TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 15/24 pp492–505
DOI: 10.26599/TST.2023.9010034
Volume 29, Number 2, April 2024

Endogenous Security Formal Definition, Innovation Mechanisms, and Experiment Research in Industrial Internet

Hongsong Chen*, Xintong Han, and Yiying Zhang

Abstract: With the rapid development of information technologies, industrial Internet has become more open, and security issues have become more challenging. The endogenous security mechanism can achieve the autonomous immune mechanism without prior knowledge. However, endogenous security lacks a scientific and formal definition in industrial Internet. Therefore, firstly we give a formal definition of endogenous security in industrial Internet and propose a new industrial Internet endogenous security architecture with cost analysis. Secondly, the endogenous security innovation mechanism is clearly defined. Thirdly, an improved clone selection algorithm based on federated learning is proposed. Then, we analyze the threat model of the industrial Internet identity authentication scenario, and propose cross-domain authentication mechanism based on endogenous key and zero-knowledge proof. We conduct identity authentication experiments based on two types of blockchains and compare their experimental results. Based on the experimental analysis, Ethereum alliance blockchain can be used to provide the identity resolution services on the industrial Internet. Internet of Things Application (IOTA) public blockchain can be used for data aggregation analysis of Internet of Things (IoT) edge nodes. Finally, we propose three core challenges and solutions of endogenous security in industrial Internet and give future development directions.

Key words: industrial Internet; endogenous security architecture; federated learning; blockchain

1 Introduction

In recent years, more and more information and communication technologies, such as the Internet of Things, cloud computing, edge computing, and mobile Internet, are emerging in front of the public, and have been widely used in industrial Internet[1]. At the same time, the industrial revolution is developing rapidly. Intelligent manufacturing will be integrated into industrial production and service, which puts forward higher requirements for traditional industrial manufacturing and products: Heterogeneous equipment and sensors are connected to the industrial network to efficiently store and analyze massive multi-source data, and to realize real-time, reliability, and security of industrial data and information[2]. industrial Internet uses big data analysis, artificial intelligence, and cloud computing technologies to conduct intelligent analysis and model construction of industrial data, which connect people, machines, and things to realize automatic control and industrial production.

Currently, the cyberspace security is easy to attack and difficult to defend. Traditional protection methods require continuous searching for all vulnerabilities in the system, patching and taking additional protection measures. Attackers only need to find one vulnerability in the system to invade, which results an extremely unequal situation of attack and defense. As a novel technique infrastructure, massive heterogeneous industrial devices connecting to industrial Internet will greatly increase the potential security vulnerabilities and

• Hongsong Chen, Xintong Han, and Yiying Zhang are with School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China. E-mail: chenhs@ustb.edu.cn; hanxt0830@163.com; s20200685@xs.ustb.edu.cn.

∗ To whom correspondence should be addressed.

Manuscript received: 2022-12-07; revised: 2023-03-21; accepted: 2023-04-26

risks[3]. Existing defense systems are essentially "static, similar, and certain", and are essentially defenseless against the ubiquitous threat of uncertainty. Besides traditional authentication and encryption techniques, there is lack of efficient countermeasures to unknown attack and security vulnerability.

Liu and Peng[4] considered identity authenticity as one of the security objectives of the 6G endogenous security system. Hu et al.[5] integrated bionic immunity and artificial intelligence into endogenous security system. Ji et al.[6] built the secure and trusted 6G endogenous security network using blockchain and zero-trust mechanism and conducted experimental analysis of lightweight encryption algorithms on resource-constrained devices. Therefore, combining the federated learning specific to this paper with the above techniques and methods, we focus on identification authentication and management, zero-trust mechanism, computer immunity, federated learning, blockchain, complexity and cost, and experiment as the comparison item. Existing comparative researches on endogenous security architecture and technologies are shown in Table 1.

Liu and Peng[4] proposed 6G endogenous security mechanism and described identity authentication technology of 6G endogenous security, such as access control technology, communication security technology, and data encryption technology. Hu et al.[5] considered endogenous security as a security defense mechanism based on the biological nervous system. The security mechanism is arranged in each security component of the network, and intelligent security center similar to the brain is constructed to make control decisions. Wei et al.[7] proposed a space-terrestrial integrated multi-identifier network. Blockchain node uses trusted cryptography module as key management and trust embedded identity authentication mechanism to achieve active immune security mechanism. Guo et al.[8] proposed the endogenous trusted resource intelligent sharing network architecture based on distributed alliance blockchain. Smart contract is used to realize on-chain identification and off-chain data scheduling and combination. Jiang et al.[9] studied human immune and neural reflex mechanisms, where neurons continuously collect information from modules. Errors are perceived through feedback, and then strategies are set at the decision level of the information system to solve or avoid the problem. Zhou et al.[10] called the ability to protect the defense mechanism itself as endogenous security and used the dynamic shuffling method of reinforcement learning to adaptively realize the unpredictability of defense strategies in the large-scale heterogeneous network environment.

However, the above works did not provide a unified and clear definition of endogenous security. They used artificial intelligence and artificial immunity to achieve endogenous security. But the technology is single and lacks authentication mechanisms, self-defense, and detection against unknown attacks, which cannot adapt to changes in the environment. With the emergence of the era of big data and industrial Internet, the large volume and high dimensionality of data make the early centralized defense methods face challenges such as high storage and high computational complexity. Large-scale data make model training slower and data security cannot be guaranteed. External detection systems and software cannot fully eliminate attacks and vulnerabilities. Firewalls cannot filter data in time.

Therefore, endogenous security defense system is needed to design threat perception and learning mechanisms through federated learning using the natural law of biological evolution. Immune feedback mechanism is obtained to maintain system stability in the complex environment. Endogenous keys and zero-trust mechanisms are used to simplify the traditional complex key management, and achieve secure and reliable distributed authentication. Users

Table 1 Research on endogenous security architecture.

Reference	Endogenous security architecture						
	Identification authentication and management[4]	Zero-trust mechanism[6]	Computer immunity[5]	Federated learning	Blockchain[6]	Complexity and cost[6]	Experiment[6]
[4]	√	×	×	×	×	Low	×
[5]	×	×	√	×	×	Medium	×
[7]	√	×	×	×	√	High	√
[8]	√	×	×	×	√	Medium	×
[9]	×	×	√	×	×	Low	√
[10]	×	×	×	×	×	High	√
Our architecture	√	√	√	√	√	Customization	√

dynamically select defense mechanisms according to their own requirements and scenarios to achieve dynamic adjustment of model complexity.

Our contributions are as follows:

• We propose the first formal definition of endogenous security in industrial Internet at the beginning of Section 2.

• A new endogenous security architecture for industrial Internet is presented and formally described. The unique feature of the architecture is the endogenous security layer, intelligent self-learning security, data full-cycle security, and interactive security. The innovation mechanism and cost analysis of endogenous security are considered simultaneously.

• On the basis of endogenous security and identify encryption, an improved clone selection algorithm based on federated learning is designed by using endogenous identify encryption and Proof of Trust (PoT) consensus mechanism. Devices that satisfy the trust level use trust attributes to implement Attribute-Base Encryption (ABE) and decryption of model parameters.

