Literature review CIBC Understanding User Behaviour in Zero Trust (ZT) Systems

I. Introduction

The concept of Zero Trust (ZT) security was introduced by Forrester in 2010 to safeguard organizational assets against cybersecurity threats [1]. In recent years, ZT has gained significant attention as a solution to address the challenges of securing modern enterprise systems. The traditional security model relied on perimeter-based defenses, where everything inside the network was trusted, and everything outside the network was considered untrusted. However, with the advent of cloud-based apps, mobile devices, Bring your own device (BYOD), and remote work, the traditional organizational resources have become less effective.

ZT security takes a different approach from traditional models and assumes that all entities, internal or external, are not trusted until proven otherwise. The objective of ZT security is to reduce security breaches by verifying the identity and access rights of all entities before granting access to resources. A recent survey reported that 78% of organizations worldwide consider ZT a priority, with 90% of them actively working to implement ZT initiatives [2]. The financial services sector has exhibited particularly high adoption rates, with 94% of organizations having or developing ZT plans [2].

ZT plans have been found to be effective in improving security threat detection and response, promoting accountability and transparency, and minimizing damage levels in the event of a breach [3]. As a result, the concept of ZT security has become an essential component of modern cybersecurity strategies, and organizations across all industries are exploring its implementation to protect their valuable assets.

II. Zero Trust Architecture Models

In recent literature, several studies have proposed various ZT architectures and implementation methods. These models address different security concerns and offer different levels of protection, but share the same core principles of ZT: never trust, always verify. For example, the National Institute of Standards and Technology (NIST) has proposed the implementation of ZT model, consisting of five core components—subject, resource, policy decision point (PDP), policy enforcement point (PEP), and supplement. These elements are crucial in making sure that access to organization resources is allowed based on the least privilege principle and is constantly monitored and managed. For instance, the subject is responsible to authenticate the user's identity and authorization through multi-factor authentication and authorization mechanisms, while the resource component grants access to resources based on the user's role and the minimum level of access required to perform their tasks. The PDP makes decisions about whether to grant or deny access to resources based on policies defined by the organization, while the PEP enforces these policies, ensuring only authorized users and devices are granted access. Finally, the supplement component

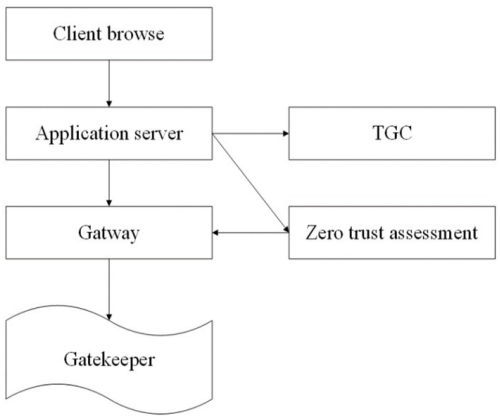

Fig. 1. Proposed Architecture of ZTA Network Access Control according to [4].

