A Survey on Zero Trust Architecture: Challenges and Future Trends

Yuanhang He,1 Daochao Huang,2 Lei Chen,1 Yi Ni,1 and Xiangjie Ma1

1No.30 Research Institute of China Electronics Technology Group Corporation, Chengdu, China
2National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), Beijing, China

Correspondence should be addressed to Lei Chen; chenl_ccsc@163.com

Received 29 March 2022; Accepted 9 May 2022; Published 15 June 2022

Academic Editor: Yan Huo

The traditional perimeter-based network protection model cannot adapt to the development of current technology. Zero trust is a new type of network security model, which is based on the concept of never trust and always verify. Whether the access subject is in the internal network or the external network, it needs to be authenticated to access resources. The zero trust model has received extensive attention in research and practice because it can meet the new network security requirements. However, the application of zero trust is still in its infancy, and enterprises, organizations, and individuals are not fully aware of the advantages and disadvantages of zero trust, which greatly hinders the application of zero trust. This paper introduces the existing zero trust architecture and analyzes the core technologies including identity authentication, access control, and trust assessment, which are mainly relied on in the zero trust architecture. The main solutions under each technology are compared and analyzed to summarize the advantages and disadvantages, as well as the current challenges and future research trends. Our goal is to provide support for the research and application of future zero trust architectures.

1. Introduction

Since human beings entered the information age, the problem of information security has been perplexing with the further development and practical application of information technology. In 2010, the Stuxnet virus attacked the supervisory control and data acquisition (SCADA) equipment designed to destroy Iran’s nuclear fuel enrichment process. The attack successfully destroyed SCADA equipment in many parts of Iran. The “Stuxnet” virus has achieved the destruction of a single piece of information and data to the actual physical facilities, which marks the entry into a new stage of cyber warfare. In 2015, a cyberattack on Ukraine’s power grid caused a massive blackout in western Ukraine, leaving a fifth of the Ukrainian capital without power. This is the first-ever cyberattack posing a danger to a nation’s critical infrastructure. Intuitively, network security issues have threatened industrial control and infrastructure.

To solve this problem, the perimeter-based security architecture is proposed, which divides the network into internal network and external network with a firewall, intrusion detection system (IDS), or intrusion prevention system (IPS) as the border. According to the physical location of the object, it is judged whether it is located in the internal network, and the object in the internal network is regarded as trusted by default. External objects, on the other hand, must be authenticated before they can be trusted. In a perimeter-based security architecture, once an object is authenticated, it is trusted for a long time. Therefore, if a malicious object is authenticated, it can continue to attack and sabotage the internal network. At the same time, with the continuous development of cloud computing and Internet of Things technologies and the popularization of telecommuting, especially since the outbreak of the new crown, telecommuting has become an indispensable way of working. Therefore, based on the
Among them, the identity authentication mainly realizes the safe and efficient access of the ZTA object in the ZTA, the access control algorithm of ZTA is the key technical cornerstones of ZTA. It is different from the perimeter-based security architectures; the trust of an object is independent of its physical location and all objects are untrusted by default. The trust of an object can only be obtained by identity authentication and trust evaluation. After the system assigns the relative permissions to the object, the object can perform related operations. In recent years, zero trust architecture has been initially applied, and the most typical example is Google’s BeyondCorp model [2]. In this model, first, users need to perform location-based identity authentication. For example, in the public network, single-point SSO is used for authentication. Authorization is also required after authentication, and the access authority can be obtained only after the authorization is successful, and the authority is obtained by granting the authorization information to the access agent through the access control engine. BeyondCorp will associate the results (people and devices) to the vnet network segment constructed based on specific services after identity authentication, forming isolation domains with different security levels. If users want to access cross-domain, they must abide by relevant security policies. In addition, zero trust also has some preliminary applications in civil aviation airport network security and virtual power grids [3].

Identity authentication, access control, and trust evaluation algorithms are the technical cornerstones of ZTA. Among them, the identity authentication mainly realizes the identification of the object in the ZTA, the access control mainly realizes the safe and efficient access of the ZTA object to the resources, and the trust evaluation algorithm realizes the evaluation of the trust degree of the ZTA object and is used as the main credential for identity authentication and access control. At present, the research of ZTA is still in the preliminary stage, and the research on the architecture, identity authentication, access control, and trust evaluation algorithm of ZTA is the key field of it. Therefore, this paper introduces the current research status of ZTA development from these four aspects and discusses the main problems they face and future research directions. Our main contributions include:

(i) This paper makes a detailed analysis and summary of the current status of zero trust research. It focuses on the current research status of ZTA architecture, identity authentication, access control, and trust evaluation algorithms, and makes a specific analysis and summary, so as to grasp the overall situation of zero trust research

(ii) Comparing the current main ZTA, identity authentication, access control, and trust evaluation algorithms, and summarize their advantages and main problems

(iii) According to the main advantages and disadvantages of current zero trust in architecture, identity authentication, access control, and trust evaluation algorithms, the main challenges they face are summarized, and the main research directions of zero trust in the future are proposed

The next arrangement of this paper: We introduce the ZTA in Section II, including its main components and operation methods. In Section III, we review the relevant literature on the research status of zero trust including zero trust control and trust evaluation algorithms. We summarize the current progress of the main schemes, and give their comparisons, as well as future research directions in Section IV. Finally, we make a summary of this paper.

