as they largely depend upon the flow characteristics of the packets, allowing the attackers to deceive the system with slight modifications to the traffic patterns (e.g., through changing packet headers to make it appear like legitimate traffic) [152]. Given this, ML approaches are deployed in SDN infrastructures to trace malicious traffic. For example, authors in [153] adopted SVM to detect DDoS in SDNs. The scheme extracted features from the traffic flow, including number of packets, bytes, packet variation, duration, etc. Through the utilisation of these features, the system is trained to serve as a classifier to classify the traffic as being normal or anomalous. Likewise, the authors of [154] adopted support vector machines (SVM) to detect anomalous traffic so as to prevent DDoS attacks in SDNs by leveraging the extracted features such as the standard deviation of the packet, bytes and flow entries. Similarly, a number of works have demonstrated the efficacy of deep-learning for intrusion detection. For example, the authors of [155] have adopted stacked-autoencoders in combination with SVM and Artificial Neural Network (ANN) for detecting impersonation attacks in Wi-Fi networks. Similarly, the authors of [156] argued that as novel cyber threats appear frequently, thereby making it difficult for conventional detection systems that rely on prior modeling of the attacks to detect them. To overcome this issue, they argued that deep-learning can help to successfully identify such novel attacks in social networks by leveraging their similarities to known attacks. This is achievable as deep-learning comprises of numerous layers of non-linear processing elements to learn the precise features that may help in attack detection in social networks, through sophisticated training on diverse mathematical models representative of the cyber threat data.

The authors of [156] demonstrated that long short-term memory recurrent neural network helps in the identification of anomalous traffic accurately in social networks as compared to other methods that rely on traditional ML algorithms.

Likewise, the authors of [157] demonstrated the efficacy of convolutional neural networks (CNNs) for attack detection. Similarly, ML (decision tree, SVM, MP) has also been shown to be useful toward detecting attacks in IoT networks [158]. The authors of [158] ran experiments on a dataset that consisted of network traffic obtained from nine IoT devices that were impacted by Mira and BASHLITE botnets. They extracted the statistical features from this data and after performing some appropriate data processing, they trained three different conventional ML algorithms and demonstrated their efficacy in accurately classifying legitimate and malicious traffic. Likewise, authors in [159] also demonstrated the effectiveness of the Random Forest algorithm for detecting different attacks in IoT such as DoS, Data type probing, and other malicious activities. Authors in [160] proposed an entropy-based method for detecting DDoS in IoT based on a stateful SDN data plane. They demonstrated that the entropy of different features such as source and destination IP addresses and ports changes significantly in the case of DoS and DDoS attacks in IoT. They also illustrated the ability of SDN to mitigate these attacks by simply adding entries to the flow tables of network switches. Likewise, the authors in [161] adopted deep learning for detecting various attacks in IoT devices. Precisely, they extracted generic features from headers of the individual packets and subsequently applied a feed-forward neural network for detecting four different types of attacks, namely, DoS, DDoS, reconnaissance, and information theft. Similarly, the authors in [162] also showed that deep-learning can be employed for online detection of network attacks on IoT gateways.

The aforementioned examples of ML application for attack detection specifically in SDN based networks can be effectively deployed as a feedback mechanism for SDN controllers which can subsequently enforce security policies on SDN-compatible network devices. An example of such a feedback system to protect IoT systems is proposed in [163], which consists of a mitigation module that takes input from the IoT communication path and feeds the output generated by the IoT device back to a detection module, which subsequently detects malicious output through application of ML techniques. The mitigation module is recommended to be a polymorphic hardware and software component which can adapt to the system security requirements. An SDN-based system can be adopted as a mitigation module which can be quickly reconfigured for meeting the stipulated security requirements.

E. POLICY ENFORCEMENT

In a ZTA, the contextual and threat intelligence feeds are leveraged to control access to the enterprise resources. As indicated in Section I, the policy decisions are taken by a PE and enforcement of policy is done by a PEP. A few representative policy enforcement frameworks for IoT networks are discussed below.

Automating access policy definition and enforcement in an IoT environment, where a large number of devices are connected to the network, is a challenging task. In addition, groups of IoT devices might work collectively for a specific workflow, which increases the complexity of manually implementing fine-grained access controls for M2M communications [164]. In consumer environments, many end-users will also lack the skills to manually configure such access policies. The heterogeneity in IoT devices and protocols will exacerbate the security policy enforcement. In such circumstances, SDN and Network Function Virtualisation (NFV) technologies can be used to orchestrate and enforce security closer to the edge. NFV is based on virtualisation technology which allows organisations to virtualise traditional hardware based network functions such as routing, switching, firewalls and many more and deploy them on commodity hardware by sharing computing resources. SDN and NFV have also become key enablers for the deployment of 5th Generation telecommunication networks which supports superior network performances to deal with billions of devices expected to be connected to the Internet [165]. Using SDN and NFV, dynamic security policies and virtual network security functions can be applied to control access to end resources [166]. Using these technologies, virtual network security functions such as virtual Firewall (vFirewall), virtual Intrusion Detection System (vIDS) and virtual Authentication and Authorization services can be quickly deployed to protect the resources. Such network security orchestrations can be achieved by combining NFV and SDN where NFV allows rapid deployment of network security services and SDN paves the way for a quicker enabling of network connectivity to network function virtualisation infrastructure (NFVI). Furthermore, to bring the NFV closer to the edge, container based network security virtualisation has been proposed wherein security functions are deployed in light-weight containers which require fewer resources compared to a virtual machine, and can be run on a single operating system [167]. With advancements in virtualisation technology, unikernels have now been proposed for virtual network function instantiation. Compared to containers, unikernels don't share the same kernel and are therefore more secure than containers running on shared kernel space [168]. However, containers seem to have better performance in terms of instantiation delays and TCP performance than unikernels, as highlighted in [169].

These technologies have an important role in realizing the ZTA model of security to deploy security functions closer to the resources. This requires PE and PEP to be SDN and NFV compatible and should be able to automate the deployment of virtual security functions in a reactive manner. In [170], authors deployed a machine learning based security orchestration function which uses monitoring agents to identify IoT attacks and dynamic reaction module that deploys a virtual security function to secure the IoT devices. In [171] authors presented a security management architecture for a NFV/SDN aware IoT systems called ANASTACIA. This framework leverages a security enforcement plane composed of NFV Management and Orchestration (MANO), SDN and IoT controllers to manage resources (computing, storage and network) for IoT devices interacting in the data plane. The framework adopts a two tier policy enforcement technique where a default preventive security policy is applied to secure IoT devices from known attacks and a reactive policy enforcement is used as an active countermeasure to prevent ongoing attack. When an attack is detected and virtual network function is deployed using NFV, SDN and IoT controllers in appropriate locations in the network. Instead of users manually selecting devices, the authors in [164] have proposed an automatic scheme to abstract the workflows and automatically select suitable IoT devices required to achieve the policy management task, which can also facilitate the specification of accurate access policies in M2M communications as well as help enforce least privilege access. The authors use search algorithms to select the most suitable devices for a particular workflow and to generate network access policies that can be implemented through SDN-based devices. In [172], the authors proposed an access control policy enforcement for zero trust networks that requires dynamic and per-connection access control. The proposed framework called Fuzzy Risk Framework (FUZRE) is based on RAdAC that makes access control decisions considering operational need and security risk posed by an access request. The FZURE architecture uses XACML, which defines the structure for an attribute based access control implementation. It comprises a PEP such as a WiFi router, which forwards the access requests to a context handler. The context handler subsequently communicates with various policy decision making modules; the environment evaluation, risk evaluation, topology awareness, access decision modules. The FZURE framework also incorporates a continuous evaluation of the access to adapt or re-evaluate access policies to changing network situations based on the balance between operational needs and posed security risks. The risk evaluation module is based on a variant of fuzzy logic that works well with subjective and vague inputs to make clear policy decisions. IoT-IDM (Intrusion Detection and Mitigation for IoT) [173], is a host-based intrusion detection and mitigation framework for smart home systems realised using the SDN OpenFlow platform. The framework consists of a device manager, sensor element, feature extraction module, detection and mitigation modules. The device manager contains the inventory of all the devices in the smart home network, their potential vulnerabilities, and any possible defence mechanisms in play. The database is based on public repositories that contain datasheets of available and new devices for smart home systems. The sensor element uses an inline sensor deployed in the SDN controller to monitor traffic of registered devices on the network. OpenFlow-based network switches are then configured to redirect the IoT device traffic to the sensor element. Based on the traffic captured by the sensor element, features are then extracted from the traffic. The detection module is responsible to detect anomalous behaviour of devices through the implementation of ML algorithms, and helps to identify the attack source specifications, which are then adopted to configure access rules on the OpenFlow switches through the mitigation module.

