CISTER

Research Centre in
Real-Time & Embedded
Computing Systems

Journal Paper

Mobile fog computing security: A user-oriented smart attack defense strategy based on DQL

Shanshan Tu Muhammad WaqasYuan MengSadaqat Ur RehmanIftekhar AhmadAnis KoubâaZahid HalimMuhammad HanifChin-Chen ChangChengjie Shi

CISTER-TR-201015

Mobile fog computing security: A user-oriented smart attack defense strategy based on DQL

Shanshan Tu Muhammad WaqasYuan MengSadaqat Ur RehmanIftekhar AhmadAnis KoubâaZahid HalimMuhammad HanifChin-Chen ChangChengjie Shi

CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail:

https://www.cister-labs.pt

Abstract

Each fog node interacts with data from multiple end-users in mobile fog computing (MFC) networks. Malicious users can use a variety of programmable wireless devices to launch different modes of smart attacks such as impersonation attack, jamming attack, and eavesdropping attack between fog servers and legitimate users. The existing research in MFC lacks in the contributions of defense of smart attack and also requires in the discussions of subjective decision making by participants. Therefore, we propose a smart attack defense scheme for authorized users in MFC in this paper. First, we construct a static zero-sum game model between smart attackers and legitimate users based on prospect theory. Second, the double Q-learning (DQL) is proposed to restrain the attack motive of smart attackers in the dynamic environment. The proposed DQL method generates the optimum defense choice of legitimate users against smart attacks so that they can efficiently determine whether to use only physical layer security (PLS) to avoid those smart attacks. We use our scheme to contrast with the basic schemes, i.e., Q-learning scheme, the Sarsa scheme, and the greedy strategy. Experiment results prove that the proposed scheme can enhance the utility of legitimate users, restrain the attack motive of smart attackers, and further provide better security protection in the MFC environment.

Computer Communications 160 (2020) 790–798

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Mobile fog computing security: A user-oriented smart attack defense strategy based on DQL

Shanshan Tu [a,b], Muhammad Waqas [a,c], Yuan Meng [a,*], Sadaqat Ur Rehman [d], Iftekhar Ahmad [e], Anis Koubaa [f,g], Zahid Halim [c], Muhammad Hanif [c], Chin-Chen Chang [h,i], Chengjie Shi [j]

[a] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[b] Beijing Electro-Mechanical Engineering Institute, Beijing, 100074, China
[c] Department of Computer Science & Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences & Technology, KPK, 23460, Pakistan
[d] Department of Computer Science, Namal Institute, Mianwali 42250, Pakistan
[e] School of Engineering, Edith Cowan University, Perth 6027, Australia
[f] Robotics and Internet-of-Things research lab, Department of Computer Science, Prince Sultan University, R&D Gai-tech Robotics, Saudi Arabia
[g] CISTER/INESC TEC and ISEP-IPP, Porto, Portugal
[h] Department of Information Engineering and Computer Science, Feng Chia University, Taichung, 40724, Taiwan
[i] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China
[j] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100195, China

ARTICLE INFO

Keywords:
Mobile fog computing
Smart attack
Prospect theory
Reinforcement learning
Game theory
Physical layer security

ABSTRACT

Each fog node interacts with data from multiple end-users in mobile fog computing (MFC) networks. Malicious users can use a variety of programmable wireless devices to launch different modes of smart attacks such as impersonation attack, jamming attack, and eavesdropping attack between fog servers and legitimate users. The existing research in MFC lacks in the contributions of defense of smart attack and also requires in the discussions of subjective decision making by participants. Therefore, we propose a smart attack defense scheme for authorized users in MFC in this paper. First, we construct a static zero-sum game model between smart attackers and legitimate users based on prospect theory. Second, the double Q-learning (DQL) is proposed to restrain the attack motive of smart attackers in the dynamic environment. The proposed DQL method generates the optimum defense choice of legitimate users against smart attacks so that they can efficiently determine whether to use only physical layer security (PLS) to avoid those smart attacks. We use our scheme to contrast with the basic schemes, i.e., Q-learning scheme, the Sarsa scheme, and the greedy strategy. Experiment results prove that the proposed scheme can enhance the utility of legitimate users, restrain the attack motive of smart attackers, and further provide better security protection in the MFC environment.

1. Introduction

Due to recent developments in the internet of things (IoT) and mobile terminals, a large number of generated data need to be processed in real-time in wireless networks. The traditional cloud computing cannot adequately meet the network requirements such as heterogeneity and low latency. Fog computing migrates some computation tasks from the cloud to the margin of the network, which may improve the system throughput by using direct transmission links. It solves the problems of reduced mobility, weak geographic information perception, and high latency of cloud computing. However, it also brings the issues of communication and data security [1–8]. In MFC, smart attackers can use smart attack methods to attack other legitimate users. Unlike other attacks, any user launches smart attacks through a smart programmable wireless device or a wireless network access platform. They destroy wireless network security between the fog layer and the user layer by acquiring wireless channel state information and defense strategy information and choosing appropriate attack modes. The attacker can launch multiple types of smart attacks, such as impersonation, jamming, replay, and eavesdropping [9–12] towards the IoT devices or keep silence in MFC networks. To ensure the security of fog computing networks and prevent threats from smart attackers, game theory is considered to be a powerful tool. It can show the incentive relationship between the subjects. As a mathematical method, it calculates the benefits of the subjects [10–12].

* Corresponding author.
E-mail addresses: sstu@bjut.edu.cn (S. Tu), engr.waqas2079@gmail.com (M. Waqas), 690146986@qq.com (Y. Meng), engr.sidkhan@gmail.com (S.U. Rehman), i.ahmed@ecu.edu.au (I. Ahmad), anis.koubaa@gmail.com (A. Koubaa), zahid.halim@giki.edu.pk (Z. Halim), muhammad.hanif@giki.edu.pk (M. Hanif), alan3c@gmail.com (C.-C. Chang), shichengjie@iie.ac.cn (C. Shi).

https://doi.org/10.1016/j.comcom.2020.06.019
Received 3 February 2020; Received in revised form 21 April 2020; Accepted 17 June 2020
Available online 22 June 2020

In MFC security, researchers believe that participants in the game are rational. They use expected utility theory (EUT) to calculate participants' utility. Participants select each step for maximum expected utility. However, in dynamic wireless networks, each participant does not know the overall network state and the accuracy of the received data packages. Therefore, the decision-making of each participant in the offensive and the defensive game has an intense subjectivity. There are differences in benefits with the results of the EUT. To deal with the above problem, we can use the prospect theory (PT) [13–16] to generate a subjective decision model and study the zero-sum game between smart attackers and legitimate users. The probability weighting function is used by PT to calculate. The Nash equilibrium (NE) of both sides of the game is analyzed to restrain the intended motive of smart attackers.

Besides, the actual MFC network is dynamic, legitimate users and smart attackers can continuously interact. Authorized users can use two different defense modes, i.e., whether to apply the higher-layer security mechanism (HLSM) [17] for better protection or only the physical layer security (PLS) mechanism to reduce system overhead. Then, the users would obtain optimal defense strategies in a dynamic environment without realizing the system model's details by using reinforcement learning methods. Reinforcement learning algorithms can interact with the environment and make dynamic decisions. They have strong adaptability to various network environments. MFC network is a dynamic environment with exposure characteristics. After completing an action, it is vulnerable to attack. And all legitimate end-users in the network can resist the attack by performing actions. Therefore, reinforcement learning algorithms can be used to optimize the decision of legitimate users, enhance the ability to solve security problems in a wireless network and reduce the attack probability. At the same time, as one of the reinforcement learning algorithms, the double Q-learning (DQL) algorithm deals with the problem of the q-learning algorithm: overestimation of Q value. It is an efficient, reinforcement learning algorithm [18]. Therefore, in this study, we propose a smart attack defense scheme for legitimate users in MFC, which uses PT and the DQL algorithm to obtain the optimal defense strategy and enhance the detection utility of authorized users. The balance between the MFC security and the system overhead caused by legitimate users' choice of defense strategies is made by reinforcement learning theory in the scheme.