2. Zero Trust Architecture

ZTA was proposed by Kindervag, principal analyst at Forrester in 2010. In a zero trust architecture, all traffic cannot be trusted, and location cannot be used as a basis for security. Instead, security measures need to be taken for all access, minimum authorization policies and strict access control are adopted, and all traffic needs to be visualized and analyzed. These concepts are significantly different from the traditional perimeter-based security architecture, and the security is stronger.

2.1. Zero Trust Basic Assumptions and Principles. ZTA is built on the following five basic assumptions:

(1) The network is in a dangerous environment all the time

(2) There are external or internal threats in the network from beginning to end

(3) The location of the network is not enough to determine the credibility of the network

(4) All devices, users, and network traffic should be authenticated and authorized

(5) Security policies must be dynamic and calculated based on as many data sources as possible

Based on the above assumptions, the zero trust model is believed to adhere to the following four basic principles

(1) Authenticate users: Assess user security based on location, device, and behavior to determine if the user is who they claim to be. Take appropriate measures (such as multifactor authentication) to ensure user authenticity

(2) Authenticate devices: Whether it is corporate devices, BYOD or public hosts, or laptops or mobile devices, implement access control policies based on device identity and security. Only trusted endpoints are allowed to access company resources
(3) Restrict access and permissions: If users and devices are authenticated, implement a role-based access control model for resources, giving them the minimum permissions to complete the work at the time.

(4) Adaptive: Various sources (such as users, their devices, all activities related to them) are always producing information continuously. Leverage machine learning to set context-sensitive access policies, automatically adjust and adapt to policies.

2.2. Zero Trust Architecture. The core goal of zero trust is to allow users in untrusted network areas to access trusted areas through authentication and policy control, as shown in Figure 1.

In order to reduce the security risk of the access process, a continuous dynamic security access control technology is required, which is not based on the network location of the access subject, but authorizes the access subject based on the security and trust status before each access object is allowed to access; continuously monitors the security of the entire access process and assesses the trust status; dynamically adjusts access rights and implements fine-grained security access control.

To achieve this goal, more network elements are needed to support the entire zero trust architecture. The ZTA architecture given by NIST is shown in Figure 2.

Among them, the identity management (ID management) system and the enterprise public key infrastructure (PKI) are mainly used for the authentication of personnel and equipment, which is the basis. The data access policy mainly provides resource access policy, and the security information and event management system (SIEM system) provides the security information and event management of the entire architecture. At the same time, to integrate capabilities such as industrial compliance policies and threat detection, more attention needs to be paid to continuous diagnostics and mitigation (CDM) systems.

In general, the ZTA is based on identity, giving digital identities to people and devices, and setting minimum permissions for access subjects; aiming at business security, realizing business concealment, transmission encryption, and fine control; with continuous trust assessment as the guarantee, including user trust assessment, environmental risk determination, and abnormal behavior discovery; using dynamic permission control as a means, including attribute-based access control baseline, trust level-based hierarchical access, and risk-aware dynamic permissions.

The zero trust architecture focuses on the security capabilities of identity, trust, access control, permissions, and other dimensions, and these security capabilities are also an indispensable part of the information-based business system, so zero trust is inherently a kind of "endogenous security." In a sense, it is a spiral sublimation of business and security. From the initial business system to complete business goals, security equipment realizes the mutual independent system of security assurance, and integrates into a close relationship between security and business, and returns to security and application again.

3. Literature Review

In this section, we review and analyze the current status of important technical research on zero trust security. It mainly includes ZTA, identity authentication, access control, and trust evaluation algorithms.

3.1. Review on Zero Trust Architecture. As early as 2010, Kindervag [4] proposed the concept of a zero trust architecture model and a method to implement it in a practical environment and innovatively proposed a zero trust architecture based on "Data Acquisition Network" (DAN) in the paper. DAN helps to extract network data to the management center and then realizes inspection and analysis of it in real time, thus realizing the concept of zero trust, but this is also accompanied by problems of higher network complexity and increased user communication delay.

After that, in 2016, DeCusatis et al. [5] proposed a zero trust method based on transport access control. This method is based on steganography and overwriting, and the authentication token is embedded in the TCP request packet and the first authentication packet. Thus realizing the concept of zero trust, this approach increases the security of enterprises in cloud computing environments and prevents unwanted fingerprinting of protected resources; this approach provides protection at layer 3/4 but not at layer 7.

Subsequently, in 2020, Rose et al. [6] summarized the existing basic zero trust architecture schemes and proposed the basic logical components of the zero trust architecture. In addition, the author paid more attention to the implementation of the zero trust architecture, considering the realization of ZTA. Rather than a massive replacement of infrastructure or processes, it is a process that proposes specific steps to apply ZTA to a perimeter-based architecture network.

Sultana [7] et al. proposed a secure medical image sharing system based on the principle of zero trust and blockchain technology. The system combines zero trust with blockchain. The blockchain is used to protect sensitive information. Comprehensive protection of medical data, but this also increases the complexity of the system and needs to be studied in terms of efficiency.

Weever et al. [8] proposed a zero trust network security model in a containerized environment, which solved how to implement zero trust for “east-west” traffic between microservices in a containerized environment, using Kubernetes and Istio service mesh to build. A zero trust model in containerized environments reduces data leakage in containerized environments, but this model does not implement behavioral analysis and data leakage detection.

In 2021, Ramezanpour and Jagannath [9] proposed an artificial intelligence-based zero trust architecture (i-ZTA), which uses artificial intelligence for intelligent detection,
Policy decision, a zero trust system control system based on identity rather than network location. The trust architecture is a key supporting component for realizing zero trust. Authentication infrastructure is for future work.

