Game Theory and Reinforcement Learning Based Secure Edge Caching in Mobile Social Networks

Qichao Xu, Zhou Su, and Rongxing Lu

***Abstract*— Edge caching has become one of promising technologies in mobile social networks (MSNs) to proximally provide popular contents for mobile users. However, since caching contents inevitably consume resources (e.g., power, bandwidth, storage, etc.), edge caching devices maybe selfish to cheat the content provider for earning service fees. In addition, due to the open access of edge caching devices, the edge caching service is vulnerable to various attacks, such as man-in-the-middle attack and content tamper attack, etc., resulting in the degradation of content delivery performance. To efficiently tackle the above problems, in this paper, we propose a secure edge caching scheme for the content provider and mobile users in MSNs. Specifically, we first develop a secure edge caching framework consisting of the content provider, multiple edge caching devices, and some mobile users. To motivate the participation of edge caching devices, Stackelberg game is exploited to model the interactions between the content provider and edge caching devices. The content provider serves as the game-leader to determine the payment strategy of secure caching service and each edge caching device is the game-follower to make the strategy on the quality of secure caching service. Especially, the zero payment mechanism is adopted to suppress the selfish behaviors of edge caching devices. Apart from this, for lack of the knowledge on interactions between the content provider and edge caching devices in dynamic network scenarios, we also employ the Q-leaning to derive the optimal payment strategy of the content provider and the security strategy of edge caching device. Extensive simulations are conducted, and results demonstrate that the proposed scheme can efficiently motivate edge caching devices to provide the content provider and mobile users with high-quality secure caching services.**

***Index Terms*— Mobile social networks (MSNs), secure edge caching service, Stackelberg game, reinforcement learning, zero payment punishment.**

I. Introduction

With the rapid advances of wireless communication technologies and smart devices (e.g., iPad, iPhone, etc.), mobile social networks (MSNs) have been pushed forward to provide ubiquitous convenient services including content sharing, delivery and exchange among mobile users [1]–[3]. Therewith, due to the proliferation of mobile phones and the dramatic growth of mobile users, the wireless data traffic in the network is expected to exponentially increase in the next few years. Related report [4] shows that the amount of global mobile data traffic in 2021 will reach to be sevenfold over that in 2015. Especially, there is an evidence that mobile multimedia contents will account for the major part of whole data traffic over cellular networks. Besides, there are numerous repeated requests of popular contents from mobile users, resulting in the redundant content transmission over backhaul networks with consuming a large amount of power and bandwidth. Meanwhile, as the content provider is located at remote area, it brings a large delivery latency in transmitting a desired content over backbone network for mobile users. Obviously, if the contents can be obtained locally from local places, mobile users' quality of experience (QoE) could be significantly improved [5], [6].

Edge caching, as a promising trend, can efficiently provide proximal caching services for mobile users, as a large number of caching-enabled intelligent devices are deployed at the edge of network and are close to mobile users [7]. The advantages brought by the edge caching devices are threefold [8], [9]. Firstly, as popular contents are stored close to mobile users when cached on edge caching devices, the transmission latency is extremely reduced. Secondly, the repeated requests of mobile users can be met by nearby edge caching devices, so that redundant content transmissions over cellular cells' backhaul links are significantly mitigated. Thirdly, the majority of data traffic is offloaded from overload cells to edge caching devices for realizing the high data rate. Overall, as the edge caching device is capable of content caching at the edge of network, all mobile users, content providers, and network operators are beneficiaries from the applications of edge caching [10].

However, the full deployment of edge caching devices in MSNs faces many challenges ahead. Specifically, the selfishness and open-access features of edge caching devices have severe effects on the content transmission for mobile users [11]–[13]. Indeed, edge caching devices are owned and deployed by rational third parties, who might be selfish to cheat content provider and provide mobile users with fake or forged contents for maximizing their individual profits. In addition, since edge caching devices are openly accessed, arbitrary mobile user could connect to edge caching devices

Manuscript received August 3, 2019; revised January 27, 2020 and February 9, 2020; accepted February 29, 2020. Date of publication March 16, 2020; date of current version May 22, 2020. This work was supported in part by NSFC under Grant U1808207 and Grant 91746114, in part by the 111 Project, and in part by the Project of Shanghai Municipal Science and Technology Commission under Grant 18510761000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Issa Traore. *(Corresponding author: Zhou Su.)*

Qichao Xu is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China (e-mail: xqc690926910@shu.edu.cn).

Zhou Su is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China, and also with the School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zhousu@ieee.org).

Rongxing Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B5A3, Canada (e-mail: rlu1@unb.ca).

Digital Object Identifier 10.1109/TIFS.2020.2980823

freely. As such, the applications with edge caching are vulnerable to the various attacks [14]–[17] (e.g., man-in-middle attack [15], DDoS attack [18], etc.) conducted by malicious entities in the networks, where the network performance is seriously degraded. Accordingly, edge caching devices may be compromised to tamper, remove, and replace cached contents, resulting in that mobile users cannot obtain requested contents and even receive malicious data, such as virus, malware, etc. Therefore, it is pressing to motivate edge caching devices to provide high-quality secure caching services for the content provider and mobile users.

Especially, with the ever-increasing popularity of smartphones and wireless applications, the number of mobile users in a community becomes larger and the required contents are more various than before. Due to the limitation of resources (e.g., bandwidth, energy, etc.), not all mobile users in the community can obtain the satisfied QoE on content delivery, i.e., they have to compete for resources and may not acquire the required contents in time [19]. As such, the content delivery performance urgently needs to be improved for the community [20]. Edge caching as a promising technique can efficiently improve the performance of content delivery in the community [21]. A large number of mobile users in the community have similar content interests and usually require the same content. If the content is cached on the edge caching devices, these mobile users can obtain the content from edge caching devices directly. However, due to the rationality of edge caching devices, they may be selfish to cheat the content providers for caching fees. In addition, by considering the diversity of mobile users, malicious mobile users may exist to attack the edge caching devices [18], [22]. Therefore, the secure caching scheme should be devised to tackle these security issues to further improve the content delivery performance for the community.

Game model is an efficient tool to formulate the content caching and provide approaches such as auction, pricing, and reputation mechanism as regard of stimulating edge caching device to contribute content caching. For example, in [23], auction based content caching is used to allocate edge caching resource to mobile users for improving the mobile users' QoE by paying the edge caching device who bids the lowest price. Besides, a commercialized edge caching scheme is proposed in [24], where the interactions between the content provider and edge caching devices are formulated as a Stackelberg game to jointly maximize the average profit of both the content provider and edge caching devices. Meanwhile, in [25], the reputation mechanism is used to evaluate the reliability of edge caching devices on providing caching service to mobile users. However, most of the current works with game theory do not have enough consideration that the selfish edge caching devices may cheat content provider and mobile users to gain profits. In addition, different from the assumption in the existing works, the parameters of network model and caching model for caching services are not readily available by content provider and edge caching devices. A learning based scheme without relying on the knowledge of network parameters is needed. Therefore, it is still an open and vital issue to provide mobile users with secure content caching services.