• We propose a blockchain-based cross-domain authentication mechanism to solve the cross-domain authentication problems between Identity-Based Encryption (IBE) domain and Public Key Infrastructure (PKI) domain in complex and heterogeneous industrial Internet scenarios. Based on the experimental results of two types of blockchains for this authentication mechanism, the on-chain and off-chain time is compared.

The rest of this article is structured as follows. Section 2 firstly gives the definition of endogenous security, and then introduces an industrial Internet endogenous security architecture proposed in this paper and functions among the various layers in the architecture. Section 3 discusses in detail the relevant technologies and mechanisms to achieve endogenous security in the architecture to provide strong support for intelligent self-learning security and interactive security. Next, Section 4 introduces the improved clone selection algorithm based on federated learning using endogenous encryption and ABE. Section 5 provides an analysis of the industrial Internet identity authentication threat model, illustrates cross-domain authentication mechanism, and conducts performance experiments based on two different blockchains. Section 6 summarizes the current challenges of industrial Internet endogenous security. Finally, Section 7 concludes the whole paper.

2 Endogenous Security Formal Definition and New Architecture in Industrial Internet

Based on the above, we give a scientific and comprehensive definition of endogenous security in industrial Internet: Endogenous security (x) is a kind of active security (a) and self-protection mechanism (y), which has diversity (d_1), built-in (i), dynamicity (d_2), and feed-back (f). It uses internal factors (p) and external factors (q) to enhance self-protection ability (T) under changeable environment (E). And it can obtain security function (z) through intelligent self-learning (m) and interactions (n), and attack and defense driving security capability enhancement (v) and evolution (r). The following formal description of the definition of endogenous security is made using propositional logic and predicate logic:

$$F(d_1, i, d_2, f) \to a,$$

$$a \wedge y \to x,$$

$$x \to E \wedge (p \vee q) \wedge T,$$

$$(m \wedge n) \vee G(v, r) \to x \wedge z.$$

where $\wedge$ denotes the conjunction, equivalent to the logical; $\vee$ denotes the disjunctive, equivalent to logical; $\to$ denotes the conditional; $F(d_1, i, d_2, f)$ is the predicate form of atomic proposition, which is the mechanism with diversity (d_1), built-in (i), dynamicity (d_2), and feed-back (f); a denotes the active security mechanism; and $G(v, r)$ is the predicate form of atomic proposition, which is the attribute of attack and defense driving security capability enhancement (v) and evolution (r).

Therefore, we propose a novel endogenous security architecture for industrial Internet. The architecture uses biological immunity mechanism and zero-trust mechanism to transform the development of real security technology into endogenous dynamic adaptation to achieve active detection, active alarm, and active removal. It avoids the risks caused by external attacks and internal vulnerabilities. The architecture can realize the access control of devices based on endogenous ID and key, monitor the whole life cycle of data, and realize lightweight security protection technology for massive heterogeneous devices in industrial Internet. Biological immune mechanism and blockchain detect unknown attacks through interactive security self-learning, ensure normal balance function, and realize self-immunity and

self-protection. This architecture is characterized by strong adaptability and good scalability, as shown in Fig. 1.

As can be seen from Fig. 1, data in the industry Internet are vulnerable to typical attacks throughout its life cycle, such as single-point fall attack, data privacy breach, and Botnet. Compared with the industrial Internet security architecture[11] proposed by the alliance of industrial Internet, the novelty of the architecture proposed in this paper is endogenous security, intelligent self-learning security, interactive security, and full-cycle data security. By self-learning and training, the architecture actively responses to security threats and adapts to complex environmental changes. There are eight layers in the architecture, which are denoted as NESA={DS, ES, CS, NS, AS, IS, ISS, FDS}. The specific description of each layer is as follows:

• Device security, DS={IS, II, IP}, includes various local devices such as Intelligent Sensor (IS), Intelligent Instrument (II), and Intelligent Product (IP). It can motor industrial equipment and industrial product changes.

• Endogenous security, ES={EKA, TC, ZTM, EIFL}, includes Endogenous Key and Authentication (EKA), Trust Computing (TC), Zero-Trust Mechanism (ZTM), and Endogenous Immunity by Federated Learning (EIFL). It can obtain perception and defense capabilities through endogenous key identity authentication and immune learning.

• Control security, CS={LI, PC, AnD} , enables Logical Isolation (LI) and Process Control (PC) between different data acquisition and monitoring control system operations, as well as Anomaly Detection (AnD) of data.

• Network security, NS={LANS, PNS}, can realize Local Area Network Security (LANS) and Public Network Security (PNS) for internal and external interconnections of the factory.

• Application security, AS={FES, FIS} ensures the security of various applications and platforms such as Factory Extranet Security (FES) and Factory Intranet Security (FIS) running in the industrial Internet.

• Interactive security, IS={DS-ES, ES-CS, CS-AS, AS-NS}, guarantees data transmission security between layers such as DS and ES (DS-ES), ES and CS (ES-CS), CS and AS (CS-AS), and AS and NS (AS-NS), and guarantees the security of all kinds of data such as interactive production data and user application data.

• Intelligent self-learning security, ISS={AtD, RL, FL}, implements Attack Detection (AtD) in industrial Internet by Reinforcement Learning (RL) and Federated Learning (FL), with endogenous key and trusted computing as the support to protect data security.

• Full-cycle data security, FDS={DC, DT, DS, DP, DED, DAD}, consists of Data Collection (DC), Data Transmission (DT), Data Storage (DS), Data Processing (DP), Data Exchange and Disclosure (DED), and Data Archiving and Destruction (DAD). The security and reliability of data at all stages are guaranteed through endogenous keys, zero-trust mechanism, blockchain, and intelligent self-learning.

The features and advantages of the architecture are the integration of bionic mechanisms, immune mechanisms, and zero-trust mechanisms to monitor real-time changes in devices and networks. The main achievements are endogenous security, intelligent self-learning security, interactive security, and data lifecycle security. The endogenous security combines blockchain and authentication technology on top of device security so that each device has an endogenous key. The zero-trust mechanism is used to identify malicious nodes, users, and traffic to ensure data security and network security. Endogenous security obtains endogenous immunity through intelligent self-learning, conducts the effective response, forms immune memory, and stores it in the blockchain distributed ledger to realize immune mechanism and ensure the reliability of the device. In the next section, this paper provides a detailed introduction to the various mechanisms implemented by endogenous security. Intelligent self-learning can enable multiple local devices to interact with the environment by combining classical reinforcement learning methods (such as Q-learning), iteratively optimize according to the obtained feedback information, change the local training objectives of the devices, and maximize the global expected return. The intelligent self-learning security conducts distributed learning models through federated learning with a large amount of industrial data to realize endogenous immunity. The intelligent self-learning security is the further enhancement of the endogenous security. There is a technology and mechanism intersection between them. Interactive security ensures data security between layers using zero-trust mechanism and blockhead. Data authentication, access control, traceability, and audit are implemented to ensure data security in all aspects of the whole life cycle. The three support and improve the security of all stages of the data lifecycle security.