Fog computing faces many security challenges. To solve the security problem between fog nodes and legitimate end-users, Hu et al. [19] studied a method to calculate the inverse of the matrix using fog nodes and guaranteed the security of user data. This method has an advantage of security and verifiability, but it lacks in what concerns user authentication before the data is transmitted. Chaudhary et al. [20] studied a 5G-based network service chain model. They implemented an end-user authentication method to deal with the distributed denial of service (DDoS) attacks for cloudlet services in the fog layer. However, this method uses several verification bills, and the verification steps are more cumbersome. Besides, Tu et al. [21] applied a Q-learning method in MFC to detect impersonation attacks. The process based on EUT can reduce an average error rate of testing impersonation attacks in MFC. But the attacking modes by the smart attackers need further study. On the other hand, Xiao et al. [22] provided NE for a mobile offloading game between smart attackers and end-users utilizing EUT, where smart attackers can carry out both impersonation attack and jamming attack at the same time. This scheme also cannot take the probability of bias brought by smart attackers into account. Besides, Yang et al. [23] proposed a jamming attack game based on PT. It describes the influence of jamming attackers and end-users on the signal to interference plus noise ratio (SINR) in the dynamic scenario with changeable channel gain. PT is a theory that describes the users' risk preferences for decision making. The theory holds that people avoid risks when they are faced with "gain" and prefer risks when they are faced with "loss." It calculates the utility of participants in the game by applying subjective probability. The proposed theory has not been implemented in the fog computing environment yet, and it may provide a useful direction towards our new research.

Also, regarding the role of reinforcement learning methods in dynamic environments, many researchers used reinforcement learning algorithms and greedy strategies to solve the optimal solution of income value, action value, or action sequence. Zhou et al. [24] proposed a delayed forward algorithm based on a greedy strategy to obtain the essential k nodes in the model, which improved the convergence efficiency. Chen et al. [25] combined genetic network planning with the Sarsa algorithm and applied it to network transaction rules, which improved transaction profits. Tu et al. [21] used the Q-learning algorithm to fog the environment to enhance the accuracy of impersonation attack detection. However, the algorithms involved in the above research works need to be more widely used in MFC networks so that researchers can comprehensively analyze their performance and effects on MFC networks.

The research mentioned above only considers the traditional key security technology and the security protection technology based on the EUT game. It does not provide any solution for fog computing networks concerning smart attackers. Meanwhile, it lacks the application of reinforcement learning algorithm in the context of fog computing. Therefore, with the help of the PT and DQL algorithm, this paper proposes a user-oriented smart attack defense scheme in MFC. The contributions of this study can be summarized as follows:

- This study proposes a security model in an MFC network. Five kinds of smart attack modes against legitimate users between the fog layer and the end-user layer are studied. The attacks include silence, jamming, eavesdropping, impersonation, and replay attack. Meanwhile, based on a higher-layer security mechanism, two defense strategies for legitimate users are proposed. One is a defense mechanism only using physical layer security; the other is a defense mechanism combining HLSM and PLS.
- Secondly, we propose the zero-sum game between smart attackers and legitimate users. Moreover, we deduct the NE of the static zero-sum game and study the influence of decision-making of attackers and authorized users for security purposes in fog computing networks.
- Finally, a method based on DQL to smart defense attacks in a dynamic environment is proposed, which can acquire the optimum defense choices for legitimate users. In this study, experiment results show that lower objective probability weight can suppress the possibility of an attack from the smart attacker. We use our scheme to contrast with the basic schemes, i.e., the Q-learning scheme, the Sarsa scheme, and the greedy strategy to resist against smart attacks. Experimental results indicate that the proposed method optimizes the decision-making process for the optimal defense strategy, improves the utility of legitimate end-users, and reduces the attack rate by adjusting the Q value. They show that the proposed method has better security protection capability.

The rest of the paper is organized as follows. Section 2 provides details about the system model and static zero-sum game, and the DQL method to prevent smart attacks is elaborated in Section 3. The summary of the proposed scheme and performance analysis are discussed in Section 4. Finally, Section 5 concludes this paper.

2. System model and methodology of static zero-sum game

This Section introduces the security model and the method of a static zero-sum game between the smart attackers and legitimate users in MFC.

Table 1
Five alternative attack modes.

Attack mode	Symbol	Interpretation
Keep silent	$SA_1^t = 0$	The attacker does not choose to launch smart attacks.
Jamming attack	$SA_1^t = 1$	The attacker sends jamming signals.
Eavesdropping attack	$SA_1^t = 2$	The attacker intercepts the information between fog plane and end-user plane.
Impersonation attack	$SA_1^t = 3$	The attacker impersonates the fog nodes to send data to make the legitimate users receive.
Replay attack	$SA_1^t = 4$	The attacker sends the data packets that the legitimate users have received.

2.1. System model

In MFC, the fog nodes apply wireless networks to communicate with the end-users. A Fog node is a device located close to the margin of IoT and is a data preprocessing unit for IoT devices [8]. Considering the communication between the fog layer and the end-user layer, any smart attacker in the MFC network may launch smart attacks against legitimate users. Thus, we define the set of smart attackers as $M = \{1, 2, \ldots, m\}, \forall m \in M$, and their attack patterns at time t are represented as SA_m^t, with the set of time values expressed as $T = \{0, 1, \ldots, t\}, \forall t \in T$. Moreover, the set of legitimate users is represented as $N = \{1, 2, \ldots, n\}, \forall n \in N$. At time t, $T = \{1, 2, \ldots, t\}, \forall t \in T$, their defense patterns are represented as EU_n^t. Suppose at a time t, and smart attacker 1 uses a smart programmable wireless device to launch a smart attack on a legitimate user when the legitimate user communicates with a fog node as it in mode SA_1^t. When $SA_1^t = 0$, it indicates that the attacker keeps silent, and $SA_1^t = 1$ indicates the attacker attacks legitimate users by sending jamming signals, which can reduce SINR. When $SA_1^t = 2$, it indicates that the attacker uses eavesdropping attack mode to intercept the information between fog plane and end-user plane. When $SA_1^t = 3$, the attacker uses a false media access control address (MAC-A) to impersonate the true fog nodes to send data to make the legitimate users receive, i.e., the attacker adopts an impersonation attack mode. When $SA_1^t = 4$, the attacker uses replay attack mode to send the data packets that the legitimate user has received, to deceive the legitimate user. Table 1 summarizes five alternative attack modes.

There are two defense modes for the attacked legitimate user EU_1^t when facing impersonation attack, jamming attack, replay attack, an eavesdropping attack. When $EU_1^t = 1$, the legitimate user only uses the physical layer to defend against smart attacks. When $EU_1^t = 2$, legitimate users will spend more system overhead. First, physical layer security technology based on channel parameters is used for preliminary detection, filtering, and eavesdropping prevention. Then the higher-layer security mechanism is used to detect data validated by the physical layer. According to the general model structure of fog computing network, we construct a security model including smart attackers, as shown in Fig. 1, where smart attackers can choose the five attack modes, while legitimate users can choose the two defense modes.

2.2. Methodology of static zero-sum game

Next, we construct a static zero-sum game between smart attackers and legitimate users, in which SA_m represents the attack mode; $Num, Num \geq 1$ represents the number of attack modes; and EU_n denotes the defense mode. Based on PT, smart attackers and legitimate users adopt a decision-making game to achieve NE. According to [26], game participants take subjective probability as the benchmark to participate in decision-making. In this study, the Prelec probability weight function is used to calculate the subjective probability, denoted by

$$W_{object}(p) = e^{-(-\ln p)^{\sigma_{object}}}, \tag{1}$$

where

- p is objective probability, $p \in (0, 1]$;
- σ_{object} represent the objective probability weight, $\sigma_{object} \in (0, 1]$;
- $object$ is the representation of the player of game, in this study, $object = attac$ or $object = user$.