In this paper, we propose a novel secure edge caching scheme to provide high-quality caching services for the content provider and mobile users. Specifically, to motivate the participation of edge caching devices, the interactions between the content provider and multiple edge caching devices are formulated based on the one-leader and multiple-followers Stackelberg game. The leader of the game is the content provider to cache contents on edge caching devices for securely delivering contents to mobile users and determine the optimal payment strategy to maximize its utility. After the content provider broadcasts the payment strategy, each edge caching device makes the strategy on the optimal quality of secure caching service to provide caching service for the content provider, and then obtains profit based on the payment strategy and the provided quality of secure caching service afterwards. Especially, the selfish edge caching devices that deliver fake contents are punished with zero payment mechanism to suppress their self-interest consciousness. Besides, for lack of the knowledge on interactions between the content provider and edge caching devices in dynamic network scenarios, the interaction process is formulated as a finite Markov decision process (MDP). The content provider can employ the Q-learning to decide the optimal payment strategy based on the observations of the qualities of secure caching services provided by edge caching devices with different payment strategies. Meanwhile, each edge caching device applies the Q-learning to achieve the optimal quality of secure caching service strategy based on the observations of the payment from the content provider. The main contribution of this paper is threefold.

- *Game Based Interaction Modelling.* We formulate the interactions between the content provider and multiple edge caching devices as a Stackelberg game and analyze the Stackelberg equilibrium (SE) of the static Stackelberg game with the interaction as the optimal strategies to maximize the utilities of all players. Each edge caching device is stimulated to provide high-quality secure caching service and the selfishness of edge caching device is suppressed.
- *Q-learning Based Optimal Strategy Decision.* We apply the Q-learning to obtain the optimal payment strategy for the content provider and the quality of secure caching service of each edge caching device via trial and error with high learning rate in dynamic network that lacks sufficient knowledge on the interactions between the content provider and edge caching devices.
- *Extensive Simulations Based Performance Evaluation.* We evaluate the performance of the proposed scheme with extensive simulations. The simulation results show that the proposed scheme can efficiently motivate edge caching devices to provide high-quality secure caching services for the content provider and mobile users.

The remainder of the paper is organized as follows. Section II reviews the related works. Section III introduces the system model. The problem formulation is presented in Section IV. The optimal strategy for static Stackelberg game is analyzed in Section V and the Q-learning based optimal strategy decision is elaborated in Section VI. Section VII

evaluates the proposed scheme and conclusion is provided in Section VIII.

II. Related Work

In this section, we first review the content delivery in MSNs and then conclude the works of edge caching, reinforcement learning and security for edge caching in wireless network.

A. Content Delivery in Mobile Social Networks

Recently, the content delivery in MSNs has been studied widely. Wang *et al*. [26] introduced a comprehensive enhanced secure instant messaging scheme, which can support denial of replaying attack and denial of forgery attack, where an offline key agreement process between users is developed by updating the ephemeral key periodically. Wang *et al*. [27] studied the opportunistic MSNs, where a dynamic trust framework is proposed to facilitate a node to obtain a trust value of another one and presented a two-hop feedback method to verify a node's honesty if they are two hops away. Li *et al*. [28] presented a communication framework by studying two tightly coupled issues, where a novel data forwarding algorithm is proposed to route the message to other access points within common connected components. Awuor *et al*. [29] proposed a mechanism to motivate content sharing in mobile social networks, which is based on the users' collective bidding, content cost sharing, and trust evaluation and proved that the mechanism can ensure the trustworthiness of their encounters' contents. Meng *et al*. [30] studied mobile communication technology in MSNs and resented a high-level distributed cooperative environmental state inference scheme, where users can exchange information with their neighbors and cooperatively infer the hidden state. However, the existing works on content delivery in MSNs should further consider the local content services for mobile users.

B. Edge Caching in Wireless Networks

The edge caching in wireless networks has attracted much attention from academia and industry. Jiang *et al.* [31] proposed a device-to-device content caching mechanism through a multi-agent reference learning approach to improve the cache byte hit rate and decrease the average downloading latency. To reduce congestion in the backhaul links in IP network and minimize the delay to fetch contents from remote servers, Chhangte *et al.* [32] presented a novel software defined networking based content caching prototype called Wi-Caching in wireless local area network. Through joint optimizing approaches for content caching, replica server placement, and request load assignment, Xu *et al.* [33] developed a mixed integer linear programming in content delivery networks to improve the overall system performance. Liu *et al.* [34] studied the joint content caching and computation offloading in wireless blockchain networks via mobile edge computing technique, and the optimization problem is solved by alternating direction method of multipliers based algorithm in a distributed way. A novel cache-aided coded content delivery mechanism in wireless networks was designed by Yang and Gndz [35], which focuses on the heterogeneous distortion requirements of users in both centralized and decentralized ways. However, to realize existing works on edge caching in reality, the security preservation for contents cached on edge caching devices should be further studied and resolved.

C. Reinforcement Learning in Wireless Network

Reinforcement learning as a typical model of machine learning technology has been deeply applied in wireless communication. Xiao *et al*. [36] studied the attack models in mobile edge computing systems and proposed security solutions to provide secure offloading to the edge caching devices against jamming attacks. In particular, the lightweight authentication and secure collaborative caching schemes were introduced to protect the data privacy. Li *et al*. [37] focused on the problem of spectrum sharing in a cognitive radio system, and developed a deep reinforcement learning-based method for the secondary user to share the common spectrum with the primary user, where users can transmit their own data successfully with required qualities of service. Shen *et al*. [38] studied a number of image-processing problems, and a deep reinforcement learning approach was developed to train a system to adjust parameters automatically in a human-like manner. Liu *et al*. [39] focused on the topic of robotic systems with cloud computing, and a resource allocation scheme was presented to make the cloud to decide whether a request should be accepted and how many resources are supposed to be allocated. Mahmud *et al*. [40] investigated the development of dedicated data-intensive machine learning techniques, and gave a comprehensive survey on the application of deep learning, reinforcement learning, and deep reinforcement learning techniques in mining biological data. However, the reinforcement learning on edge caching for security of contents needs to be further analyzed.

D. Security for Edge Caching

The security for edge caching in wireless networks has been widely studied. Araldo *et al.* [41] proposed an encrypted content caching scheme in which the network provider partitions the caching space into slices and assigns slices to different content providers. Tiburski *et al.* [42] designed a security architecture that integrates trust mechanisms with embedded virtualization, which can efficiently prevent unauthorized access to cached contents on edge devices. Ma *et al.* [43] introduced a blockchain-based trusted data management scheme for edge caching to guarantee content trust and security, where a flexible and configurable blockchain architecture is devised. Dbouk *et al.* [44] proposed an ad-hoc mobile edge cloud that utilizes Wi-Fi Direct as means of achieving connectivity, sharing resources, and integrating security services among nearby mobile devices to defend devices against malware and other intrusions. Wu *et al.* [45] presented an edge-caching-based content-aware filtering method for security services in information-centric social networks, where the assessment and content-matching mechanism is presented for security services. Although the above works have made efforts on secure caching, the selfishness and open access features of edge caching devices and the diversified caching demands of the content provider should be further taken into account when devising the secure edge caching scheme.

