Behavioral Analysis Zero-Trust Architecture Relying on Adaptive Multifactor and Threat Determination

Chit-Jie Chew[1], Po-Yao Wang[1], and Jung-San Lee[1*]
[1]Department of Information Engineering and Computer Science,
Feng Chia University, Taichung, Taiwan
[e-mail: leejs@fcu.edu.tw]
[*]Corresponding author: Jung-San Lee

Received June 5, 2023; accepted August 19, 2023; published September 30, 2023

Abstract

For effectively lowering down the risk of cyber threating, the zero-trust architecture (ZTA) has been gradually deployed to the fields of smart city, Internet of Things, and cloud computing. The main concept of ZTA is to maintain a distrustful attitude towards all devices, identities, and communication requests, which only offering the minimum access and validity. Unfortunately, adopting the most secure and complex multifactor authentication has brought enterprise and employee a troublesome and unfriendly burden. Thus, authors aim to incorporate machine learning technology to build an employee behavior analysis ZTA. The new framework is characterized by the ability of adjusting the difficulty of identity verification through the user behavioral patterns and the risk degree of the resource. In particular, three key factors, including one-time password, face feature, and authorization code, have been applied to design the adaptive multifactor continuous authentication system. Simulations have demonstrated that the new work can eliminate the necessity of maintaining a heavy authentication and ensure an employee-friendly experience.

Keywords: Zero-trust, Cyber threat, Machine Learning, Behavior Analysis, Authentication

1. Introduction

As the explosive development of network technology, the threatens of cybersecurity have become increasingly pervasive with a wide range of intentional attacks. Industries have spent a large amount of overhead on building robust information barriers such as privilege management [1], virtual private networks [2], firewalls [3], and intrusion detection systems [4]. However, hackers are always able to exploit new breaches to bypass these measures through techniques like phishing [5], identity spoofing [6], and zero-day attacks [7]. The primary trigger is the "mutual trust" structure between the internal environment. Once hackers successfully pass the authentication system with a legitimate or camouflage identity, they are regarded as the trusted nodes by the entire internal environment; thus, rendering the protection equipment ineffective. This has raised the risk of lateral movement [8].

Actually, CISCO incidents [9] have demonstrated the susceptibility of authentication systems. In this case, an attacker may steal a certificate and spoof a one-time password to plant a backdoor in the system. The root cause of this incident is that an authentication system only uses the certificate issuance to demonstrate the trust in the logger. As a certificate is essentially considered a "long-time continuous pass," the hacker who has stolen the certificate could gain access without obtaining the password for the employee account. Undoubtedly, all personnel with high authority have become the prominent target. Hackers can even frequently customize the advanced persistent threat (APT) attack or phishing emails to acquire privilege. In addition, the ordinary employee has not immune to being a target. The reason is that it is quite easier for malicious attackers to compromise this group by impersonating trustworthy identities.

Recently, how to deploy the zero-trust architecture (ZTA) [10-12] has become a critical cybersecurity issue in the worldwide. The concept of ZTA is to maintain a distrustful attitude towards all devices, identities, and communication requests, which provides only the minimum access and validity required. Referring to the SP 800-207 zero-trust architecture [11] proposed by the National Institute of Standards and Technology (NIST) in 2020, employees must complete an authentication through the policy enforcement point whenever they access company resources, as indicated in **Fig. 1**. The detail of each component is described as follows.

- Policy enforcement point (PEP) – As the core point between employees and company resources, PEP is mainly responsible for terminating, monitoring, and allowing connections. In addition, PEP has to update the connection strategy promptly according to PA instructions.
- Policy engine (PE) – PE evaluates the request based on the implemented policy, continuous diagnostics and mitigation (CDM), security information and event management, activity log, and threat intelligence, which deciding whether to authorize the permission.
- Policy administrator (PA) – PA establishes or denies the connection according to the judgment result of PE, which generating the authorized token to establish a session between employee and resource.

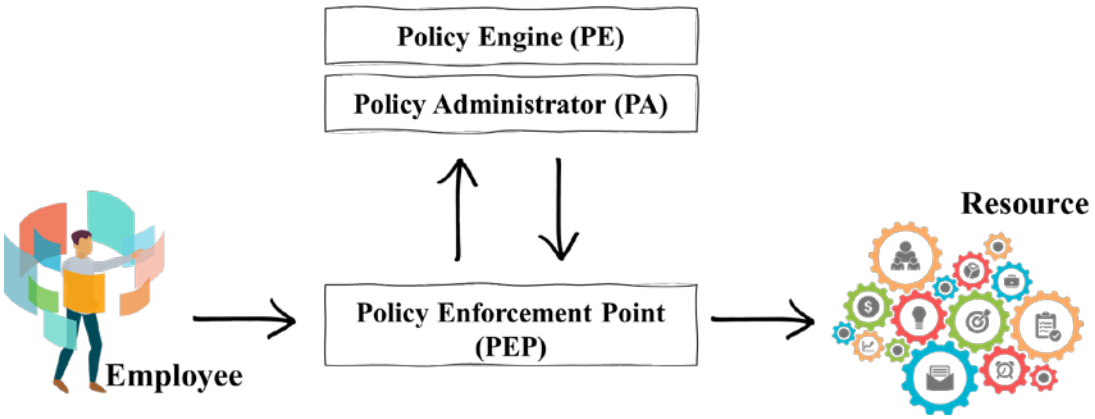


Fig. 1. Components of ZTA

In fact, ZTA is a questioning attitude toward all resource requests. Conversely, the meaning of trust in the past network architecture was meant to provide convenience for employees to use various identity recognition [13] technologies to realize friendly authentication, such as trusted IP sources, certificates, devices, and one-time passwords. Once successfully accessing the system, there is no need to repeat identity authentication for a short period. Therefore, ZTA possesses robust defenses against traditional "mutual trust" and "long-time continuous pass." Unfortunately, implementing ZTA requires a significant equipment investment and a complicated authentication process, which absolutely daunting the burden on equipment and decreasing the efficiency of employees.

So far, researchers have gradually discussed the application of ZTA in various scenarios, such as smart cities [14], the Internet of Things [15], and cloud computing [16]. Regrettably, the proposed methods are only reliable and efficient for specific environments. Moreover, prior works have used a single authentication mechanism without considering the risks associated with the behavior. Considering the sophisticated and unpredictable resource requests of employees, the authentication method requires more stringent requirements, such as the potential risks caused by various impersonation attacks or types of access resources. The challenges of implementing ZTA for company operating structure are described as below.

- Construction fee: Numerous studies have proposed robust continuous authentication technology [17], including physiological and behavioral authentication methods. However, the premise of solid identity verification relies on various identity extraction devices to support it, which dramatically increasing the cost of equipment purchase. This results in a harsh limitation when constructing an employee authentication mechanism.
- Employee impact: Performing a previous identity verification for each request leads to a significant degradation in work efficiency. Even the employee resource access behavior no longer poses only a single degree of risk, and the threat arises differently as the request changes. In order to guarantee the security, adopting the most secure and complex multifactor authentication brings a more troublesome unfriendly operation and burden.
- System burden: In a ZTA environment, all resource requests must be validated, which means that the system is under an incalculable burden. With the required performance of the authentication mechanism, the capacity of the system becomes an critical and unignorable issue.

To address the challenges, we incorporate machine learning technology to construct an employee behavior analysis zero-trust architecture (BA-ZTA) according to adaptive multifactor authentication and threat determination, which can eliminate the necessity of

maintaining a heavy authentication and ensure an employee-friendly experience. BA-ZTA is characterized by the ability of adjusting the difficulty of identity verification through the user behavioral patterns and the risk degree of the resource. Significantly, the adaptive multifactor continuous authentication system is designed with three factors: one-time password, face feature, and authorization code. Without loss of generality, an employee is supposed to equip with the device containing a camera, such as laptop or mobile phone. Thus, the face feature is the critical feature of the BA-ZTA, which enabling employees to process an unconscious authentication. This can substantially reduce the complexity and inconvenience for employees. The contribution of this paper is summarized as follows.