And the traditional relationship between the other layers is that device security adopts strategies such

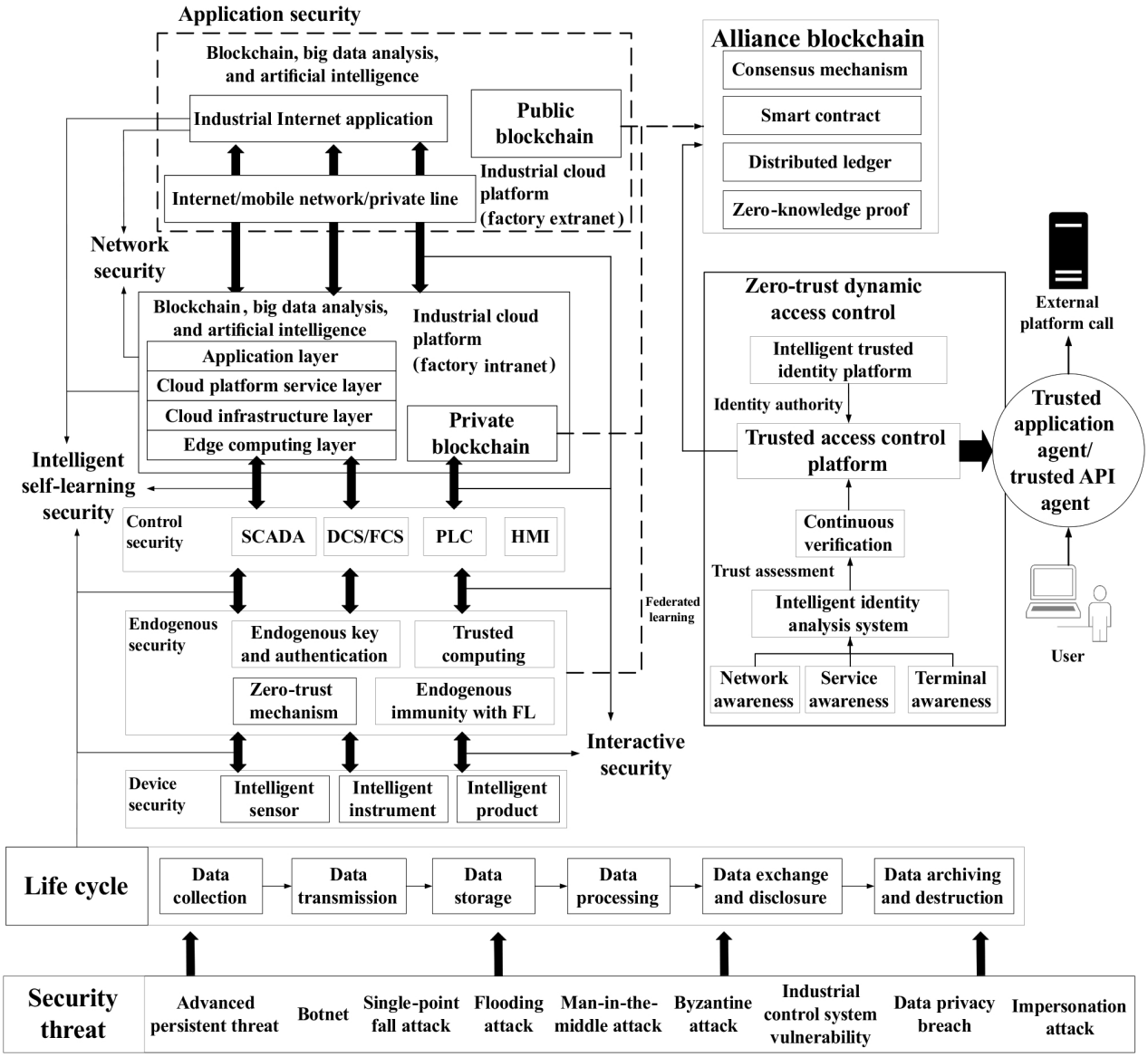

Fig. 1 A novel endogenous security architecture for the industrial Internet.

as firmware security enhancement and vulnerability repair to ensure the security of intelligent terminals and production equipment. Control security further protects equipment and communication security through various control security mechanisms and protocols. Network security is based on device security and control security to achieve communication and transmission boundary isolation, access authentication, and identity resolution security. Application security ensures device application security and cloud security through user authorization management, code security, and other policies.

The specific interactions between the seven layers can be done in accordance with international, national, or industry-related standards and white papers, such as security framework of industrial Internet of Things[12] and the architecture of industrial Internet of Things v2.0[11] published by the Alliance of industrial Internet. International standards include SM9 cryptographic algorithms Information Technology—Amendment 1: SM9 Mechanism[13].

To the cost analysis of endogenous security, we mainly analyze the four components of endogenous security. Endogenous key and authentication mainly use Static Random Access Memory, Physical Unclonable Function (SRAM PUF) to generate unique ID and key for each device, which can also be configured for most low-cost industrial devices. Trusted computing is implemented by measurement and verification through modules such as Trusted Cryptography Module (TCM) and Trusted Platform Module (TPM), but the hardware

cost is high. The zero-trust mechanism evaluates the trust level of the device for calculation, resulting in corresponding calculation cost. Both trusted computing and zero-trust mechanisms are suitable for edge devices and gateways with high computing power and storage capacity. Endogenous immune self-learning uses blockchain for consensus to generate communication costs. Encryption algorithms generate computation costs, which is more suitable for edge computing servers, intelligent gateways, and high-performance sensors.

3 Endogenous Security Innovation Mechanism in the Industrial Internet

In this section, we introduce the endogenous security layer in the novel architecture of endogenous security for the industrial Internet in Fig. 1, located between the device security layer and the control security layer. It provides key technical support for intelligent self-learning security and interactive security.

The structure of the endogenous security innovation mechanism for the industrial Internet is illustrated in Fig. 2.

The endogenous security innovation mechanism is divided into five main parts: device-based identification key and authentication, zero-trust mechanism, trusted computing and proof, blockchain supporting technologies, and endogenous immunity.

The device-based identification key and authentication section is applied to various industrial production devices mentioned in the device security hierarchy, including identification key generation and identification password algorithms. Trusted node devices use random physical differences within entities, such as the SRAM-based storage mechanism of Physical Unclonable Function (PUF) in combination with software identity to generate their own identity keys. SRAM PUF has the advantages of small cell area, low power consumption, and mobility, which can be realized simply by loading software onto the chips of different devices. Meanwhile, the identity passwords and encrypted state data are generated by the lightweight R-ate pair algorithm in the National Identity Cryptography Standard SM9.