The Adaptive Risk-Based Access Control Model for IoT proposed in [174] adopts a risk estimation process depending on the IoT environment features, to calculate the risk associated with an access request. The risk estimation algorithm considers the context information of the user, the sensitivity of the resource being accessed, the severity of the requested action, and past risk scores associated with the user, in order to estimate risk. The quantified risk is calculated by the likelihood of a security incident to occur and the potential impact of such an incident. In addition, an adaptive approach is incorporated to evaluate risk continuously through the session. The resource sensitivity module assigns a risk metric associated with individual resources based on the sensitivity of the data or resource being requested. The risk estimator is responsible for analysing all the possible risk features to estimate the risk of granting resource access, whilst adhering to policy decisions. For continuous evaluation of user behaviour, the authors propose the use of smart contracts to ensure that the terms of the contract are adhered to throughout a given communication session.

F. TRUST COMPUTATION

As indicated in Section II-A, a trust algorithm is considered to be the cornerstone of the PE. Since there is no direct literature available relating to trust computation for enterprise networks or CI, we describe a few trust computing techniques employed in the AdHoc IoT networks which can be extended for trust calculation in CI. For example, the authors in [175] proposed a trust-based access control and management technique for fog computing platforms. For computing the trust, they use a score-based method which assigns different weights to a user's historic behavior, the type of device used, and its location, and then compares the computed score against a threshold, thereby making it a contextual score-based algorithm, as discussed in Section I. Likewise, the authors in [176] also proposed a weight-based trust evaluation technique that makes use of information entropy and demonstrates its appropriateness as compared to traditional weight-based methods. In [177], the authors applied a Fuzzy-logic in Trust Based Access Control (FTBAC) mechanism, which incorporates contextual experience, knowledge and recommendations about the device to calculate trust. According to the authors, the access control is directly proportional to the trust established for the device requesting access.

In [178] the authors present a trust management system for securing the data plane of an ad-hoc network. The proposed method incorporates a fuzzy logic and graph theory based trust calculation model which allows individual nodes to calculate trust scores for all the other nodes in the network. The authors use average delay (AD) and packet delivery ratio (PDR) as the statistical parameters for trust calculation, as data plane attacks can affect the delay of critical packets or even target the successful delivery of packets. The PDR and the AD are calculated based on the packet sequence numbers and the packet timestamp, related to the acknowledgement packets received on the path. First, the path trust value is determined and then the node trust value is calculated. The idea is that all the paths that traverse the malicious node will have lower trust scores when compared to the paths that pass through the normal node. Considering the network as a connected graph, matrices of path trust values and vertices of the graph are obtained and flooded on the ad-hoc network by the threat actor, which enables all the participating nodes to calculate the node trust values for the entire network.

Issues in Integration of Individual Components for Trust Computation to Automate the Security Actions: Trust algorithms have been applied in peer-to-peer (p2p) and ad-hoc communications, to take the feedback of neighbouring nodes, central trust controllers and historical transactions. However, a trust algorithm that takes threat intelligence, device security status, contextual information, and cyber activity logs into consideration is lacking. Such a trust algorithm will encounter numerous challenges when compared to the p2p and ad-hoc trust algorithms, as it needs to handle heterogeneous data sources that generate data comprising numerical and imprecise information. The solution proposed in [172] which considers contextual information and situation awareness, feeds in a RAdAC system using XACML access control policy language. However, the proposed work lacks the implementation details of the risk calculating function and the inner workings on how the access control policies are implemented at both the application and network layers.

VIII. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

This section summarises our findings on successful implementation of ZTA, with a particular emphasis on identified knowledge gaps in the various state of the art methodologies.

Discussions outlined in the previous sections indicate that there are numerous challenges in accomplishing zero trust. All the techniques considered essential for achieving zero trust, such as authentication, access-control, encryption, and security automation have shortcomings. In this concluding section of the paper, we pinpoint the weaknesses of these techniques and also identify directions for future work.

A. AUTHENTICATION

Despite the universal deployment and use of authentication technologies, there are still several dimensions of authentication that are yet to be fully realised. Below, we detail some of these as potential *future research directions*.

- *User Authentication*: As indicated in section III-A, almost all user authentication mechanisms such as passwords, fingerprints, facial recognition and iris scans have vulnerabilities. Multi-factor authentication (MFA) is a solution endorsed by many but most of these solutions rely on a secondary-device such as a mobile phone and also demand a significant user interaction for accomplishing the MFA. These requirements deter the widespread usage of MFA solutions. In view of this, solutions that demand minimum participation from users are needed. For example, the proximity of two devices (i.e. a login device and a pre-registered mobile phone) established using the ambient sound recorded by them has been proposed as a potential second-factor in [179]. However, this solution is still reliant upon a secondary device, thereby rendering it useless if the mobile phone is stolen, discharged, or lost. Therefore, MFA solutions that are not dependent upon a secondary device and which require minimal interaction from the user are needed. In view of this, entirely new directions for authentication are being pursued. For example, [180] recently proposed to use the frequency response of the person's ear canal as a distinctive feature for authentication. However, this solution demands an earpiece with a microphone for sending the sound signal into the ear cavity and recording its reflections simultaneously, which will be onerous for users. Similarly, authors in [181] proposed to use the whizzing sound of human breath as a feature for authentication. However, this solution necessitates a deliberate action from the users, i.e. it demands a user to place the microphone very close to the nose and to perform a deep-breath or sniff. This discussion reveals that even the most recent advancements in the field of user authentication come with drawbacks, thereby demanding significant research efforts on user authentication.
- *Continuous User Authentication*: In addition to entry point authentication, continuous (or active) authentication is also an active area of research. However, most continuous authentication mechanisms rely on user's behavioural biometric features associated with typing, tapping, and gait patterns. The problem with this approach is that they rely on sensors embedded in the devices. For continuous authentication across numerous devices (e.g., mobile-phone, laptops, tablets), the reliance upon sensors is an issue, as not all the devices possess resembling embedded sensors. Besides this, behavioral biometrics are also dependent upon the situation in which they are captured. For example, a user's tapping or typing patterns may be different while walking at different speeds. This necessitates a thorough processing and machine-learning pipeline that can extract the discriminating features of a user.
- *Context-Aware User Authentication*: Context-aware authentication is also being frequently adopted for user authentication. We observe that most context-aware authentication mechanisms rely on location information that is concatenated with some other information such as device time or proximity. However, some of the sensors that are deployed for capturing the contextual information (e.g., accelerometer or gyroscope) may not be available on some devices, and thus may not be usable across all sorts of devices. Therefore, to carry out context-aware user authentication, a suitable alternative could be to pair the user's location with his daily activity on the mobile phone and to generate questions from that activity (e.g., calls and message logs). However, this solution

will demand a series of questions by taking into account the user's location information, which may pose privacy concerns for users. Nevertheless, this can be a suitable option for context-aware fallback authentication, which is only rarely required. Identifying the appropriate set of questions and crafting a location-based contextual framework are other open research directions.

- *Device Authentication*: In addition to user authentication, device authentication is also required for M2M communication and CPS. The popular mechanisms as indicated in the literature include Physically Unclonable Functions (PUF) that are built in these devices to generate an unique identifier by leveraging the hardware characteristics of the device. However, the literature suggests that PUFs are vulnerable to modelling attacks to clone the device identifier. This vulnerability can be addressed by using cryptographic methods, which are also widely adopted for mutual authentication in ad hoc and opportunistic networks such as VANETs. However, cryptographic methods are at a greater risk of subversion with advances made in quantum computing. Some recent research on quantum computing reveals that most current cryptosystems could become vulnerable in the future to exhaustive key search attacks. Therefore, to counter these issues, post-quantum cryptography is being actively researched (see section Post-Quantum Cryptography detailed ahead). A suitable alternate for device authentication can be inspired from context-aware authentication, wherein a device can leverage the sensors or other peripherals to capture the information related to its ambience or other parameters to enable authentication.