Tian and Song [10] proposed a zero trust approach based on BLP and BIBA models, which conducts comprehensive trust scores for system components such as users, terminals, channels, files, and applications, and requires confidentiality and integrity, setting different weights to achieve better confidentiality and integrity protection of objects, but does not consider the initial trust value granting method of users, terminals, environments, objects, and other entities, and cannot effectively avoid human-factor errors in initial trust. The assignment, as well as the completeness and rationality of the list weight assignment, is for further study.

Ghate et al. [11] proposed an advanced zero trust architecture that leverages generalized attribute relation extraction to automate fine-grained access control to achieve low-cost fine-grained access control, performance, scalability, and security for real enterprise networks. The evaluation is for future work.

A comparison of these ZTAs is shown in Table 1.

3.2. Review on Identity Authentication. Authentication infrastructure is a key supporting component for realizing zero trust architecture’s identity-based capabilities. As an access control system based on identity rather than network location, a zero trust system first needs to give digital identities to people and devices in the network, and combines the identifiable people and devices at runtime to construct access subjects. Access subjects need to authenticate themselves before being granted access to specific resources. Identity authentication technology is a method or means used in the process of confirming the real identity of the visitor. Under the ZTA, the system continuously monitors the user’s behavior during access, and can adjust policies in real time according to the user’s behavior. The process of proving identity is primarily based on information obtained from the user, which can be described by a set of characteristics. The characteristics should be unique and permanent. The multifactor authentication technology based on multifeature identification will be widely used because of its stronger security.

According to the different types of features, we divide authentication into two categories: user-to-device authentication and device-to-device authentication.

User-to-device authentication: Methods of authentication based on biometrics have been widely proposed, and these biometrics can be used to prove the identity of users due to their uniqueness and persistence. Many researchers use human physiological attributes for identity authentication. In 2016, Kindervag et al. [4] proposed to use sensors and applications in Android smartphones to collect fingerprint information for identity authentication, but this method is only for some series of mobile phones and lacks comprehensiveness. The security of single-factor authentication is weak. In 2019, Henderson et al. [12] proposed a two-factor authentication that combines a fingerprint sensor and an LED pulse oximeter, aiming to optimize the shortcomings of a single fingerprint scan with an LED pulse oximeter. In 2015, Matsuyama et al. [13] proposed a method for continuous authentication using low-frequency brain signals, which measured changes in oxyhemoglobin in the brain by near-infrared spectroscopy (NIRS). But the study required placing probes on test subjects’ foreheads, which is not suitable for real-life use. In 2016, Mahbub et al. [14] proposed to utilize facial attributes captured by smart devices for continuous identity authentication. There are also studies on identity authentication based on user behavior characteristics. In 2018, Ehatisham-ul-Haq et al. [15] used behaviors such as walking, running, sitting, standing, and going upstairs and downstairs to distinguish different users. In 2020, Abuhmad et al. [16] used sensors in smartphones to capture user behavior patterns such as jogging and exercising while holding a smartphone, combined with contextual information such as text messages, voice, and video chat for authentication. And the method does not require any sensitive software or hardware permissions that could violate user privacy.

Device-to-device authentication: Due to the lack of biometrics, there are relatively few features available for authentication. But this approach assumes that device power consumption is linear and that battery capacity is also affected by the environment. Due to the heterogeneity of
IoT devices, not all devices are battery powered. In 2018, Chuang et al. [17] proposed to utilize the remaining battery capacity of the sensor device as a dynamic feature for authentication. In 2019, Wang et al. [18] proposed the use of electromagnetic radiation (EMR) for identity authentication between devices. However, electromagnetic radiation is easily disturbed by the external environment, so the characteristic factor selected by this scheme also has defects. Meng et al. [19] propose a D2D continuous authentication protocol combining blockchain and trust assessment, where the time interval of authentication is dynamic. The trust assessment center evaluates every device and outputs the trust level; the higher the trust level, the longer the interval of authentication is dynamic. The trust scheme is dynamic and flexible for each device with different trust levels.

To sum up, the user-to-device authentication can be verified by collecting the user’s biometric information by means of sensors in wearable devices such as mobile phones and watches. And using a variety of information for multifactor authentication can achieve higher security. However, due to the heterogeneity between devices, the identity authentication method suitable for device-to-device has few common features for authentication. It is necessary to further explore the unique and persistent features between devices.

After obtaining the identity information, both parties need to conduct a session to transmit information, and compare the obtained information with the information stored in the database to confirm the identity. According to the different transmission protocols, it can be divided into certificate-based authentication, encryption-based authentication, and nonencryption-based authentication.

Certificate-based Authentication: In 2013, Kothmayer et al. [20] proposed a two-way authentication security scheme for IoT based on the Datagram Transport Layer Security (DTLS) protocol, which uses RSA-based asymmetric encryption and X.509 authentication. However, the protocol requires 8 handshakes to establish a session, so the device needs to have higher computational cost and storage space to implement this solution. In 2016, Verma et al. [21] proposed a certificate-based protocol for node authentication in mobile ad hoc networks. The protocol utilizes trust management mechanisms to keep track of certificate operations and authentication operations and uses digital signatures with hash functions to maintain certificate authenticity. The protocol performs well in terms of robustness. In 2020, Kumar and Gandhi [22] proposed an enhanced DTLS based on smart gateways to overcome denial of service attack servers. The protocol uses packet loss rate to evaluate performance and based on data transfer and handshake time to evaluate protocol efficiency.