The zero-trust mechanism is divided into control plane and data plane to establish trust on untrusted access with dynamic changes. The data plane transmits data through trusted application agents. The control plane provides authentication, trust assessment, continuous authentication, and access control functions to establish new security boundaries. In particular, continuous authentication means that after the device has been authenticated, it is authenticated irregularly to ensure the reliability of the equipment. The trusted computing and proof module can allow for the setting of binary or multi-valued trust assessment levels, working with zero-trust mechanism to complete the trust building and quantification. The supporting technologies of blockchain include smart contracts, distributed ledgers, heterogeneous consensus mechanisms, and zero-knowledge proofs. For the consensus mechanism, we propose PoT. PoT dynamically forms the final trust level through the identity trust level and behavior trust level. The identity trust level is obtained by continuous verification under zero-trust mechanism. Transactions stored in the blockchain evaluate the trust level based on the past behavior of the device. The higher the trust level of the node, the higher the probability of obtaining

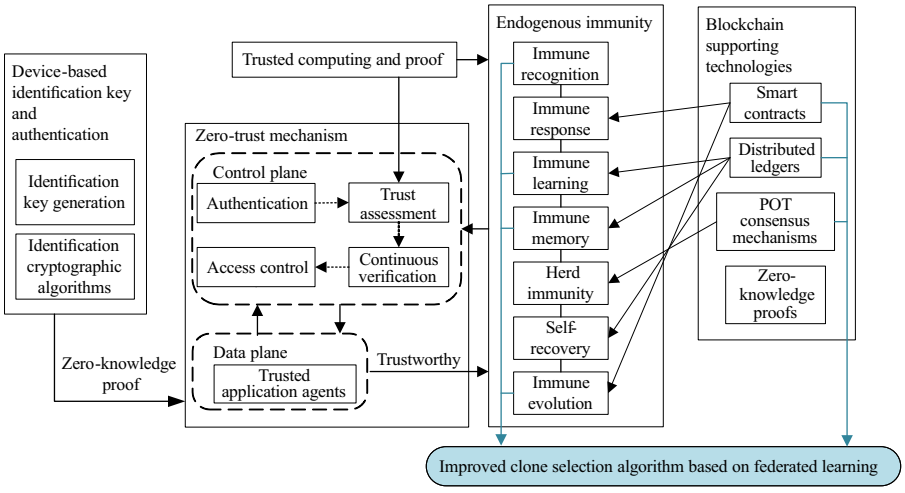

Fig. 2 Structure of the endogenous security mechanism for the industrial Internet.

the accounting right.

The node connects to zero trust mechanism through zero knowledge proof. And the node passes the identity password and encrypted state data to the trusted application agent of the zero-trust data plane. The zero-trust mechanism control plane uses a smart contract to compare the device identity password with the data in the distributed ledger for authentication. If the authentication is passed, the cooperative trusted computing and proof module obtains the trust assessment level updates. The trust level of the device (cumulative points) is updated, and the encrypted status data are uploaded to the blockchain through PoT consensus. Access control is performed on authenticated devices based on the trust level.

Endogenous immunity is based on blockchain technology for malicious node device identification and subsequent processing and recovery to provide exponential defense benefits. It includes immune recognition, immune response, immune learning, immune memory, herd immunity, self-recovery, and immune evolution. Immune recognition detects the anomalies and attacks in the system. If immune memory exists, immune response is immediately carried out; otherwise, immune learning learns and trains to make effective responses. Antibody measures with high affinity for attacking antigen are remembered to form herd immunity and realize self-recovery. Minimal antibody model parameters cover the attack antigen space and enable immune evolution.

In the industrial Internet, according to trust level, endogenous immunity identifies malicious node devices. Smart contracts are invoked for immune response, and immune learning is performed to take effective method, which is stored in a distributed ledger to form an immune memory for rapid response. Other trusted node devices can obtain herd immunity through heterogeneous PoT consensus mechanism. If the device suffers from attack and loses data, it can be self-recovery by the encrypted state data recorded on the distributed ledger after trusted proof. The smart contract is used to carry out the clonal selection and high-frequency mutation of antibodies during the above immunization process to realize immune evolution.

Endogenous immunity is the innovation mechanism of endogenous security. In the industrial Internet, facing the large-scale unknown attacks, the combination of intelligent self-learning and interactive security is required. Therefore, an improved clone selection algorithm based on federated learning is proposed. The improved clone selection algorithm plays a role in the endogenous immunity.

4 Improved Clone Selection Algorithm Based on Federated Learning

The clonal selection algorithm proposed by Castro and Zuben[14] was an artificial immunization algorithm based on the principle of clonal selection, which imitates the process of antigen recognition and antibody production in the immune system in order to select the best antibody. However, the traditional clonal selection algorithm is not suitable for the large-scale heterogeneous device diversity in the industrial Internet. The lack of communication between groups leads to the lack of appropriate and effective guidance for groups with low adaptive value.

Federated learning is used to realize the communication and information exchange of heterogeneous device clusters in industrial Internet. Each learner trains a model locally and transmits its local model parameters to a central server for aggregation[15]. Trusted local devices are selected to participate in training through blockchain PoT consensus. This essentially protects privacy better than sharing data.

At the same time, the traditional clonal selection algorithm is combined with federated learning to protect the transmission of antibody model through endogenous ID encryption and control the variation degree of the model according to the detection effect of the model. The smart contract replicates the model and selects the client for the next round. Through multiple rounds of training, the defense strategy is updated to realize automatic intelligence. Blockchain retains model parameters and has security and tamper-proof capability.

In this section, we propose an improved clone selection algorithm based on federated learning, and it is shown in Fig. 3.

As shown in Fig. 3, the affinity refers to the accuracy of the local model to defend against attacks, and the diversity refers to the similarity between multiple local models.

In Step 0, the central server mainly completes the following three tasks. Firstly, it initializes the model structure and parameters. Secondly, the observation period parameters T are initialized to compare the performance of multiple models within the period.