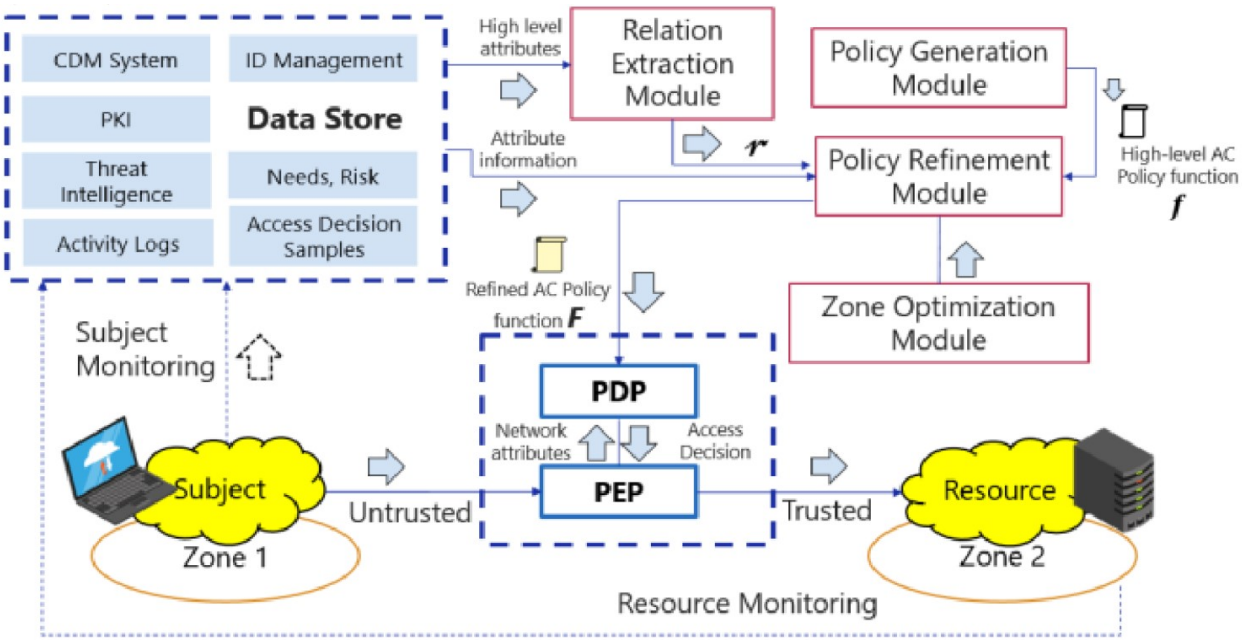

Fig. 2. Block Diagram of ZT Architecture in [5].

provides additional information to the PDP, leading to better accuracy in descision making and reducing false positives and false negatives.

In a different study [4], the authors propose a new ZT approach that ensures security by requiring verification for every access request, without assuming any implicit trust. The proposed ZT approach comprises several components, as shown in figure 1. These components work together to create a secure and efficient system for businesses to carry out their operations. The core elements of this approach include:

- The Application Server: This server houses critical business applications and reports security information to the Evaluation Server.
- The Evaluation Server: This server assesses the installation status of the Application Server and sends the results to the Gateway, which regulates access to the Gatekeeper.
- The Token Generate Center (TGC): This center generates and manages tokens for secure access to the Gatekeeper.

The interdependence of these components ensures that businesses have a secure, efficient, and reliable system in place to carry out their operations. By requiring verification for every access request, this ZT approach minimizes the risk of security breaches and data loss, which can have catastrophic consequences for businesses in today's digital world. This approach provides a valuable contribution to the field of ZT, and its implementation gives an insight for organizations to enhance their security posture.

In another study [5], *Ghate et. al* propose a new ZT architecture that leverages Generalized Attribute Relation Extraction (GARE) and Machine Learning (ML) techniques to automate fine-grained access control for enterprise networks. The proposed architecture consists of four core modules: the Relation Extraction Module (REM), the Policy Generation Module (PGM), the Policy Refinement Module (PRM), and the Zone Optimization Module (ZOM), as illustrated in Figure 2.

The REM is responsible for extracting attributes from the network and correlating them with high-level attributes such as user roles and device types. The PGM then generates a high-level policy based on the relationships identified by the REM. This policy captures the intent of the access control policy without being overly specific, which reduces the cost of policy definition and allows for more fine-grained access control. The PRM refines the high-level policy into an executable network access policy using interpretable rules generated by the REM. The ZOM uses ML techniques to analyze the network topology and identify areas where zone definitions can be optimized for more efficient policy enforcement and reduced risk of conflicts.