The Prelec probability weight function [26] describes the result that the objective probability of decision-making is adjusted by the game object participating in a game because of the influence of the weight. Enlightened by PT, when faced with high probability events, decision-makers underestimate the corresponding objective probability. Conversely, when faced with low probability events, decision-makers overestimate the similar objective likelihood. Therefore, PT can be mathematically described by Prelec probability weight function. In a static zero-sum game, for the legitimate users, the gain from detecting smart attack SA_m, in defensive mode EU_n, is denoted by $G_{SA_m}^{EU_n}$. If no smart attack is detected, the security loss suffered is indicated by $L_{SA_m}^{EU_n}$.

In any defense mode, there is the rate of false positives and miss detection. The rate of false positives refers to the probability that the legitimate data sent by the legitimate node detected as illegal data. The rate of miss detection indicates the probability that the illegal data detected as legitimate data, which is the reason for the security loss. Therefore, combining these two ratios, the error rate of legitimate users detecting smart attack SA_m in defense mode EU_n is denoted by $R_{SA_m}^{EU_n}$. According to the system model, there are five attack modes of smart attackers and two defense modes of legitimate users. The utility of smart attackers and legitimate users are expressed as:

$$U_{user}(SA_m, EU_n) = -U_{attac}(SA_m, EU_n) = G_{SA_m}^{EU_n} - R_{SA_m}^{EU_n} L_{SA_m}^{EU_n}, \tag{2}$$

where $R_{SA_m}^{EU_n} \in [0, 1]$, and is quantified to C non-zero levels. $P_c^{SA_m, EU_n}$ is the probability of $R_{SA_m}^{EU_n} = c/C$, and satisfies the probability distribution of $[P_c^{SA_m, EU_n}]_{0 \leq c \leq C}$, where $P_c^{SA_m, EU_n} \geq 0$, $p = \sum_{c=0}^{C} P_c^{SA_m, EU_n} = 1$. According to (1), when both sides of the game calculate utility based on EUT, the formula is

$$U_{user}^{EUT}(SA_m, EU_n) = -U_{attac}^{EUT}(SA_m, EU_n) = G_{SA_m}^{EU_n} - \sum_{c=0}^{C} c P_c^{SA_m, EU_n} L_{SA_m}^{EU_n}/C. \tag{3}$$

When both game participants use PT to calculate utility, they make decisions based on probability, but not on objective average detection error rate. Therefore, according to (1) and (2), the utility of both sides are as follows.

$$U_{user}^{PT}(SA_m, EU_n) = G_{SA_m}^{EU_n} - \sum_{c=0}^{C} c W_{attac}(P_c^{SA_m, EU_n}) L_{SA_m}^{EU_n}/C, \tag{4}$$

$$U_{attac}^{PT}(SA_m, EU_n) = \sum_{c=0}^{C} c W_{user}(P_c^{SA_m, EU_n}) L_{SA_m}^{EU_n}/C - G_{SA_m}^{EU_n}. \tag{5}$$

In the process of the game, both sides of the game change their probability by adjusting the objective weight and pursue the maximum of their own utility to achieve NE. There are two kinds of NE, one is pure strategy NE, and the other is mixed strategy NE. The pure strategy NE is a definite NE point, and the mixed strategy NE makes the strategy of any participant the best strategy relative to the strategy of other participants because of the decision probability. When a smart attacker holds the view that the defender can detect attacks accurately, the smart attacker will choose to stop attacks. When a legitimate user subjectively holds that he can use the HLSM mechanism to achieve more utility, it will let $EU_n = 2$ [17]. This paper summarizes the conditions for the emergence of two NE strategies in the zero-sum game. The pure strategy NE combination is expressed in this paper as (SA_m^*, EU_n^*), which is the way to maximize the utility of players in the game and should satisfy the following conditions.

$$U_{user}^{PT}(SA_m^*, EU_n^*) \geq U_{user}^{PT}(SA_m^*, EU_n), EU_n = 1, 2, \tag{6}$$

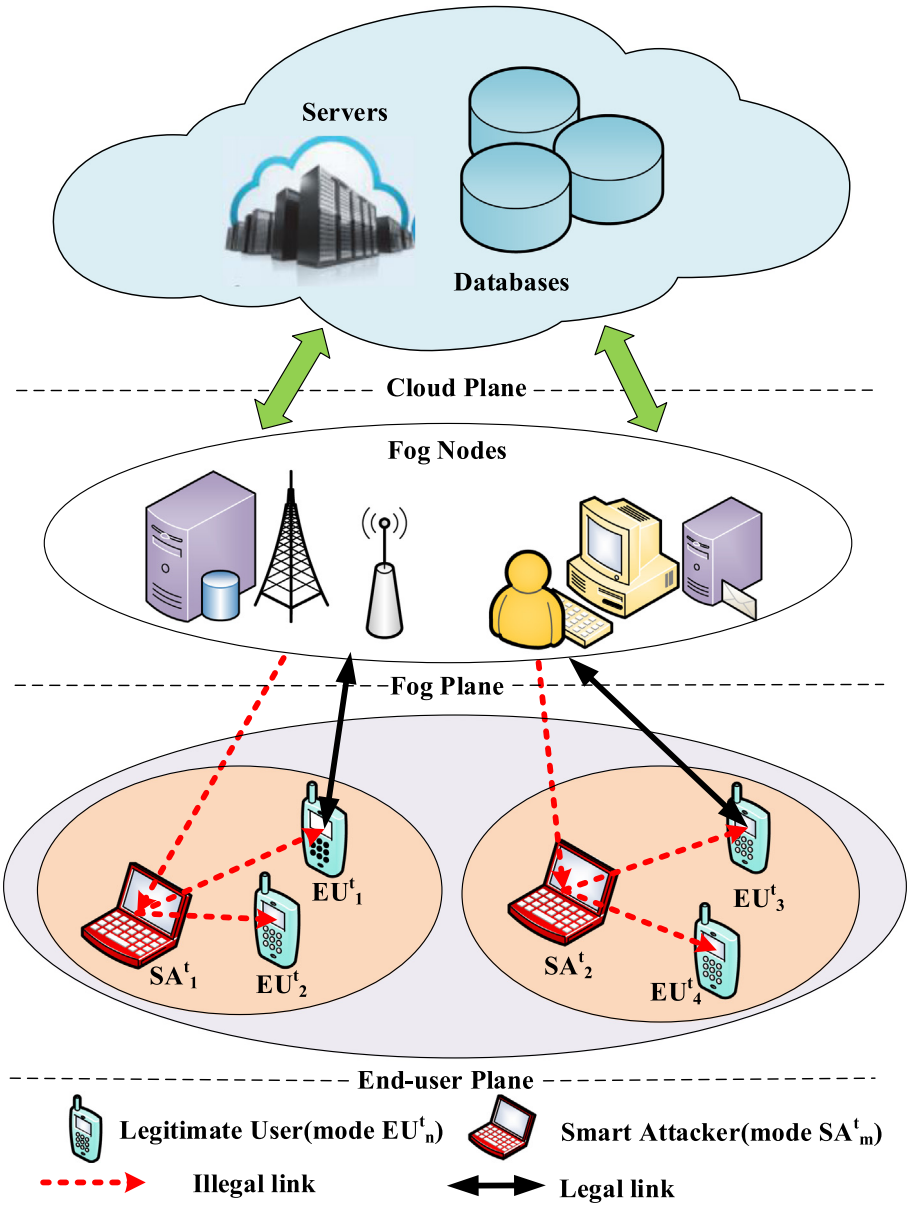

Fig. 1. Smart attack security model for fog computing.