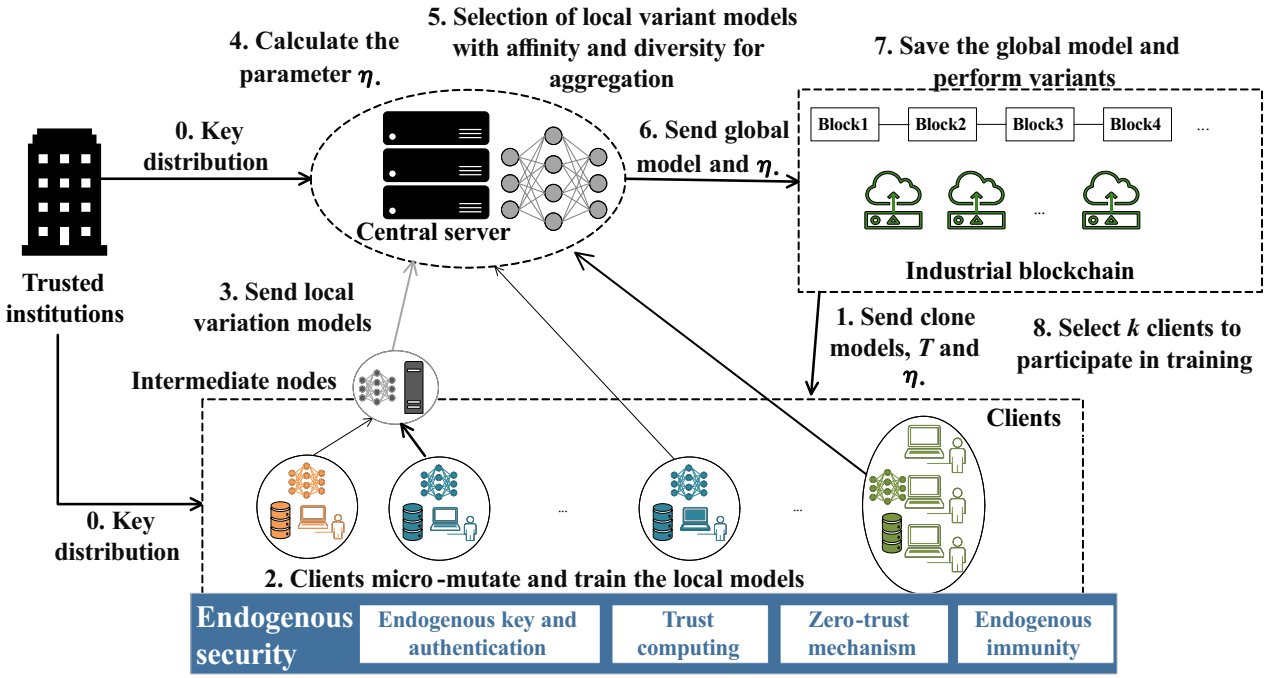

Fig. 3 Structure of the endogenous security innovation mechanism with federated learning for the industrial Internet.

Thirdly, it sets the initial parameter $\eta = 1$ for the calculation of model variation. In Step 1, the above three parameters encrypted with the lowest trusted level as an attribute are respectively sent to the client for training and industrial blockchain for storage. In addition, the trusted institutions issue digital certificates for the central server and each client by the private key. The clients verify the certificate of the central server before the next round of model training to ensure that the endogenous ID of the server is secure.

After receiving the global model, the client uses parameters η and Eq. (1) for micro-variation in Step 2.

$$P^{\text{tv}} = P^t + \eta \text{Cauchy}(0, 1) \tag{1}$$

where P^{tv} is the variant of antibody P^t, and Cauchy (0, 1) is a Cauchy random variable. If $t\%T = 0$, the performance of the model after mutation and the training models within the cycle will be compared, and the optimal model will be selected for the next training.

The client uses private data for training to get local model updates and sends model parameter to the central server, encrypted with the server's endogenous ID in Step 3. The central server uses its own endogenous key to decrypt the local model. In Step 4, η is determined by the similarity between local models, as shown in Eq. (2). η is used to control the intensity of Cauchy variation and inversely related to the magnitude of the affinity. The larger the affinity, the smaller the value of η. The smaller the affinity, the larger the value of η.

$$\eta = \frac{1}{N} \sum_{i,j \in N} \sqrt{\| P_i^t - P_j^t \|^2} \times (\text{Rand}(0, 1) \times 2 - 1) \tag{2}$$

Then in Step 5, the server sorts the local models according to the affinity and diversity. It uses N local devices to calculate the trust level and uses the trust level as the attribute to encrypt the global model. The global model and η are sent to the industrial blockchain in Step 6. In Step 7, the smart contracts verify global model updates through test data, preventing the central server from maliciously modifying parameters. Only model updates that pass the verification will be saved and cloned by the industrial blockchain. According to the PoT consensus mechanism, it selects K local clients with high score from N local devices in Step 8 and sends the global model to K clients. The clients that meet the trust level can decrypt the global model by trust attribute and then train it. The above steps are repeated until the target convergence requirements are met. The specific algorithm is shown in Algorithm 1.

The model is optimized based on the affinity selection of antibodies to antigens in the clonal selection algorithm. To achieve security protection of the interaction data, the antibody model is encrypted and decrypted by the SM9 algorithm when uploading the model for aggregation calculation; the antibody model is encrypted and decrypted by the trust Attribute-Based Encrpytion (ABE) when downloading the model. In order to calculate the

Algorithm 1 Improved clone selection algorithm based on federated learning

Input: A: the data collected by industrial device (attack (antigen)). S: the set of attack signature patterns. n: the number of aggregated models (antibodies). d: the number of new models (antibodies) generated per iteration. N: the number of clients. η: the Cauchy parameter.

Output: M: the set of memorized antibody models.

1: Initialize the client-side model score set $G_N = \{0, 0, \ldots, 0\}$, with iterative training rounds $t = 1$.
2: The central server captures and analyses the attack (antigen) A and extracts the feature pattern $S' \in S$. The initial global model P^1 and the observation period parameter T is generated. P^1, T, and $\eta = 1$ encrypted with the lowest trusted level as an attribute are saved to the blockchain and sent to clients.
3: **while** global model P^t does not satisfy the termination condition **do**
4: The client conforming to the trust level uses the trust level as the attribute to decrypt P^t.
5: According to Eq. (1), the client micro-mutates the global model P_i^{tv}.
6: If $t\%T = 0$, the client selects the best local model among $P_i^{\text{tv}}, P_i^{t-1}, \ldots, P_i^{t-T}$.
7: The client discriminates potential attacks (antigens) A in the local data and train them, generating a local variant model P_i^t.
8: Using SM9 identification cryptography algorithm to encrypt the local variant model P_i^t by the server's endogenous ID. The ciphertext of P_i^t is sent to the server.
9: The central server uses endogenous key to decrypt the ciphertext of local model and ranks local models based on accuracy and similarity.
10: According to Eq. (2), the server calculates the parameter η.
11: The central server selects the top N local models for aggregation to obtain the global model P^{t+1}.
12: The central server calculates the trust level and encrypts the global model P^{t+1} by attribute-base encryption algorithm.
13: The central server transfers the ciphertext of P^{t+1} and η to the blockchain for preservation to M.
14: The blockchain uses the smart contract to decrypt and clone global model P^t. According to PoT consensus mechanism and the combined score G_N, the top k clients C_i are selected to distribute antibody P_i^t, where $i = 1, 2, \ldots, k$.
15: $t = t + 1$.
16: **end**
17: **return** M.

extra cost of privacy protection, we mainly calculate the time of model encryption and decryption by SM9 and ABE, and node signature and validation by SM9. The SM9 algorithm is tested using the open source GMSSL-Python library. ABE is implemented by Ciphertext Policy Attribute Based Encryption (CPABE) software library (http://acsc.cs.utexas.edu/cpabe/index.html). The experimental results are shown in Tables 2 and 3. Through the time comparison, it can be seen that ABE is more efficient in encryption and decryption than SM9. The time of encryption and decryption by SM9 increases linearly with the size of the model, while the time required for SM9 signature and verification is minimal.

We improve the clone selection algorithm based on federated learning, which is suitable for massive heterogeneous devices in industrial Internet. Each of clients uses local data to train the model and encrypts the parameters of the model through endogenous ID and key, which increases the communication security among the heterogeneous devices and protects the privacy of the model data. At the same time, the trust level of trusted computing is used as the attribute to describe the key. The ciphertext can be decrypted only if the attribute contained in the key conforms to the trust level. In addition, the combination of blockchain cryptographic storage and smart contract variation makes the improved algorithm more secure and efficient.

5 Experimental Research

Aiming at the industrial Internet under the environment of massive heterogeneous nodes and frequent cross-domain data exchange, it is vulnerable to typical attacks. As shown in Fig. 4, we take the identity authentication scenario as an example for threat modeling. The Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Elevation of privilege (STRIDE) threat modeling framework is applied to identify potential threats against the system. With the device as the entry point, possible threats

Table 2 Comparison time of encryption, decryption, signature, and verification based on SM9.