B. ACCESS CONTROL

Zero Trust necessitates a fine-grained and context-aware access control. The literature suggests that these access control requirements can be accomplished by attributes-based and usage-based access control. However, different IoT-enabled environments such as smart homes, smart grids, healthcare IoT, and smart buildings have entirely different access control requirements, requiring different arrangements of access control components. Most of the current frameworks for the aforementioned scenarios do not identify the appropriate requirements required by the corresponding scenarios, thereby leading to inadequate access control instantiations. Recently, blockchain is also emerging as a candidate for distributed access control. However, blockchain is still immature for use within this domain with its dependence on a consensus mechanism, making it less attractive than traditional centralised systems. Risk-aware access control incorporating capabilities of a fine-grained access control scheme such as Function-Based Access Control (FBAC) is recommended. This will enable evaluation of the risk posed for an access request, by leveraging information derived from diverse sources including threat-intelligence and CPS situational awareness systems, and granting access decisions at a high-level of granularity, as highlighted in [172]. However, such risk evaluation frameworks will face challenges due to the numerous sources of inputs and the volume of contextual-data collected. Similar challenges will be encountered by frameworks that use trust levels to grant or deny access to resources. The main challenges that such frameworks will encounter are summarized below that may be addressed in future:

- Converting numerous sources of threat intelligence and SSA into actionable security policies is challenging due to varying data formats and heterogeneity of devices such as firewalls, IPS, SDN controllers and other network appliances. Interactions between such frameworks and the devices necessitates compatible communication standards. With no specific standards in place, realising such interactions is challenging.
- Contextual and behaviour data can be obtained from various security and logging appliances as well as from network hosts. Combining and correlating these large volumes of heterogeneous data threat identification and risk quantification in a reasonable time is a complex task. Besides, TA will also face challenges when combining different forms of information obtained from TI, SSA, SIEM, contextual and behavioural data sources, to making decisions.
- Risk evaluation frameworks or TA that incorporate security state of devices will face challenges in acquiring such information from constrained IoT and CPS devices used in critical infrastructure.
- Current TAs are designed for adhoc or P2P networks and are not directly applicable for trust evaluation in enterprise networks or CPS with heterogeneous devices and applications since the input parameters for calculating trust in adhoc and P2P networks differ considerably from the parameters that exist in enterprise networks and those specified in a ZTA.
- Current risk based access control techniques do not adequately enforce access control policies at both the network and application level as applying controls only at a particular level (layer) will lead to resources becoming vulnerable to cyber attacks at other protocol layers.
- Adoption of a central risk score or trust evaluation framework will make the central PE node vulnerable to cyber attacks such as DoS attacks. This requires either a robust PE calculation engine or a distributed PE to be in place, which can prevent such attacks.

Countering such challenges requires the use of technologies that enable the processing of large volumes of data and correlation of events from multiple sources to perceive the threat to the system. The recent advancements in deep learning techniques have a potential use in such scenarios where useful intelligence is to be obtained from such data sources.

C. MICRO-SEGMENTATION AND SDP

Micro-segmentation has been proposed as an effective strategy for implementing ZTA, as it allows network perimeters to be positioned close to the application servers, enabling simpler enforcement of fine-grained access controls. It also enables creating and managing policy enforcement per access group, for preventing unauthorised access to resources within the network perimeter. However, some associated challenges in implementing micro-segmentation are described below that may be addressed in future:

- The effective segmentation of applications is difficult due to the complex workflow that exists in the interactions and dependencies for large networks.
- Legacy, monolithic and non-virtualised applications are not suitable for micro-segmentation.
- Effectively translating workflow access requirements into network or application level access control policies is a challenging task. For example, network level access restrictions might not prevent malicious activities at the application layer. This has been addressed in limited manner through the use of micro-services in granular applications running in container services, as highlighted in [119].
- Managing and maintaining various access control policies is another challenge which is exacerbated through constantly changing application requirements and with the introduction of new applications, thus leading to misconfiguration and errors. In order to counter this issue, dynamic workflow access detection techniques and access control policies that can identify the workflow interactions and update access policies accordingly, are needed.
- As micro-segmentation is only applicable for application level access control, protection of physical devices and resources necessitates the consideration of SDN-based mechanisms such as those proposed in [131] and [132].

In addition to micro-segmentation, SDP has been adopted for implementing ZTA. Even though, SDP provides increased security to networks, challenges exist in successful implementation of ZTA. One of the challenges of SDP is that it requires comprehensive changes to the network as these networks differ significantly from traditional networking practices. In addition, clients and servers require major modifications in the communication process to be compatible with SDP requirements. Furthermore, the central SDP controller can become a target for malicious cyber attacks, which can adversely affect operations of the SDP-based network [134].

D. POST-QUANTUM CRYPTOGRAPHY

Recently, quantum-computing is progressing at a rapid pace. However, with all the anticipated benefits of quantum-computing, there are some issues that require careful thought. For example, quantum computers are likely to subvert current cryptographic defences such as RSA and ECC. In [182] it was demonstrated that 2048-bit RSA can be easily cracked in around eight hours by a quantum computer consisting of 20 million qbits. Although quantum computers of such enormous power are currently non-existent (i.e. only 128 qubits presently), they are likely to have such potential in the future. Similarly, the authors of [183] showed that almost all of the current public-key mechanisms can be broken by Shor's algorithms [184]. Similarly, for hash functions, Grover's algorithm [185] can facilitate quantum birthday attack to find collision, thereby necessitating an output of larger size. Therefore, to safeguard against such advancements in the foreseeable future, post-quantum cryptography is being actively researched. Post-quantum cryptography is basically a new approach to cryptography that can be implemented on legacy computers–one which can withstand attacks launched by a powerful quantum computer. One of the obvious choices is to increase the length of encryption keys which effectively will hyper-exponentiate the number of permutations that a quantum computer algorithm (e.g. Grover's algorithm) will have to run. Authors in [183] have shown five key approaches for which no known way of applying the Shor's algorithm exists. These approaches are code-based encryption, lattice-based encryption/signature, multivariate-quadratic-equation signatures, and hash-based signatures. However, numerous shortcomings of the aforementioned approaches are mentioned in [183]. The standardization of post-quantum cryptosystems is currently underway at NIST [186]. The candidates (for encryption and digital signatures) included as finalists in Round 3 of NIST's standardization project are summarized below in Table 13 (interested readers are referred to [187] for more details). However, even without standardization, Google recently conducted some experiments with Lattice-based cryptosystems and demonstrated their appropriateness in large scale encryption cracking applications [188], [189]. As pointed out in [183], post-quantum cryptosystems still need varying designs alongside techniques for optimization and implementation. Besides, attack analysis of currently available approaches is also somewhat limited and their incorporation into existing (legacy) systems and applications is challenging and is expected to take more time [190]. All of these dimensions are active research areas which need thorough investigation for continued and sustained realization of zero trust in the future.

TABLE 13. NIST round 3 finalist candidates for PQC.

Proposal	For	Approach
Classic McEliece	Enc	Code-Based
Crytals-Kyber	Enc	Lattice-Based
NTRU	Enc	Lattice-Based
Saber	Enc	Lattice-Based
Crystals-Dilithium	Sig	Lattice-Based
Falcon	Sig	Lattice-Based
Rainbow	Sig	Multivariate-Based

IX. CONCLUSION

This paper presents a comprehensive description of the new security paradigm, zero trust. We present the basic tenets of ZTA along with its logical components and an analysis of a range of implementation techniques. As there is no single technology or architecture that can successfully implement a ZTA, the paper reviews various techniques and approaches that have been identified as essential to realise its adoption. Based on this, we highlight the role of authentication and access control techniques that take into consideration an organisation's context, behaviour and perceived threats, so as to constantly re-evaluate the trust in ongoing connections for a successful realization of ZTA. In addition, encryption, micro-segmentation, and software defined perimeters are also essential components in realising a comprehensive ZTA. The paper articulates state-of-the-art approaches to instantiate the identified security techniques for adoption in various deployment scenarios and finally concludes with a description of various challenges that are posed by contemporary authentication mechanisms, access control schemes, trust and risk computation techniques, micro-segmentation approaches, and SDP. The identified challenges may serve as potential future research directions for instantiating ZTA in its true spirit.

REFERENCES

[1] X. Yan and H. Wang, *Survey on Zero-Trust Network Security*. Singapore: Springer, 2020.

[2] B. Embrey, ''The top three factors driving zero trust adoption,'' *Comput. Fraud Secur.*, vol. 2020, no. 9, pp. 13–15, Jan. 2020.

[3] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, ''Zero trust architecture,'' NIST, Gaithersburg, MD, USA, Tech. Rep. NIST 800-207, 2019.

[4] C. Buck, C. Olenberger, A. Schweizer, F. Völter, and T. Eymann, ''Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust,'' *Comput. Secur.*, vol. 110, Nov. 2021, Art. no. 102436, doi: 10.1016/j.cose.2021.102436.

[5] S. Teerakanok, T. Uehara, and A. Inomata, ''Migrating to zero trust architecture: Reviews and challenges,'' *Secur. Commun. Netw.*, vol. 2021, pp. 1–10, May 2021.

[6] C. Cunningham, J. Blankenship, S. Balaouras, R. Murphy, M. Cyr, and P. Dostie, ''The zero trust eXtended (ZTX) ecosystem,'' Forrester, Cambridge, Ma, USA, Tech. Rep., 2018.

[7] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, ''When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2016, pp. 1068–1079.

[8] A. Kelly, ''Cracking passwords using keyboard acoustics and language modeling,'' M.S. thesis, School Inform., Univ. Edinburgh, Edinburgh, U.K., 2010.

[9] S. Shah and S. Kanhere, ''Recent trends in user authentication—A survey,'' *IEEE Access*, vol. 7, pp. 112505–112519, 2019.