$$U_{attac}^{PT}(SA_m^*, EU_n^*) \geq U_{attac}^{PT}(SA_m, EU_n^*), 0 \leq SA_m \leq 4. \quad (7)$$

Therefore, according to (4)–(7), for example, when $SA_m = 0, 1, 2, 3$ and smart attackers decide to adopt stop attacking mode or impersonation attack mode, the pure strategy NE conditions for impersonation attack are summarized.

$$U_{user}^{PT}(0,1) \geq U_{user}^{PT}(0,2), \quad (8)$$

$$U_{attac}^{PT}(0,1) \geq \max\{U_{attac}^{PT}(1,1), U_{attac}^{PT}(2,1), U_{attac}^{PT}(3,1)\}. \quad (9)$$

NE condition (1) - when (8) and (9) are satisfied, pure strategy NE is (0,1).

$$U_{user}^{PT}(0,2) \geq U_{user}^{PT}(0,1), \quad (10)$$

$$U_{attac}^{PT}(0,2) \geq \max\{U_{attac}^{PT}(1,2), U_{attac}^{PT}(2,2), U_{attac}^{PT}(3,2)\}. \quad (11)$$

NE condition (2) - when (10) and (11) are satisfied, pure strategy NE is (0,2).

$$U_{user}^{PT}(3,1) \geq U_{user}^{PT}(3,2), \quad (12)$$

$$U_{attac}^{PT}(3,1) \geq \max\{U_{attac}^{PT}(0,1), U_{attac}^{PT}(1,1), U_{attac}^{PT}(2,1)\}. \quad (13)$$

NE condition (3) - when (12) and (13) are satisfied, pure strategy NE is (3,1).

$$U_{user}^{PT}(3,2) \geq U_{user}^{PT}(3,1), \quad (14)$$

$$U_{attac}^{PT}(3,2) \geq \max\{U_{attac}^{PT}(0,2), U_{attac}^{PT}(1,2), U_{attac}^{PT}(2,2)\}. \quad (15)$$

NE condition (4) - when (14) and (15) are satisfied, pure strategy NE is (3,2).

In addition, we use a table to state the conditions for the establishment of mixed strategy NE when $SA_m = 0, 1, 2, 3$. In Table 2, *Pro* represents the probability that the user or attacker selects the corresponding defense mode or attack mode. a_1, a_2, etc. represent the utility of both sides. For game players, the mixed strategy NE needs to make it impossible for both sides to make a targeted decision to gain the upper hand. Therefore, for the user, when mixed strategy NE occurs, the probability of selecting defense mode 1 and mode 2 should equal

Table 2
Mixed strategy NE utility.

User	Attacker			
	$Pro_C(0)$	$Pro_D(1)$	$Pro_E(2)$	$Pro_F(3)$
$Pro_A(1)$	a_1, a_2	c_1, c_2	e_1, e_2	g_1, g_2
$Pro_B(2)$	b_1, b_2	d_1, d_2	f_1, f_2	h_1, h_2

the utility gained by the attacker choosing mode 0, 1, 2 and 3. For the attacker, it is similar. The mixed strategy NE conditions are as follows.

$$a_2 Pro_A + b_2 Pro_B = c_2 Pro_A + d_2 Pro_B = e_2 Pro_A + f_2 Pro_B = g_2 Pro_A + h_2 Pro_B \tag{16}$$

$$a_1 Pro_C + c_1 Pro_D + e_1 Pro_E + g_1 Pro_F = b_1 Pro_C + d_1 Pro_D + f_1 Pro_E + h_1 Pro_F \tag{17}$$

When considering the zero-sum game of multiple smart attackers and legitimate users, the principle of NE conditions is the same as above. The NE strategy must exist, and maybe one of the above two approaches. Next, taking a smart attacker and a legitimate user as an example, this paper describes the dynamic approach of resisting smart attacks in the zero-sum game.

3. Methodology of preventing smart attack in dynamic environment

In practical MFC networks, the game between smart attackers and legitimate users is dynamic. The players in the game do not understand the channel information, attack rate, and the overall network environment model. So they can continue to play games to prevent smart attacks, restrain the motive of smart attackers, and enhance the utility of legitimate users. The Q-learning algorithm is a kind of reinforcement learning method. It can acquire an optimal scheme in a dynamically varying environment with inadequate information [27]. Suppose that in the Q-learning algorithm, $Q(s,a)=0$, and at any action a', $Q(s',a')=0$, with estimated error satisfying $N(0,\sigma^2)$ distribution. Therefore, the value of $Q(s,a)$ is an immediate earning value r, and is updated according to $Q(s,a)=\mu(r+\delta\max_{a'}Q(s',a'))$. Since $Q(s,a)=E[r_{t+1}+\delta r_{t+2}+\delta^2 r_{t+3}+\cdots|A_t=a,S_t=s]$, A_t and S_t are the set of action value and state value respectively, $E[\max_{a'}Q(s',a')]\geq\max_{a'}E[Q(s',a')]$, then $Q(s,a)\geq r$, i.e., Q-learning algorithm has the problem of an overestimation of Q value. To overcome this problem, Hasselt proposed the DQL method [18]. In this Section, the DQL algorithm is applied to the game between smart attackers and legitimate users to acquire the optimum defense scheme for legitimate users, in which SA_i^t denotes attack mode, and EU_n^t denotes defense mode.

In DQL, there are two Q-value tables. They are used to evaluate the benefits of adopting an action in a certain state, corresponding to the game scenario proposed. State refers to the attack mode selected by the smart attacker in a period before a certain time, and an action refers to the defense pattern selected by the legitimate user at time t. The two Q-value tables can make up for each other's Q-value calculation errors and make the calculation results of Q-value updating more accurate. Besides, according to the methodology of a static zero-sum game, legitimate users calculate the utility value according to PT. We can use this value as an immediate earning value in the DQL algorithm. The calculation formulae of updating two Q-value tables are as follows.

$$Q_1(s^t,EU_n^t)\leftarrow(1-\mu)Q_1(s^t,EU_n^t)+\mu(U_{user}^{PT}(s^t,EU_n^t)+\delta Q_2(s^{t+1},\lambda_1^*)), \tag{18}$$

$$Q_2(s^t,EU_n^t)\leftarrow(1-\mu)Q_2(s^t,EU_n^t)+\mu(U_{user}^{PT}(s^t,EU_n^t)+\delta Q_1(s^{t+1},\lambda_2^*)), \tag{19}$$

where

- s^t is the system state at time t;
- μ is incentive decay coefficient, $\mu\in(0,1]$;
- δ is learning efficiency, $\delta\in[0,1]$;
- λ_1^* and λ_2^* are the defense modes resulting in maximum Q-value in tables Q_1 and Q_2 in state s^{t+1}, with calculation formulae as follows.

$$\lambda_1^*=\arg\max_{EU_n^{t+1}}Q_1(s^{t+1},EU_n^{t+1}), \tag{20}$$

$$\lambda_2^*=\arg\max_{EU_n^{t+1}}Q_2(s^{t+1},EU_n^{t+1}), \tag{21}$$

$$V(s^t)=\max[\frac{Q_1(s^t,EU_n^t)+Q_2(s^t,EU_n^t)}{2}],\ EU_n^t\in\{1,2\}. \tag{22}$$

$V(s^t)$ represents the maximum mean value of Q_1+Q_2 in current state corresponding to each defense mode. Therefore, the optimal defense mode λ^* is given by:

$$\lambda^*=\arg\max_{EU_n^t}[\frac{Q_1(s^t,EU_n^t)+Q_2(s^t,EU_n^t)}{2}],\ EU_n^t\in\{1,2\}. \tag{23}$$

Legitimate users apply $\varepsilon-greedy$ strategy to choose defense modes and update Q-value tables in each state. By using $\varepsilon-greedy$ strategy, we can select the sub-optimal defense mode with probability ε. And the defense mode satisfying $V(s^t)$ is selected with probability $1-\varepsilon$ with $\varepsilon\in(0,1)$. The game method based on DQL algorithm to prevent smart attack is summarized as 1.