Encryption-based Authentication: In 2015, Shivraj et al. [23] proposed one-time password (OTP) authentication for IoT infrastructure. The protocol employs Identity-Based Elliptic Curve Cryptography (IBE-ECC) to provide lightweight end-to-end authentication between devices. The scheme is lightweight with smaller key size and smaller infrastructure, but performs poorly in terms of increased

<table>
<thead>
<tr>
<th>Literature</th>
<th>Implementation</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinnervag [4]</td>
<td>Extract network data to the management center using DAN</td>
<td>Inspect and analyze data in real time</td>
<td>Network complexity increases and user communication delay increases</td>
</tr>
<tr>
<td>DeCusatis et al. [5]</td>
<td>Using steganography and overwriting methods, the authentication token is embedded in the TCP request packet and the authentication header</td>
<td>Increased security for businesses in cloud computing environments and prevents unwanted fingerprinting of protected resources</td>
<td>No protection at layer 7, not comprehensive enough</td>
</tr>
<tr>
<td>Rose et al. [6]</td>
<td>Automatically link new APIs to existing service mesh categories by using machine learning-based smart association models</td>
<td>Simplify the creation, management, and monitoring of APIs</td>
<td>Difficult to achieve in real environment</td>
</tr>
<tr>
<td>Sultana et al. [7]</td>
<td>The system combines zero trust with blockchain, the blockchain is used to protect sensitive information, and zero trust realizes comprehensive protection of medical data</td>
<td>Combining blockchain with zero trust</td>
<td>Low efficiency</td>
</tr>
<tr>
<td>Weever et al. [8]</td>
<td>A zero trust network security model in a containerized environment</td>
<td>Reduced data leakage in containerized environments</td>
<td>Behavioral analysis and data leak detection are not implemented</td>
</tr>
<tr>
<td>Ramezanpour and Jagannath [9]</td>
<td>Using artificial intelligence for intelligent detection, assessment, and decision-making</td>
<td>Improve the efficiency of ZTA components in processing big data</td>
<td>Only at the theoretical level</td>
</tr>
<tr>
<td>Tian et al. [10]</td>
<td>Zero trust approach based on BLP and BIBA models</td>
<td>Set different weights based on confidentiality and integrity requirements</td>
<td>The weight distribution is not reasonable enough</td>
</tr>
<tr>
<td>Ghat et al. [11]</td>
<td>Automate fine-grained access control with generalized attribute relation extraction</td>
<td>Low cost</td>
<td>Failed to measure performance in a real environment</td>
</tr>
</tbody>
</table>
OTP size, increased computational complexity, and time-consuming performance. In 2016, Kumar et al. [24] proposed a lightweight authentication-based session key establishment protocol for smart home. The protocol requires a security service provider, which is a trusted server. The security service provider assigns important parameters, generates tokens, and distributes the tokens to communication devices. The protocol has high computational and storage efficiency and can defend against a variety of common attack behaviors, but feasibility monitoring is carried out through proof-of-concept implementations. In 2021, Syed et al. [25] proposed a lightweight continuous device-to-device authentication (LCDA) protocol that utilizes communication channel properties and a tunable mathematical function to generate dynamically changing session keys for continuous device-to-device authentication. An evaluation of the protocol using the Scyther tool shows that both the mutual authentication and the continuous authentication phases comply with security properties such as integrity, confidentiality, freshness, and resistance to protocol attacks. However, the effectiveness of this protocol on various constrained devices requires further research in the future.

Nonencrytpion-based Authentication: In 2015, Gope and Hwang [26] proposed a nontraceable authentication protocol in distributed IoT architectures. This scheme only uses hash functions and bitwise XOR operations to construct a lightweight authentication mechanism. The method is light in calculation and consumes less resources of the device. Taking sensor device movement into account, the scheme not only guarantees the legitimacy of sensor nodes but also supports identity anonymity and untraceability. In 2017, Ying and Nayak [27] proposed an anonymous and lightweight authentication based on the smart card (ASC) protocol to solve the authentication problem in in-vehicle ad hoc networks. ASC uses operations such as hash function and XOR to verify the legitimacy of the user (vehicle) and the verification of data messages. Utilizing a trusted authority to send anonymous certificates and keys to the vehicle, vehicle users must first authenticate and obtain a session key before they can interact securely with each other. The protocol has better efficiency in terms of communication/computational overhead, end-to-end delay, and packet loss rate. However, there are serious security problems in offline identity guessing attacks, session linking attacks, and replay attacks. In 2019, Chen et al. [28] aimed at the problem of offline identity guessing attack and time-consuming authentication phase in literature [27]. To improve security and reduce the time required for authentication, they proposed a patch on the protocol of Ying et al. The patched protocol performs better in terms of security and performance than the original protocol. The comparison of the methods used in the above papers is shown in Table 2.