[10] T. Brewster. (2018). *We Broke Into a Bunch of Android Phones With a 3D-Printed Head*. [Online]. Available: https://www.forbes.com

[11] D. Goodin. (2016). *Breaking the Iris Scanner Locking Samsung's Galaxy S8 is Laughably Easy*. [Online]. Available: https://arstechnica.com/

[12] S. W. Shah, S. S. Kanhere, J. Zhang, and L. Yao, ''VID: Human identification through vein patterns captured from commodity depth cameras,'' *IET Biometrics*, vol. 10, no. 2, pp. 142–162, Mar. 2021.

[13] S. W. Ali Shah, A. Shaghaghi, S. S. Kanhere, J. Zhang, A. Anwar, and R. Doss, ''Echo-ID: Smart user identification leveraging inaudible sound signals,'' *IEEE Access*, vol. 8, pp. 194508–194522, 2020.

[14] RSA. (2020). *RSA SecureID*. [Online]. Available: https://www.rsa.com

[15] Yubico. (2020) *YubiKeys*. [Online]. Available: https://www.yubico.com

[16] S. W. A. Shah and S. Kanhere, ''Wi-auth: WiFi based second factor user authentication,'' in *Proc. 14th EAI Int. Conf. Mobile Ubiquitous Systems: Comput., Netw. Services*, New York, NY, USA, 2018, pp. 393–402.

[17] S. W. A. Shah, J. J. Jeong, and R. Doss. (2021). *How Hackers Can Use Message Mirroring Apps to See All Your SMS texts-and Bypass 2FA Security*. [Online]. Available: https://theconversation.com/how-hackers-can-use-message-mirroring-apps-to-see-all-your-sms-texts-and-bypass-2fa-security-165817

[18] S. W. Shah and S. S. Kanhere, ''Wi-sign: Device-free second factor user authentication,'' in *Proc. 15th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, New York, NY, USA, 2018, pp. 135–144.

[19] S. W. Shah and S. S. Kanhere, ''Wi-access: Second factor user authentication leveraging WiFi signals,'' in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2018, pp. 330–335.

[20] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, ''Towards a better understanding of context and context-awareness,'' in *Proc. CHI Workshop What, Who, When, How Context-Awareness*, 2000, pp. 304–307.

[21] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley, ''CASA: Context-aware scalable authentication,'' in *Proc. 9th Symp. Usable Privacy Secur. (SOUPS)*, 2013, pp. 1–10.

[22] S. Buthpitiya, Y. Zhang, A. K. Dey, and M. Griss, ''n-gram geo-trace modeling,'' in *Proc. 9th Int. Conf. Pervasive Comput.*, 2011, pp. 97–114.

[23] M. Jakobsson, ''Implicit authentication for mobile devices,'' in *Proc. 4th USENIX Conf. Hot Topics Secur.*, 2009, pp. 25–27.

[24] K. Benzekki, A. El Fergougui, and A. ElBelrhiti ElAlaoui, ''A context-aware authentication system for mobile cloud computing,'' *Proc. Comput. Sci.*, vol. 127, pp. 379–387, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050918301479

[25] S. H. Kim, D. Choi, S. H. Kim, S. Cho, and K. S. Lim, ''Context-aware multimodal FIDO authenticator for sustainable IT services,'' *Sustainability*, vol. 10, no. 5, p. 1656, 2018.

[26] Y. Ashibani, D. Kauling, and Q. Mahmoud, ''Design and implementation of a contextual-based continuous authentication framework for smart homes,'' *Appl. Syst. Innov.*, vol. 2, no. 1, p. 4, Jan. 2019.

[27] K. Olejnik, I. Dacosta, J. S. Machado, K. Huguenin, M. E. Khan, and J.-P. Hubaux, ''SmarPer: Context-aware and automatic runtime-permissions for mobile devices,'' in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 1058–1076.

[28] P. Gupta, T. K. Wee, N. Ramasubbu, D. Lo, D. Gao, and R. K. Balan, ''HuMan: Creating memorable fingerprints of mobile users,'' in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2012, pp. 479–482.

[29] S. K. Dandapat, S. pradhan, B. Mitra, R. R. Choudhury, and N. Ganguly, ''ActivPass: Your daily activity is your password,'' in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 2325–2334.

[30] K. Niinuma, U. Park, and A. K. Jain, ''Soft biometric traits for continuous user authentication,'' *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 771–780, Dec. 2010.

[31] J. Roth, X. Liu, and D. Metaxas, ''On continuous user authentication via typing behavior,'' *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4611–4624, Oct. 2014.

[32] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, ''Continuous user authentication using multi-modal biometrics,'' *Comput. Secur.*, vol. 53, pp. 234–246, Sep. 2015.

[33] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, ''Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication,'' *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 136–148, Jan. 2013.

[34] L. Li, X. Zhao, and G. Xue, ''Unobservable re-authentication for smartphones,'' in *Proc. 20th Netw. Distrib. Syst. Secur. Symp.*, 2014, pp. 1–16.

[35] C. Nickel, C. Busch, S. Rangarajan, and M. Mobius, ''Using hidden Markov models for accelerometer-based biometric gait recognition,'' in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl.*, Mar. 2011, pp. 58–63.

[36] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. Makela, and H. Ailisto, ''Identifying users of portable devices from gait pattern with accelerometers,'' in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, pp. ii/973–ii/976.

[37] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides, ''Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics,'' in *Proc. IEEE 5th Int. Conf. Biometrics: Theory, Appl. Syst. (BTAS)*, Sep. 2012, pp. 8–15.

[38] H. M. Thang, V. Q. Viet, N. Dinh Thuc, and D. Choi, ''Gait identification using accelerometer on mobile phone,'' in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*, Nov. 2012, pp. 344–348.

[39] A. Almehmadi and K. El-Khatib, "On the possibility of insider threat prevention using intent-based access control (IBAC)," *IEEE Syst. J.*, vol. 11, no. 2, pp. 373–384, Jun. 2017.

[40] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1DMRLBP) in ECG biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2818–2832, Dec. 2016.

[41] K. PG and M. S. Holi, "Electromyography analysis for person identification," *Int. J. Biometrics Bioinf. (IJBB)*, vol. 5, no. 3, p. 172, 2011.

[42] A. Mosenia, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, "CABA: Continuous authentication based on bioaura," *IEEE Trans. Comput.*, vol. 66, no. 5, pp. 759–772, May 2017.

[43] M. Shahzad and M. P. Singh, "Continuous authentication and authorization for the Internet of Things," *IEEE Internet Comput.*, vol. 21, no. 2, pp. 86–90, Mar./Apr. 2017.

[44] S. W. Shah and S. S. Kanhere, "Smart user identification using cardiopulmonary activity," *Pervas. Mobile Comput.*, vol. 58, Aug. 2019, Art. no. 101024.

[45] I. Ali, S. Sabir, and Z. Ullah, "Internet of Things security, device authentication and access control: A review," 2019, *arXiv:1901.07309*.

[46] K.-Y. Lam and C.-H. Chi, "Identity in the Internet-of-Things (IoT): New challenges and opportunities," in *Information and Communications Security*. Cham, Switzerland: Springer, 2016, pp. 18–26.

[47] A. L. M. Neto, A. L. F. Souza, I. Cunha, M. Nogueira, I. O. Nunes, L. Cotta, and N. Gentille, "AoT: Authentication and access control for the entire IoT device life-cycle," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst. (CD-ROM)*, 2016, pp. 1–15.

[48] U. Chatterjee, R. S. Chakraborty, and D. Mukhopadhyay, "A PUF-based secure communication protocol for IoT," *ACM Trans. Embed. Comput. Syst.*, vol. 16, no. 3, pp. 1–25, Apr. 2017.

[49] Z. Huang and Q. Wang, "A PUF-based unified identity verification framework for secure IoT hardware via device authentication," *World Wide Web*, vol. 23, pp. 1057–1088, Apr. 2019.

[50] A. Kanuparthi, R. Karri, and S. Addepalli, "Hardware and embedded security in the context of Internet of Things," in *Proc. ACM Workshop Secur., Privacy Dependability Cyber Vehicles (CyCAR)*, 2013, pp. 61–64.

[51] A. Braeken, "PUF based authentication protocol for IoT," *Symmetry*, vol. 10, no. 8, p. 352, 2018, doi: 10.3390/sym10080352.

[52] M. Á. Prada-Delgado, I. Baturone, G. Dittmann, J. Jelitto, and A. Kind, "PUF-derived IoT identities in a zero-knowledge protocol for blockchain," *Internet Things*, vol. 9, Mar. 2020, Art. no. 100057.

[53] U. Ruhrmair, J. Sölter, F. Sehnke, and X. Xu, "PUF modeling attacks on simulated and silicon data," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1876–1891, Nov. 2013.