Algorithm 1 Defense Strategy Based on DQL

Step ⋆1 Initial value:
μ, δ, SA_m^0, ε, $Q_1(s^t,EU_n^t)=Q_2(s^t,EU_n^t)=0$, $V(s^t)=0$;
Step ⋆2 Algorithm steps:
for $t=1,2,3,\ldots$ **do**
  $s^t=SA_m^{t-1}$;
  Use $\varepsilon-greedy$ strategy to select defense mode EU_n^t;
  Discover next state SA_m^t;
  Calculate and obtain $U_{user}^{PT}(s^t,EU_n^t)$;
  Update $Q_1(s^t,EU_n^t)$ by (18) and (20) with probability 0.5;
  Otherwise, update $Q_2(s^t,EU_n^t)$ by (19) and (21) ;
  Update $V(s^t)$ by (22);
end for

4. Performance evaluation

Based on PT, it is assumed that the zero-sum game between legitimate users and smart attackers is carried out by selecting defend modes and attack modes, respectively. Firstly, we construct four attack modes in a static zero-sum game: silence, jamming attack, eavesdropping attack, and impersonation attack, and analyze the influence of objective probability weight on NE. Secondly, in the dynamic game, five attack modes are constructed: silence, jamming attack, eavesdropping attack, impersonation attack, and replay attack. To acquire the optimum defense strategy and increase the utility of legitimate users, a DQL game strategy to prevent smart attacks is proposed. And four different methods are compared in 300 slots [17] under four evaluation indicators.

4.1. NE proof

Taking four attack modes as examples, in Section 2, we evaluate the formation conditions of pure strategy NE, which is proved as follows:

NE condition (1) - If (8) and (9) are satisfied, the pure strategy NE combination is (0,1).

Proof. When (8) is satisfied, according to (4), there exists

$$U_{user}^{PT}(0,1) = G_0^1 - \sum_{c=0}^{C} cW_{user}(P_c^{0,1})L_0^1/C \geq G_0^2 - \sum_{c=0}^{C} cW_{user}(P_c^{0,2})L_0^2/C = U_{user}^{PT}(0,2), \tag{24}$$

which indicates that for legitimate users, the utility value of defense mode 2 is lower than that of defense mode 1. Therefore, legitimate users will choose the defense mode that maximizes their utility, which satisfies (6). When (9) is satisfied, according to (5), there exists

$$U_{attac}^{PT}(0,1) = \sum_{c=0}^{C} cW_{attac}(P_c^{0,1})L_0^1/C - G_0^1 \geq \sum_{c=0}^{C} cW_{attac}(P_c^{1,1})L_1^1/C - G_1^1 = U_{attac}^{PT}(1,1), \tag{25}$$

$$U_{attac}^{PT}(0,1) = \sum_{c=0}^{C} cW_{attac}(P_c^{0,1})L_0^1/C - G_0^1 \geq \sum_{c=0}^{C} cW_{attac}(P_c^{2,1})L_2^1/C - G_2^1 = U_{attac}^{PT}(2,1), \tag{26}$$

$$U_{attac}^{PT}(0,1) = \sum_{c=0}^{C} cW_{attac}(P_c^{0,1})L_0^1/C - G_0^1 \geq \sum_{c=0}^{C} cW_{attac}(P_c^{3,1})L_3^1/C - G_3^1 = U_{attac}^{PT}(3,1). \tag{27}$$

The above equations indicate that smart attackers think that legitimate users will detect any attack. The utility value of attack mode 0 is higher than that of launching a jamming attack, eavesdropping attack, or impersonation attack. Therefore, smart attackers will choose the attack mode that maximizes their utility, which satisfies (7). Since (6) and (7) are met, for game participants, to maximize their utilities, they will not change their decision-making if their counterpart does not change their decision, forming a pure strategy NE combination (0,1).

The proof of the pure strategy NE conditions labeled as NE condition (2), NE condition (3), and NE condition (4) in Section 2 are similar as that of NE condition (1). The method in Section 3 is based on Section 2. And the formula for calculating the legitimate user utility and the way to obtain NE condition remains unchanged. Therefore, the proof process of the method in Section 3 is the same as that of NE condition (1). The mixed strategy NE proof is more complicated. In this paper, the mathematical proof is not described too much.

4.2. Parameter setup and evaluation index

In this study, based on pulse code modulation (PCM), 300-time slots are set up in the simulation experiment, with each time slot representing 12500/32 μs. for the convenience of calculation. The next time intervals are expressed in microseconds. We use a computer with an Intel i5 processor and MatLab software for simulation under the Windows operating system. And we consider the corresponding fog nodes, smart attackers, and legitimate users in the system model. For a smart attacker and a legitimate user, from the perspective of the legitimate user, we generate the objective probability matrix that follows $[P_c^{SA_m,EU_n}]_{0\leq c\leq C}$ distribution about the attack detection error rate randomly. Table 3 shows the symbols and initial values used in the experiment. The initial values and the matrix are used as input data for calculating $U_{user}^{PT}(s^t, EU_n^t)$

In the dynamic game of withstanding smart attack, the indices of evaluating the four methods are as follows.

Indicator (1) - The utility of legitimate users: the average PT-based utility of authorized users in each time slot.

Indicator (2) - Attack rate: the rate of the number of attack modes selected by smart attackers in each time slot to all modes.

Table 3
List of symbols and parameters.

Parameter	Meaning	Value
Num	Number of attack modes	4 in Figs. 2 and 5 in Figs. 3–6.
SA_m or SA_m^t	Attack mode	0,1, …, $Num-1$
EU_n or EU_n^t	Defense mode	1,2
σ_{object}	Objective probability weight	[0,1]
ε	Strategy selection rate	0.9
μ	The learning efficiency	0.9
δ	Incentive decay coefficient	0.6
$G_{SA_m}^{EU_n}$	Benefits of the legitimate user in attack mode SA_m	[0.99, 0.81; 1.2, 1.4; 1.5, 1.6; 1.71, 2.19] in Fig. 2. [3.6, 3.1; 1.6, 5.2; 1.5, 5.3; 1.4, 5.5; 1.3, 5.7] in Figs. 3–6.
$L_{SA_m}^{EU_n}$	Security loss of the legitimate user in attack mode SA_m	[0.5, 0.8; 0.7, 0.6; 1.1, 0.5; 1.3, 0.3] in Fig. 2. [0.2, 0.1; 0.6, 0.3; 0.7, 0.4; 0.8, 0.5; 0.9, 0.6] in Figs. 3–6.
$R_{SA_m}^{EU_n}$	Rate of detection error	–
U_{user}^{EUT}	Expected utility of the legitimate user	–
U_{user}^{PT}	PT-based utility of the legitimate user	–
$C, 0 \leq c \leq C$	Probability quantization level	10
$P_c^{SA_m,EU_n}$	Detection error rate distribution	–

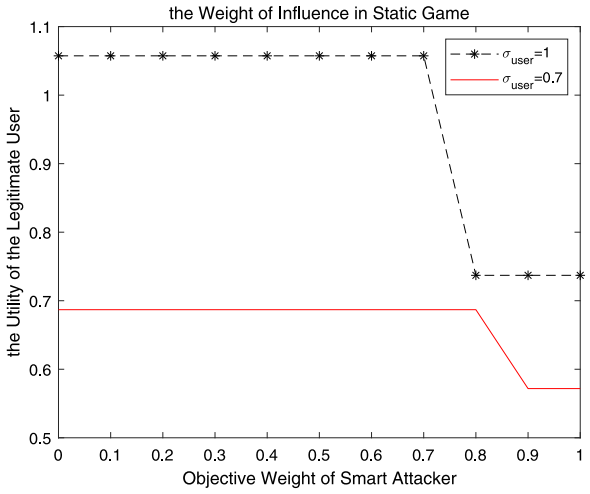

Fig. 2. Weight of Influence in the static zero-sum game under uncertain detection error rate at the NEs, with $Num = 4$, $G_{SA_m}^{EU_n} = [0.99, 0.81; 1.2, 1.4; 1.5, 1.6; 1.71, 2.19]$, $L_{SA_m}^{EU_n} = [0.5, 0.8; 0.7, 0.6; 1.1, 0.5; 1.3, 0.3]$, $C = 10$, 10 objective weights and the attacker launches jamming attacks if $SA_m^t = 1$, eavesdropping attacks if $SA_m^t = 2$, impersonating attacks if $SA_m^t = 3$, and does not attack if $SA_m^t = 0$.