In addition to these core modules, the proposed

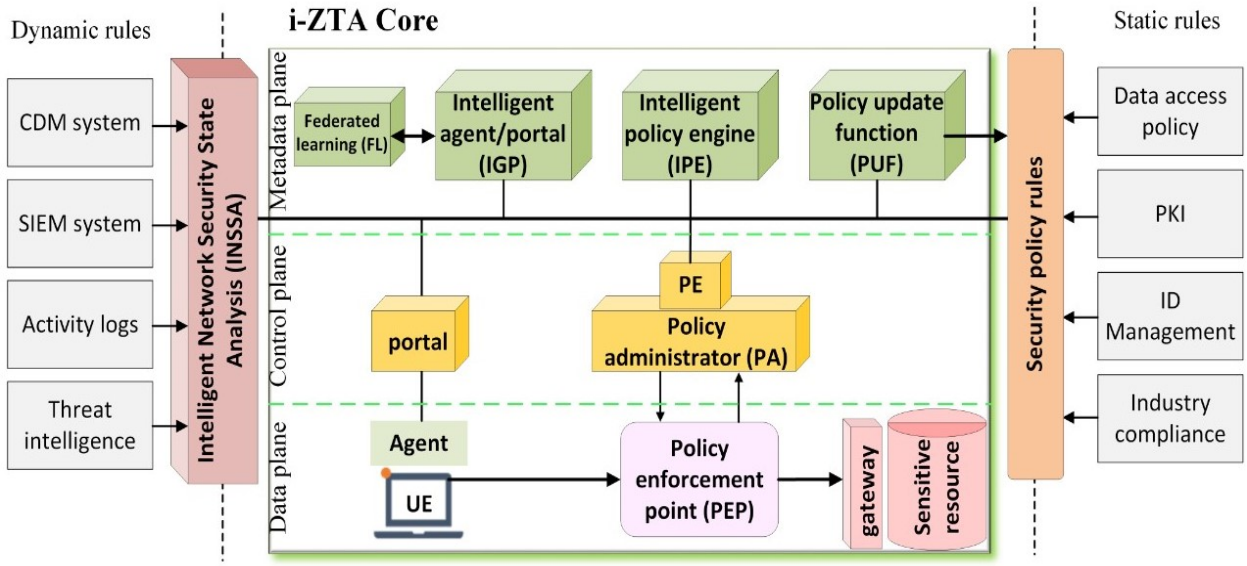

Fig. 3. Logical Components of i-ZTA as presented in [6].

architecture includes other components such as subject, resource, PDP, and PEP, as suggested by NIST, to define the scope of the access control policy and enforce it at the network level. The primary objective of this proposed architecture is to reduce the complexity and cost of policy creation while improving network security. By automating the process of access control using GARE and ML, the proposed architecture significantly reduces the workload of security administrators. The proposed architecture presents a promising approach to automating fine-grained access control for enterprise networks, which can help organizations better manage their security posture and reduce the risk of security breaches.

To ensure seamless processing of data, the components of the ZT architecture must be equipped to handle big data. With an increase in the volume of users and network assets, it has become imperative to incorporate intelligent monitoring, evaluation, and decision-making capabilities, all of which rely heavily on the use of artificial intelligence (AI) as a critical enabler. The authors in [6] propose an intelligent ZTA (i-ZTA) architecture that aims to handle big data and enable intelligent monitoring through the use of AI. The i-ZTA architecture consists of various components, including: Data access policy , Public key infrastructure (PKI), Identity (ID) management and Industry compliance. These peripheral modules are incorporated into the i-ZTA architecture to ensure comprehensive network security. The i-ZTA architecture also includes core components of NIST, PEP and PDP. The PEP serves as the first point of contact for access requests, while the PDP is responsible for making access decisions based on internal and external information about the security state of the subject and network assets. To enable effective decision-making, the i-ZTA architecture incorporates various AI engines, including:

- Intelligent Policy Engine (IPE): IPE uses an AI trust algorithm to authorize access requests based on subject privileges, security state, policy rules, network state, and confidence level of access. Reinforcement Learning (RL) is used to maximize usability with the least privileges.
- Intelligent Network Security State Analysis (INSSA): INSSA employs a Graph Neural Network (GNN) model to assess network security state and identify potential attacks.
- Intelligent Agent/Portal (IGP): IGP analyzes the security posture of network traffic to provide environmental awareness and maintain a high confidence level of subject access.
- Intelligent Policy Engine (IPE): IPE makes decisions on granting access based on all information provided by INSSA, IGP, and policy rules.

The i-ZTA components are divided into static and dynamic modules, with the dynamic modules being the distinct features of i-ZTA. The i-ZTA architec-

ture also divides the network into three logical and possibly physical planes to ensure comprehensive and secure protection of network resources against unauthorized access and potential attacks. These features ensure that the i-ZTA can effectively protect network resources against unauthorized access and potential attacks, even as network assets and users continue to grow.