(1) BA-ZTA contains a reliable constant behavior analysis module (CBA) and threat determination module (TD). In particular, CBA constructs constant analysis for the corresponding employee through their activity logs. Thereby, TD can accurately analyze the threat level of the access request of employees.
(2) Adaptive multifactor continuous authentication system (AMCA) dynamically adjusts the feature that employees must provide to be identified based on the threat level, which optimizing employee-friendly and mitigating the dilemma of repeated authentication. Simultaneously, BA-ZTA needs not continuously perform the most complex authentication, which reducing the system burden.
(3) BA-ZTA offers practicality and scalability in the real world. Notably, the slight authentication burden facilitates the construction of BA-ZTA without additional equipment and incurring the high costs. Additionally, despite the increasing number of employees, BA-ZTA has the ability to handle these requests in real time.

The rest of this paper is organized as follows. In section 2, the prior authentication approaches are discussed in section 2. The detail design of BA-ZTA are presented in section 3, followed by the experimental results to discourse the achievement of BA-ZTA in section 4. Finally, the contributions of this research are concluded in section 5.

2. Related Works

Various methods integrated with machine learning have been developed to achieve verification aside from account password and one-time password (OTP). Nowadays, the leading identity certification can be classified as mode-based, physiological, and behavioral authentication methods, as depicted in **Fig. 2**.

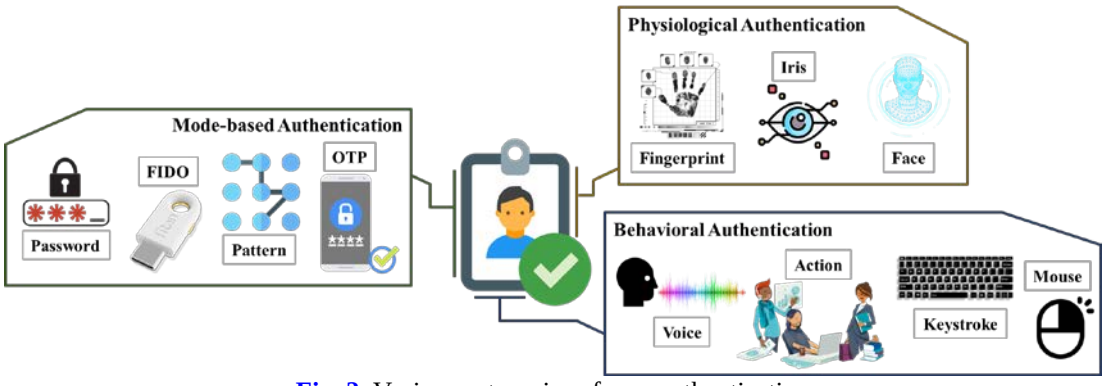


Fig. 2. Various categories of user authentication

Currently, mode-based is the most widely adopted identity proofing method in arbitrary systems, which incorporating password [18], FIDO [19], pattern [20], and one-time password [21]. Although it offers users convenience, the authentication process is potentially under risk of rainbow tables [22], device vulnerability password leakage [23-24], and OTP interception malware [25].

Physiology and behavior are served as the most representative biological characteristics of individual personnel, which significantly enhancing the holistic security of identity authentication. As in the fingerprint research, Aseel Bedari et al. [26] have proposed an effective biometric authentication system for the Internet of Things (IoT) framework. In addition to achieving reliable architecture, the system can ensure the revocability, diversity, unlinkability, and noninvertibility properties. Furthermore, Juan E. Tapia et al. [27] have designed an iris liveness detection technique concentrating on bona fide images, which can enhance the verification ability for the recognition system. Also, Jing Lei et al. [28] presented a fast privacy-preserving face authentication (PrivFace). In particular, PrivFace is able to ensure the revocable and reusable biometric credentials through a lightweight random masking technique.

As to the field of behavioral authentication, Juan Manuel Espín López et al. [29] first developed a supervised machine learning (ML) biometric continuous authentication system to verify staff and avoid security breaches using sensors, applications statistics, or speaker data in Industry 4.0. Md L. Ali et al. [30] then designed a hybrid model with a partially observable hidden Markov model and support vector machine (POHMM/SVM) for keystroke biometric authentication, which is able to layout excellent performance and precisely handle the missing or irregular data. Subsequently, Chao Shen et al. [31] have constructed a mouse-interaction authentication system with a pattern-growth-based mining method. This method can offer a more stable and discriminative frequent behavior segmentation, which resulting in a higher accuracy.

Undoubtedly, all these works have highlighted the effectiveness of biometric features in providing an accurate and reliable authentication. However, it requires specific equipment for such feature extraction, which brings the challenge in deploying ZTA over the real environment due to the high construction costs and complexity. In light of this, the existence of real-life equipment tools for authentication has been seriously considered in BA-ZTA, which importing a comparative strength authentication to discriminate employees based on the threat risk level. This can ensure the system reliability while minimizing any negative impact on staff efficiency and construction overheads.

3. The Description of BA-ZTA

Before accessing company resources, users are instructed to attend the company to accomplish initialization in person. Firstly, the employee registers a clear photo with facial features to complete the biometric creation. Afterward, the user has to finish the device configuration, such as mobile phone and laptop. Thus, the corresponding certificate can be generated once the device is subject to the verification [32]. Finally, the company binds an authorized licensee for the user to authenticate for the highest risk. Whenever the user needs to access resources or any services, the process is carried out according to BA-ZTA framework, as shown in **Fig. 3**.

At the outset, the access request generated by the employee undergoes sequential evaluation through the constant behavior analysis module (CBA) and threat determination module (TD) to determine the corresponding threat value. Based on the outcomes, the adaptive

multifactor continuous authentication system (AMCA) dynamically adjusts verification difficulty and prompts the user to complete identity validation. Subsequently, users provide authentication factors, such as OTP, biometrics, and authorization code, according to the level of identity proof required by the system. Thereafter, the authentication system identifies the user using the provided factors according to the maximum adoption scores AS_{otp}, AS_{bio}, and AS_{auth} that can be obtained for each feature, where the adoption score is preset by BA-ZTA. Especially, the biometric features are assessed by face recognition for the confidence level. Upon successful authentication, the user is granted specific access privileges to resources or services, whereas if the authentication fails, the system prompts the user for authentication proof once more. In the event of three consecutive errors, the user is temporarily restricted from requesting the resource in BA-ZTA, and subsequent authentication requirements become more stringent than usual. The notation used in BA-ZTA is described in **Table 1**.

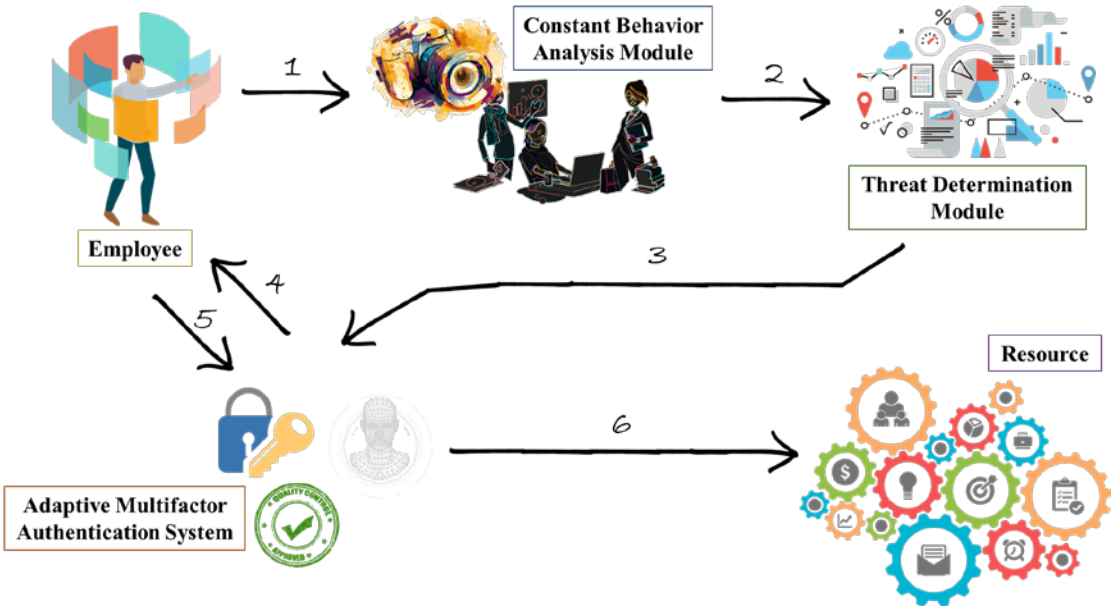


Fig. 3. The authentication process of BA-ZTA

Table 1. Notation Definition

Sign	Definition
AS_{otp}	The maximum adoption score of the one-time password.
AS_{bio}	The maximum adoption score of biometrics.
AS_{auth}	The maximum adoption score of the authentication code.
$Geoloc$	The geographical coordinates of the packet.
$Risk$	The sensitivity of the target access resources.
$DevCert$	Validity of device certificate.
t	Request time.
Δt	Time interval with t.
WD	The day is a working day or a non-working day.