[54] J. R. Wallrabenstein, "Practical and secure IoT device authentication using physical unclonable functions," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud (FiCloud)*, Aug. 2016, pp. 99–106.

[55] D. Chen, N. Zhang, Z. Qin, X. Mao, and Z. Qin, "S2M: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 88–100, Feb. 2017.

[56] S. W. A. Shah, N. F. Syed, A. Shaghaghi, A. Anwar, Z. Baig, and R. Doss, "Towards a lightweight continuous authentication protocol for device-to-device communication," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 1119–1126.

[57] I. ID. (2020). *SRAM PUF Technology by Intrinsic ID*. [Online]. Available: https://www.intrinsic-id.com/sram-puf/

[58] Ericsson. (2020). *IoT Platform*. [Online]. Available: https://www.ericsson.com/en/internet-of-things/iot-security/identity-management

[59] S. W. Shah, N. F. Syed, A. Shaghaghi, A. Anwar, Z. Baig, and R. Doss, "LCDA: Lightweight continuous device-to-device authentication for a zero trust architecture (ZTA)," *Comput. Secur.*, vol. 108, Sep. 2021, Art. no. 102351.

[60] P. Grassi, "Digital identity guidelines," NIST Special Publication, Gaithersburg, MD, USA, Tech. Rep. 800-63B, Jun. 2022.

[61] F. Alliance. *Fido UAF Architectural Overview*. Accessed: Mar. 6, 2022. [Online]. Available: https://fidoalliance.org/specs/fido-uaf-v1.2-rd-20171128/FIDO-UAF-COMPLETE-v1.2-rd-20171128.pdf

[62] R. Feng, X. Xu, X. Zhou, and J. Wan, "A trust evaluation algorithm for wireless sensor networks based on node behaviors and D-S evidence theory," *Sensors*, vol. 11, no. 2, pp. 1345–1360, Jan. 2011.

[63] F. Alliance. *Universal 2nd Factor (U2F) Overview*. Accessed: Mar. 6, 2022. [Online]. Available: https://fidoalliance.org/specs/fido-u2f-v1.2-ps-20170411/FIDO-U2F-COMPLETE-v1.2-ps-20170411.pdf

[64] W3C. *Web Authentication: An API for Accessing Public Key Credentials*. Accessed: Mar. 6, 2022. [Online]. Available: https://w3c.github.io/webauthn/

[65] NIST. *Federal Information Processing Standards (FIPS)*. Accessed: Mar. 6, 2022. [Online]. Available: https://csrc.nist.gov/publications/fips

[66] I. for Open AuTHentication. *Oath Reference Architecture, Release 2.0*. Accessed: Mar. 6, 2022. [Online]. Available: https://openauthentication.org/wp-content/uploads/2015/09/ReferenceArchitectureVersion2.pdf

[67] O. W. Group. *The Oauth 2.1 Authorization Framework*. Accessed: Mar. 6, 2022. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-ietf-oauth-v2-1-00

[68] V. C. Hu, "Guide to attribute based access control (ABAC) definition and consideration," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. NIST 800-162, 2014.

[69] D. Ferraiolo, R. Chandramouli, R. Kuhn, and V. Hu, "Extensible access control markup language (XACML) and next generation access control (NGAC)," in *Proc. ACM Int. Workshop Attribute Based Access Control (ABAC)*, 2016, pp. 13–24.

[70] S. Monir, "A lightweight attribute-based access control system for IoT," Ph.D. dissertation, Dept. Comput. Sci., Univ. Saskatchewan, Saskatoon, SK, Canada, 2017.

[71] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan, "ConXsense: Automated context classification for context-aware access control," in *Proc. Asia CCS*, 2014, pp. 293–304.

[72] J. Wang, H. Wang, H. Zhang, and N. Cao, "Trust and attribute-based dynamic access control model for Internet of Things," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery*, Oct. 2017, pp. 342–345.

[73] B. Bezawada, K. Haefner, and I. Ray, "Securing home IoT environments with attribute-based access control," in *Proc. 3rd ACM Workshop Attribute-Based Access Control*, Mar. 2018, pp. 43–53.

[74] Q. Ni, E. Bertino, and J. Lobo, "Risk-based access control built on fuzzy inferences," in *Proc. ASIACCS*, 2010, pp. 250–260.

[75] J. Li, Y. Bai, and N. Zaman, "A fuzzy-modelling approach for access control in ehealth," in *Proc. IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. Trust*, Jun. 2013, pp. 17–23.

[76] H. F. Atlam and G. B. Wills, "An efficient security risk estimation technique for risk-based access control model for IoT," *Internet Things*, vol. 6, Jun. 2019, Art. no. 100052.

[77] J. L. Hern, A. Ramos, A. J. Jara, L. Marin, and A. F. Skarmeta1, "Distributed capability-based access control for the Internet of Things," *J. Internet Services Inf. Secur.*, vol. 3, pp. 1–6, Nov. 2013.

[78] J. L. Hernández-Ramos, A. J. Jara, L. Marín, and A. F. S. Gómez, "DCapBAC: Embedding authorization logic into smart things through ECC optimizations," *Int. J. Comput. Math.*, vol. 93, no. 2, pp. 345–366, 2016.

[79] A. L. Marra, F. Martinelli, P. Mori, and A. Saracino, "Implementing usage control in Internet of Things: A smart home use case," in *Proc. IEEE Trustcom/BigDataSE/ICESS*, Aug. 2017, pp. 1056–1063.

[80] N. Shalev, I. Keidar, Y. Weinsberg, Y. Moatti, and E. Ben-Yehuda, "WatchIT: Who watches your IT guy?" in *Proc. 26th Symp. Operating Syst. Princ.*, 2017, pp. 515–530.

[81] Y. Desmedt and A. Shaghaghi, *Function-Based Access Control (FBAC): Towards Preventing Insider Threats Organizations*. Cham, Switzerland: Springer, 2018, pp. 143–165, doi: 10.1007/978-3-030-04834-1_8.

[82] U. Maroof, A. Shaghaghi, and S. Jha, "PLAR: Towards a pluggable software architecture for securing IoT devices," in *Proc. 2nd Int. ACM Workshop Secur. Privacy Internet–Things (IoT S&P)*, New York, NY, USA, 2019, pp. 50–57, doi: 10.1145/3338507.3358619.

[83] S. Ravidas, A. Lekidis, F. Paci, and N. Zannone, "Access control in Internet-of-Things: A survey," *J. Netw. Comput. Appl.*, vol. 144, pp. 79–101, Oct. 2019.

[84] Kantarainitiative. (2017). *User-Managed Access (UMA) 2.0 Grant for OAuth 2.0 Authorization*. [Online]. Available: https://kantarainitiative.org/file-downloads/rec-oauth-uma-grant-2-0-pdf/

[85] A. Ouaddah, H. Mousannif, A. A. Elkalam, and A. A. Ouahman, "Access control in the Internet of Things: Big challenges and new opportunities," *Comput. Netw.*, vol. 112, pp. 237–262, Jan. 2017.

[86] M. Jemel and A. Serhrouchni, "Decentralized access control mechanism with temporal dimension based on blockchain," in *Proc. IEEE 14th Int. Conf. e-Business Eng. (ICEBE)*, Nov. 2017, pp. 177–182.

[87] Z. Abaid, A. Shaghaghi, R. Gunawardena, S. Seneviratne, A. Seneviratne, and S. Jha, "Health access broker: Secure, patient-controlled management of personal health records in the cloud," in *Proc. Comput. Intell. Secur. Inf. Syst. Conf.* Cham, Switzerland: Springer, 2019, pp. 111–121.

[88] D. Di Francesco Maesa, P. Mori, and L. Ricci, "Blockchain based access control," in *Distributed Applications and Interoperable Systems*. Cham, Switzerland: Springer, 2017, pp. 206–220.

[89] D. Di Francesco Maesa, P. Mori, and L. Ricci, "Blockchain based access control services," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul./Aug. 2018, pp. 1379–1386.

[90] S. Ding, J. Cao, C. Li, K. Fan, and H. Li, "A novel attribute-based access control scheme using blockchain for IoT," *IEEE Access*, vol. 7, pp. 38431–38441, 2019.

[91] S. Rouhani and R. Deters, "Blockchain based access control systems: State of the art and challenges," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Oct. 2019, pp. 423–428.

[92] P. Zhang, L. Wang, W. Wang, K. Fu, and J. Wang, "A blockchain system based on quantum-resistant digital signature," *Secur. Commun. Netw.*, vol. 2021, pp. 1–13, Mar. 2021.

[93] W. Yin, Q. Wen, W. Li, H. Zhang, and Z. Jin, "An anti-quantum transaction authentication approach in blockchain," *IEEE Access*, vol. 6, pp. 5393–5401, 2018.