Indicator (3) - Max Q-value: the max Q-value updated in each time slot during the update process of the Q table.

Indicator (4) - Average action value: average defense mode value adopted by legitimate users in each time slot during Q table updating [28].

4.3. Simulation results

Based on the NE conditions of the static zero-sum game described in Section II, this Section describes the effect of objective probability weight in (2) of this scheme on the utility of legitimate users and NE condition. In Fig. 2, the utility of legitimate users remains unchanged as the objective probability weight of smart attackers increases. This is because the objective probability weight is low. As long as the attacker holds the view that the legitimate user will detect the attack, even if the probability is small, the smart attacker will not launch the

Table 4
Results and comparison of four methods in terms of four indicators.

Methods	Indicators			
	The utility of the legitimate users	Attack rate	Max Q-value	Average action value
Proposed method	3.7352	0.4029	11.4984	1.5971
Q-learning [21]	3.7181	0.3575	11.9810	1.6382
Greedy [24]	3.2239	0.5195	–	–
Sarsa [25]	3.6422	0.3564	11.8589	1.6436

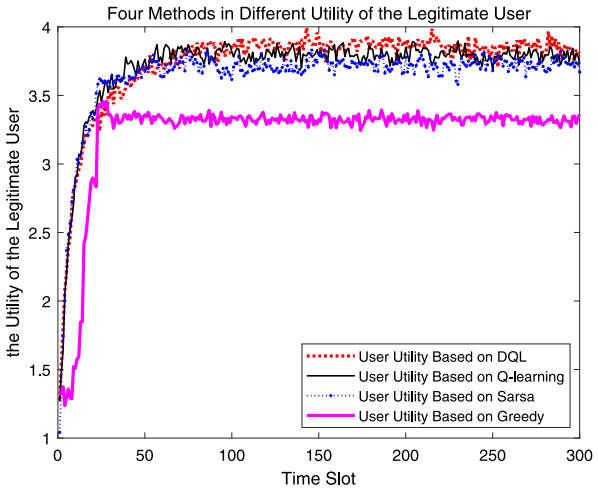

Fig. 3. Four methods (DQL, Q-learning, Sarsa, Greedy) in different utility of the legitimate user about dynamic game, with $Num = 5$, $G_{SA_m}^{EU_n} = [3.6, 3.1; 1.6, 5.2; 1.5, 5.3; 1.4, 5.5; 1.3, 5.7]$, $L_{SA_m}^{EU_n} = [0.2, 0.1; 0.6, 0.3; 0.7, 0.4; 0.8, 0.5; 0.9, 0.6]$, $C = 10$, $\sigma_{user} = 1$, $\sigma_{attac} = 0.7$ and 300 time slots.

attack, and the attack motivation is suppressed. When $\sigma_{attac} = 0.7$ and $\sigma_{user} = 1$, the utility of legitimate users begins to decrease, attackers launch impersonation attack, attack mode changes from 0 to 3, and there is no pure strategy NE. Likewise, if $\sigma_{user} = 0.7$, when $\sigma_{attac} = 0.8$, attackers begin to attack, reducing the utility of legitimate users from 0.6869 to 0.5718. The mixed strategy NE is changed. Therefore, the probability of attack can be reduced by reducing the weight to restrain the attack motive. Meanwhile, if legitimate users believe that using HLSM mode can increase utility value, then this large system overhead defense mode will be adopted.

To better illustrate the advantages of the DQL-based smart attack defense strategy, we apply three representative algorithms to the selection of defense strategy for the smart attack and analyze them in the four evaluation aspects as above. The three algorithms are the Q-learning algorithm [21], Greedy algorithm [24], Sarsa algorithm [25], respectively. Among which, Q-learning algorithms, the Sarsa algorithm, and the DQL algorithm all belong to the reinforcement learning algorithm. Since the Greedy algorithm does not involve the Q table, so only indices (1) and (2) are considered. The results of four methods based on four indicators are shown in Table 4, which represents the average of 300 time-slots.

As is shown in Fig. 3, it depicts the legitimate user utility results of four algorithms in 300-time slots. In terms of the utility of authorized users, the four algorithms have the same trend. The proposed method converges after the 70th time slot. The utility of legitimate users in our approach increases by 0.46% on an average compared with the process based on Q-learning, and by 2.55% on an average compared with the method based on Sarsa. This indicates that our approach optimizes the estimation of Q value and enhances the correctness of decision-making. The three methods based on the reinforcement learning algorithm using

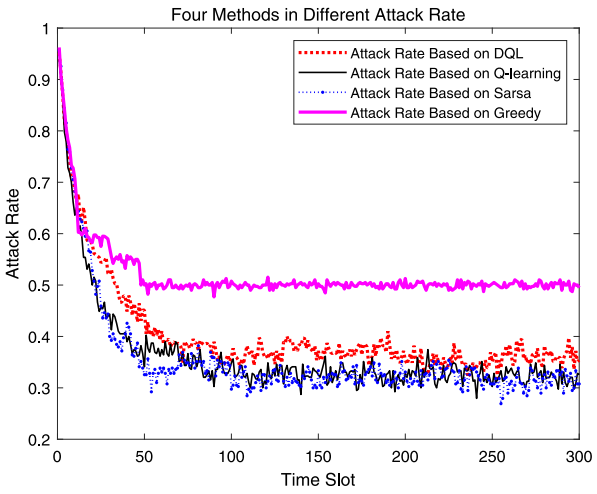

Fig. 4. Four methods (DQL, Q-learning, Sarsa, Greedy) in different attack rate (four kinds of attack) about dynamic game, with $Num = 5$, $G_{SA_m}^{EU_n} = [3.6, 3.1; 1.6, 5.2; 1.5, 5.3; 1.4, 5.5; 1.3, 5.7]$, $L_{SA_m}^{EU_n} = [0.2, 0.1; 0.6, 0.3; 0.7, 0.4; 0.8, 0.5; 0.9, 0.6]$, $C = 10$, $\sigma_{user} = 1$, $\sigma_{attac} = 0.7$ and 300 time slots.

Q-table have higher legitimate user utility. This is because they continuously learn the experience of decision-making before choosing the optimum defense scheme. Each decision determines a more appropriate defense strategy and realizes the whole process through the Q table. This makes a trade-off between immediate utility and future utility. Conversely, the method based Greedy maximizes the quick utility and quickly leads to locally optimal solutions for legitimate users.

In terms of convergence speed of the proposed method, about the proposed method, the convergence rate is the slowest because of the other renewal of two Q tables. However, considering the overall results, the proposed method has the best performance. It can help legitimate users to increase the utility of detection attacks, reduce the detection error rate, and improve the security between the fog layer and user layer in the MFC network.