$\boldsymbol{GDS}$	Gaussian distribution score evaluated by one-class SVM.
$\boldsymbol{AD}$	The abnormal degree of employee.
$\boldsymbol{l}$	The layer length of threat determination module, where $l \in \{1,2,3, \ldots 6\}$.
$\boldsymbol{n_j^l}$	the j-th neuron calculation result of l-th layer.
$\boldsymbol{\sigma}$	Sigmoid function.
$\boldsymbol{d^l}$	The neuron number of l-th layer.
$\boldsymbol{\omega_{jk}^l}$	The weight for j-th neuron in l-th layer corresponds to the k-th neuron in the previous layer.
$\boldsymbol{b_j^l}$	Bias value of n_j^l.
$\boldsymbol{n_1^6}$	Probability value for low threat.
$\boldsymbol{n_2^6}$	Probability value for medium threat.
$\boldsymbol{n_3^6}$	Probability value for high threat.
$\boldsymbol{M}$	The highest probability value among n_1^6, n_2^6, and n_3^6.
$\boldsymbol{IL}$	The impact level of the request.
$\boldsymbol{RS}$	The risk score of the request.
$\boldsymbol{rs_L}$	The minimum recognition score an employee must achieve for low risk.
$\boldsymbol{rs_M}$	The minimum recognition score an employee must achieve for medium risk.
$\boldsymbol{rs_H}$	The minimum recognition score an employee must achieve for high risk.
$\boldsymbol{IS}$	The identification score an employee must achieve to access a resource.
$\boldsymbol{TH}$	The threshold value of the time range that can be recognized.
$\boldsymbol{FS'_{bio}}$	The score was obtained for previous facial features.
$\boldsymbol{F'_{otp}}$	One-time password issued by AMCA.
$\boldsymbol{F'_{auth}}$	Authentication code issued by AMCA.
$\boldsymbol{F_{otp}}$	One-time password provided by the employee.
$\boldsymbol{F_{bio}}$	The biometrics feature provided by the employee.
$\boldsymbol{F_{auth}}$	The authentication code provided by the employee.
$\boldsymbol{FS_{otp}}$	Feature score of one-time password.
$\boldsymbol{FS_{bio}}$	Feature score of biometrics.
$\boldsymbol{FS_{auth}}$	Feature score of authentication code.
$\boldsymbol{n}$	The total number of feature values.
$\boldsymbol{x_i}$	The i-th value of F_{bio}.
$\boldsymbol{y_i}$	The i-th value of registered personal biometrics.
$\boldsymbol{TS}$	Trust score of the employee.

Subsection 3.1 elaborates on the constant behavior analysis module, while subsection 3.2 outlines the development of the threat determination module. The adaptive multifactor continuous authentication system and its implementation are explained in subsection 3.3.

3.1 Constant Behavior Analysis Module (CBA)

When an employee requests access to company resources or services, the constant behavior analysis module evaluates the request against habitual behavior. The detailed process is illustrated in **Fig. 4**.

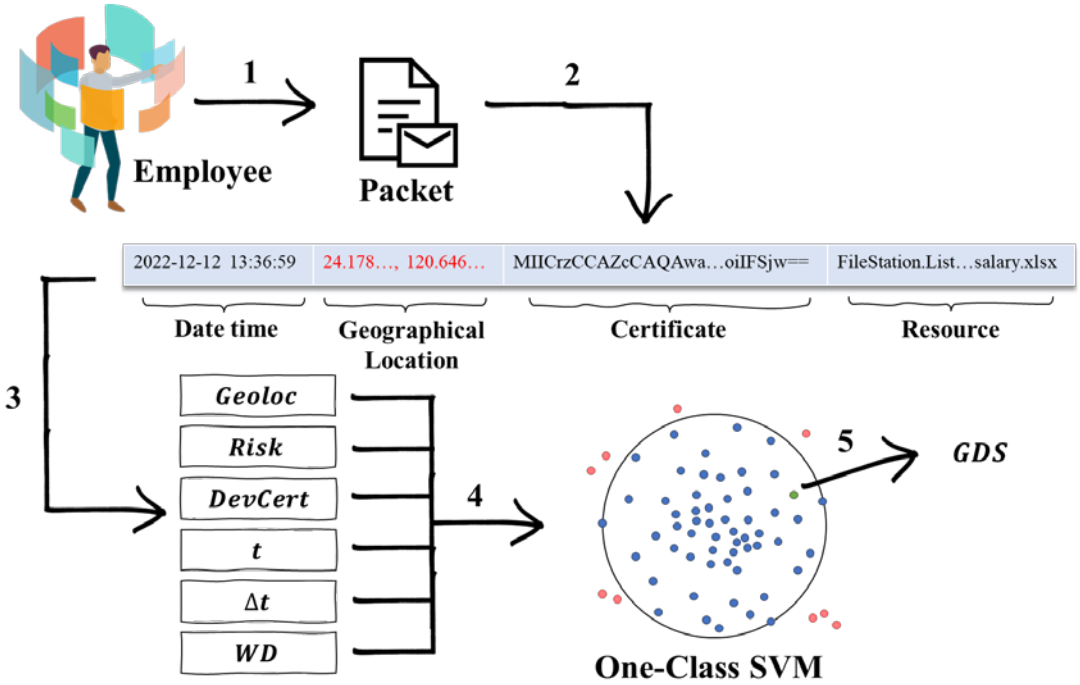


Fig. 4. The operation of the constant behavior analysis module

Step 1: Whenever an employee interacts with a company resource, a packet is generated to facilitate communication.
Step 2: CBA extracts critical information from the packet, such as date & time, geographical location, certificate, and target resource.
Step 3: According to the extracted information, pre-processing is performed to analyze $Geoloc$, $Risk$, $DevCert$, t, Δt, and WD. The following is the corresponding processing method for the respective feature.

- $Geoloc$ set by geographical coordinates, which determines whether the user's current location is within the company, inside the country, or abroad, and set it to 0, 1, or 2, respectively.
- $Risk$ is the sensitivity of the target access resources, as the risk level of resources varies by industry type. Therefore, the risk reference belongs to the company's ISO/IEC 27001 information security management system (ISMS) [33], which defines the risk threat level as 0, 1, or 2. The degree of data sensitivity increases with the numerical value.
- $DevCert$ represents whether the equipment certificate has been registered as a company asset, with 1 or 0 denoting belonging and non-belonging individually.
- t indicates the time at which the user requests to the resource.
- Δt calculates the time difference between the time t' of the previous user's request and the current t.

- WD decides whether it is a business day according to the date, which indicates a business day with 0 and a non-business day with 1.

Step 4: $Geoloc$, $Risk$, $DevCert$, t, Δt, and WD import to one-class SVM [34] as the analysis data.

Step 5: Based on the results of one-class SVM, the gaussian distribution score GDS is determined for this request.

3.2 Threat Determination Module (TD)

This stage involves analyzing the threat level of the request traffic to the company. **Fig. 5** shows the evaluation process, and the varying degrees of threat impact generated by DNN [35] is determined.