[94] T. M. Fernandez-Carames and P. Fraga-Lamas, "Towards post-quantum blockchain: A review on blockchain cryptography resistant to quantum computing attacks," *IEEE Access*, vol. 8, pp. 21091–21116, 2020.

[95] Y. Li, N. S. Dhotre, Y. Ohara, T. M. Kroeger, and E. Miller, "Horus: Fine-grained encryption-based security for large-scale storage," in *Proc. 11th USENIX Conf. File Storage Technol.*, 2013, pp. 147–160.

[96] S. LABS. (Mar. 2020). *Decentralized Cloud Storage is Here*. [Online]. Available: https://storj.io/index.html

[97] CryptoMove. (Mar. 2020). *The New Standard in Data Security*. [Online]. Available: https://www.cryptomove.com

[98] (Mar. 2020). *Homomorphic Encryption Standardization*. [Online]. Available: https://homomorphicencryption.org/introduction

[99] C. Zhao, S. Zhao, M. Zhao, Z. Chen, C.-Z. Gao, H. Li, and Y.-A. Tan, "Secure multi-party computation: Theory, practice and applications," *Inf. Sci.*, vol. 476, pp. 357–372, Feb. 2019.

[100] M. S. Turan, K. A. McKay, Ç. Çalik, D. Chang, and L. Bassham, "Status report on the first round of the NIST lightweight cryptography standardization process," NIST, Gaithersburg, MD, USA, Tech. Rep. NISTIR 8268, 2019.

[101] O. Lo, W. J. Buchanan, and D. Carson, "Power analysis attacks on the AES-128 S-box using differential power analysis (DPA) and correlation power analysis (CPA)," *J. Cyber Secur. Technol.*, vol. 1, no. 2, pp. 88–107, 2017.

[102] W. Buchanan. (Feb. 2020). *PHOTON*. [Online]. Available: http://asecuritysite.com/encryption/photon

[103] W. J. Buchanan, S. Li, and R. Asif, "Lightweight cryptography methods," *J. Cyber Secur. Technol.*, vol. 1, nos. 3–4, pp. 187–201, Sep. 2017.

[104] A. BogdanovL, R. KnudsenG, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe, "PRESENT: An ultra-lightweight block cipher," in *Cryptographic Hardware and Embedded Systems*, vol. 1. Berlin, Germany: Springer, 2007, pp. 450–466.

[105] B. WJ. (Mar. 2020). *XTEA (eXtended TEA)*. [Online]. Available: http://asecuritysite.com/ encryption/xtea

[106] *CLEFIA*. Accessed: Mar. 6, 2022. [Online]. Available: https://asecuritysite.com/encryption/clefia

[107] K. McKay, L. Bassham, M. Turan, and N. Mouha, "Report on lightweight cryptography," NIST, Gaithersburg, MD, USA, Tech. Rep. NISTIR 8114, 2017.

[108] N. S. Agency. (2021). *Quantum Computing and Post-Quantum Cryptography: Frequently Asked Questions*. [Online]. Available: https://media.defense.gov/2021/Aug/04/2002821837/-1/-1/1/Quantum_FAQs_20210804.PDF

[109] R. Perlner and D. Cooper, "Quantum resistant public key cryptography: A survey," in *Proc. 8th Symp. Identity Trust Internet (IDtrust)*, Gaithersburg, MD, USA, 2009, pp. 85–93.

[110] B. Gilles, H. Peter, and T. Alain, "Quantum cryptanalysis of hash and claw-free functions," in *Proc. Latin Amer. Symp. Theor. Inform.* Berlin, Germany: Springer, 1998, pp. 163–169.

[111] A. C. Randy Kuang, D. Lou, and A. He, "Quantum secure lightweight cryptography with quantum permutation pad," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 6, no. 4, pp. 790–795, 2021.

[112] R. Kuang, D. Lou, A. He, and A. Conlon. (2021). *Quantum Safe Lightweight Cryptography With Quantum Permutation Pads*. [Online]. Available: https://www.quantropi.com/wp-content/uploads/2021/05/QPP-AES-icccs21-final-1.1.pdf

[113] I. Anshel, M. Anshel, D. Goldfeld, and S. Lemieux, *Key Agreement, the Algebraic Eraser, and Lightweight Cryptography*, vol. 418. Providence, RI, USA: American Mathematical Society, Jan. 2006, pp. 1–34.

[114] A. Ben-Zvi, S. Blackburn, and B. Tsaban, "A practical cryptanalysis of the algebraic eraser," in *Advances in Cryptology–(CRYPTO)*, M. Robshaw and J. Katz, Eds. Berlin, Germany: Springer, 2016, pp. 179–189.

[115] J. Hoffstein, J. Pipher, and J. H. Silverman, "NTRU: A ring-based public key cryptosystem," in *Algorithmic Number Theory*, J. P. Buhler, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 267–288.

[116] N. Li, D. Liu, and S. Nepal, "Lightweight mutual authentication for IoT and its applications," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 4, pp. 359–370, Oct./Dec. 2017.

[117] X.-W. Wu, E.-H. Yang, and J. Wang, "Lightweight security protocols for the Internet of Things," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–7.

[118] A. Lerner. (2017). *Microsegmentation*. [Online]. Available: https://blogs.gartner.com/andrew-lerner/2017/03/21/microsegmentation/

[119] Z. Zaheer, H. Chang, S. Mukherjee, and J. Van der Merwe, "eztrust: Network-independent zero-trust perimeterization for microservices," in *Proc. ACM Symp. SDN Res.*, 2019, pp. 49–61.

[120] C. DeCusatis, P. Liengtiraphan, A. Sager, and M. Pinelli, "Implementing zero trust cloud networks with transport access control and first packet authentication," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Nov. 2016, pp. 5–10.

[121] illumio. (2020). *Illumio*. [Online]. Available: https://www.illumio.com/

[122] cilium. (2020). *Cilium*. [Online]. Available: https://cilium.io/

[123] Cisco. (2020). *Application Centric Infrastructure*. [Online]. Available: https://www.cisco.com/c/en_au/solutions/data-center-virtualization/application-centric-infrastructure/index.html

[124] S. Staff. (2014). *What is VMware NSX and VMware SDN?* [Online]. Available: https://www.sdxcentral.com/networking/sdn/

[125] VArmour. (2020). *Helping Security and IT Teams Navigate Risk*. [Online]. Available: https://www.varmour.com//

[126] unisys. (2020). *Microsegmentation Drives Network Isolation*. [Online]. Available: https://www.unisys.com/offerings/security-solutions/unisys-stealth-products-and-services/microsegmentation

[127] Consul. (2020). *Connect*. [Online]. Available: https://www.consul.io/docs/connect/index.html

[128] Istio. *Micro-Segmentation With Istio Authorization*. Accessed: Mar. 6, 2022. [Online]. Available: https://istio.io/blog/2018/istio-authorization/

[129] J. Li, E. Altman, and C. Touati, "A general SDN-based IoT framework with NVF implementation," *ZTE Commun.*, vol. 13, no. 3, pp. 42–45, 2015. [Online]. Available: https://hal.inria.fr/hal-01197042

[130] A. Shaghaghi, M. A. Kaafar, S. Scott-Hayward, S. S. Kanhere, and S. Jha, "Towards policy enforcement point as a service (PEPS)," in *Proc. IEEE Conf. Netw. Function Virtualization Softw. Defined Netw. (NFV-SDN)*, Nov. 2016, pp. 50–55.

[131] S. Abiezzi and G. Bollela, "Fine-grained IoT access control via device proxies and SDN-based micro-segmentation," U.S. Patent App. 15 962 849, Oct. 31, 2019.

[132] L. Deri and A. Del Soldato, "An architecture for distributing and enforcing IoT security at the network edge," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2018, pp. 211–218.

[133] D. Puthal, S. P. Mohanty, P. Nanda, and U. Choppali, "Building security perimeters to protect network systems against cyber threats [Future directions]," *IEEE Consum. Electron. Mag.*, vol. 6, no. 4, pp. 24–27, Oct. 2017.

[134] A. Moubayed, A. Refaey, and A. Shami, "Software-defined perimeter (SDP): State of the art secure solution for modern networks," *IEEE Network*, vol. 33, no. 5, pp. 226–233, Sep./Oct. 2019.

[135] H. Zorkta and B. Almutlaq, "Harden single packet authentication (HSPA)," *Int. J. Comput. Theory Eng.*, vol. 4, no. 5, p. 717, 2012.

[136] J. Hemmes and J. Dressler, "Work-in-progress: IoT device signature validation," in *Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2019, pp. 0043–0048.

[137] J. O'Raw, D. Laverty, and D. J. Morrow, "Securing the industrial Internet of Things for critical infrastructure (IIoT-CI)," in *Proc. IEEE 5th World Forum Internet Things (WF-IoT)*, Apr. 2019, pp. 70–75.