TABLE I
SUMMARY OF NOTATIONS

Notations	Description
$\mathcal{I}$	Set of edge caching devices.
$q_{i,m}$	Quality of secure caching service from edge caching device i for content m .
$N_i(t)$	The number of mobile users in the coverage of edge caching device i at time slot t.
$\mathcal{M}$	Set of contents.
f_m	Popularity of content m.
p_m	Importance of content m.
$r_{i,m}$	Ratio of mobile users requesting content m via edge caching device i.
$\mathbf{g}$	Payment strategy matrix of the content provider.
$\mathbf{q}$	Security quality strategy matrix of all edge caching devices.
$g_{i,m}$	Payment for the secure caching service from edge caching device i on content m.
$\mathbf{q}_i$	Security quality strategy vector of edge caching device i.
$\mathbf{g}_i$	Payment strategy vector for edge caching device i.
$\mathbf{s}_o^t$	Quality state of edge caching device at time slot t.
$\mathbf{g}_i$	Payment strategy vector for edge caching device i.
$Q(s_{i,m}, g_{i,m})$	Q function of Q-learning for the content provider.
$\hat{s}_{i,m}^t$	Payment state of the content provider for content m edge caching device i at time slot t.
$\hat{Q}(\hat{s}_{i,m}, q_{i,m})$	Q function of Q-learning for the edge caching device i on content m.

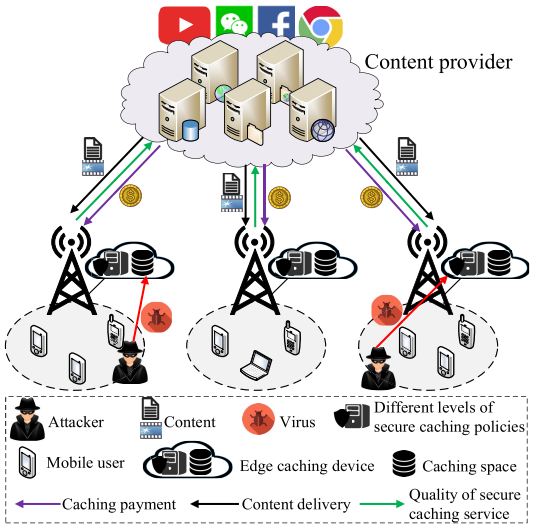


Fig. 1. An illustration of MSN model. The MSN is composed of a content prover, multiple edge caching devices, and a mass of mobile users.

In this paper, different from existing works, the proposed scheme studies the secure edge caching in MSNs. The selfishness and open-access feature of edge caching device are considered to improve the security of content delivery. In addition, the Q-learning as a reinforcement mode is used to obtain the optimal strategies of both the content provider and edge caching devices in dynamic Stackelberg game, without the whole knowledge of accurate network parameters.

III. SYSTEM MODEL

In this section, we introduce the system model including network model, content model, and threat model

A. Network Model

In this paper, we consider a typical MSN model, as shown in Fig. 1, which includes a content provider, multiple edge caching devices, and some mobile users.

1) *Content Provider:* The content provider, denoted by o, stores the original content files, where the copyrights of the contents are bought from content producers. In order to attract more users for potential benefits, the content provider hopes that it can securely deliver the desired popular contents to mobile users as soon as possible. However, since the content provider is deployed at some remote areas far away from mobile users, mobile users have to retrieve contents with a long latency, resulting in the degrading QoE of the content delivery. To improve mobile users' QoE, the content provider has high willingness to securely cache contents on edge caching devices with an acceptable payment.
2) *Edge Caching Devices:* Edge caching devices equipped with the edge caching technology can provide caching service for content provider and mobile users. Specifically, an edge caching device is placed at the edge of backhaul network between the content provider and mobile users to provide proximate caching services for mobile users. As such, if contents are cached on edge caching devices, mobile users can retrieve the requested contents with a short latency by accessing nearby edge caching devices. For simplicity, to reduce the overlapping coverage area among edge caching devices, each edge caching device is placed at different locations such as campus, hospital, etc. Each edge caching device has a certain selfishness and open-access feature. The selfishness means that edge caching device cheats content provider to gain profits, while this node does not cache contents in its caching space and delivers fake contents to mobile users. Open-access means that each edge caching device can be accessed by any entities including malicious parties, where cached contents may be tampered, removed, and replaced due to attacks. To avoid the negative effect of the open-access feature on content caching, each edge caching device is also capable of providing security mechanism to protect cached contents for the content provider and mobile users. For instance, edge caching device can employ disaster backup mode to store multiple copies of contents at different places, whereby the integrities and confidentialities of contents can be preserved [14]. Let $\mathcal{I} = \{1, 2, \cdots i, \cdots, I\}$

denote the set of edge caching devices in the network. By controlling the quality of secure caching service, each edge caching server can execute different levels of security mechanisms during content caching. The higher the quality of secure caching service is, the higher the level of the adopted security mechanism becomes. Here, the quality of secure caching service from edge caching device i for caching content m is denoted by $q_{i,m}$, which satisfies the following constraint condition:

$$q_{i,m} = \begin{cases} -1, & \text{edge caching device } i \text{ is selfish,} \\ [0,1], & \text{otherwise.} \end{cases} \quad (1)$$

Here, $q_{i,m} = -1$ represents that edge caching device i is selfish and delivers fake contents to requesters. $q_{i,m} = 1$ is the highest quality of secure caching service from edge caching device i and edge caching device i provides low quality of secure caching service when $0 < q_{i,m} < 1$. If $q_{i,m} = 0$, edge caching device i does not participate in content caching for the content provider and only forwards contents as a relay.

3) *Mobile Users:* In the network, mobile users require contents from nearby edge caching devices. If contents are cached on edge caching devices, the corresponding contents are returned to mobile users directly with negligible latencies. The random walk model is utilized to describe mobile user's mobility. The moving velocity of each mobile user is randomly in $\left[\tilde{V}_{\min}, \tilde{V}_{\max}\right]$. The moving direction of each mobile user is randomly in $[0, 2\pi]$. The time of a mobile user's one moving is randomly in $[0, \hat{t}_{\max}]$. Besides, after moving, each mobile user can stay at a location for a certain time. The staying time of one mobile user is randomly in $[0, \tau_{\max}]$. According to mobility model, the number of mobile users within the coverage of each edge caching device is variable over time. As such, the set of mobile users within the coverage of edge caching device i at time slot t is denoted as $\mathcal{N}_i(t) = \left\{n_{i,1}, n_{i,2}, \cdots, n_{i,N_i(t)}\right\}$, where $N_i(t)$ is the number of mobile users within the coverage of edge caching device i at time slot t. Here, mobile users can send experience reports about the quality of secure caching service from associated edge caching devices to the content provider and then the content provider evaluates the qualities to pay for secure caching services. For simplicity, it is assumed that the content provider can accurately evaluate the quality of secure caching service from edge caching devices. For example, the content provider can deliberately launch various different degrees of malicious attacks to an edge caching device for evaluating the quality of secure caching service, when contents are cached on the edge caching device.