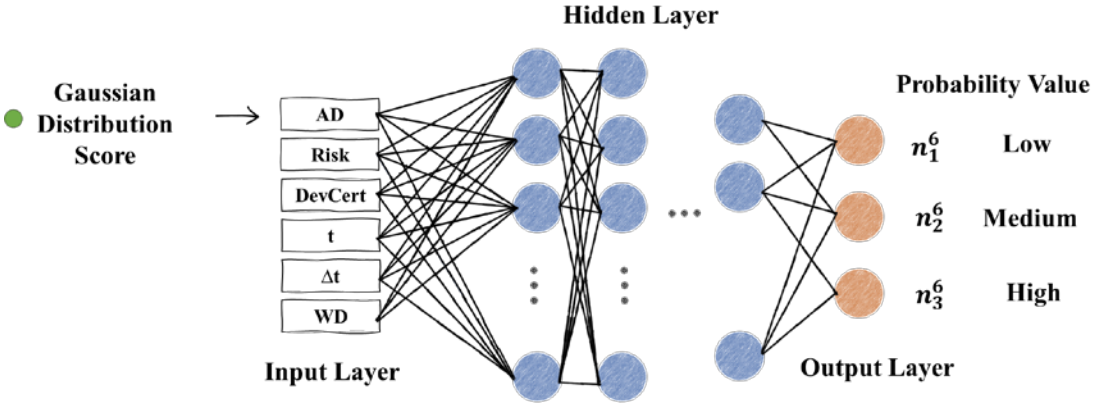


Fig. 5. The process of the threat determination module

Step 1: GDS is normalized to obtain the abnormal degree AD, as shown in (1),

$$AD = \begin{cases} 0, & GDS \geq 0 \\ 1, & 0 \geq GDS \geq -10 \\ 2, & otherwise \end{cases} \tag{1}$$

Step 2: AD, $Risk$, $DevCert$, t, Δt, and WD are input to the hidden layer as neurons in the input layer.

Step 3: TD is divided into seven layers, where $l \in \{0,1,2,\ldots 6\}$. The first layer is the input layer, consisting mainly of 6 features. The subsequent five layers are the hidden layer, which consists of 512, 512, 256, 256, and 128 neurons, respectively. The last layer is the output layer, which is composed of three neurons, each indicating low, medium, and high risk evaluation results. The corresponding value of each neuron in the hidden layer and output layer is calculated according (2),

$$n_j^l = \sigma\left(\sum_k^{d^{l-1}} \omega_{jk}^l n_k^{l-1} + b_j^l\right) \tag{2}$$

Step 4: The analysis result of TD is input to the following AMCA module.

3.3 Adaptive Multifactor Continuous Authentication System (AMCA)

Depending on the threat assessment of TD, AMCA adaptively adjusts the corresponding identity proof requirement to guarantee access security, as shown in **Fig. 6**.

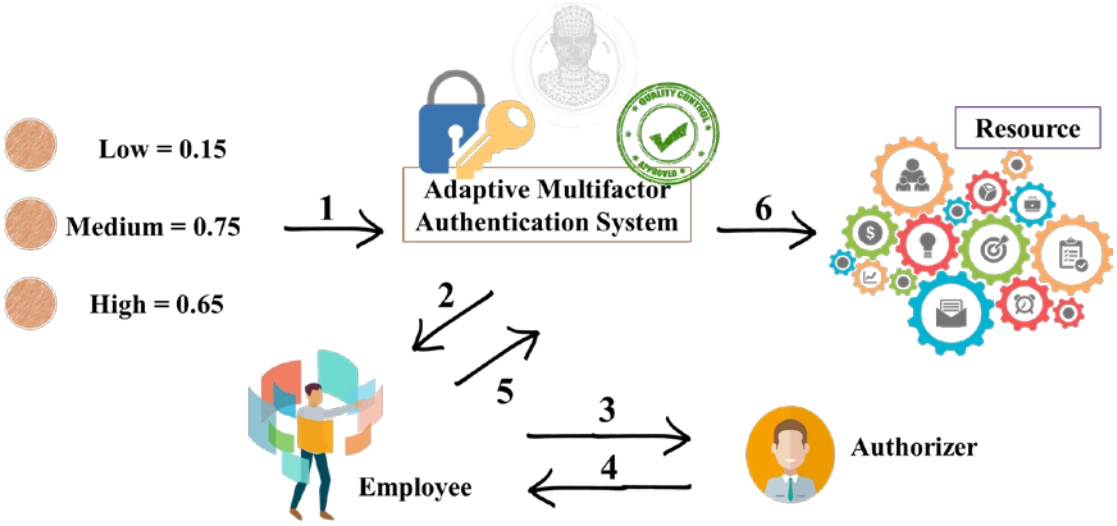


Fig. 6. The authentication process of AMCA

Step 1: n_1^6, n_2^6, and n_3^6 are imported to the AMCA as an evaluation reference.
Step 2: AMCA judges the risk level according to (3) to acquire the maximum value M with the highest analysis neuron and sets the impact level IL by (4),

$$M = max(n_1^6, n_2^6, n_3^6) \tag{3}$$

$$IL = \begin{cases} 0, M = n_1^6 \\ 1, M = n_2^6 \\ 2, M = n_3^6 \end{cases} \tag{4}$$

where 0, 1, and 2 indicate the low, medium, and high impact levels. Subsequently, AMCA evaluates the risk score RS that the employee has to satisfy relying on (5),

$$RS = \begin{cases} rs_L + (1 - (M - n_2^6)) * (rs_M - rs_L), IL = 0 \\ rs_M + (1 - (M - n_3^6)) * (rs_H - rs_M), IL = 1 \\ rs_H + (M - n_2^6) * (1 - rs_H) \quad , IL = 2 \end{cases} \tag{5}$$

where rs_L, rs_M, and rs_H are the minimum recognition scores for low, medium, and high risk. Afterward, AMCA determines whether the date of the current request time t and the previous request time t' are the same. If it is the same, the corresponding identification score IS is calculated as in (6); otherwise, the module sets the IS to RS.

$$IS = \begin{cases} RS - \left(1 - \frac{\Delta t}{TH} * FS'_{bio}\right), \Delta t < TH \\ RS \quad , otherwise \end{cases} \tag{6}$$

After completing the evaluation, AMCA informs the employee of IS, AS_{otp}, AS_{bio}, AS_{auth}, and F'_{otp}, where AS_{otp}, AS_{bio}, and AS_{auth} are the maximum score that can prove the identification of the employee. If $IS > AS_{otp} + AS_{bio}$, an authorization code F'_{auth} is transmitted to the authorized who bonded with the employee.
Step 3: Upon receiving the IS, the employee returns the corresponding features for authentication according to the requirement in Table 2 The situation required for identity proof mainly depends on whether the current device has a webcam, thus, the authentication modes for individual risk levels are divided into the corresponding two categories. Based on the device and the risk level, the user is asked to provide either one time password F_{otp}, biometrics feature F_{bio}, or authorization code F_{auth} to prove that the identity has not been forged. In case

the identity proof has possessed F_{auth}, the employee sends a request to his authorizer; otherwise, the operation turns to step 5.

Table 2. The requirement for authentication
(✓ means the device equipped with webcam, ● presents that it is a default necessary options、○ means the user can decide whether to add additional feature)

Scenario			Authentication Feature		
Risk	**Risk condition**	**Webcam**	F_{otp}	F_{bio}	F_{auth}
Low	$IS < AS_{otp}$	✓	○	●	-
		-	●	-	-
Medium	$AS_{otp} < IS < AS_{otp} + AS_{bio}$	✓	●	●	-
		-	○	-	●
High	$AS_{otp} + A_{bio} < IS < AS_{otp} + AS_{bio} + AS_{auth}$	✓	○	○	●
		-	●	-	●
	$AS_{otp} + AS_{auth} < IS < AS_{otp} + AS_{bio} + AS_{auth}$	✓	○	○	●
		-	●	-	●
	$AS_{bio} + AS_{auth} < IS < AS_{otp} + AS_{bio} + AS_{auth}$	✓	○	●	●