In Fig. 4, the total attack rates of four different methods are compared. With the time, the total attack rate decreases and then fluctuates around a certain value due to the convergence. The total attack rate of the proposed method is reduced from 0.95 to 0.35 in the 100th time slot. Our method increases the total attack rate by an average of 12.7% compared with the method based on Q-learning. This is because the DQL algorithm contributes to the solution of an overestimation of Q value. Firstly, when the Q table is updated, the Q-value is smaller than the method based on Q-learning, thus correcting the legitimate user's choice of defense mode at each moment, increasing the immediate utility of the legitimate users. Secondly, compared with the method based on Q-learning, the smart attacker in the proposed method will pursue risk-benefit and will consider the mode of increasing its utility when choosing to attack. Therefore, the total attack rate of the proposed method is higher compared with the method based on Q-learning. The three methods based on the reinforcement learning algorithm have lower the total attack rate than the Greedy method. Therefore, the three methods based on the reinforcement learning algorithm can better restrain the aggressive motive of attackers. Still, the total attack rate of the proposed method is higher than that of the other two methods.

As is shown in Fig. 5, it depicts the change of the maximum Q in three defense strategies based on the reinforcement learning algorithm in 1–300 time slots. Max Q-value increases with time. After convergence, the max Q-value of the proposed method is the smallest. The max Q-value decreases by 4.2% compared with the method based on Q-learning. The reason is that the DQL compensates for the estimation error of Q-value with two Q tables, which makes the estimated gain

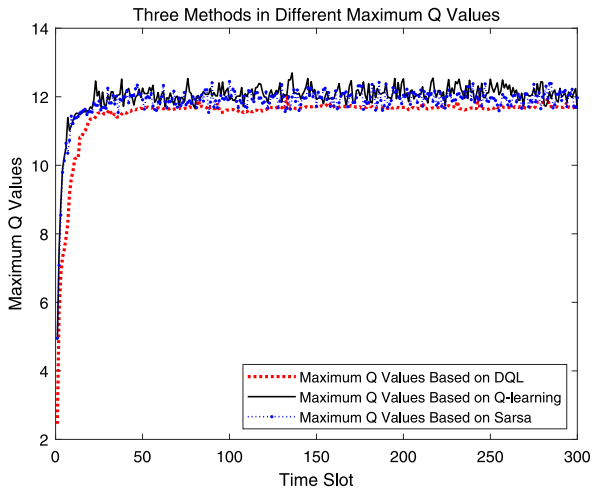

Fig. 5. Three methods (DQL, Q-learning, Sarsa) in different maximum Q values about dynamic game, with $Num = 5$, $G_{SA_m}^{EU_n} = [3.6, 3.1; 1.6, 5.2; 1.5, 5.3; 1.4, 5.5; 1.3, 5.7]$, $L_{SA_m}^{EU_n} = [0.2, 0.1; 0.6, 0.3; 0.7, 0.4; 0.8, 0.5; 0.9, 0.6]$, $C = 10$, $\sigma_{user} = 1$, $\sigma_{attac} = 0.7$ and 300 time slots.

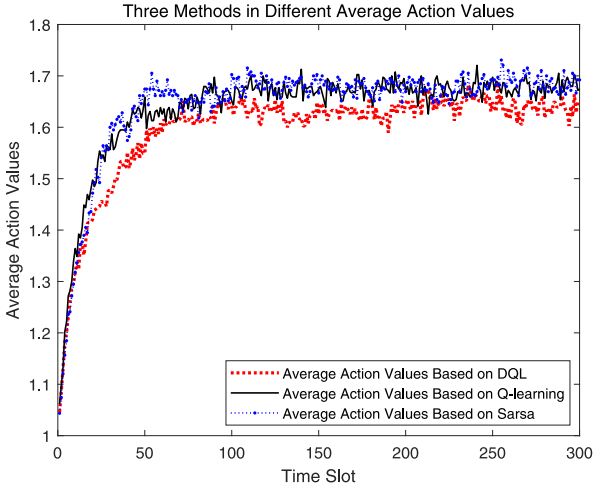

Fig. 6. Three methods (DQL, Q-learning, Sarsa) in different average action values about dynamic subjective game, with $Num = 5$, $G_{SA_m}^{EU_n} = [3.6, 3.1; 1.6, 5.2; 1.5, 5.3; 1.4, 5.5; 1.3, 5.7]$, $L_{SA_m}^{EU_n} = [0.2, 0.1; 0.6, 0.3; 0.7, 0.4; 0.8, 0.5; 0.9, 0.6]$, $C = 10$, $\sigma_{user} = 1$, $\sigma_{attac} = 0.7$ and 300 time slots.

value in each state more accurate. Also, since the actual value of the current state and the next state are taken into account in the calculation of Q-value by Sarsa algorithm, the average max Q-value in 300-time slots is reduced by about 1.19%.

Fig. 6 depicts average action values in three defense strategies based on the reinforcement learning algorithm in 1–300 time slots. Legitimate users can use defense mode value as an action value. The action value of the three methods has the same trend, and the convergence speed is similar. It converges after the 80th slot. After convergence, the average action value of the proposed method is about 1.60. It is reduced by about 2.51% compared with the process based on Q-learning. Since the action value can only be 1 or 2, the smaller the action value, the more likely the legitimate user is to use the PLS mechanism. Likewise, compared with the Sarsa-based method, the DQL-based process reduces the action value selected in each time slot by an average of 2.83%. After correcting the estimation error of Q-value and enhancing the correctness of decision-making, our proposed method is more inclined to choose the defense mode using the PLS mechanism compared with the other two techniques. The defense mode with action value-1 can cost less system overhead. Therefore, the proposed method for preventing smart attacks can optimize the decision-making of defense mode. It has excellent performance and enhances the security of the MFC network.

4.4. Discussions

In summary, in the DQL method that prevents smart attacks, the key is to (1) restrain the motivation of smart attackers, (2) reduce the attack rate, (3) optimize the decision-making process of the legitimate user to obtain the optimal defense strategy, (4) improve the utility of the authorized user. The reinforcement learning algorithm can explore the environment and benefit through executing actions until reaching goals. The simulation results show that the method based on reinforcement learning has better performance than the greedy strategy in increasing the utility of legitimate users and reducing the attack rate. Meanwhile, the proposed method makes authorized users have an excellent utility compared with the method based on Q-learning and the technique based on Sarsa. Moreover, as can be seen from Fig. 5, and Fig. 6, the proposed method can solve three problems (1) overcomes the problem of over-estimation of Q value in the Q-learning-based method, (2) optimizes the decision-making process of obtaining the optimal defense strategy, (3) reduces the detection error rate of legitimate users. Therefore, in MFC, the proposed method is more appropriate for the dynamic environment of the game between smart attackers and authorized users.

5. Conclusions

This paper presented a model on the smart attacks for end-users in MFC. It proposed a DQL defense scheme that can restrain the attack motive of smart attackers in the dynamic environment and generate the optimum defense choice of legitimate users against smart attacks. The scheme first addressed a static zero-sum game between the smart attacker and the authorized user. The utility of players in the game and NE conditions was calculated. Then, the DQL algorithm was used to defend against smart attacks, and the optimal defense strategy based on a zero-sum game in a dynamic environment was achieved. The experiment results indicate that our method can effectively enhance the utility of legitimate users and reduce the attack rate. The proposed method improves the utility of the authorized user by an average of 6.29% and decreases the average action value about reinforcement learning methods by an average of 2.69% compared with the methods based on Q-learning algorithm, Sarsa algorithm and greedy strategy to prevent smart attacks. It restrains the total attack rate by 22.44% compared with the method based on the greedy approach. Therefore, the proposed method has better security protection capability. In future work, with the help of other reinforcement learning algorithms, the specific techniques to detect attacks about various security threats will be continuously investigated in a fog computing environment.