B. Content Model

Within the time period $\{1, 2, \cdots, T\}$, a certain number of contents are requested by mobile users. The set of contents in time period $\{1, 2, \cdots, T\}$ is denoted as $\mathcal{M} = \{1, 2, \cdots, M\}$. The popularity distribution of contents $\mathcal{M}$ are represented by vector $\mathbf{f} = [f_1, f_2, \cdots, f_M]$. Here, each mobile user makes an independent request of content $m \in \mathcal{M}$, with a probability of f_m. With the descending order of requesting times during a certain period (e.g., one day or one week), the popularity of content $m \in \mathcal{M}$ is

$$f_m = \left[(\tau(m))^{\gamma} \sum_{m=1}^{M} m^{-\gamma} \right]^{-1}, \quad (2)$$

where γ is a positive number to govern the skewness of the popularity. The popularity is uniformly distributed when $\gamma = 0$. If γ is high, popular contents will account for the majority of requests. $\tau(m)$ is the index of content m with the decreasing order of requesting times among all contents. Eq. (2) implies that content with a smaller index has a larger popularity.

In addition, the importance of different contents should be considered. For example, a content is about the national economic news generated by the national government department. Another content is world cup news from a sport newspaper. Although the world cup news is more popular than the national economic news, the national economic news is more important. The importance distribution among all contents is denoted by $\mathbf{p} = [p_1, p_2, \cdots, p_M]$, where p_m represents the importance degree of content m. The importance of a content can be determined by the priority of the content source. For example, the priority of the national government department is higher than the sport newspaper. The importance distribution $\mathbf{p}$ can be also modeled by the Zipf distribution. As such, we have

$$p_m = \left[(\kappa(m))^{\beta} \sum_{m=1}^{M} m^{-\beta} \right]^{-1}, \quad (3)$$

where β is a positive value for characterizing the importance of the contents and $\kappa(m)$ is the index of content m with the decreasing order of the priority among all contents.

Apart from the popularity and importance, the requests of mobile users within different edge caching devices are different. For edge caching device i, the request distribution is denoted by $\mathbf{r}_i = [r_{i,1}, r_{i,2}, \cdots, r_{i,M}]$, where $r_{i,m}$ is the ratio of mobile users requesting content m within edge caching device i. The number of mobile users who request content m from edge caching device i is $r_{i,m}N_i$, which can be seen as the priority of the edge caching device. If a content is requested by a large number of mobile users under an edge caching device, this content has high priority to be cached.

C. Threat Model

There are two types of threats in the system to affect the security of the caching services.

1) Selfishness of Edge Caching Devices: Since edge caching devices are deployed by multiple rational third parties, they may be selfish to cheat the content provider for caching service fees, where fake or forged contents are delivered back to content requesters.

2) Open Access of Edge Caching Devices: Since edge caching devices can be accessed openly, malicious mobile users can conduct several cyber-attacks (e.g., man-in-middle

attack, black hole attack, etc). The cached contents are tampered, removed or replaced, and the requesters cannot obtain the desired contents from edge caching devices.

D. Design Goals

Our design goals have two desirable objectives as follows: on one hand, our scheme should be effective to motivate edge caching devices to provide high-quality secure caching services for content provider and mobile users. On the other hand, the content provider should be fast to find the optimal payment strategy without the information of network model and caching model parameters in dynamic network scenarios.

IV. Problem Formulation

In the network, if the mobile user's QoE is improved with the high-quality secure caching services offered by edge caching devices, the content provider can also gain profit since it can attract more mobile users to request contents. As such, both the content provider and mobile users can be seen as the same interest group. For stimulating edge caching devices to provide high-quality secure caching services and suppressing edge caching devices' selfishness, the content provider chooses a payment strategy to award edge caching devices' contributions. Then, given payment strategy, each edge caching device determines the quality of secure caching service. Both the content provider and edge caching devices aim to maximize their profits. The content provider hopes that it can enjoy high-quality services with low payments, whereas each edge caching device hopes that the price can be as high as possible. Therefore, there is a competition between the content provider and each edge caching device. The competition interaction between the content provider and edge caching devices is formulated as a Stackelberg game to analyze the optimal strategies of all players. To formulate the Stackelberg game, the utility functions of the content provider and edge caching devices need to be designed, respectively.

A. Utility Function of Content Provider

The utility function of the content provider is first introduced. As the content provider wants to cache contents on edge caching devices, the utility of the content provider is constituted by the utility regarding on each content cached on one edge caching device. As such, the utility of the content provider can be obtained by

$$U_o(\mathbf{g}, \mathbf{q}) = \sum_{i=1}^{I} \sum_{m=1}^{M} u_o(g_{i,m}, q_{i,m}), \tag{4}$$

where $u_o(g_{i,m}, q_{i,m})$ is the content provider's utility to cache content m on edge caching device i. $\mathbf{g}$ is the payment strategy of the content provider. $\mathbf{q}$ is the vector on qualities of secure caching services from all edge caching devices. Here, the content provider adopts non-uniform payment policy, where the payments for different contents on different edge caching devices are different. Thus, the payment strategy can be defined as

$$\mathbf{g} = \begin{bmatrix} g_{1,1}, & g_{1,2}, & \cdots, & g_{1,M} \\ g_{2,1}, & g_{2,2}, & \cdots, & g_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ g_{I,1}, & g_{I,2}, & \cdots, & g_{I,M} \end{bmatrix} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_I]^{\mathbf{T}}, \tag{5}$$

where $\mathbf{T}$ means matrix transposition and $g_{i,m}$ is the payment for the secure caching service from edge caching device i on content m. The vector on qualities of secure caching services from all edge caching devices can be given by

$$\mathbf{q} = \begin{bmatrix} q_{1,1}, & q_{1,2}, & \cdots, & q_{1,M} \\ q_{2,1}, & q_{2,2}, & \cdots, & q_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ q_{I,1}, & q_{I,2}, & \cdots, & q_{I,M} \end{bmatrix} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_I]^{\mathbf{T}}, \tag{6}$$

As the utility function of the content provider is defined as the difference between the satisfaction and payment for secure caching service, we have

$$u_o(g_{i,m}, q_{i,m}) = F_{i,m}(q_{i,m}) - C_{i,m}(g_{i,m}, q_{i,m}), \tag{7}$$

where $F_{i,m}(q_{i,m})$ is the satisfaction of the content provider with the quality of secure caching service $q_{i,m}$. $C_{i,m}(g_m, q_{i,m})$ is the cost of the content provider to pay for the secure caching service from edge caching device i on content m. Here, we use the typical logarithmic function as the satisfaction, which has been widely used in the resource allocation schemes [46]. $F_{i,m}(q_{i,m})$ can be defined by

$$F_{i,m}(q_{i,m}) = \begin{cases} \alpha r_{i,m} N_i(t) f_m p_m \log(1 + q_{i,m}), & q_{i,m} \geq 0, \\ \varsigma r_{i,m} N_i(t) f_m p_m q_{i,m}, & q_{i,m} = -1, \end{cases} \tag{8}$$