[138] P. Amthor, D. Fischer, W. E. Kühnhauser, and D. Stelzer, "Automated cyber threat sensing and responding: Integrating threat intelligence into security-policy-controlled systems," in *Proc. 14th Int. Conf. Availability, Reliability Secur.*, 2019, pp. 1–10.

[139] N. Moustafa, E. Adi, B. Turnbull, and J. Hu, "A new threat intelligence scheme for safeguarding industry 4.0 systems," *IEEE Access*, vol. 6, pp. 32910–32924, 2018.

[140] E. Bou-Harb, W. Lucia, N. Forti, S. Weerakkody, N. Ghani, and B. Sinopoli, "Cyber meets control: A novel federated approach for resilient CPS leveraging real cyber threat intelligence," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 198–204, May 2017.

[141] S. Appala, N. Cam-Winget, D. McGrew, and J. Verma, "An actionable threat intelligence system using a publish-subscribe communications model," in *Proc. 2nd ACM Workshop Inf. Sharing Collaborative Secur.*, Oct. 2015, pp. 61–70.

[142] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Comput. Secur.*, vol. 72, pp. 212–233, Jan. 2018.

[143] S. Zhou, Z. Long, L. Tan, and H. Guo, "Automatic identification of indicators of compromise using neural-based sequence labelling," in *Proc. 32nd Pacific Asia Conf. Lang., Inf. Comput.*, Hong Kong, Dec. 2018, pp. 1–3.

[144] S. Yoon and J. Kim, "Remote security management server for IoT devices," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2017, pp. 1162–1164.

[145] E. Lear, R. Droms, and D. Romascanu, *Manufacturer Usage Description Specification*, IETF Draft, Internet Engineering Task Force, Fremont, CA, USA, document RFC 8520, 2017.

[146] A. Hamza, D. Ranathunga, H. H. Gharakheili, M. Roughan, and V. Sivaraman, "Clear as MUD: Generating, validating and applying IoT behavioral profiles," in *Proc. Workshop IoT Secur. Privacy*, 2018, pp. 8–14.

[147] B. Lee, R. Vanickis, F. Rogelio, and P. Jacob, "Situational awareness based risk-adapatable access control in enterprise networks," 2017, *arXiv:1710.09696*.

[148] U. Franke and J. Brynielsson, "Cyber situational awareness—A systematic review of the literature," *Comput. Secur.*, vol. 46, pp. 18–31, Oct. 2014.

[149] J. Wu, K. Ota, M. Dong, J. Li, and H. Wang, "Big data analysis-based security situational awareness for smart grid," *IEEE Trans. Big Data.*, vol. 4, no. 3, pp. 408–417, Sep. 2016.

[150] R. Montesino, S. Fenz, and W. Baluja, "SIEM-based framework for security controls automation," *Inf. Manage. Comput. Secur.*, vol. 20, no. 4, pp. 24–263, Oct. 2012.

[151] A. Stern. (Mar. 2018). *What is Security Automation?* [Online]. Available: https://www.siemplify.co/blog/what-is-security-automation/

[152] M. S. Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Machine-learning techniques for detecting attacks in SDN," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 277–281.

[153] M. M. Oo, S. Kamolphiwong, T. Kamolphiwong, and S. Vasupongayya, "Advanced support vector machine-(ASVM-) based detection for distributed denial of service (DDoS) attack on software defined networking (SDN)," *J. Comput. Netw. Commun.*, vol. 2019, May 2019, Art. no. 8012568.

[154] J. Ye, X. Cheng, J. Zhu, L. Feng, and L. Song, "A DDoS attack detection method based on SVM in software defined network," *Secur. Commun. Netw.*, vol. 2018, Apr. 2018, Art. no. 9804061.

[155] M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Deep abstraction and weighted feature selection for Wi-Fi impersonation detection," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 621–636, Mar. 2018.

[156] Y. Fu, F. Lou, F. Meng, Z. Tian, H. Zhang, and F. Jiang, "An intelligent network attack detection method based on RNN," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2018, pp. 483–489.

[157] D. Kwon, K. Natarajan, S. C. Suh, H. Kim, and J. Kim, "An empirical study on network anomaly detection using convolutional neural networks," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2018, pp. 1595–1598.

[158] I. Kotenko, I. Saenko, A. Kushnerevich, and A. Branitskiy, "Attack detection in IoT critical infrastructures: A machine learning and big data processing approach," in *Proc. 27th Euromicro Int. Conf. Parallel, Distrib. Netw.-Based Process. (PDP)*, Feb. 2019, pp. 340–347.

[159] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. Hashem, "Attack and anomaly detection in iot networks using machine learning techniques: A review," *Internet Things*, vol. 7, Feb. 2019, Art. no. 100059.

[160] J. Galeano-Brajones, J. Carmona-Murillo, J. F. Valenzuela-Valdés, and F. Luna-Valero, "Detection and mitigation of dos and DDoS attacks in IoT-based stateful SDN: An experimental approach," *Sensors*, vol. 20, no. 3, p. 816, 2020.

[161] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, "Deep learning-based intrusion detection for IoT networks," in *Proc. IEEE 24th Pacific Rim Int. Symp. Dependable Comput. (PRDC)*, Dec. 2019, pp. 256–25609.

[162] B. Olivier, Y. Yin, A.-G. Javier, R. Manuel, and G. Erol, "IoT attack detection with deep learning," in *Proc. ISCIS Secur. Workshop*, 2018. [Online]. Available: https://hal.laas.fr/hal-02062091

[163] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the Internet of Things in the age of machine learning and software-defined networking," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4829–4842, Dec. 2018.

[164] M. Al-Shaboti, A. Chen, and I. Welch, "Automatic device selection and access policy generation based on user preference for IoT activity workflow," in *Proc. 18th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun./13th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE)*, Aug. 2019, pp. 769–774.

[165] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN—Key technology enablers for 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2468–2478, Nov. 2017.

[166] A. Molina Zarca, M. Bagaa, J. Bernal Bernabe, T. Taleb, and A. F. Skarmeta, "Semantic-aware security orchestration in SDN/NFV-enabled IoT systems," *Sensors*, vol. 20, no. 13, p. 3622, Jun. 2020.

[167] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging SDN and NFV security mechanisms for IoT systems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 812–837, 1st Quart., 2019.

[168] T. Kurek, "Unikernel network functions: A journey beyond the containers," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 15–19, Dec. 2019.

[169] J. B. Filipe, F. Meneses, A. U. Rehman, D. Corujo, and R. L. Aguiar, "A performance comparison of containers and unikernels for reliable 5G environments," in *Proc. 15th Int. Conf. Design Reliable Commun. Netw. (DRCN)*, Mar. 2019, pp. 99–106.

[170] M. Bagaa, T. Taleb, J. B. Bernabe, and A. Skarmeta, "A machine learning security framework for iot systems," *IEEE Access*, vol. 8, pp. 114066–114077, 2020.

[171] A. Molina Zarca, J. B. Bernabe, R. Trapero, D. Rivera, J. Villalobos, A. Skarmeta, S. Bianchi, A. Zafeiropoulos, and P. Gouvas, "Security management architecture for NFV/SDN-aware IoT systems," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8005–8020, Oct. 2019.

[172] R. Vanickis, P. Jacob, S. Dehghanzadeh, and B. Lee, "Access control policy enforcement for zero-trust-networking," in *Proc. 29th Irish Signals Syst. Conf. (ISSC)*, Jun. 2018, pp. 1–6.

[173] M. Nobakht, V. Sivaraman, and R. Boreli, "A host-based intrusion detection and mitigation framework for smart home IoT using OpenFlow," in *Proc. 11th Int. Conf. Availability, Rel. Secur. (ARES)*, Aug. 2016, pp. 147–156.

[174] H. F. Atlam, A. Alenezi, R. J. Walters, G. B. Wills, and J. Daniel, "Developing an adaptive risk-based access control model for the Internet of Things," 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. 655–661.

[175] W. B. Daoud, M. S. Obaidat, A. Meddeb-Makhlouf, F. Zarai, and K.-F. Hsiao, "TACRM: Trust access control and resource management mechanism in fog computing," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 28, Dec. 2019.

[176] P. Sun, "Research on cloud computing service based on trust access control," *Int. J. Eng. Bus. Manage.*, vol. 12, Jan. 2020, Art. no. 1847979019897444.

[177] P. N. Mahalle, P. A. Thakre, N. R. Prasad, and R. Prasad, "A fuzzy approach to trust based access control in Internet of Things," in *Proc. Wireless VITAE*, Jun. 2013, pp. 1–5.

[178] S. Tan, X. Li, and Q. Dong, "A trust management system for securing data plane of ad-hoc networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7579–7592, Sep. 2016.