III. User Behaviour within ZT systems

ZT technology has gained significant attention in recent years as an effective way to enhance system security. While research efforts have primarily focused on the architectural aspects of ZT, the importance of user behavior in this technology has often been overlooked. According to Gartner [7], User and Entity Behavior Analysis (UEBA) is a critical component in the success of ZT technology and is expected to become a standard element of enterprise security systems in the future. UEBA plays a fundamental role in ZT models by enabling early detection of potential security threats through the identification of anomalies in user activity data [8]. The process involves feeding user activity data into a ML system that uses customized models to analyze patterns, correlations, and anomalies in user behavior. Anomalies can manifest as significant deviations in user download or upload patterns, indicating unauthorized access or control over device accessibility services. Upon the detection of an anomaly, the ”respond” stage of ZT models is activated, and an alert is sent to an agent, triggering an automated response such as suspending the suspicious activity or isolating the user account. In this matter, the incorporation of UEBA in ZT models allows for the early identification and mitigation of potential security threats, thereby enhancing enterprise system security [8].

IV. Understanding User Behaviour

The importance of user behaviour in a robust security system cannot be overstated in today's digital age. To illustrate this point, let's consider the case of Bob, an engineer working for a financial organization. While working on a new database system, Bob attempted to save sensitive information to a thumb drive on his local computer. This prohibited action triggered an alert on the organization's behavioral analytics system.

Behavioral analytics systems like the one used by Bob's organization are designed to detect potential insider threats by analyzing various data points in real-time. These data points include metadata such as Bob's IP address, operating system, session time, and other relevant information. By analyzing this data, the system can determine whether Bob's actions are legitimate or potentially harmful to the organization's security. In the event of a detected insider threat, the behavioral analytics system responds promptly and automatically. It disconnects the user from the network, blocks further attempts to access sensitive data, and notifies the appropriate personnel of the situation. This quick response helps minimize the potential damage caused by insider threats.

In an ideal security scenario, the behavioral analytics system would implement a highly sophisticated approach, leveraging more data points and automated responses. For instance, it could leverage advanced ML algorithms to more accurately detect and pinpoint potential insider threats. This could entail analyzing data from a more extensive range of sources, such as anomalies and patterns within the location, browser history, device types and others. However, simply identifying suspicious behavior alone is insufficient to conclude that an individual is an insider threat. There may be valid reasons for such behavior, such as an employee being fatigue or under the influence of alcohol. Therefore, to avoid erroneously identifying a legitimate employee as a potential threat, security experts must collect additional data to validate their suspicions.

ML tests related to user focus can be highly effective in identifying insider threats. Such tests can be conducted using a variety of data sources, including wearable devices like smartwatches or fitness trackers, as well as employee email or chat logs. Assuming that the collection of such data is both ethical and legal, and the data itself is reliable, an ML model can be trained to accurately predict whether an employee is currently under focus or not.

To train such a model, various features can be analyzed to determine an employee's mental state. For instance, the model could assess an employee's keyboard and mouse activity, eye-tracking data, heart rate, breathing rate, and audio data. These features are all considered part of multi-factor authentication (MFA), which research studies have shown

can reduce the risk of attacks and improve user identification accuracy [9], [10]. By analyzing such data, patterns may indicate distraction, drowsiness, intoxication, or fatigue. For instance, a decrease in activity or eye movement, abnormal heart rate or breathing patterns, slurred speech, or changes in tone can all be potential indicators of an employee's lack of focus. If the ML model detects such patterns, it can trigger an alert of potential insider threats. This technology might be an essential tool for organizations looking to safeguard their sensitive information and protect more against insider threats.

To further enhance security, the organization could implement strict access controls that limit access to sensitive information. For instance, it could enforce role-based access controls that restrict access to sensitive data based on an employee's job function and level of clearance. The organization could provide regular cyber-security awareness training for all employees. This training would cover topics such as how to identify and report suspicious activity, how to use secure passwords, and how to avoid phishing attacks. By doing so, employees would have a better understanding of the potential risks and how to avoid them.

In this matter, user behavioral analytics system is a critical component of any organization's security strategy. By monitoring and assessing user behavior using various data points and automated responses, it can help detect and prevent insider threats. Organizations should continuously monitor and assess user behavior as part of a ZT security strategy to enhance the overall security and resilience of their IT systems [11].