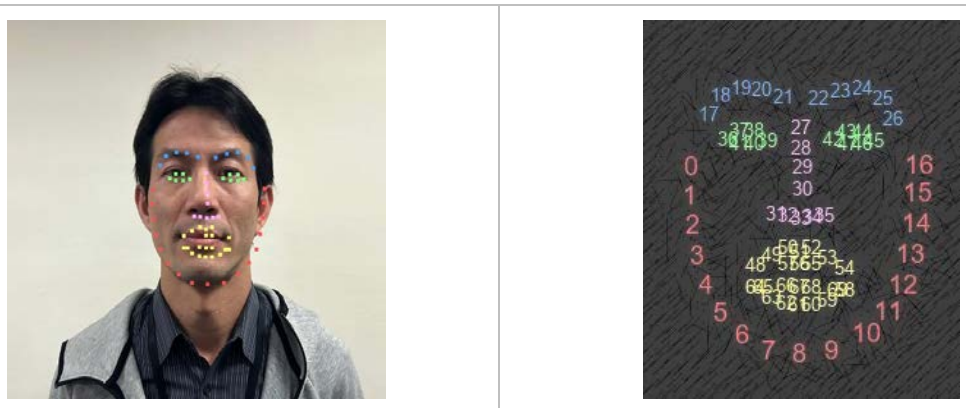
Step 4: After receiving the request from the user, the authorizer confirms whether the behavior of the employee is appropriate. If so, the authorizer transmits F'_{auth} to the employee; otherwise, the current access request fails, and the subsequent steps need not to carry on.
Step 5: The employee submits the features directly to AMCA for authentication.
Step 6: AMCA measures the corresponding feature scores FS_{otp}, FS_{bio}, and FS_{auth} Among them, if FS_{otp} and FS_{auth} are equal to the code issued by the system, they get one point; otherwise, the score is 0. As for FS_{bio}, AMCA acquires 68 feature coordinates from F_{bio} by Dlib [36], as shown in **Fig. 7(a)**. In **Fig. 7(b)**, 1~17 points represent chin; 18~27 display eyebrows; 28~36 denote nose; 37~48 mean eyes; 49~68 indicate lips. Subsequently, those coordinates are converted into 128 feature vectors, and the confidence score FS_{bio} obtained from this biometric is shown in (7),

$$FS_{bio} = 1 - \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{7}$$

Eventually, the employee trust score TS is calculated through (8),

$$TS = FS_{otp} * AS_{otp} + FS_{bio} * AS_{bio} + FS_{auth} * AS_{auth} \tag{8}$$

If $TS \geq IS$, AMCA authorizes the employee to access resources. On the contrary, the authentication fails, and the employee must return to step 3 for re-authentication. If the number of failure trial reaches three continuously, AMCA rejects the request, and the employee is classified as at-risk personnel and has been denied access to resources.



(a) Facial features (b) Coordinates corresponding to the feature

Fig. 7. Labeling of employee

4. Evaluation Results

Prior works [10-12] have conducted numerous trials on incorporating ZTA in various scenarios. Here, we focus specifically on a human behavior-based authentication framework. To demonstrate the practicability of BA-ZTA, a description of the realistic dataset utilized for the simulation and validation based on the behavior of company employees is presented in subsection 4.1. Accuracy of the threat determination module (TD) in estimating user behavior anomalies is conducted in subsection 4.2. At the same time, the efficiency of the adaptive multifactor continuous authentication system (AMCA) is analyzed in subsection 4.3. Finally, the suitability of the BA-ZTA for real-world scenarios is examined in subsection 4.4.

4.1 Construction of Simulation Dataset

To construct a zero-trust environment that is close to the real-world scenario, this study collaborated with 20 employees $U_1, U_2, \ldots U_{20}$ from the information-related company. A behavior dataset is generated based on their individual work habits, consisting of 10 regular employees, 7 long-term overtime employees, and 3 managers. The behavioral log is collected for a period of 14 days, including four holidays. The followings are the basic work principles for employees.

- Working hours: Employees have the flexibility of working hours between 8:00 and 9:00, lunch breaks between 12:00 and 13:30, and off hours between 17:00 and 18:00.
- Overtime: Overtime hours are recorded for employees who work after 18:00.
- Office locations: Some employees have fieldwork responsibilities, so office locations are categorized as internal company locations, domestic business locations, and overseas business locations.
- Threat level of accessing services and resources: Threat levels are defined as high, medium, and low based on the company ISMS [33] definitions. In addition, the potential threat to the overall operation, assets, and reputation of the desired access service or resource is classified according to the grading principles of Annex 9 of the Taiwan Information Security Management System Protection Requirements [37], shown as follows.
  - Extreme risk: Threats of catastrophic proportions include research and development information or operational secrets.

- Moderate risk: Potential threats impact the company but stay manageable, such as the system maintenance information.
- Low risk: Threats can only have a very slight impact on the company, such as a publicly available document.

The dataset is collected according to the above principles, as shown in **Table 3** There are 20 employees with different personal characteristics, which includes 10 general employees, 7 overtime employees, and 3 managers. Among whom, 50% of the general employees are occasionally traveling. All overtime employees belong to the type that may travel occasionally. Noteworthy, manager U_{18} who spends more time in the company than on business trips, and manager U_{19} and U_{20} who travel almost domestically and internationally. **Fig. 8** illustrates the average number of access request by individual types of employees during a period of 24 hours. First of all, it could be observed that the resource request of general employees are relatively stable, which gradually appear starting from 8 am, and the number of requests ranges between 140 to 160 around 9 am. During the lunchtime, it decreases by about 50%, and it keeps a gradually reducing after the flexible off hours at 5 pm. For the overtime employees, request times are similar to general employees during regular working hours, except for that the number of requests during overtime periods is significantly higher than the other categories. Finally, it is worth noting that managers typically have lots of business trips and are usually not required to access company resources, resulting in more fluctuation in the overall number of requests.

Table 3. Work model of company members

Employee	***Category***			***Business trip***				
	General Employees	Overtime Employees	Manager	None	Domestic		Overseas	
					Occasional	Frequent	Occasional	Frequent
U_1	✓				✓			
U_2	✓				✓			
U_3	✓				✓			
U_4	✓				✓			
U_5	✓				✓			
U_6	✓			✓				
U_7	✓			✓				
U_8	✓			✓				
U_9	✓			✓				
U_{10}	✓			✓				
U_{11}		✓			✓			
U_{12}		✓			✓			
U_{13}		✓			✓			
U_{14}		✓			✓			
U_{15}		✓			✓			
U_{16}		✓			✓			
U_{17}		✓			✓		✓	
U_{18}			✓		✓		✓	
U_{19}			✓			✓		✓
U_{20}			✓			✓		✓

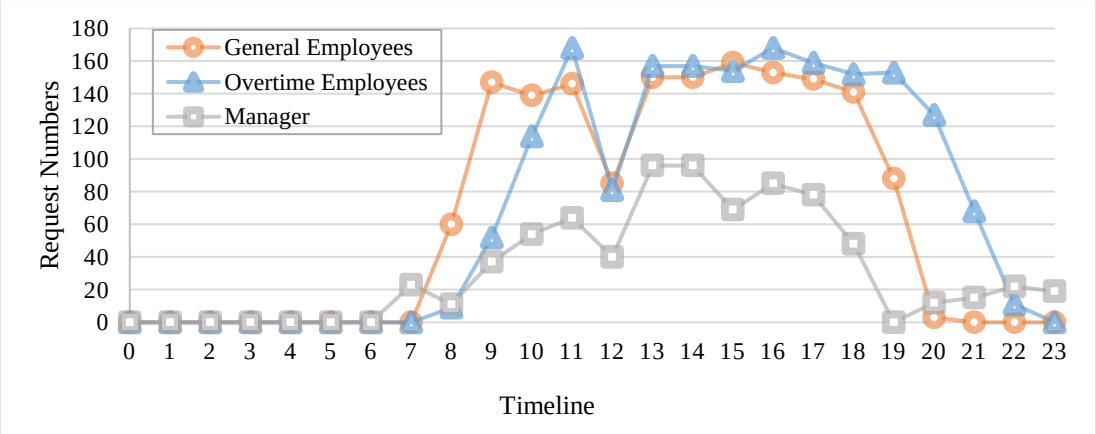


Fig. 8. Average number of accesses in 24 hours

4.2 Analysis of Threat Determination Module

To examine the accuracy of the predicted threat level, we have manually labeled and categorized each resource request of the dataset into high, medium, and low. Consequently, we constructed the experiment to determine whether the risk with the highest probability is the same as the risk label based on the results of TD. In the measurement, accuracy, precision, recall, and F1-score [38] are indicators used to certify the reliability of user behavior assessment. According to the assessment results presented in **Table 4**, it could be observed that the overall accuracy for different categories of employees reaches at least 0.925, which indicates BA-ZTA possessing a certain degree of threat determination. The performance of all categories of employees is depicted in **Fig. 9**.