3.3. Review on Access Control. In the early traditional access control model, role-based access control (RBAC) [29] was viewed as a task performed by a user, assigning him/her one or more roles to indirectly associate permissions with the user. It is considered an alternative to mandatory access control (MAC) and discretionary access control (DAC), which can realize centralized management of role membership and access control, but it only describes the characteristics of the subject and lacks the description of the characteristics of the object, the permissions cannot be dynamically changed, and the constraint particles are large, which can be. The scalability is not strong, and it is difficult to apply in a distributed environment. Attribute-based access control (ABAC) [30] is similar to RBAC in that it mainly grants or denies user requests based on arbitrary attributes of users and globally identifiable attributes of objects, but the disadvantage is that all elements need to be described in the form of attributes. Some relationships are not easily described with basic properties. Thomas et al. designed a task-based access control model (TBAC) [31] and proposed the concept of task-oriented. The model is to establish a secure access mechanism in the workflow, making it widely used in workflow systems and distributed computing systems. However, TBAC is not suitable for complex network environments. It does not involve the issue of user rights assignment, but simply introduces a set of trustees to represent the executors of tasks, and does not discuss how to determine trustees in the actual environment.

Attribute-based encryption (ABE) scheme to achieve access control is mainly studied from three aspects: the first is fine-grained access control, the second is the problem of user attribute revocation, and the third is the multi-authorization center scheme. In 2007, Oh and Park [31] proposed a detailed CP-ABE scheme, which embeds the access structure into the ciphertext and the attribute set of the data user into the private key, only when the attribute set satisfies the access structure, to decrypt the ciphertext. However, the security of this scheme is not ideal, and it is easy to be broken. In the same year, the nonmonotonic access policy ABE of Bethencourt et al. [32] allows the data owner to insert the revoked user ID into the ciphertext in the form of “non” when the data is confidential, so as to realize the revocation of the user’s access right to the ciphertext, but this scheme is less flexible. In 2009, in the scheme of Ostrovsky et al. [33], the data owner decides and manages the authorized user list of each attribute, and realizes the revocation of user attributes by sending out the user representation from the authorized list of attributes, using the idea of encrypting the broadcast. Direct revocation is introduced into ABE, which ensures that unrevoked users are not affected and do not need to update keys for users regularly. In 2017, Attrapadung and Imai [34] et al. proposed a generalized software-defined storage (SDS) constrained access control method for cloud storage; the method is based on CP-ABE with hidden access policy. SDS constraints are handled through the participation of additional entities and additional human attributes. But it does not completely hide the constraint policy structure from all entities without affecting SDS constraint enforcement. Constraints such as “and,” “or,” and “threshold” are not considered. In 2018, Nurmat et al. [35] proposed a fine-grained access control scheme based on location server for the mobile cloud environment, which takes the dynamic location of the mobile user as the user’s information and the location range as the access policy. And will satisfy the demands for the dynamic location of
the access policy is authorized to users for privacy protection. However, this solution introduces a third-party location-based services (LBS). If the LBS is maliciously damaged, the entire system will crash.

Due to the diversity of user devices and the consistency of access policies, the zero trust access control model requires support for dynamic network access. The first zero trust architecture was proposed by Google. After Google suffered a highly sophisticated APT attack in 2009, it began to redesign the security architecture for employees and devices to access internal applications, so that employees can achieve secure access at any location. No traditional VPN required.

<table>
<thead>
<tr>
<th>Literature</th>
<th>Methods</th>
<th>Continuous authentication</th>
<th>Multifactor authentication</th>
<th>Strengths</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kothmayr et al. [20]</td>
<td>Datagram Transport Layer Security (DTLS) protocol, RSA-based asymmetric encryption, X.509 authentication</td>
<td>No</td>
<td>No</td>
<td>The system architecture follows the IoT model and inherits the security properties of UDP.</td>
<td>The protocol has eight handshakes, which is computationally expensive</td>
</tr>
<tr>
<td>Verma et al. [21]</td>
<td>Certificate</td>
<td>Yes</td>
<td>No</td>
<td>The protocol has better performance in terms of throughput, end-to-end delay, and packet loss. Has a small amount of computation and communication overhead</td>
<td>No discussion of resilience to foreign attacks</td>
</tr>
<tr>
<td>Kumar and Gandhi [22]</td>
<td>Certificate, Advanced Encryption Standard Counter and Cipher Block Chain Message Authentication Code (AESCCM), Elliptic Curve Digital Signature Algorithm (ECDSA)</td>
<td>No</td>
<td>No</td>
<td>Overcome the denial of service attack server vulnerable to DTLS protocol</td>
<td></td>
</tr>
<tr>
<td>Shivraj et al. [23]</td>
<td>Elliptic Curve Cryptographic (ECC)</td>
<td>No</td>
<td>No</td>
<td>The protocol is scalable, with small keys and robustness</td>
<td></td>
</tr>
<tr>
<td>Kumar et al. [24]</td>
<td>Symmetric key, hash function</td>
<td>No</td>
<td>No</td>
<td>The scheme provides important security properties, including protection against a variety of common attacks, such as denial of service attacks and eavesdropping attacks</td>
<td>Preliminary evaluation and feasibility testing was carried out through the implementation of the proof of concept</td>
</tr>
<tr>
<td>Syed et al. [25]</td>
<td>Cryptography</td>
<td>Yes</td>
<td>No</td>
<td>The protocol can be adapted to devices with limited computing and storage resources</td>
<td>Difficulties in measuring Channel State Information (CSI) for heterogeneous IoT devices</td>
</tr>
<tr>
<td>Gope and Hwang [26]</td>
<td>Hash function, XOR</td>
<td>Yes</td>
<td>No</td>
<td>The protocol provides more security features under the premise of ensuring less computational overhead, with anonymity and nontraceability</td>
<td>Security analysis is just a proof by means of theoretical analysis</td>
</tr>
<tr>
<td>Ying and Nayak [27]</td>
<td>Hash function, XOR</td>
<td>No</td>
<td>No</td>
<td>An efficient password modification phase that does not rely on TA (trusted authority) and third-party servers is proposed, which can resist offline password guessing attacks.</td>
<td>There is no reasonable extension of the protocol, and the protocol is insecure against offline identity guessing attacks, session link attacks, and replay attacks</td>
</tr>
<tr>
<td>Chen et al. [28]</td>
<td>Hash function, XOR</td>
<td>No</td>
<td>No</td>
<td>Fixed the security vulnerability found in [27]</td>
<td>The protocol only uses the iPhone as a test platform.</td>
</tr>
</tbody>
</table>
In order to meet the needs of the company's internal mobile office, Google designed and implemented a relatively stable zero trust network model BeyondCorp [36], in which the authorized access rights after identity authentication need to be obtained by granting the access agent authorization information through the "access control engine." It adopts the ZTN method for access control, but does not describe in detail the implementation of policy language, risk management, or decision-making continuity, and does not fully consider the inheritance and reuse of existing networks, just visualize its security capabilities as a product.