Model size	Time (s)			
	Encryption	Decryption	Signature	Verification
1 KB	0.804	0.982	0.335	0.276
1 MB	7.52	7.64	0.373	0.316
10 MB	75.10	77.70	0.756	0.709

Table 3 Comparison time of encryption and decryption based on ABE.

Model size	Time (s)	
	Encryption	Decryption
1 KB	0.11	0.0469
1 MB	0.14	0.0625
10 MB	4.851	2.0488

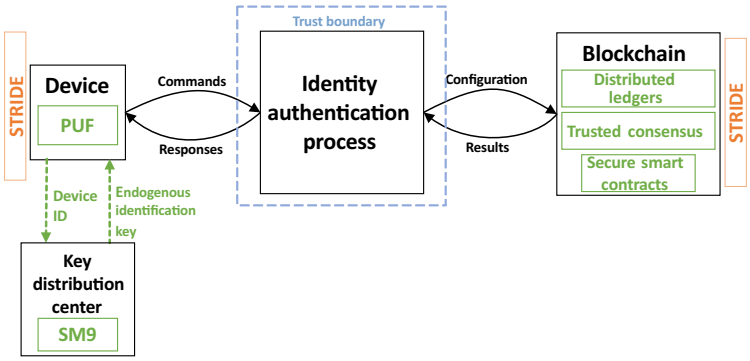

Fig. 4 Threat model of identity authentication scenario and proposed endogenous security solutions.

include spoofing, tampering, information disclosure, and elevation of privilege. For storing and reading database information, the threats considered include tampering and information disclosure.

We propose endogenous security solutions to mitigate the above threats as shown in Fig. 4 marked in green, and the results of comparison with traditional non-endogenous security solutions are shown in Table 4.

The green section on the left side of Fig. 4 is the solution for device-side threats, where the device cooperates with the key distribution center to generate the device ID and the endogenous identification key based on PUF. Device ID and the endogenous identification key are generated based on PUF. Compared with other methods that require plaintext storage of keys, trusted certificate authority, key injection, and additional hardware, PUF does not require key storage and can be implemented simply by loading software onto the chip, making it more suitable for industrial Internet.

As shown in the right part of Fig. 4, we achieve cross-domain authentication in the industrial Internet by storing information on the blockchain and using its supported technologies such as distributed ledgers, trusted consensus, and secure smart contracts for automatic authentication. The cross-domain authentication mechanism is shown in Fig. 5.

This mechanism consists of IBE domain and PKI domain, with PoT node in the blockchain set as proxy nodes. The IBE domain uses PUF to automatically generate device ID and SM9 algorithm to generate endogenous key. PKI domain manages public keys and issues certificates through the Certification Authority (CA). Nodes in the IBE domain are device nodes, and nodes in the PKI domain are blockchain nodes or gateway nodes. PoT node, as a proxy node for

Table 4 Non-endogenous security solutions vs. our solutions.

Threat type	Non-endogenous security solution	Our solution
Spoofing	Authentication mechanism	Endogenous identification key based on PUF
Tampering	Verify that all input is verified for correctness	Blockchain supporting technologies
Repudiation	Tamper-proof logs	Digital signature with endogenous identification keys
Information disclosure	Encrypting the data flow	Encrypt the data with endogenous identification keys
Denial of service	Set firewall	Isolate/shield attacker though identity authentication
Elevation of privilege	Enforcing principle of least privilege	Perform continuous authentication

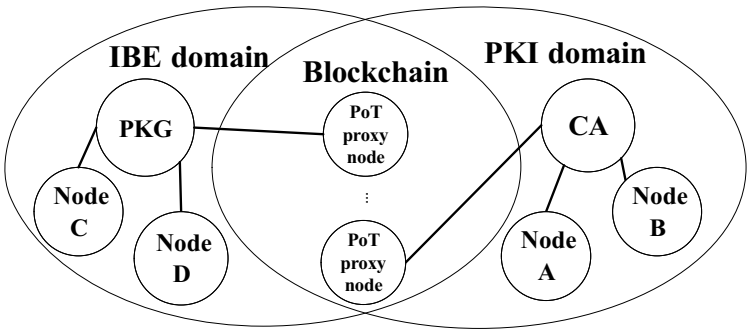

Fig. 5 Cross-domain authentication mechanism.

cross-domain authentication, uses non-interactive zero-knowledge proof to authenticate devices. After verifying the reasonable identity of nodes, smart contracts can be used to upload the authentication results to blockchain. The proxy nodes reach an agreement on the authentication result through the PoT consensus mechanism. PoT node adds trusted nodes and expiration dates to the trusted whitelist to implement the secure transfer of identity information of users in the IBE domain and PKI domain. We mainly implement the left part of Fig. 5, which is the user registration process and the identity verification process using zero-knowledge proof, as shown in Figs. 6 and 7, respectively.

We select and implement two blockchain technologies in this section: Ethereum alliance blockchain and Internet of Things Application (IOTA)[16] public blockchain to compare and analyze the time cost of user identity authentication based on blockchain and identification key. The respective applicability scenarios in the industrial Internet are then explored based on experimental performance validation.

The registration process is shown in Fig. 6.

In Steps (1)–(3), we use SRAM PUF and the improved R-ate bilinear pair algorithm of SM9 to generate industrial device ID as the public key PK and the endogenous identity key, as the secret key sk. In Steps (5)–(9), the hash of the ID and the signature of the hash are uploaded to the blockchain, and registration

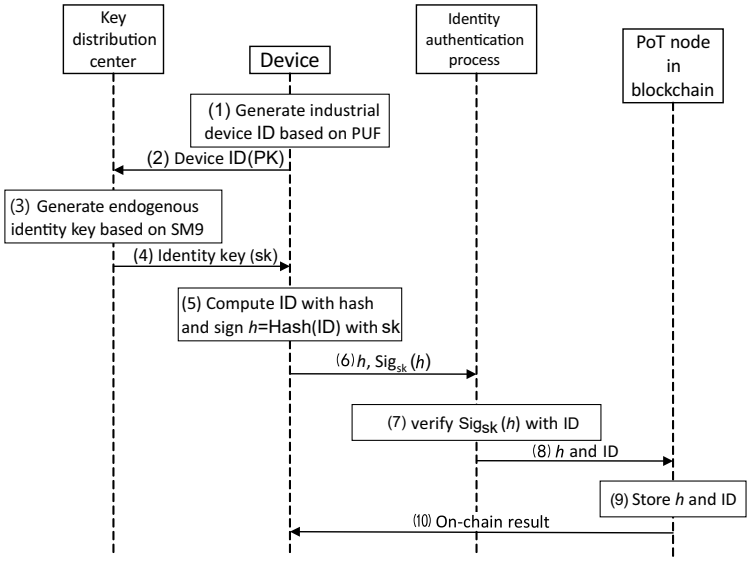

Fig. 6 User registration process.