CRediT authorship contribution statement

Shanshan Tu: Data curation, Validation, Writing - review & editing, Funding acquisition. **Muhammad Waqas:** Writing - review & editing, Supervision. **Yuan Meng:** Conceptualization, Methodology, Software, Data curation, Validation, Visualization, Writing - original draft. **Sadaqat Ur Rehman:** Writing - review & editing. **Iftekhar Ahmad:** Writing - review & editing. **Anis Koubaa:** Writing - review & editing. **Zahid Halim:** Writing - review & editing. **Muhammad Hanif:** Writing - review & editing. **Chin-Chen Chang:** Writing - review & editing. **Chengjie Shi:** Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the editor-in-chief, the associate editor, and the anonymous reviewers for their valuable comments and suggestions. This work was partially supported by The China National Key R&D Program (No. 2018YFB0803600), National Natural Science Foundation of China (No. 61801008), Beijing Municipal Natural Science Foundation, China (No. L172049), Scientific Research Common Program of Beijing Municipal Education Commission, China (No. KM201910005025) and Defense Industrial Technology Development Program, China (No. JCKY2016204A102) sponsored this research in parts.

References

[1] M. Mukherjee, S. Lei, W. Di, Survey of fog computing: Fundamental, network applications, and research challenges, IEEE Commun. Surv. Tutor. 20 (3) (2018) 1826–1857, http://dx.doi.org/10.1109/COMST.2018.2814571.

[2] L.F. Bittencourt, J.J. Diaz-Montes, R. Buyya, O.F. Rana, M. Parashar, Mobility-aware application scheduling in fog computing, IEEE Cloud Comput. 4 (2) (2017) 26–35, http://dx.doi.org/10.1109/MCC.2017.27.

[3] M. Mukherjee, R. Matam, L. Shu, L. Maglaras, V. Kumar, Security and privacy in fog computing: Challenges, IEEE Access 5 (2017) 19293–19304, http://dx.doi.org/10.1109/ACCESS.2017.2749422.

[4] M. Huss, M. Waqas, A.Y. Ding, Y. Li, J. Ott, Security and privacy in device-to-device (d2d) communication: A review, IEEE Commun. Surv. Tutor. 19 (2) (2017) 1054–1079, http://dx.doi.org/10.1109/COMST.2017.2649687.

[5] M. Waqas, M. ahmed, L. Yong, D. Jin, S. chen, Social-aware secret key generation for secure device-to-device communication via trusted and non-trusted relays, IEEE Trans. Wireless Commun. 17 (6) (2018) 3918–3930, http://dx.doi.org/10.1109/TWC.2018.2817607.

[6] J. Fu, Y. Liu, H. Chao, B.K. Bhargava, Z. Zhang, Secure data storage and searching for industrial iot by integrating fog computing and cloud computing, IEEE Trans. Ind. Inf. (ISSN: 1551-3203) 14 (10) (2018) 4519–4528, http://dx.doi.org/10.1109/TII.2018.2793350.

[7] A. Alrawais, A. Alhothaily, C. Hu, X. Xing, X. Cheng, An attribute-based encryption scheme to secure fog communications, IEEE Access 5 (2017) 9131–9138, http://dx.doi.org/10.1109/ACCESS.2017.2705076.

[8] A.M. Rahmani, T.N. Gia, B. Negash, A. Anzanpour, I. Azimi, M. Jiang, P. Liljeberg, Exploiting smart e-health gateways at the edge of healthcare internet-of-things: A fog computing approach, Future Gener. Comput. Syst. 78 (2017) 641–658, http://dx.doi.org/10.1016/j.future.2017.02.014.

[9] L. Xiao, G. Han, D. Jiang, H. Zhu, Y. Zhang, H.V. Poor, Two-dimensional anti-jamming mobile communication based on reinforcement learning, IEEE Trans. Veh. Technol. 67 (2018) 9499–9512, http://dx.doi.org/10.1109/TVT.2018.2856854.

[10] L. Xiao, Y. Li, G. Han, H. Dai, H.V. Poor, A secure mobile crowdsensing game with deep reinforcement learning, IEEE Trans. Inf. Forensics Secur. 13 (2017) 35–47, http://dx.doi.org/10.1109/TIFS.2017.2737968.

[11] L. Xiao, J. Liu, Q. Li, N.B. Mandayam, H.V. Poor, User-centric view of jamming games in cognitive radio networks, IEEE Trans. Inf. Forensics Secur. 10 (12) (2015) 2578–2590, http://dx.doi.org/10.1109/TIFS.2015.2467593.

[12] L. Xiao, T. Chen, J. Liu, H. Dai, Anti-jamming transmission stackelberg game with observation errors, IEEE Commun. Lett. 19 (6) (2015) 949–952, http://dx.doi.org/10.1109/LCOMM.2015.2418776.

[13] C.L. Min, J. Park, There is no perfect evaluator: An investigation based on prospect theory, Hum. Factors Ergon. Manuf. Serv. Ind. 28 (2) (2018) 383–392, http://dx.doi.org/10.1002/hfm.20748.

[14] W.F. Dai, Q.Y. Zhong, C.Z. Qi, Multi-stage multi-attribute decision-making method based on the prospect theory and triangular fuzzy multimoora, Soft Comput. (2) (2018) 1–12, http://dx.doi.org/10.1007/s00500-018-3017-0.

[15] A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty, J. Risk Uncertain. 5 (4) (1992) 297–323, http://dx.doi.org/10.2307/41755005.

[16] D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk, Econometrica 47 (2) (1979) 263–291, http://dx.doi.org/10.2307/1914185.

[17] C. Xie, L. Xiao, User-centric view of smart attacks in wireless networks, in: IEEE International Conference on Ubiquitous Wireless Broadband, 2016, http://dx.doi.org/10.1109/ICUWB.2016.7790439.

[18] H.V. Hasselt, Double q-learning, in: NIPS, 2010.

[19] P. Hu, H. Ning, T. Qiu, H. Song, Y. Wang, X. Yao, Security and privacy preservation scheme of face identification and resolution framework using fog computing in internet of things, IEEE Internet Things J. 4 (5) (2017) 1143–1155, http://dx.doi.org/10.1109/JIOT.2017.2659783.

[20] R. Chaudhary, N. Kumar, S. Zeadally, Network service chaining in fog and cloud computing for the 5g environment: Data management and security challenges, IEEE Commun. Mag. 55 (11) (2017) 114–122, http://dx.doi.org/10.1109/MCOM.2017.1700102.

[21] S. Tu, M. Waqas, S.U. Rehman, M. Aamir, C.C. Chang, Security in fog computing: A novel technique to tackle an impersonation attack, IEEE Access 6 (2018) 74993–75001, http://dx.doi.org/10.1109/ACCESS.2018.2884672.

[22] L. Xiao, C. Xie, T. Chen, H. Dai, H.V. Poor, Mobile offloading game against smart attacks, in: Computer Communications Workshops, 2016, http://dx.doi.org/10.1109/INFCOMW.2016.7562110.

[23] Y. Yang, L.T. Park, N.B. Mandayam, I. Seskar, A. Glass, N. Sinha, Prospect pricing in cognitive radio networks, IEEE Trans. Cogn. Commun. Netw. 1 (1) (2015) 56–70, http://dx.doi.org/10.1109/TCCN.2015.2488636.

[24] C. Zhou, Z. Peng, W. Zang, G. Li, On the upper bounds of spread for greedy algorithms in social network influence maximization, IEEE Trans. Knowl. Data Eng. 27 (10) (2015) 2770–2783, http://dx.doi.org/10.1109/TKDE.2015.2419659.

[25] C. Yan, S. Mabu, K. Hirasawa, J. Hu, Genetic network programming with sarsa learning and its application to creating stock trading rules, in: IEEE Congress on Evolutionary Computation, 2008, http://dx.doi.org/10.1109/CEC.2007.4424475.

[26] D. Prelec, The probability weighting function, Econometrica 66 (3) (1998) 497–527, http://dx.doi.org/10.2307/2998573.

[27] C.J.C.H. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (3–4) (1992) 279–292, http://dx.doi.org/10.1007/BF00992698.

[28] Y. Meng, S. Tu, J. Yu, F. Huang, Intelligent attack defense scheme based on dql algorithm in mobile fog computing, J. Vis. Commun. Image Represent. 65 (2019) 1–7, http://dx.doi.org/10.1016/j.jvcir.2019.102656.