V. User Behaviour Solutions

UEBA technology has been widely adopted in cutting-edge research to minimize cybersecurity risks and fraud [12]. For example, some studies focus on predicting the behavior of malicious users by continuously monitoring file and directory locations for suspicious changes [13]. Others use user analytics such as mouse speed, distance, angles, and the number of clicks during a user session for user identification and masquerade detection [14]. In addition, some studies profile user behavior by considering not only computer usage but also network resources, resulting in improved generalization performance of decision tree classification models [15]. Another research work monitors application usage, application performance, such as CPU, memory, the websites visited, the number of windows opened, and typing habits to detect anomalies based on the normal usage patterns of profiled users during different sessions [16]. Furthermore, anomaly detection models have been used to analyze three types of user log data sets: the user's daily activity summary, email contents topic distribution, and the user's weekly email communication history [17]. The authors of [18] present a user anomaly detection system that incorporates contextual and behavioral aspects of data from multiple dimensions, including network traffic and user activities. The system utilizes contextual data such as user event logs, login and logout times, source and destination IP addresses, user ID, and process ID accessed for effective detection of anomalies in a stream of events and network traffic. In this respect, adopting UEBA solutions in an organization's security strategy can help enhance its security posture and mitigate the risk of insider threats. The examples demonstrate the value of UEBA technology in preventing and detecting security incidents and fraudulent activities.

VI. User Behaviour Challenges

UEBA plays a vital role in this paradigm by acting as the first line of defense against security threats. However, there are several challenges that must be addressed to fully realize the potential of UEBA technology.

One of the primary challenges of UEBA is the quality and integration of data from multiple sources [19]. UEBA solutions rely on data from various sources, such as log files, network traffic, and user activity, to identify potential security threats. Integrating and normalizing data from multiple sources can be challenging, and poor data quality can lead to false positives and false negatives. To address this challenge, UEBA solutions need to incorporate data integration and normalization capabilities to ensure that high-quality data is available for analysis.

Another key challenge of UEBA technology is scalability and performance. UEBA solutions need to process large volumes of data in real-time to identify potential security threats. Processing large datasets in real-time can be challenging, and solutions need to be scalable and performant to handle the volume of data [20]. To address this challenge,

UEBA solutions need to incorporate scalable and performant architecture and data processing capabilities to ensure that they can handle large volumes of data in real-time.

A third challenge of UEBA technology is explainability and interpretability [21]. UEBA solutions use ML algorithms and statistical models to analyze data and identify anomalies. However, these models can be difficult to interpret, and it may be challenging to explain the reasoning behind a particular detection or alert. To address this challenge, UEBA solutions need to incorporate explainability and interpretability features, such as visualization tools or natural language processing, to enable security analysts to understand the reasoning behind a particular detection or alert.

Finally, privacy and compliance is another challenge of UEBA technology [22]. UEBA solutions may collect and analyze sensitive data about users and entities, such as personally identifiable information or login credentials. This raises concerns about privacy and compliance with regulations. To address these concerns, UEBA solutions need to incorporate privacy and compliance features, such as data encryption, access controls, and auditing, to ensure that sensitive data is protected and that the solution complies with applicable regulations.

UEBA technology is an essential component of a ZT security paradigm. However, several challenges must be addressed to fully realize its potential. By addressing these challenges, UEBA technology can continue to play a critical role in protecting organizations against security threats in today's ever-evolving threat landscape.

VII. Design plan

The proposed intelligent ZT architecture is a comprehensive security model that encompasses several engines and components that work together to provide the highest level of protection within a system. The architecture comprises three main components: user-centric, device-centric, and network-centric.

A. User-Centric Components

The user-centric components of the ZT model ensure that users must authenticate and prove their identity before being granted access to network resources. This authentication can include multi-factor authentication, device posture checks, and behavioral analytics to ensure that users are who they claim to be. The user-centric components of the ZT model include the Software-Defined Perimeter (SDP), Continuous Authentication and Authorization (CAA), Decision-making process of Privacy and Data Protection (PDP), and Intelligent Policy Engine (IPE).