where α is the satisfaction parameter of secure content caching and ς is the satisfaction loss parameter. Here $\alpha > 0$ and $\varsigma > 0$. With Eq. (8), the content satisfaction parameter is a piecewise function. If the quality of secure caching service is larger than zero, the content provider can obtain positive satisfaction. Otherwise, when the quality of secure caching service is equal to -1, the content provider is cheated by the edge caching device and has the negative satisfaction. Since the content provider should pay for the secure caching services of edge caching devices, the cost function $C_{i,m}(g_m, q_{i,m})$ is defined by

$$C_{i,m}(g_{i,m}, q_{i,m}) = \begin{cases} g_{i,m} \theta q_{i,m}, & q_{i,m} \in [0, 1] \\ 0, & q_{i,m} = -1, \end{cases} \tag{9}$$

where $g_{i,m}$ is payment for the highest-quality caching service from edge caching device i on content m, i.e., $q_{i,m} = 1$. θ is the payment adjust parameter. From Eq. (9), the payment to the edge caching device that does not participate content caching or cheats the content provider is zero. With consideration of all contents and edge caching devices, the utility of

the content provider at time slot t is

$$\begin{aligned} &U_o(\mathbf{g},\mathbf{q}) \\ &= \sum_{i=1}^{I}\sum_{m=1}^{M} u_o(g_{i,m},q_{i,m}) \\ &= \sum_{i=1}^{I}\sum_{m=1}^{M}\Big\{x_{i,m}(\alpha r_{i,m}N_i(t)f_m p_m \log(1+q_{i,m}) - g_m\theta q_{i,m}) \\ &\quad + (1-x_{i,m})\varsigma r_{i,m}N_i(t)f_m p_m q_{i,m}\Big\}, \end{aligned} \tag{10}$$

where $x_{i,m}$ is the indictor function of edge caching device i on content m, which can be obtained by

$$x_{i,m} = \begin{cases} 1, & q_{i,m} \in [0,1] \\ 0, & q_{i,m} = -1. \end{cases} \tag{11}$$

B. Utility Function of Edge Caching Device

The utility function of each edge caching device is based on the payment offered from the content provider and the cost to provide secure caching service for the content provider. As such, the utility function of edge caching device is the difference between the payment and the cost on secure caching service. The utility function of the edge caching device $i, \forall i \in \mathcal{I}$, can be obtained by

$$U_i(\mathbf{q}_i,\mathbf{g}) = L_i(\mathbf{q}_i,\mathbf{g}) - \Phi_i(\mathbf{q}_i), \tag{12}$$

where $L_i(\mathbf{q}_i,\mathbf{g})$ is the payment offered by the content provider with the payment strategy $\mathbf{g}$ and the vector on qualities of secure caching services on M contents $\mathbf{q}_i$. $\Phi_i(\mathbf{q}_i)$ is the cost function of the edge caching device i to provide secure caching service for the content provider. If the edge caching device provides high-quality secure caching service, it will receive large payment. As such, the payment function is defined by

$$L_i(\mathbf{q}_i,\mathbf{g}) = \sum_{m=1}^{M} C_{i,m}(g_{i,m},q_{i,m}), \quad \forall i \in \mathcal{I}. \tag{13}$$

The cost function is also related to the quality of secure caching service. If the quality of secure caching service is high, the edge caching device needs much effort to guarantee the securities of cached contents. As such, the cost function of the edge caching device i can be obtained by

$$\Phi_i(\mathbf{q}_i) = \sum_{m=1}^{M} \varphi_{i,m}(q_{i,m}), \quad \forall i \in \mathcal{I}, \tag{14}$$

where $\varphi_{i,m}(q_{i,m})$ is the cost of the edge caching device i with quality of secure caching service on content m, which can be written by

$$\varphi_{i,m}(q_{i,m}) = \begin{cases} c_i \upsilon {q_{i,m}}^2, & q_i \in [0,1] \\ \psi_i, & q_i = -1, \end{cases} \tag{15}$$

where c_i is the cost parameter of the edge caching device i with the highest quality secure caching service. It means the total cost of the edge caching device i to provide the highest-quality secure caching service. υ is the adjustment parameter for the edge caching device i. ψ_i is a fixed value that the resource (e.g., power and bandwidth, etc.) consumption of edge caching device i when it cheats the content provider. Then, the utility of edge caching device i can be rewritten by

$$\begin{aligned} U_i(\mathbf{q}_i,\mathbf{g}) &= L_i(\mathbf{q}_i,\mathbf{g}) - \Phi_i(\mathbf{q}_i) \\ &= \sum_{m=1}^{M} u_{i,m}(q_{i,m},g_{i,m}), \end{aligned} \tag{16}$$

where $u_{i,m}(q_{i,m},g_{i,m})$ is the utility of edge caching device i to securely cache content m. Here, we have

$$u_{i,m}(q_{i,m},g_{i,m}) = \begin{cases} g_{i,m}\theta q_{i,m} - c_i \upsilon {q_{i,m}}^2, & q_{i,m} \in [0,1] \\ -\psi_i, & q_{i,m} = -1. \end{cases} \tag{17}$$

C. Optimization Problems

With the above description, the interactions between the content provider and edge caching devices are formulated as a one-leader and multiple-followers Stackelberg game denoted by $\mathbb{G} = \{(o,1,2,\cdots,I);(\mathbf{g},\mathbf{q});(U_o,U_{1\le i\le I})\}$, where each player wants to select the optimal strategy for maximizing its utility. The content provider chooses the optimal payment strategy $\mathbf{g}$ for maximizing its profit and edge caching device $i, \forall i \in \mathcal{I}$ determines the optimal qualities of secure caching services $\mathbf{q}_i$ to maximize its utility. As such, two optimization problems are introduced.

Problem 1: The optimization problem for maximizing the profit of the content provider can be formulated as

$$\begin{aligned} &\max_{\mathbf{g}} \quad U_o(\mathbf{g},\mathbf{q}), \\ &s.t.\ g_{i,m} \ge 0, \quad \forall i \in \mathcal{I}, \forall m \in \mathcal{M}. \end{aligned} \tag{18}$$

Problem 2: The optimization problem for maximizing edge caching device i's utility can be written as

$$\begin{aligned} &\max_{q_{i,1},q_{i,2},\cdots,q_{i,M}} \quad U_i(\mathbf{g},\mathbf{q}_i), \\ &s.t.\ q_{i,m} = -1 \text{ or } q_{i,m} \in [0,1], \\ &\quad \forall i \in \mathcal{I}, \quad \forall m \in \mathcal{M}. \end{aligned} \tag{19}$$

The objectives of both the content provider and edge caching devices are to maximize their utilities. As such, the problem defined in formula (18) is to determine the optimal payment strategy for maximizing the utility of the content provider in each time slot. It takes the number of mobile users that require cached contents, content popularity, content importance, and the qualities of caching services provided by edge caching devices into consideration. The problem defined in formula (19) is exploited to maximize the utility of each edge node by determining the optimal strategy on the quality of secure edge caching service, with the consideration of payment and cost. The restriction of the problem in formula (19) is that the quality of secure caching service cannot exceed the range defined in the system model.