In 2016, Ward and Beyer used transmission access control and first packet authentication to realize zero trust cloud network. Based on the principle of zero trust network, they redesigned the network architecture of data center and demonstrated the principle of zero trust through transmission access control system. One of the most important principles is steganographic overlay, which embeds authentication tokens into TCP packet requests and first packet authentication. The system can be used as part of a defense-in-depth strategy to strengthen the security of protected resources in enterprise computing and cloud environments, preventing protected resources from being unnecessary. However, it has not conducted penetration testing and cannot guarantee the emergence of other vulnerabilities.

In 2016, DeCusatis et al. [5] took advantage of the advantages of software-defined network (SDN) to centrally control traffic, designed a trust-based network access path dynamic authorization technology, established a user's trust degree hierarchy, analyzed the user's trust degree in real time, and based on the user's trust degree. Trust and security make real-time adjustments to defense paths. However, it only studies the target user behavior measurement indicators and measurement algorithms, and does not provide a feasible zero trust system implementation architecture, and the proposed indicators and algorithm measurements are not detailed enough, and do not consider highly concealed foul behaviors.

In 2018, Vanickis et al. [37] developed the FURZE system, a risk-adaptive access control policy implementation framework based on Kandalba's policy modeling method, to support future security requirements. In FURZE, the application of decision continuity imposes requirements on control functions to maintain session state information so that access control can adapt to the environment or other influencing factors, thereby changing the balance between operational requirements and security risks, and triggering policy re-opening and evaluate. However, it does not include development language tools and is not integrated into the existing professional dynamic program (PDP) system, so the stability of the runtime mechanism cannot be guaranteed.

In 2020, Yao et al. [38] proposed a dynamic access control and authorization system based on a zero trust security architecture. The system uses the TBAC model, and its user portrait and user trust are generated according to user behavior. The system adopts real-time hierarchical control in different scenarios to achieve dynamic and fine-grained access control and authorization. However, due to the influence of the TBAC model, tasks and roles cannot be clearly separated, and passive access control and role hierarchy are not supported.

In 2021, da Silva et al. [39] proposed zero trust access control with context awareness and behavior-based continuous authentication for smart homes. A zero-aware smart home system is proposed to provide access control to the smart home system using zero trust continuous identity authentication to continuously verify the authenticity of the user, powered by edge computing to eliminate unreliable service providers and access from any means. However, it has not been applied to the actual environment, the impact of latency and concurrency has not been tested, and the accuracy cannot be guaranteed. In the same year, Hatakeyama et al. [40] proposed a new access control model for zero trust networks, which does not assume trusted properties such as source networks. And to verify and evaluate whether the user requesting access is worth relying on each access request, based on the evaluation results, consider the decision of whether to allow access. However, it does not standardize the format and semantics of the context in ZTF and cannot operate the authorization server and the identifier used when the context cannot be shared. In the same year, Mandal [41] et al. proposed a cloud-based zero trust access control strategy by establishing a MAC spoofing defense mechanism in the SDN framework of the cloud architecture to support the work-from-home approach driven by COVID-19. When changes for the access control strategy of the enterprise structure are required, it shows greater accuracy by examining source TCP/IP traffic and corresponding MAC addresses, collecting individual network traffic from untrusted zones. Its AI-based models help lower thresholds and normalize traffic when the network is growing rapidly. However, under the security threat of advanced attackers, the optimal security of lowering the threshold and cloud resources cannot be guaranteed, and the time-consuming nature of analyzing traffic and removing deceived users has not been resolved. Yang et al. [42] proposed an adaptive dynamic access control model based on blockchain and short-term tokens, introduced user trust assessment into the role-based access control model, and used a deep learning-based user abnormal behavior detection algorithm to dynamically evaluate user behavior and update trust, and realize corresponding access rights adjustment on the basis of dynamically updating short-term tokens, but it also has the common problems of RABC: it is difficult to establish an initial role structure and lack of flexibility in IT technology.

Early security policies were divided into two types: discretionary access control (DAC) and mandatory access control (MAC) [43]. However, with the development of computer and network technology, DAC and Mac can no longer meet the needs of practical applications. Therefore, role-based access control (RBAC), object-based access control (OBAC), and task-based access control (TBAC) have emerged. However, with the emergence of new computing environments such as cloud computing and Internet of things, some of its characteristics have brought great challenges to the application of access control technology, which
makes the traditional access control model for closed environment difficult to apply to the new computing environment. Facing the new computing environment with massive, dynamic, and strong privacy [44], the efficiency is very low. The subsequent access control model based on attribute [30] (ABAC) takes the attributes of the subject and object as the basic decision-making elements, and can flexibly use the attribute set of the requester to decide whether to grant its access rights. In addition, the strong expansibility of ABAC enables it to be combined with data privacy protection mechanisms such as encryption mechanism. On the basis of realizing fine-grained access control, protect user data from analysis and disclosure, such as attribute-based encryption (ABE) [45]. The detailed comparison of common access control models is shown in Table 3.