Table 4. Threat assessment results of each employee

Category	Employee	Accuracy	Precision	Recall	F1
General Employee	U_1	0.962	0.970	0.932	0.951
	U_2	0.994	0.996	0.990	0.993
	U_3	0.988	0.987	0.987	0.987
	U_4	0.969	0.979	0.944	0.961
	U_5	0.991	0.981	0.994	0.987
	U_6	0.986	0.987	0.984	0.985
	U_7	0.955	0.960	0.952	0.956
	U_8	0.994	0.993	0.993	0.993
	U_9	0.988	0.992	0.982	0.987
	U_{10}	0.945	0.944	0.940	0.942
Average		**0.977**	**0.979**	**0.970**	**0.974**
Overtime Employee	U_{11}	0.931	0.941	0.921	0.931
	U_{12}	0.925	0.930	0.930	0.930
	U_{13}	0.924	0.925	0.924	0.924

	U_{14}	0.933	0.944	0.927	0.935
	U_{15}	0.967	0.975	0.956	0.965
	U_{16}	0.949	0.942	0.954	0.948
	U_{17}	0.939	0.924	0.939	0.931
Average		**0.938**	**0.940**	**0.936**	**0.938**
Manager	U_{18}	0.987	0.987	0.987	0.987
	U_{19}	0.944	0.915	0.916	0.915
	U_{20}	0.946	0.910	0.949	0.929
Average		**0.959**	**0.937**	**0.951**	**0.944**

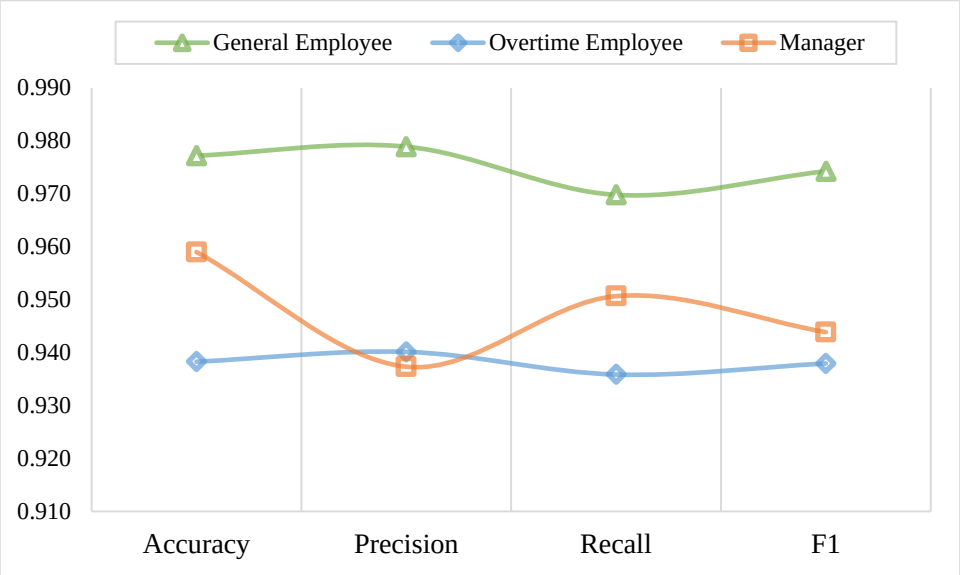


Fig. 9. Average evaluation results of different categories of employee

The simulation outcomes have significantly demonstrated that the general employee achieves the best result in the overall indicator The main reason is that the employees basically access resources during the normal commuting hours. Once employee deviates from their regular work patterns, BA-ZTA can accurately determine the risk. However, the performance of U_1, U_4, U_7 and U_{10} is poor compared to other general employees, as the recall values have declined to 0.932, 0.944, 0.952, and 0.940, respectively. The underlying instability is the occasional overtime, but BA-ZTA relies on the constant behavior of the employee to obtain the evaluation. Therefore, the system considers the risk value as low when the request time is very close to the overtime period. In fact, the access risk is more dangerous during the non-working hours.

Regrettably, an overtime employee is the weakest adaptability of our method. Since the working hours of this category are unstable and unpredictable, the overall assessment of overtime employees is slightly lower than others. In particular, the perception of U_{15} is exceptionally better because U_{15} is the most frequent overtime worker, which has convinced the system to consider overtime as the norm. In the evaluation of the managers, we can observe that the performance of accuracy, recall and F1 stay in mid-range, except for the value of

precision, which is the worst performance among all types. This phenomenon is because managers often perform the high-risk behaviors, such as accessing confidential or sensitive data. Thus, BA-ZTA accidentally determines the general access request to be risky. Additionally, U_{18} is a person who travels infrequently and mostly spends time in the company. Hence, BA-ZTA can precisely assess each request. Although the effectiveness of the risk assessment varies for different types of employees, the BA-ZTA can maintain a satisfactory level of performance in general.

4.3 Efficiency evaluation of AMCA

Adaptive multifactor continuous authentication model is designed to optimize employee-friendly and mitigate the dilemma of repeated authentication within zero trust architecture. AMCA adopts face as the core authentication feature, and employees can be unconsciously authenticated, while BA-ZTA obtains the most accurate and representative information. One significant rationale for friendly authentication is that there is a high likelihood of the employee being in front of the device with a webcam when they request company resources; thus, making BA-ZTA be able to capture a face shot easily for authentication. According to the design shown in **Table 1**, as long as IS is within the acceptable S_{bio}, the employee could pass the authentication directly. Furthermore, sensitive data with potential extreme risk is insufficient solely on one-time passwords and facial recognition. AMCA has devised an authorization code in steps 3 and 4 of subsection 3.2 to achieve multi-person authentication. This mechanism is a twofold authentication safeguard, where a binding authorization code protector can verify the normalcy of the employee behavior in person, even if the other two characteristics are imitated by an adversary.

Significantly, the facial can be considered as the core of BA-ZTA, which is used to solve the difficulties of complicated authentication and dramatically improve the feasibility of zero trust framework. In BA-ZTA, the human face is the most convenient core among all the verification factors. However, it is also unstable because the employee cannot always face the camera positively while validation is in progress. Therefore, we have simulated the unintentional twist of the employee to resemble the real-life situation more closely. The usual left and right swing amplitude are set to +45°~-45°, as shown in **Fig. 10**. Subsequently, AMCA performs fractional authentication of facial features to obtain F_{bio}, where three different graphic size of 128*128, 256*256, and 512*512 are used for simulation. **Fig. 11** presents the experimental findings, which indicating that BA-ZTA flawlessly recognize F_{bio} at 0°. Regardless of the graphic size, the same facial characteristics can be precisely extracted as those previously recorded in BA-ZTA. As previously noted, it is impractical to maintain a fixed orientation of the face toward the camera. The face swinging from side to side often leads to the feature and quality degradation in AMCA capturing. Despite the facial degradation, BA-ZTA can still recognize faces successfully due to incomplete acquisition of facial features. Moreover, AMCA provides an appropriate score based on the acquired features; thus, demonstrating the practicality of adaptive BA-ZTA in real-life scenarios.