- Software-Defined Perimeter (SDP): The SDP is a user-centric security technology because it prioritizes the user's identity when granting access to an organization's resources. Unlike traditional network-centric security measures, SDP solutions create a dynamic, identity-based perimeter around resources. This approach provides a more granular and flexible approach to security, as it allows organizations to control access to specific resources based on contextual information based on user's identity, such as biometrics.
- Continuous Authentication and Authorization (CAA): The CAA component constantly re-verifies the user's identity and authorization by monitoring and evaluating their behavior to detect potential security threats. This process uses advanced technologies such as behavior analysis and machine learning to analyze data from various sources. By continuously monitoring user behavior, CAA ensures that access to network resources remains secure and is only granted to authorized users.
- Privacy and Data Protection (PDP): PDP evaluates access requests based on the user's identity, role, and permissions, as well as the policies defined by the organization regarding the handling of personal data. It can also take into account the device or network components involved in the access request when determining the appropriate level of data protection measures to apply. While PDP is primarily user-centric, it can also be device or network-centric to some extent depending on the specific context and implementation.
- Intelligent Policy Engine (IPE): IPE evaluates the risk assessment of subjects' Environmental Awareness (ENA) scores to provide access accordingly. ENA takes into account factors such as the user's identity, location, device, and other relevant information to evaluate the risk level of granting access to network resources. The

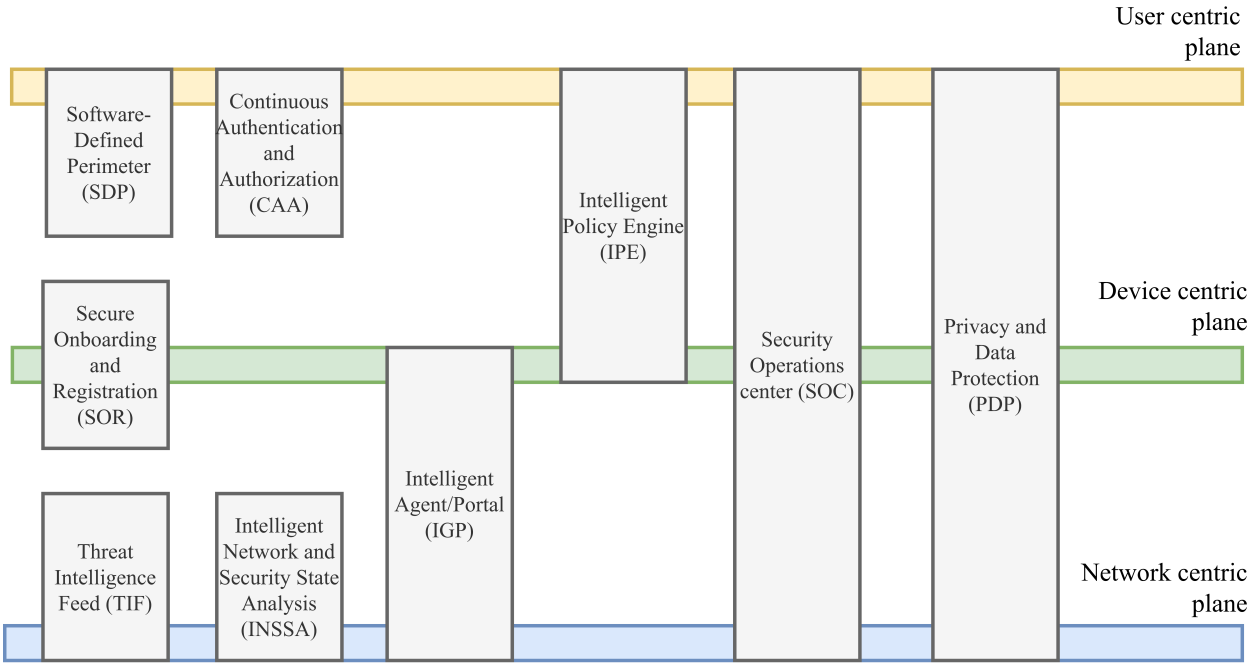

Fig. 4. ZT Network Architectural Design.

IPE approach applies to both user-centric and device-centric approaches, such as assessing the context and behavior of a user's device before granting access to network resources.