3.4. Review on Trust Assessment Algorithm. A continuous, fine-grained evaluation model is an important part of implementing a zero trust system. The trust evaluation module accepts all kinds of security data monitored and collected by the auxiliary platform, analyzes and judges the data, and forms the trust value of the access request. This trust value will serve as the key basis for the authorization mechanism. The process of quantifying trust will be time-shifted to meet the requirements of high dynamics. Moreover, the level division of trust assessment is also more refined than the traditional model.

Scholars have greatly improved the performance of various aspects of trust assessment in combination with emerging technologies in recent years.

As early as 2018, Jayasinghe U and others [46] and others applied machine learning technology to the node trust evaluation framework, implemented a data credibility labeling method based on unsupervised learning technology, and based on this node trust labeling method. The corresponding trust prediction model is established, which not only improves the accuracy of the trust evaluation algorithm but also improves the ability of the trust evaluation technology to identify trusted interactions. On the basis of the success of this technology, the author also proposes map reduction and data parallelism as research directions, trying to solve the scalability problem existing in the current model.

In 2019, Gao Z and others [47] designed a multidimensional adaptive trust evaluation mechanism using edge computing for nodes with weak computing power that widely exist in the network (especially local edge computing networks), which improved node trust. The robustness of the evaluation, but the flexibility of the model, is not high and should be improved in future research. In the same year, Boussard M and others [48] applied blockchain technology to trust evaluation and proposed a specific trust evaluation framework suitable for home IoT networks, realizing efficient trust evaluation of smart products in home IoT networks. But the single blockchain that this research relies on may be deployed on multiple chains or implemented through channels, and this technology has not yet been successfully developed, so scalability issues remain to be solved.

Using the concept of fuzzy logic, Guleng S and others [49] designed a direct trust evaluation method and further proposed a multiagent trust evaluation scheme, considering the characteristics of the entire trust forwarding chain (trust value, forwarding hops, etc.), which improves the accuracy of trust assessment. However, the survivability of the model under abnormal conditions is weak, especially considering the complexity of routing decisions with various constraints such as mobility, bandwidth, link quality, and reliability, and the robustness of this design needs to be improved. In the same year, Rani R and others [50] used game theory to design a lightweight trust evaluation scheme, which not only improved the success rate of malicious node detection but also improved the energy efficiency of trust evaluation. However, the types of external attacks that can be resisted are relatively single, and the comprehensiveness of the protection of trust assessment is relatively poor.

In 2020, Chuan T et al. [51] proposed a method for implementing the concept of zero trust, which described seven evaluation elements of zero trust evaluation: required procedures (including weak password detection procedures, website detection procedures, configuration detection procedures, host vulnerability detection programs, brute force protection programs, hardening programs, mandatory access control programs, and micro-isolation control programs), operating system security vulnerabilities, network security vulnerabilities, weak passwords, high-risk ports, sensitive information protection, and accounts and passwords.

In 2021, Basta N and others [52] adopted microsegmentation technology, which limited the attacker's ability to move laterally in the network by binding fine-grained security policies to a single workload, and initially realized trust assessment under the concept of zero trust. Furthermore, the authors develop an analytical framework to describe and quantify the effectiveness of microsegmentation in enhancing network security. In the same year, in 2021, Zhang Yi and others [53] proposed a trust evaluation optimization mechanism using the fuzzy network analysis method, which effectively refined the evaluation granularity, scientifically quantified the behavioral trust value of users in the cloud computing environment, and improved the evaluation of the objectivity of trust assessment techniques. In the same year, Papakonstantinou N and others [54] used the concept of zero trust to provide a multidisciplinary early design risk assessment framework for early joint safety and security assessment based on the system interdisciplinary dependency model, which more accurately estimated the probability of successful attacks on system components. In the same year, Ramezanpour K and Jagannath [9] adopted reinforcement learning to perform three tasks of trust assessment (i.e., providing an initial network environment risk assessment, learning unnecessary communication flows in devices, and providing models for device communication patterns) and used graph neural networks. The zero trust assessment is modeled by simulating the state of the 5G network, and the application of artificial intelligence in the zero trust assessment is discussed. The application of the intelligent zero trust architecture mentioned by the author in the
Table 3: Comparison of access control models.