(a) -45°	(b) -15°	(c)0°	(d)15°	(e)45°

Fig. 10. Schematic diagram of facial features

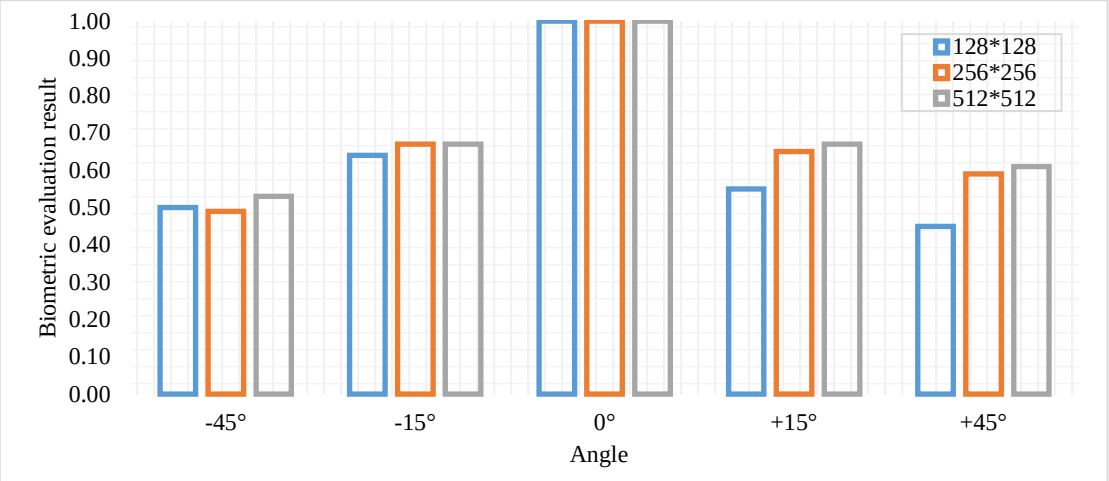


Fig. 11. Facial feature scores in various angles and image quality

4.4 Examination of BA-ZTA

Authentication efficiency is crucial when implementing the zero trust framework, as it involves continuously authenticating users before granting access to resources. In a real-life workplace, the stochastic nature of authentication numbers poses a challenge to system practicability and scalability. Simulations offer valuable insights into BA-ZTA under different scenarios and loads, which are constructed in Python with a personal computer running Windows 10 64-bit. It is merely equipped with an Intel Core i5-9400 2.9-GHz, NVIDIA RTX 2060, and 16GB RAM. Regrettably, it is important to acknowledge that the experimental results may not fully represent the performance of the actual server host. This is because of a significant performance gap between the experimental equipment and the actual server host. Notwithstanding, if the experimental result can perform well with relatively inferior equipment, BA-ZTA definitely renders excellent performance and scalability when being adopted in an actual server.

In order to measure realism, the experiment evaluates the computation time required for the BA-ZTA to authenticate multiple requests simultaneously, as illustrated in **Fig. 12**. The simulation considers a scenario with 0 to 2000 requests to evaluate the scalability of BA-ZTA. Meanwhile, each validation must determine the biometric score FS_{bio}, which is undoubtedly the most burdensome case for the proposed system to carry out the assessment. Based on the experimental data, it is apparent that BA-ZTA demonstrates undeniable efficiency in the 128*128 and 256*256 authentication, where 2000 requests can be handled with a mere 0.7s and 1.3s, respectively. However, BA-ZTA fails to achieve the real-time performance when asking the employee to submit a 512*512 biometric. The time required to accomplish 700 requests is substandard to the preceding performance. Compared with a 256*256 scenario, 2000 requests cost nearly three times longer to complete, reaching 3.8s. Although the performance of BA-ZTA is limited in high-quality images with the current restricted equipment, we can actually discover from the measurement in subsection 4.3 that the quality of biometrics has no critical impact on the recognition of BA-ZTA. Conclusively, BA-ZTA has fully demonstrated its practicality and scalability to implement in an authentic environment.

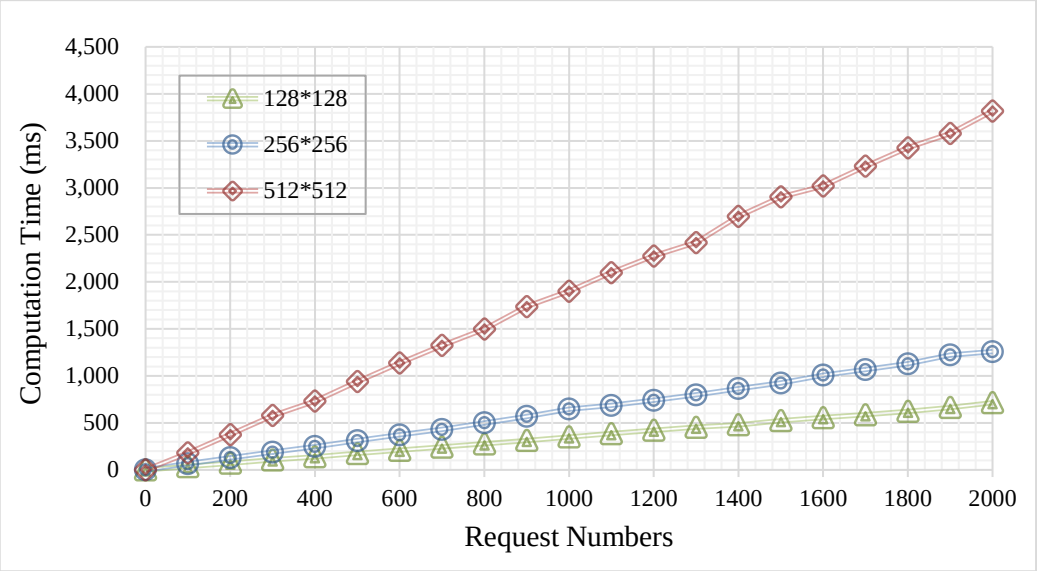


Fig. 12. Performance of BA-ZTA

5. Conclusion

The adoption of ZTA has brought enterprise stronger support in the protection against cyber-attack. However, the process of continuous and complex authentication also leads to a troublesome and unfriendly burden for enterprise and employee. In this article, we have imported the machine learning technique to design a brand-new ZTA based on the behavior analysis of employees. Specifically, AMCA, which is the core subsystem, can dynamically adjust the feature that employees have to provide for identification according to the threat level; thus, optimizing employee-friendly and mitigating the dilemma of repeated authentication. Without the help of complex verification equipment, diverse simulation outcomes have demonstrated that BA-ZTA can effectively eliminate the cost of maintaining a heavy authentication and ensure an employee-friendly experience.

References

[1] M. Lorch, D. B. Adams, D. Kafura, M. S. R. Koneni, A. Rathi, and S. Shah, "The PRIMA System for Privilege Management, Authorization and Enforcement in Grid Environments," in *Proc. of First Latin American Web Congress*, pp. 109-116, Jan. 2004. Article (CrossRef Link).

[2] S. Khanvilkar and A. Khokhar, "Virtual Private Networks: An Overview with Performance Evaluation," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 146-154, Oct. 2004. Article (CrossRef Link).

[3] R. Oppliger, "Internet Security: Firewalls and Beyond," *Communications of The ACM*, vol. 40, no. 5, pp 92-102, May 1997. Article (CrossRef Link).

[4] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An Intelligent Intrusion Detection System (IDS) for Anomaly and Misuse Detection in Computer Networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713-722, Nov. 2005. Article (CrossRef Link).

[5] A. Aleroud and L. Zhou, "Phishing Environments, Techniques, and Countermeasures: A Survey," *Computers & Security*, vol. 68, pp. 160-196, Jul. 2017. Article (CrossRef Link).