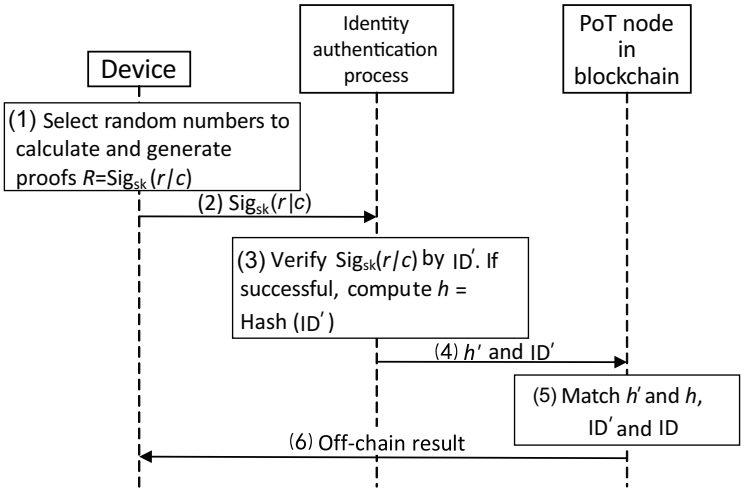

Fig. 7 User zero-knowledge proof verification process.

is completed.

The process of verification is shown in Fig. 7.

The legitimacy of the user's identity is verified by non-interactive zero-knowledge proof. In Steps (1)–(3), the user generates an identify proof R, chooses a random number r, and computes c=Hash(r+ID′). R is denoted as identify proof. If the identity proof R is verified by the identity authentication process, the hashes of ID and ID are sent to the blockchain. In Steps (4) and (5), the blockchain performs the comparative search of the hash and ID in the block to verify whether the user's identity is legitimate.

Based on the Ethereum platform, the Ethereum alliance blockchain experiment uses the Geth client to configure and start an Ethereum network node and opens the Remote Produce Call (RPC) interface to create an account. Smart contracts are written in Solidity language for identity authentication. The authentication contract is packaged into Java code through the Maven project and deployed to Ethereum as a transaction using the Web3.js library.

Based on Python 3.7, IOTA public blockchain experiment calls the PyOTA library to create transactions, read transactions, sign transactions, generate addresses, connect to a public node, and send transactions to it to interact with the ledger.

The deployment process of the smart contract is shown in Fig. 8. The computer running environment is as follows: CPU (Core i7, 2.4 GHz), RAM (8 GB), hard disk 1 TB, and 64 bit operating system Win10. SRAM PUF is used to generate endogenous ID. SM9 bilinear pairwise algorithm generates endogenous key based on endogenous ID. Both can be used as an initial measure of trust for user authentication. Subsequently, the trust level can be calculated according to the user's behavior. Zero-knowledge proof is used to verify the identity of the user, and the logic of user registration and verification is automatically executed by the smart contract.

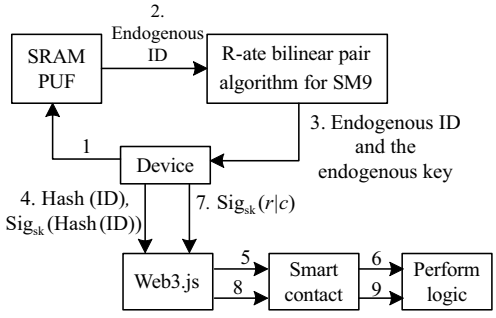

Fig. 8 Process of zero-knowledge proof on Ethereum private blockchain.

The number of users is set to 1, 10, 100, 1000, and 5000. The on-chain time, match time, and off-chain time of the simulated authentication are taken as the final result. The on-chain time indicates the user registration time, which is Steps (6)–(8) of Fig. 6. The match time represents the comparison of h and h', which is Steps (4) and (5) of Fig. 7. The off-chain time is the process of transferring data to the local area after matching, which is Step (6) in Fig. 7. The two types of blockchain experimental results are shown in Figs. 9–11.

The comparison of the on-chain time of the two kinds of blockchain technology is shown in Fig. 9. As the number of users increases, the on-chain time increases the fastest in the Ethereum alliance blockchain, and the IOTA public blockchain is shorter than it. The comparison of the verification time of the two kinds of blockchain technology is shown in Fig. 10. The match time for both blockchains does not increase with the number of users, and the time for the Ethereum alliance blockchain is much less than that of the IOTA public blockchain. The comparison of the off-chain time of the two kinds of blockchain technology is shown in Fig. 11. As the number of users increases, the off-chain time of the two blockchains showed a slow upward trend with little change. But the off-chain time of the IOTA public

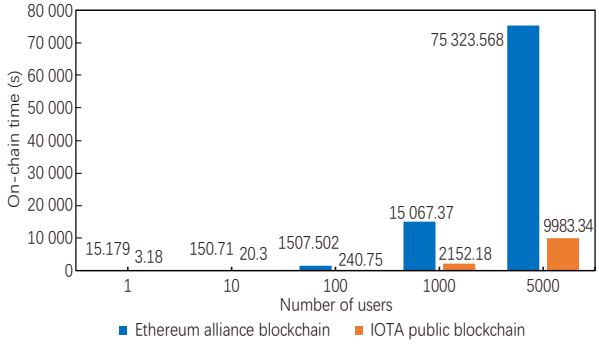

Fig. 9 Two types of blockchain on-chain time experimental results.

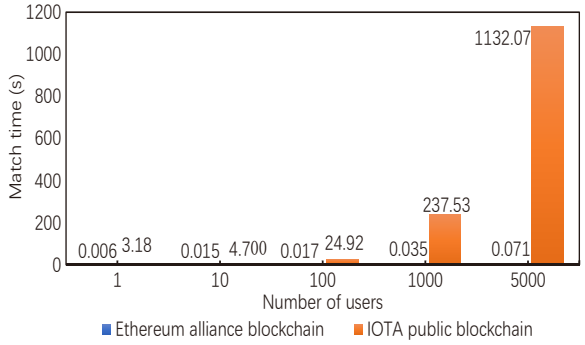

Fig. 10 Two types of blockchain match time experimental results.

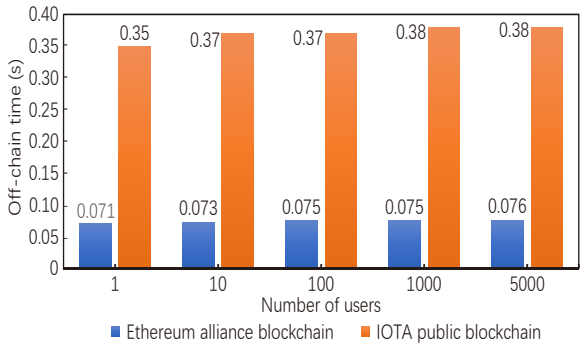

Fig. 11 Two types of blockchain off-chain time experimental results.

blockchain is much longer than another one.

Because the on-chain members of the Ethereum alliance blockchain need to reach a consensus, so that the on-chain time increases. However, IOTA uses Tangle structure as distributed ledger, and only two blocks need to be selected for verification. The on-chain time is short. Ethereum alliance blockchain uses smart contracts to traverse the list of users stored in the block and then authenticate the users, which is faster. IOTA needs to connect to Devnet public network for verification, which increases the match time. Since the amount of stored data is small, the off-chain time of Ethereum and IOTA does not change obviously as the user increases. The off-chain time of the Ethereum alliance blockchain deployed locally is shorter than that of IOTA public network.