B. Network-Centric Components

From a network perspective, the ZT model involves continuously monitoring network traffic, analyzing network state, and assessing the risk level of connected devices and users. This monitoring ensures that any suspicious activity is detected and mitigated before it can cause harm. To achieve a secure and resilient network, a network-centric security architecture comprises several components including the Intelligent Network and Security State Analysis (INSSA), Intelligent Agent/Portal (IGP), The Threat Intelligence Feed (TIF) and Security Information and Event Management (SIEM).

- Intelligent Network and Security State Analysis (INSSA): INSSA is a network-centric approach that uses a graph neural network to model the state of the network and assign risk scores to the nodes. This approach is focused on detecting and preventing potential attacks, such as DoS and DDoS, and assigning risk scores to the nodes accordingly. For instance, the approach analyzes the network for unusual traffic patterns that may indicate a potential attack and assigns risk scores to the nodes that are more susceptible to the attack.
- Intelligent Agent/Portal (IGP): IGP is a network-centric approach that uses a reinforcement learning engine to analyze traffic in the network and provide a model for the communication pattern of devices. This approach optimizes the network's communication pattern to enhance security. For instance, the approach monitors the communication between devices on a network and creates a model that identifies normal traffic patterns. If any traffic pattern deviates from the model, the network security solution will trigger an alert to investigate the issue. This approach is also reffered as device-centric, as it focuses on monitoring the communication patterns of devices on the network.
- Security Information and Event Management (SIEM): SIEM is a network-centric component that collects and analyzes security data from across the network to provide real-time insights into potential security threats. SIEM solutions are typically deployed at the network level to monitor and analyze log data from various sources, such as servers, network devices, and endpoints. The SIEM correlates and aggregates this data to detect and alert on potential security incidents, such as unauthorized access at-

tempts, malware infections, or other suspicious activities within the network.

- Threat Intelligence Feed (TIF): TIF is a network-centric tool that provides real-time information about potential threats that may impact an organization's network. The TIF is designed to help security teams monitor and analyze threat intelligence data related to external threats, such as malware, phishing campaigns, and other types of attacks. This component is focused on providing network-centric threat intelligence to help organizations identify and respond to potential security incidents promptly. IGP and PDP components mentioned before, are also network-centric components that play a critical role in a network-centric security architecture.

C. Device-Centric Components

A device-centric approach to network security places devices at the center of its security strategy, verifying their identity and authorizing access to network resources. The Secure Onboarding and Registration (SOR) component ensures only authorized devices can access the network by using multi-factor authentication and background checks. This approach is used because devices are the primary entry point for attackers and are more vulnerable to security threats than other network components. SOR, along with other components such as the IGP, IPE and PDP, provide an effective way to manage and secure the increasing number of connected devices in modern networks.

Finally, a Security Operations Center (SOC) can adopt one of three approaches, namely user-centric, device-centric, or network-centric, based on the particular requirements and emphasis of the organization. A user-centric SOC concentrates on monitoring user activity and access to resources within the organization's network. It utilizes methodologies such as user behavior analytics (UBA) and identity and access management (IAM) to identify and react to possible security breaches associated with user activity. In contrast, a device-centric SOC emphasizes securing and monitoring individual devices within the organization's network, such as servers, workstations, and mobile devices. This may entail implementing endpoint security measures, such as antivirus and intrusion prevention systems, to identify and respond to potential threats aimed at specific devices. In comparison, a network-centric SOC focuses on scrutinizing network traffic and activity to identify and respond to potential security incidents. This can entail techniques such as network flow analysis, packet inspection, and intrusion detection and prevention systems (IDPS). Ultimately, the SOC's emphasis is determined by the organization's specific cybersecurity needs and priorities. In practice, many SOCs combine aspects of all three approaches to provide comprehensive security coverage across the organization's infrastructure.

To establish a secure and trustworthy network environment in today's ever-evolving threat landscape, it is essential to incorporate user-centric, device-centric, and network-centric components into a ZT model. The inclusion of user-centric elements, like multi-factor authentication and device posture assessments, ensures that only authorized users are granted access to network resources. Device posture assessments restrict or block access to network resources for devices that do not comply with established security standards, such as outdated software or unpatched systems. On the other hand, network-centric components provide network protection against potential cyber-attacks by creating secure perimeters dynamically based on several factors. Thus, the implementation of a ZT model that incorporates both user, device and network-centric components is crucial in achieving a secure and trusted network environment. Organizations can safeguard their assets and maintain a high level of security against threats that can harm the organization's reputation, intellectual property, or financial stability.