[6] J. R. V. d. Merwe, X. Zubizarreta, I. Lukčin, A. Rügamer, and W. Felber, "Classification of Spoofing Attack Types," in *Proc. of 2018 European Navigation Conference*, pp. 91-99, Aug. 2018. Article (CrossRef Link).
[7] L. Bilge and T. Dumitraş, "Before We Knew It: An Empirical Study of Zero-day Attacks in The Real World," in *Proc. of The 2012 ACM Conference on Computer and Communications Security*, pp. 833-844, Oct. 2012. Article (CrossRef Link).
[8] Z. H. Tian, W. Shi, Y. H. Wang, C. S. Zhu, X. J. Du, S. Su, Y. B. Sun, and N. Guizani, "Real-Time Lateral Movement Detection Based on Evidence Reasoning Network for Edge Computing Environment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4285-4294, Jul. 2019. Article (CrossRef Link).
[9] S. Gatlan, "Cisco Hacked by Yanluowang Ransomware Gang, 2.8GB Allegedly Stolen," *Bleeping Computer*, Aug. 2022. [Online]. Available: https://www.bleepingcomputer.com/news/security/cisco-hacked-by-yanluowang-ransomware-gang-28gb-allegedly-stolen/.
[10] J. Kindervag, "Build Security into Your Network's DNA: The Zero Trust Network Architecture," *Forrester Research*, Nov. 2010. [Online]. Available: https://www.virtualstarmedia.com/downloads/Forrester_zero_trust_DNA.pdf.
[11] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "NIST Special Publication 800-207 Zero Trust Architecture," *NIST*, Aug. 2020. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-207/final.
[12] National Center for Cyber Security Technology, "Government Zero Trust Network Description," Jul. 2022. [Online]. Available: https://download.nics.nat.gov.tw/UploadFile/zerotrustnetworks/%e6%94%bf%e5%ba%9c%e9%9b%b6%e4%bf%a1%e4%bb%bb%e7%b6%b2%e8%b7%af%e8%aa%aa%e6%98%8e_V1.9_1110712.pdf.
[13] D. Pohn and W. Hommel, "IMC: A Classification of Identity Management Approaches," *Computer Security*, vol. 12580, pp. 3-20, Dec. 2020. Article (CrossRef Link).
[14] F. Wang, G.S. Li, Y.L. Wang, W. Rafique, M. R. Khosravi, G.F. Liu, Y.W. Liu, and L.Y. Qi, "Privacy-aware Traffic Flow Prediction based on Multi-party Sensor Data with Zero Trust in Smart City," *ACM Transactions on Internet Technology*, vol. 23, no. 3, pp. 1-19, 2023. Article (CrossRef Link).
[15] L. Meng, D.C. Huang, J.H. An, X.W. Zhou, and F.H. Lin, "A Continuous Authentication Protocol without Trust Authority for Zero Trust Architecture," *China Communications*, vol. 19, no. 8, pp. 198-213, Aug. 2022. Article (CrossRef Link).
[16] L. Ferretti, F. Magnanini, M. Andreolini, and M. Colajanni, "Survivable Zero Trust for Cloud Computing Environments," *Computers Security*, vol. 110, pp.1-18, Nov. 2021. Article (CrossRef Link).
[17] Z. M. Zhang, H. S. Ning, F. Farha, J. G. Ding, and K. K. R. Choo, "Artificial Intelligence in Physiological Characteristics Recognition for Internet of Things Authentication," *Digital Communications and Networks*, pp. 1-23, Oct. 2022. Article (CrossRef Link)
[18] M. M. Taha, T. A. Alhaj, A. E. Moktar, A. H. Salim, and S. M. Abdullah, "On Password Strength Measurements: Password Entropy and Password Quality," in *Proc. of 2013 International Conference on Computing, Electrical and Electronic Engineering*, pp. 497-501, Aug. 2013. Article (CrossRef Link).
[19] D. Balfanz, "FIDO U2F Implementation Considerations," *FIDO Alliance Proposed Standard*, pp. 1-5, 2015. [Online]. Available: https://fidoalliance.org/specs/fido-u2f-implementation-considerations-ps-20150514.pdf.
[20] O. Alpar, "Intelligent Biometric Pattern Password Authentication Systems for Touchscreens," *Expert Systems with Applications*, vol. 42, no. 17-18, pp. 6286-6294, Oct. 2015. Article (CrossRef Link).

[21] V. L. Shivraj, M. A. Rajan, M. Singh, and P. Balamuralidhar, "One Time Password Authentication Scheme based on Elliptic Curves for Internet of Things (IoT)," in *Proc. of 2015 5th National Symposium on Information Technology: Towards New Smart World*, pp. 1-6, Feb. 2015. Article (CrossRef Link).

[22] H. Kumar, S. Kumar, R. Joseph, D. Kumar, S. K. S. Singh, A. Kumar, and P. Kumar, "Rainbow Table to Crack Password using MD5 Hashing Algorithm," in *Proc. of 2013 IEEE Conference on Information & Communication Technologies*, pp. 433-439, Apr. 2013. Article (CrossRef Link).

[23] S. A. Baho and J. Abawajy, "Analysis of Consumer IoT Device Vulnerability Quantification Frameworks," *Electronics*, vol. 12, no. 5, Feb. 2023. Article (CrossRef Link).

[24] M. La Polla, F. Martinelli, and D. Sgandurra, "A Survey on Security for Mobile Devices," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 446-471, 2013. Article (CrossRef Link).

[25] A. Bhardwaj and S. Goundar, "Keyloggers: Silent Cyber Security Weapons," *Network Security*, vol. 2020, no. 2, Nov. 2021. Article (CrossRef Link).

[26] A. Bedari, S. Wang and J. Yang, "A Two-Stage Feature Transformation-based Fingerprint Authentication System for Privacy Protection in IoT," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2745-2752, Apr. 2022. Article (CrossRef Link).

[27] J. E. Tapia, S. Gonzalez, and C. Busch, "Iris Liveness Detection Using a Cascade of Dedicated Deep Learning Networks," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 42-52, 2022. Article (CrossRef Link).

[28] J. Lei, Q. Pei, Y. Wang, W. Sun, and X. Liu, "PrivFace: Fast Privacy-Preserving Face Authentication with Revocable and Reusable Biometric Credentials," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 5, pp. 3101-3112, Sep.-Oct. 2022. Article (CrossRef Link).

[29] J. M. E. López, A. H. Celdrán, F. Esquembre, G. M. Pérez, and J. G. Marín-Blázquez, "A Supervised ML Biometric Continuous Authentication System for Industry 4.0," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 9132-9140, Dec. 2022. Article (CrossRef Link).

[30] M. L. Ali, K. Thakur, and M. A. Obaidat, "A Hybrid Method for Keystroke Biometric User Identification," *Electronics*, vol. 11, no. 17, Sep. 2022. Article (CrossRef Link).

[31] C. Shen, Y. Chen, X. Guan, and R. A. Maxion, "Pattern-Growth Based Mining Mouse-Interaction Behavior for an Active User Authentication System," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 2, pp. 335-349, Mar.-Apr. 2020. Article (CrossRef Link).

[32] B. Liu, L.J. Xiao, J. Long, M.D. Tang, and O. Hosam, "Secure Digital Certificate-based Data Access Control Scheme in Blockchain," *IEEE Access*, vol. 8, pp. 91751-91760, May. 2020. Article (CrossRef Link).

[33] ISO/IEC JTC 1/SC 27 Information Security, Cybersecurity and Privacy Protection, "ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection - Information Security Management Systems - Requirements," Oct. 2022. [Online]. Available: https://www.iso.org/standard/82875.html.

[34] D. Tax and R. Duin, "Support Vector Data Description," *Machine Learning*, vol. 54, pp. 45-66, Jan. 2004. Article (CrossRef Link).

[35] W. S. McCulloch and W. Pitts, "A Logical Calculus of The Ideas Immanent in Nervous Activity," *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, Dec. 1943. Article (CrossRef Link).

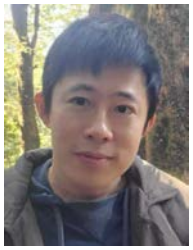
[36] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755-1758, Dec. 2009. Article (CrossRef Link).

[37] Ministry of Digital Affairs, "Security Responsibility Level Classification Method," Aug. 2021. [Online]. Available: https://www-api.moda.gov.tw/File/Get/3OcL9qFXX69sUwP.

[38] D. Powers, "Evaluation: from Precision Recall and F-measure to ROC Informedness Markedness and Correlation," *International Journal of Machine Learning Technology*, vol. 2, pp. 37-63, Jan. 2011. Article (CrossRef Link).

Chit-Jie Chew is pursuing his PhD degree in information engineering and computer science in Feng Chia University, Taichung, Taiwan. His current research interests include information security and blockchain applications.

Po-Yao Wang is pursuing his MS degree in information engineering and computer science in Feng Chia University, Taichung, Taiwan. His current research interests include zero trust architecture and network security.



Jung-San Lee received his Ph.D. degree in computer science and information engineering in 2008 from National Chung Cheng University, Chiayi, Taiwan. Since 2017, he has worked as a professor in the Department of Information Engineering and Computer Science at Feng Chia University, Taichung, Taiwan. His current research interests include network management, electronic commerce, and blockchain.