Aiming to the problem of experiment scale, we can refer from the above experimental data: The on-chain time of Ethereum blockchain is longer than that of IOTA blockchain, but the off-chain time is shorter than that of IOTA blockchain.

Based on the analysis of the experimental results, Ethereum alliance blockchain can be used to provide the secure identity resolution services for universal objects query on the industrial Internet. IOTA public blockchain can be used for data analysis and aggregation of edge nodes.

6 Future Challenges of Industrial Internet Endogenous Security

(1) Integration challenge of different endogenous security mechanisms

The implementation of endogenous security including different technologies, such as blockchain, trusted computing, artificial intelligence, computer immunity, and zero trust mechanism. The integration of technologies will also bring conflicts and problems.

We think the integration order of various technologies can be proved to be effective by formal description method.

(2) Dependable challenge of endogenous security under unknown cyber threats

Most of the network information systems lack the endogenous ability to defend unknown cyber-attacks. The attacks are becoming more intelligent with the development of artificial intelligent and big data technology.

We propose to improve intelligent security from endogenous security, so that endogenous security has self-learning ability to detect unknown attacks.

(3) Lack of endogenous security proof ability and evaluation standards for the industrial Internet

How to evaluate the quality of endogenous security system in different application scenarios is a problem. The quality of endogenous security should consider the balance between security and cost. However, there is a lack of evaluation standards for endogenous security in the industrial Internet.

We suggest that international industry standards can be based on the joint universities and enterprises to develop endogenous security evaluation standards. Security formal methods can be used to prove endogenous security mechanism.

7 Conclusion

In this paper, we research on endogenous security in industrial Internet. The endogenous security in industrial Internet is firstly defined scientifically; then we propose a novel endogenous security architecture for industrial Internet, including an innovative endogenous security layer, intelligent self-learning, interactive security, and data lifecycle security. It is combined in zero-trust mechanism, endogenous immunity, trust blockchain, and federated learning. In addition, we propose an improved clone selection algorithm based on federated learning and trust blockchain, and achieve endogenous immune learning ability. Experiments are conducted to evaluate the performance of SM9 identification cryptography and attribute based encryption algorithm under clone select algorithm. We propose a threat model for the industrial Internet authentication scenario, design a cross-domain authentication mechanism using endogenous security, and conduct performance experiments based on blockchains. Based on the experimental analysis, Ethereum alliance blockchain can be used to provide the identity resolution services on the industrial Internet.

IOTA public blockchain can be used for data aggregation analysys of IoT edge nodes. Finally, we propose three core challenges on industrial Internet endogenous security and give them solutions.

In the future, we will devote to designing and realizing endogenous security mechanisms according to different industrial application scenarios of the industrial Internet.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (No. 2018YFB0803403) and Fundamental Research Funds for the Central Universities (Nos. FRF-AT-19-009Z and FRF-AT-20-11) from the Ministry of Education of China.

References

[1] H. Song, J. Bai, Y. Yi, J. Wu, and L. Liu, Artificial intelligence enabled Internet of Things: Network architecture and spectrum access, *IEEE Comput. Intell. Mag.*, vol. 15, no. 1, pp. 44–51, 2020.

[2] F. Foukalas and A. Tziouvaras, Edge artificial intelligence for industrial Internet of Things applications: An industrial edge intelligence solution, *IEEE Ind. Electron. Mag.*, vol. 15, no. 2, pp. 28–36, 2021.

[3] D. Wei, H. Ning, F. Shi, Y. Wan, J. Xu, S. Yang, and L. Zhu, Dataflow management in the Internet of Things: Sensing, control, and security, *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 918–930, 2021.

[4] Y. Liu and M. Peng, 6G endogenous security: Architecture and key technologies, (in Chinese), *Telecommunications Science*, vol. 36, no. 1, pp. 11–20, 2020.

[5] A. Hu, L. Fang, and T. Li, Research on bionic mechanism based endogenous security defense system, (in Chinese), *Chinese Journal of Network and Information Security*, vol. 7, no. 1, pp. 11–19, 2021.

[6] X. Ji, J. Wu, L. Jin, K. Huang, Y. Chen, X. Sun, W. You, S. Huo, and J. Yang, Discussion on a new paradigm of endogenous security towards 6G networks, *Front. Inf. Technol. Electron. Eng.*, vol. 23, no. 10, pp. 1421–1450, 2022.

[7] G. Wei, H. Li, Y. Bai, G. Li, and K. Xing, Space-terrestrial integrated multi-identifier network with endogenous security, (in Chinese), *Space-Integrated-Ground Information Networks*, vol. 1, no. 2, pp. 66–72, 2020.

[8] S. Guo, Y. Qi, M. Dai, X. Qiu, F. Qi, and P. Zhang, Endogenous trusted network architecture for intelligent sharing, (in Chinese), *Journal on Communications*, vol. 41, no. 11, pp. 86–98, 2020.

[9] Z. Jiang, T. Li, and A. Hu, Research on endogenous security methods of embedded system, in *Proc. 2020 IEEE 6th Int. Conf. Computer and Communications* (*ICCC*), Chengdu, China, 2020, pp. 1946–1950.

[10] Z. Zhou, X. Kuang, L. Sun, L. Zhong, and C. Xu, Endogenous security defense against deductive attack: When artificial intelligence meets active defense for online service, *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 58–64, 2020.

[11] Alliance of industrial Internet, The architecture of industrial Internet of Things v2.0, http://www.aii-alliance.org/upload/202004/0430_162140_875.pdf, 2020.

[12] Alliance of industrial Internet, Security framework of industrial Internet of Things, http://www.aii-alliance.org/upload/202002/0228_140108_424.pdf, 2018.

[13] Information Technology—Security Techniques—Encryption Algorithms—Part 5: Identity-Based Ciphers—Amendment 1: SM9 Mechanism, ISO/IEC 18033-5:2015/Amd 1:2021, 2021-02.

[14] L. N. D. Castro and F. J. V. Zuben, Learning and optimization using the clonal selection principle, *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 239–251, 2002.

[15] J. Pang, Y. Huang, Z. Xie, J. Li, and Z. Cai, Collaborative city digital twin for the COVID-19 pandemic: A federated learning solution, *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 759–771, 2021.

[16] S. Popov, The tangle, http://iotatoken.com/IOTA_Whitepaper.pdf, 2016.

Xintong Han received the master degree from University of Science and Technology Beijing, China in 2023. Her research areas include industrial Internet and blockchain.

Yiying Zhang received the master degree from University of Science and Technology Beijing, China in 2023. Her research interests include network and information security, blockchain technology, and artificial intelligence.

Hongsong Chen is currently a professor at School of Computer and Communication Engineering, University of Science and Technology Beijing, China. He was a visiting scholar in Purdue University from 2013 to 2014. His current research interests include network and information security, industrial Internet security, and artificial intelligence security. He is a member of IEEE.