REFERENCES

[1] J. Kindervag, S. Balaouras *et al.*, "No more chewy centers: Introducing the zero trust model of information security," *Forrester Research*, vol. 3, 2010.

[2] O. Inc, "The state of zero trust security," 2021. [Online]. Available: https://www.okta.com/sites/default/files/2021-07/WPR-2021-ZeroTrust-070821.pdf?o=4677

[3] C. Cunningham and J. Pollard, "The eight business and security benefits of zero trust," *Forrester Reseach November*, 2017.

[4] T. Chuan, Y. Lv, Z. Qi, L. Xie, and W. Guo, "An implementation method of zero-trust architecture," in *Journal of Physics: Conference Series*, vol. 1651, no. 1. IOP Publishing, 2020, p. 012010.

[5] N. Ghate, S. Mitani, T. Singh, and H. Ueda, "Advanced zero trust architecture for automating fine-grained access control with generalized attribute relation extraction," *IEICE Proceedings Series*, vol. 68, no. C1-5, 2021.

[6] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5g/6g networks: Principles, challenges, and the role of machine learning in the context of o-ran," *Computer Networks*, p. 109358, 2022.

[7] Gartner, "Ueba will become a standard component of enterprise security," https://blogs.gartner.com/anton-chuvakin/2016/12/12/ueba-clearly-defined-again/, 2018.

[8] B. Hale, D. L. Van Bossuyt, N. Papakonstantinou, and B. O'Halloran, "A zero-trust methodology for security of complex systems with machine learning components," in *International design engineering technical conferences and computers and information in engineering conference*, vol. 85376. American Society of Mechanical Engineers, 2021, p. V002T02A067.

[9] D. Greenwood, "Applying the principles of zero-trust architecture to protect sensitive and critical data," *Network Security*, vol. 2021, no. 6, pp. 7–9, 2021.

[10] C. A. Iordache, A. V. Dragomir, and C. V. Marian, "Public institutions updated enhanced biometric security, zero trust architecture and multi-factor authentification," in *2022 International Symposium on Electronics and Telecommunications (ISETC)*. IEEE, 2022, pp. 1–4.

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

[12] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[13] G. Magklaras and S. Furnell, "Insider threat prediction tool: Evaluating the probability of it misuse," *Computers Security*, vol. 21, no. 1, pp. 62–73, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404802001098

[14] A. Garg, R. Rahalkar, S. Upadhyaya, and K. Kwiat, "Profiling users in gui based systems for masquerade detection," in *Proceedings of the 2006 IEEE Workshop on Information Assurance*, vol. 2006, 2006, pp. 48–54.

[15] K. Tabia and S. Benferhat, "On the use of decision trees as behavioral approaches in intrusion detection," in *2008 Seventh International Conference on Machine Learning and Applications*. IEEE, 2008, pp. 665–670.

[16] G. Pannell and H. Ashman, "User modelling for exclusion and anomaly detection: a behavioural intrusion detection system," in *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings 18*. Springer, 2010, pp. 207–218.

[17] J. Kim, M. Park, H. Kim, S. Cho, and P. Kang, "Insider threat detection based on user behavior modeling and anomaly detection algorithms," *Applied Sciences*, vol. 9, no. 19, p. 4018, 2019.

[18] T. Prarthana and N. Gangadhar, "User behaviour anomaly detection in multidimensional data," in *2017 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*. IEEE, 2017, pp. 3–10.

[19] L. Cai and Y. Zhu, "The challenges of data quality and data quality assessment in the big data era," *Data science journal*, vol. 14, 2015.

[20] L. Jun and Z. Peng, "Mining explainable user interests from scalable user behavior data," *Procedia Computer Science*, vol. 17, pp. 789–796, 2013.

[21] S. Khanna, "Computer vision user entity behavior analytics," *arXiv preprint arXiv:2111.13176*, 2021.

[22] J. Kaur, K. Kaur, S. Kant, and S. Das, "Ueba with log analytics," in *2022 3rd International Conference on Computing, Analytics and Networks (ICAN)*. IEEE, 2022, pp. 1–7.