The objective of game $\mathbb{G}$ is to find the SE, from which neither the leader (i.e., content provider) nor the followers (i.e., edge caching devices) have incentives to deviate.

Definition: Let $\mathbf{g}^* = \{\mathbf{g}_1^*,\mathbf{g}_2^*,\cdots,\mathbf{g}_I^*\}$ be a solution for problem 1, and $\mathbf{q}^* = \{\mathbf{q}_1^*,\mathbf{q}_2^*,\cdots,\mathbf{q}_I^*\}$ be a solution for

problem 2. Then the point $(\mathbf{g}^*, \mathbf{q}^*)$ is an SE for the proposed Stackelberg game $\mathbb{G}$ if the following conditions are satisfied:

$$U_o(\mathbf{g}^*, \mathbf{q}^*) \geq U_o(\mathbf{g}, \mathbf{q}^*), \tag{20}$$

$$u_i(\mathbf{g}^*, \mathbf{q}_i^*) \geq u_i(\mathbf{g}^*, \mathbf{q}_i), \ \forall i \in \mathcal{I}. \tag{21}$$

Formula (20) and formula (21) mean that the optimal strategy $\mathbf{g}^*$ can offer the maximum utility for the content provider when each edge caching device has determined its optimal strategy. Similarly, with the optimal strategy $\mathbf{g}^*$ given by the content provider, the edge caching device i can obtain the largest utility with the optimal quality of secure caching service $\mathbf{q}_i^*$.

V. Static Stackelberg Game Analysis

In this section, we analyze the optimal strategies of both the content provider and edge caching devices with the static Stackelberg game with one interaction, where the parameters of the game are public knowledge to all players. The goal of the Stackelberg analysis is to find the SE, where both the content provider and edge caching devices determine the optimal strategies and achieve the maximum utilities. The backward induction method is exploited to analyze the proposed game, where each follower is first analyzed, under which the strategy of the leader is obtained. Namely, we first analyze the edge caching devices' decision processes to obtain the optimal strategies on qualities of secure caching services. Then, we investigate the optimal payment strategy of the content provider. Indeed, we consider commercial caching framework consisting of some content providers and multiple edge caching devices, where the content provider purchases secure caching services provided by these edge caching devices. Specifically, the content provider first determines the service payment to compensate for the cost of the associated edge caching device, and each edge caching device chooses the corresponding quality of secure caching service to make profit. Apparently, the above iteration between the content provider and edge caching devices conforms to the feature of the Stackelberg game, where the content provider acts as the leader to determine the payments and edge caching devices are followers to make strategies on the qualities of secure caching services. Especially, with the analysis of the Stackelberg game, strategies of both the content provider and edge caching devices can be obtained to maximize the profits of both players.

A. The Optimal Strategy of Edge Caching Device

As contents are independent with each other in the network, we first analyze the optimal decision of edge caching device i on content m. From Eq. (9) and Eq. (15), $C_{i,m}(0,-1) = C_{i,m}(0,0) = 0$, whereas $\varphi_{i,m}(-1) > \varphi_{i,m}(0) = 0$. Therefore, if the quality of secure caching service of each edge caching device is evaluated, the utility of edge caching device with cheating behavior is smaller than that without participating in the secure content caching, i.e., $u_{i,m}(0,-1) = -\psi_i < u_{i,m}(0,0)$. Accordingly, with zero-payment punishment, the selfishness of each node can be efficiently avoided.

Then, we analyze the optimal strategy of each edge caching device when the quality of secure caching service is larger than zero. Here, the utility function of the edge caching device i to securely cache content m can be rewritten by

$$u_{i,m}(g_{i,m}, q_{i,m}) = g_m \theta q_{i,m} - c_i \upsilon {q_{i,m}}^2, \quad 0 < q_{i,m} \leq 1. \tag{22}$$

The first order differential of edge caching device i's utility with respect to $q_{i,m}$ is

$$\frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}} = g_{i,m}\theta - 2c_i \upsilon q_{i,m}, \quad 0 < q_{i,m} \leq 1. \tag{23}$$

The second order differential of the edge caching device i's utility with respect to $q_{i,m}$ is

$$\frac{\partial^2 u_{i,m}(g_{i,m}, q_{i,m})}{\partial {q_{i,m}}^2} = -2c_i \upsilon < 0, \quad 0 < q_{i,m} \leq 1. \tag{24}$$

Based on Eq. (24), since the second order differential of edge caching device i's utility $\frac{\partial^2 u_{i,m}(g_{i,m}, q_{i,m})}{\partial {q_{i,m}}^2}$ is smaller than zero, the first order differential of edge caching device i's utility $\frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}}$ is a monotonic decreasing function, where we can have

$$\lim_{q_{i,m} \to 0} \frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}} = g_{i,m}\theta > 0. \tag{25}$$

$$\lim_{q_{i,m} \to 1} \frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}} = g_{i,m}\theta - 2c_i \upsilon. \tag{26}$$

Here we consider two cases.

Case 1: High payment. If the payment of the content provider is high, i.e., $g_{i,m} \geq \frac{2c_i \upsilon}{\theta}$, we can have

$$\lim_{q_{i,m} \to 1} \frac{\partial u_{i,m}(g_m, q_{i,m})}{\partial q_{i,m}} = g_m \theta - 2c_i \upsilon \geq 0. \tag{27}$$

The first order differential of edge caching device i's utility is not less than zero with $0 < q_{i,m} \leq 1$, so that the utility function of edge caching device $u_{i,m}(g_{i,m}, q_{i,m})$ is a monotone increasing function. The maximum utility can be obtained at the end of $q_{i,m}$ interval and the optimal strategy of the edge caching device i is ${q_{i,m}}^* = 1$. The maximum utility of the edge caching device i can be obtained by

$${u_{i,m}}^*(g_{i,m}, q_{i,m}) = g_{i,m}\theta - c_i \upsilon, \quad g_{i,m} \geq \frac{2c_i \upsilon}{\theta}. \tag{28}$$

Case 2: Low payment. If the payment of the content provider is low, i.e., $0 < g_m < \frac{2c_i \upsilon}{\theta}$, we can have

$$\lim_{q_{i,m} \to 1} \frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}} = g_{i,m}\theta - 2c_i \upsilon < 0. \tag{29}$$

Based on Eq. (29), the first order differential of edge caching device i's utility is first larger than zero and then smaller than zero. Accordingly, the utility function of edge caching device i first increases and then decreases with $q_{i,m}$, i.e., $u_{i,m}(g_{i,m}, q_{i,m})$ is a strictly concave function, where there exists a maximum utility for edge caching device i. The optimal strategy of edge caching device i can be obtained by solving the following equation.

$$\frac{\partial u_{i,m}(g_{i,m}, q_{i,m})}{\partial q_{i,m}} = g_{i,m}\theta - 2c_i \upsilon q_{i,m} = 0. \tag{30}$$