[179] N. Karapanos, C. Marforio, C. Soriente, and S. Capkun, "Sound-proof: Usable two-factor authentication based on ambient sound," in *Proc. 24th USENIX Secur. Symp. (USENIX Secur.)*, Aug. 2015, pp. 483–498.

[180] A. Takayuki, "Ear acoustic authentication technology: Using sound to identify the distinctive shape of the ear canal," *NEC Tech. J.*, vol. 13, no. 2, pp. 87–90, 2018.

[181] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing acoustics-based user authentication," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2017, pp. 278–291.

[182] C. Gidney and M. Ekerå, "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits," *Quantum*, vol. 5, p. 433, Apr. 2021, doi: 10.22331/q-2021-04-15-433.

[183] D. J. Bernstein and T. Lange, "Post-quantum cryptography," *Nature*, vol. 549, pp. 188–194, Sep. 2017.

[184] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Nov. 1994, pp. 124–134.

[185] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput. (STOC)*, 1996, pp. 212–219.

[186] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, R. Perlner, A. Robinson, and D. Smith-Tone, "Status report on the first round of the NIST post-quantum cryptography standardization process," NIST, Gaithersburg, MD, USA, Tech. Rep. NISTIR 8240, 2019.

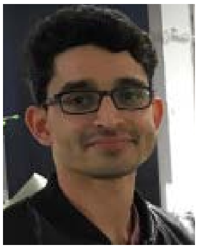
[187] W. Beullens, J.-P. D'Anvers, A. Hülsing, T. Lange, L. Panny, C. de Saint Guilhem, and N. P. Smart, "Post-quantum cryptography," Eur. Union Agency Cybersecur., Enisa, Athens, Tech. Rep., 2021, doi: 10.2824/92307.

[188] M. Braithwaite. (2016). *Experimenting with Post-Quantum Cryptography*. [Online]. Available: https://security.googleblog.com/2016/07/experimenting-with-post-quantum.html

[189] E. Alkim, L. Ducas, T. Pöppelmann, and Schwabe, "Post-quantum key exchange—A new hope," in *Proc. 25th USENIX Secur. Symp.*, 2016, pp. 327–343.

[190] W. Barker, W. Polk, and M. Souppaya, "Getting ready for post-quantum cryptography: Explore challenges associated with adoption and use of post-quantum cryptographic algorithms," NIST Cyber Security White Paper (DRAFT), Gaithersburg, MD, USA, White Paper 26, 2020.

NAEEM FIRDOUS SYED (Member, IEEE) received the B.E. degree from Anna University, Chennai, India, the M.Sc. degree from the King Fahd University of Petroleum and Minerals, Dharan, Saudi Arabia, and the Ph.D. degree in cyber security from Edith Cowan University (ECU) with a focus on the message queuing telemetry transport protocol. He is currently a Lecturer in cyber security with the School of Information Technology, Deakin University, Australia. Prior to this, he was a Postdoctoral Research Fellow with Edith Cowan University (ECU). He has over seven years of industry experience in software and network engineering. He also has strong cyber security research experience, with contributions that include journal articles, conference papers, and book chapters published by high-quality peer-reviewed outlets. He has received several awards and fellowships. His research interests include the Internet of Things (IoT) in general (and IoT security in particular), network security, network forensics, intrusion detection systems, machine learning, and deep learning techniques in network forensics.



SYED W. SHAH (Member, IEEE) received the M.Sc. degree in electrical and electronics engineering from the University of Bradford, U.K., and the Ph.D. degree in computer science and engineering from The University of New South Wales, Sydney, Australia. He is currently a Research Fellow with Deakin University, Australia. His research interests include pervasive/ubiquitous computing, user authentication/identification, the Internet of Things, signal processing, data analytics, privacy, and security. He has received multiple awards, scholarships, and fellowships. His research work has been featured by numerous leading media outlets, including CNN, the Australian Broadcasting Corporation, *New Scientist* magazine, and CJAD 800 Montreal.

ARASH SHAGHAGHI (Member, IEEE) received the B.Sc. degree from Heriot-Watt University, the M.Sc. degree in information security from University College London, and the Ph.D. degree in computer science and engineering from The University of New South Wales (UNSW), Sydney, Australia. He is currently a Senior Lecturer in cyber security with the Royal Melbourne Institute of Technology, Australia. He is also a Visiting Fellow with the School of Computer Science and Engineering, UNSW Sydney. He has previously been affiliated with Deakin University, Data61 (CSIRO), The University of Melbourne, and The University of Texas at Dallas. He is a multi-award winning cyber security educator and researcher with a track record of publications at competitive international conferences and journals. To date, his cyber security research has garnered over AU$300,000 (across Chief Investigator and Partner Investigator roles) from various internal and external sources, including the Australian Government and the Cyber Security Cooperative Research Centre. He has reviewed numerous journal article submissions and has served as a technical program committee member, organizing member, and reviewer roles at various prestigious security and networking conferences. He currently serves as an Associate Editor for the journal *Ad Hoc Networks*. He is a member of the Australian Information Security Association. His research activities have been covered by the Australian Broadcasting Corporation and other media outlets. Visit arashshaghaghi.com for more information.

ADNAN ANWAR (Member, IEEE) received the master's degree (by Research) and the Ph.D. degree from UNSW. He has previously worked as a Data Scientist and Analytics Team Leader at Flow Power. He has over ten years of industrial, research, and teaching experience in universities and research laboratories, including La Trobe University, Deakin University, The University of New South Wales (UNSW), and National Information and Communications Technology, Australia (now merged with Commonwealth Scientific and Industrial Research Organisation as Data61). He is currently a Lecturer and the Deputy Course Director of Postgraduate Cyber Security Education with the School of Information Technology, Deakin University, Australia. He has had over 60 published works, including articles in Q1-ranked journals and conference papers and book chapters published by prestigious outlets. His broad research interests include the security of sensor-connected Internet of Things devices, cloud security, the security of supervisory control and data acquisition systems in critical infrastructure, data-driven intelligent techniques, and other data science applications. He was a recipient of several awards, such as the University Postgraduate Award, the UNSW Tuition Fee Remission Scholarship, and the Best Paper Award. He has received industry funding and several travel grants through the Association of Computing Machinery and the Postgraduate Research Support Scheme. He is active in the IEEE Computer Society Technical Committee on Data Engineering as well as in the IEEE Cybersecurity Committee.

ZUBAIR BAIG (Senior Member, IEEE) is currently a Senior Lecturer with Deakin University's School of Information Technology and the Division Lead for Cyber Physical Systems and Internet of Things (IoT) Cyber Security at Deakin's Strategic Centre for Cyber Security Research and Innovation. He has authored over 95 journal articles, conference papers, and book chapters, and five white papers. He is also the Inventor of two Cyber Security Technologies granted patents by the U.S. Patent and Trade Office. He is currently serving as an Editor for three international journals: *IET Wireless Sensor Systems* (the Institute of Engineering and Technology in partnership with Wiley), *PSU Research Review* (Prince Sultan University in partnership with Emerald Publishing), and the *Journal of Information and Telecommunication* (Taylor & Francis). He has served on the technical program committees of numerous international conferences and has delivered over 20 keynote talks on cyber security. His research interests include cyber security, artificial intelligence, critical infrastructure (CI), and the IoT. He has a broad risk assessment skill set which extends to the IoT, CI, and other sensor network contexts. He has successfully secured funding from the Cyber Security Cooperative Research Centre, the Australian Department of Home Affairs, and the Australian Department of Defence to conduct cutting-edge research in the field of cyber security. He falls within the top 2% of Scientists in Stanford University's 2021 ''Updated science-wide author databases of standardized citation indicators.''

ROBIN DOSS (Senior Member, IEEE) is currently a Full Professor and the Director of the Strategic Centre for Cyber Security Research and Innovation at Deakin University—a multidisciplinary research center whose projects span the technical, business, human, policy, and legal dimensions of cyber security. He also leads the Next Generation Authentication Technologies theme for the Critical Infrastructure Security Research Program at the National Cyber Security Cooperative Research Centre. Prior to this role, he was the Deputy Head of School for the School of Information Technology, Deakin University. He has an extensive research publication portfolio, and in 2019, he was a recipient of the Cyber Security Researcher of the Year Award from the Australian Information Security Association. His research interests include the broad areas of systems security, protocol design and security analysis, with a focus on smart cyber-physical critical infrastructure. His research program has been funded by the Australian Research Council; government agencies, such as the Defense (now Australian) Signals Directorate; the Department of Industry, Innovation and Science; and various industry partners. He has contributed to large multi-year projects under the European Union's Framework Program and has been funded by the Indian Government under its Scheme for Promotion of Academic and Research Collaboration. He is a member of the executive council of the IoT Alliance Australia; the Founding Chair of the Future Network Systems and Security conference series; and an Associate Editor for the journal *Cyber Physical Systems*.