<table>
<thead>
<tr>
<th>Name</th>
<th>Model introduction</th>
<th>Advantage</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomous access control (DAC)</td>
<td>User centered, allowing users to control file access without specifying rules in advance</td>
<td>It is very flexible and can assign access rights between principals and objects</td>
<td>System maintenance and verification of safety principles are very difficult</td>
</tr>
<tr>
<td>Mandatory access control (MAC) [44]</td>
<td>Users cannot customize permissions, and access control policies are managed in a centralized manner</td>
<td>Limitations of customer service DAC model</td>
<td>Rely on trusted components</td>
</tr>
<tr>
<td>Role-based access control (RBAC) [29]</td>
<td>Assign multiple roles to users and give them permissions and responsibilities as principals</td>
<td>Central management with role members and ACS</td>
<td>Difficult to establish initial role structure and lack of flexibility in IT technology</td>
</tr>
<tr>
<td>Organization-based access control (ORBAC)</td>
<td>A more abstract control strategy. It is designed to address topics, objects, and actions. Policies determine which subjects have some actions to access certain objects</td>
<td>Eliminate conflicts between security rules</td>
<td>Vulnerabilities vulnerable to certain types of attacks</td>
</tr>
<tr>
<td>Task-based access control (TBAC) [31]</td>
<td>Implement different access control policies for different workflows or different tasks that agree to workflows</td>
<td>When a task is introduced, it can be authorized actively and represent the change of task status</td>
<td>Tasks and roles cannot be clearly separated, and passive access control and role hierarchy are not supported</td>
</tr>
<tr>
<td>Attribute-based access control (ABAC) [30]</td>
<td>Approve or reject user requests based on any attributes of the user and selected attributes of objects that may be globally recognized</td>
<td>Subjects can access a wider range of objects and flexibly assign policies and security features</td>
<td>It is difficult to calculate the final permission set of a given user effectively</td>
</tr>
<tr>
<td>Policy-based access control (PBAC)</td>
<td>A method of combining roles and attributes with logic to create flexible dynamic control strategies</td>
<td>Flexibility with fine-grained or coarse-grained</td>
<td>Imperfect conflict detection mechanism</td>
</tr>
<tr>
<td>Use control (UCON)</td>
<td>It contains three basic elements: subject, object, authority, and three other elements related to authorization: authorization rules, conditions, and obligations</td>
<td>Support trust management and digital rights management, add subject and object attributes, and control them in the process of topic access</td>
<td>Delegation without permission description, explicit management description, and temporal description</td>
</tr>
</tbody>
</table>

In identity authentication, because single-factor authentication has only one unique factor for identity authentication, once the unique password or biometrics is stolen, it will collapse completely, while multifactor authentication can improve the defects of single-factor and greatly reduce the threat of network attacks. Because even if the attacker intercepts the password information, the difficulty of obtaining the authorization of the second or third factor is greatly increased. However, the incomplete authentication information is not enough to access. Continuous authentication changes the way that visitors can access system information for a long time after one-time authentication in the initial stage. It continuously grants user resource access rights before and during the session, reducing the security risks caused by attackers in the middle of the session, to enhance the security of the system. From single-factor authentication to two-factor and multifactor authentication, from one-time authentication to continuous authentication, security continues to improve. In the future, multifactor authentication methods and continuous authentication methods will be widely used in zero trust architectures because of their better security. Whether based on certificates, encrypted authentication protocols, or nonencrypted protocols, different protocols have trade-offs between security and resource consumption. It aims to reduce resource consumption as much as possible in the authentication process under the premise of ensuring the security of the zero trust system, which is also the direction of the identity authentication part of the future zero trust architecture.

In recent years, the number and complexity of security attacks against enterprises have risen sharply and will continue to grow in the next few years, which will only increase the complexity of the computing environment. Therefore, the access control system needs to be dynamically adjusted, and the risk assessment has been incorporated into the access control process. The access control decision will include many factors: such as the trust degree of users and devices and the situational environment of users and
devices, i.e., location, time, task type, and the current security threat level in the user’s direct environment. In addition, the access level assigned to devices or users may change over time. The access control system must be able to judge the current trust level by consulting various data sources and making corresponding decisions. Therefore, risk-based access control should be used in many areas. The main result of this trend is to shift from the traditional perimeter-based security model to the application of the so-called zero trust network security model, and treat the enterprise intranet and the Internet to the same extent, that is, lack of trust. Under the zero trust model, we need to solve the problems of minimizing authorization and dynamic authorization control for users. The current zero trust access control model should not be limited to a certain access control model. Only by using a variety of technologies together can we achieve the required level and requirements of access control. For a period of time, RBAC and ABAC will still play an important role in the zero trust model.

Judging from the history of the development of trust assessment technology, the assessment has gradually changed from one-sided to comprehensive, the trust factors considered are more and more extensive, and the theories used in the assessment are gradually enriched, ranging from subjective logic to Bayesian, entropy theory, etc. It makes the trust evaluation close to the subjective evaluation from all directions. Accuracy and fine-grainedness are eternal topics of trust assessment in a zero trust security environment. Not only that, but in different network environments, trust evaluation algorithms also face various requirements. For example, for weak nodes in the network, improving the evaluation efficiency and saving computing resources are as important as ensuring the accuracy of the evaluation. In many applications in social network environments, it is also a trade-off between ensuring the comprehensiveness and accuracy of the evaluation and protecting the user’s private information. To sum up, in future research, not only will it be the consistent direction of efforts to continuously improve the comprehensiveness and subjectivity of trust assessment, but also to improve the dynamics and accuracy of trust assessment under the zero trust theory and to conduct the assessment according to the characteristics of different network environments. Adaptive adjustment is also an issue that cannot be ignored.

5. Conclusion

In this article, we expound the concept of zero trust and introduce the background and development of this technology. The core technologies relied on in the ZTA are analyzed in detail: identity authentication, access control, and trust assessment. We analyze and summarize the advantages and disadvantages of existing research on identity authentication, access control, and trust assessment, summarize the urgent problems and challenges of each technology, and propose future efforts and development trends. The work of this paper has guiding significance for the future migration of perimeter-based network security structure to ZTA.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