Then, the optimal strategy of edge caching device i on quality of secure caching service is

$$q_{i,m}{}^* = \frac{g_{i,m}\theta}{2c_i\upsilon}. \tag{31}$$

Therefore, the optimal strategy of edge caching device i on content m is

$$q_{i,m}{}^* = \begin{cases} 1, & g_{i,m} \geq \dfrac{2c_i\upsilon}{\theta}, \\ \dfrac{g_m\theta}{2c_i\upsilon}, & 0 < g_{i,m} \leq \dfrac{2c_i\upsilon}{\theta}, \\ 0, & g_{i,m} = 0. \end{cases} \tag{32}$$

The maximal utility of edge caching device i with the optimal strategy is

$$u_{i,m}{}^*(g_{i,m}, q_{i,m}) = \begin{cases} g_{i,m}\theta - c_i\upsilon, & g_{i,m} \geq \dfrac{2c_i\upsilon}{\theta}, \\ \dfrac{(g_{i,m}\theta)^2}{4c_i\upsilon}, & 0 < g_{i,m} < \dfrac{2c_i\upsilon}{\theta}, \\ 0, & g_{i,m} = 0. \end{cases} \tag{33}$$

B. The Optimal Strategy of Content Provider

Since the content provider adopts non-uniform payment strategy, we can analyze the policy for each content with one edge caching device. Based on the optimal quality of caching service strategy in Eq. (32), if $g_{i,m} > \frac{2c_i\upsilon}{\theta}$, the utility of the content provider with $q_{i,m}{}^* = 1$ can be rewritten as

$$u_o(g_{i,m}, q_{i,m}) = \alpha r_{i,m} N_i(t) f_m p_m \log 2 - g_{i,m}\theta, \ g_{i,m} \geq \frac{2c_i\upsilon}{\theta}. \tag{34}$$

Since the utility in Eq. (34) is a monotonic decreasing function, the optimal payment strategy of the content provider can be obtained by

$$g_{i,m}{}^* = \frac{2c_i\upsilon}{\theta}. \tag{35}$$

The maximum utility of the content provider is

$$u_o{}^*(q_{i,m}, g_{i,m}) = \alpha r_{i,m} N_i(t) f_m p_m \log 2 - 2c_i\upsilon. \tag{36}$$

If $0 < g_{i,m} < \frac{2c_i\upsilon}{\theta}$, the utility of the content provider with $q_{i,m}{}^* = \frac{g_{i,m}\theta}{2c_i\upsilon}$ can be rewritten as

$$u_o(q_{i,m}, g_{i,m}) = \alpha r_{i,m} N_i(t) f_m p_m \log\left(1 + \frac{g_{i,m}\theta}{2c_i\upsilon}\right) - g_{i,m}\theta\frac{g_{i,m}\theta}{2c_i\upsilon}, \quad s.t.\ 0 < g_{i,m} < \frac{2c_i\upsilon}{\theta}. \tag{37}$$

The first order differential of the content provider's utility with respect to $g_{i,m}$ is

$$\frac{\partial u_o(q_{i,m}, g_{i,m})}{\partial g_{i,m}} = \frac{\alpha r_{i,m} N_i(t) f_m p_m \theta}{2c_i\upsilon + g_{i,m}\theta} - \frac{g_{i,m}\theta^2}{c_i\upsilon}. \tag{38}$$

The second order differential of the content provider's utility with respect to $g_{i,m}$ is

$$\frac{\partial^2 u_o(q_{i,m}, g_{i,m})}{\partial {g_{i,m}}^2} = -\frac{\alpha r_{i,m} N_i(t) f_m p_m \theta^2}{(2c_i\upsilon + g_{i,m}\theta)^2} - \frac{\theta^2}{c_i\upsilon} < 0. \tag{39}$$

Since the second order differential in Eq. (39) is smaller than zero, the utility function of the content provider is a strict concave function, where the maximum value can be obtained by solving $\frac{\partial u_o(q_{i,m}, g_m)}{\partial g_m} = 0$ with KKT conditions. Then, the optimal payment strategy of content provider on the content m for edge caching device i is

$$g_{i,m}{}^* = \begin{cases} \dfrac{2c_i\upsilon}{\theta}, & \alpha r_{i,m} N_i(t) f_m p_m - 8c_i\upsilon\theta \geq 0, \\ \dfrac{\sqrt{\Theta_{i,m}} - c_i\upsilon\theta}{\theta^2}, & \alpha r_{i,m} N_i(t) f_m p_m - 8c_i\upsilon\theta < 0. \end{cases} \tag{40}$$

Here, $\Theta_{i,m} = {c_i}^2\upsilon^2\theta^2 + c_i\upsilon\theta\alpha r_{i,m} N_i(t) f_m p_m$. The maximum utility of the content provider with the optimal payment strategy in Eq. (40) can be calculated by

$$u_o{}^*(g_m, q_{i,m}) = \begin{cases} Z_{i,m}\log 2 - 2c_i\upsilon, \ Z_{i,m} - 8c_i\upsilon\theta \geq 0, \\ Z_{i,m}\log\left(1 + \dfrac{\sqrt{\Theta_{i,m}} - c_i\upsilon\theta}{2c_i\upsilon\theta}\right) - \dfrac{Z_{i,m} + 2c_i\upsilon\theta - 2\sqrt{\Theta_{i,m}}}{2\theta}, \\ Z_{i,m} - 8c_i\upsilon\theta < 0. \end{cases} \tag{41}$$

where $Z_{i,m} = \alpha r_{i,m} N_i(t) f_m p_m$. $Z_{i,m}$ can been seen as the value of the content m under the coverage of edge caching device i. From Eq. (40) and Eq. (32), if the value of the content is high or the cost parameter is low, the content provider has a high willingness to motivate edge caching devices to provide the highest-quality secure caching service. Otherwise, the content provider will select payment strategy based on the cost of the edge caching device to obtain a satisfied quality of secure caching service.

With the above game analysis, the SE between the content provider and edge caching device i on content m can be given by

$$(g_{i,m}{}^*, q_{i,m}{}^*) = \begin{cases} \left(\dfrac{2c_i\upsilon}{\theta}, 1\right), & Z_{i,m} - 8c_i\upsilon\theta \geq 0, \\ \left(\dfrac{\sqrt{\Theta_{i,m}} - c_i\upsilon\theta}{\theta^2}, \ \dfrac{\sqrt{\Theta_{i,m}} - c_i\upsilon\theta}{2\theta c_i\upsilon}\right), & Z_{i,m} - 8c_i\upsilon\theta < 0. \end{cases} \tag{42}$$

VI. Q-Learning Based Optimal Strategy Decision for Dynamic Stackelberg Game

In this section, we analyze the dynamic Stackelberg game between the content provider and edge caching devices, where the interactions between the content provider and edge caching devices are repeatedly conducted over time. In the static Stackelberg game, the parameters (e.g., satisfied parameters, cost parameter, etc.) for utilities of both the content provider and edge caching devices should be public knowledge, while these parameters, in reality, are private and cannot be fully known by all players. Alternatively, the content provider can conduct several interactions with the edge to search the optimal strategies for them through trial and error. The Q-learning can

Algorithm 1 Q-Learning Based Payment Strategy Searching Algorithm

Initialize $\mathbf{q} = \mathbf{0}$, $Q(\mathbf{s}, \mathbf{q}) = \mathbf{0}$, $V(\mathbf{s}) = \mathbf{0}$, ϑ, ϖ, $\mathbf{g} = \{hg_{\max}/H\}_{0\le h\le H}$
for $t = 1, 2, 3, \cdots$ **do**
  Evaluate the number of available edge caching devices $I(t)$ and that of mobile users within the coverage of each available edge caching device.
  for $i = 1 : I(t)$ **do**
    for $m = 1 : M$ **do**
      $s_{i,m}^t = q_{i,m}^{t-1}$.
      Select and perform $g_{i,m}^t \in \mathbf{g}$ via the ε-greedy algorithm.
      Send payment message with $g_{i,m}^t$ to edge caching device i.
      Observe and evaluate the quality of secure caching service $q_{i,m}^t$.
      Pay edge caching device i with $g_{i,m}^t \theta q_{i,m}^t$.
      Calculate $u_o(q_{i,m}^t, {g_{i,m}}^t)$.
      Update $Q(s_{i,m}^t, g_{i,m}^t)$ via Eq. (43).
      Update $V(s_{i,m}^t)$ via Eq. (44).
    end for
  end for
end for

Algorithm 2 Q-Learning Based Quality of Secure Caching Service Strategy Searching Algorithm

Initialize $\mathbf{g} = \mathbf{0}$, $\hat{Q}(\mathbf{s}, \mathbf{q}) = \mathbf{0}$, $\Upsilon(\mathbf{s}) = \mathbf{0}$, ϱ, ρ, $\mathbf{q} = \{k/K\}_{0\le k\le K}$
for $t = 1, 2, 3, \cdots$ **do**
  Evaluate the number of available edge caching devices $I(t)$ and that of mobile users within the coverage of each available edge caching device.
  for $i = 1 : I(t)$ **do**
    for $m = 1 : M$ **do**
      $\hat{s}_{i,m}^t = g_{i,m}^{t-1}$.
      Select and perform $q_{i,m}^t \in \mathbf{q}$ via the ϵ-greedy algorithm.
      Provide secure caching service with $q_{i,m}^t$ for the content provider.
      Observe the payment of the content provider $g_{i,m}^t$.
      Calculate $u_{i,m}(q_{i,m}^t, \hat{s}_{i,m}^t)$.
      Update $\hat{Q}(\hat{s}_{i,m}^t, q_{i,m}^t)$ via Eq. (46).
      Update $\Upsilon(\hat{s}_{i,m}^t)$ via Eq. (47).
    end for
  end for
end for

be utilized to model the content provider and edge caching devices under multiple interactions.

A. Q-Learning Based Payment Strategy Decision

In general, a high payment for the secure caching service decreases the content provider's immediate utility but it stimulates more edge caching devices to provide secure caching services in the future. Apparently, the current payment strategy of the content provider affects the future secure caching services and the future profit. The payment decision in the dynamic game can be formulated as a MDP, where reinforcement learning algorithm is an effective approach to achieve the optimal strategy. The Q-learning is a typical reinforcement learning mode that can be employed by the content provider to derive the optimal payment strategy without knowing the caching parameters of edge caching devices for the game. The content provider applies Q-learning process to obtain the optimal payment strategy, where the security quality state of the edge caching device is observed by the content provider at time slot t. Here, the state is denoted by $\mathbf{s}_o^t$, where $t \in \{1, 2, \cdots, T\}$. The system state consists of the previous quality of secure caching service from each edge caching device, i.e., $\mathbf{s}_o^t = \mathbf{q}^{t-1}$. The Q-learning based payment decision process for the content provider is shown in Algorithm 1. For simplicity, the content provider quantizes the payment into $H + 1$ level and chooses the payment level $g \in \{hg_{\max}/H\}_{0\le h\le H}$, where $g_{\max}$ is the maximum payment. In the Algorithm 1, the Q function in the Q-learning is denoted by $Q(s_{i,m}, g_{i,m})$, which means the expected long-term discounted utility for the state-action pair $(s_{i,m}, g_{i,m})$.

To obtain the optimal strategy, the Q function is updated over time slot according to the iterative Bellman equation:

$$Q(s_{i,m}^t, {g_{i,m}}^t) = (1-\varpi)Q(s_{i,m}^t, g_{i,m}^t) + \varpi\left(u_o(g_{i,m}^t, {q_{i,m}}^t) + \vartheta V(s_{i,m}^{t+1})\right) \quad (43)$$

where $s_{i,m}^{t+1}$ is the new state of the edge caching device i on content m at time slot $t+1$ from $s_{i,m}^t$ with payment action $g_{i,m}^t$. ϑ is the discount factor indicating the myopic views of the content provider regarding the future reward, and $\varpi \in (0, 1]$ is the learning rate. $u_o(q_{i,m}{}^t, g_{i,m}^t)$ is the utility of the content provider with the payment $g_{i,m}^t$ and the quality of secure caching service ${q_{i,m}}^t$, which is also the immediate reward of the payment action $g_{i,m}^t$. The $V(\cdot)$ denotes the highest value of the Q function, which is calculated by

$$V(s_{i,m}^{t+1}) = \max_{g_{i,m}} Q(s_{i,m}^{t+1}, g_{i,m}^{t+1}). \quad (44)$$

Since the tradeoff between the exploitation and exploration has an important effect on the convergence performance, the ε-greedy algorithm is applied for the content provider to choose the optimal strategy and avoid staying in the local optimization. More specifically, the "optimal" payment ${g_{i,m}}^*$ is chosen with a high probability $1-\varepsilon$, and the other payment strategies are randomly chosen with a very small probability ε. Thus the payment action of the content provider for content m with edge caching device i at time t is given by

$$\Pr\{{g_{i,m}}^t = g_{i,m}^*\} = \begin{cases} 1-\varepsilon, & g_{i,m}^* = \arg\max\ Q(s_{i,m}^t, g_{i,m}) \\ \varepsilon, & \text{otherwise.} \end{cases} \quad (45)$$

B. Q-Learning Based Quality of Secure Caching Service Strategy Decision

Since each edge caching device has no knowledge on the payment parameter of the content provider, the edge caching device cannot immediately find the optimal strategy to maximize its utility. Therefore, the edge caching devices also apply Q-learning to derive the optimal strategy on quality of secure caching service in the dynamic Stackelberg game consisting of the repeated interactions between the content provider and edge caching devices. The state in the Q-learning of each edge caching device is the payment of the content provider. Namely, the state for the edge caching device i on content m consists of previous payment offered by the content provider at time t, i.e., $\hat{s}_{i,m}^t = g_{i,m}^{t-1}$. As summarized in Algorithm 2, the Q-learning based optimal quality of secure caching service strategy is achieved. For simplicity, the quality of secure caching service is quantized into K level by the edge caching devices, and the quality of secure caching service is chosen with $q_{i,m} \in \{k/K\}_{0 \le k \le K}$. The Q function in the Q-learning for edge caching device i on securely caching content m is denoted by $\hat{Q}(\hat{s}_{i,m}^t, {q_{i,m}}^t)$, which is updated based on the iterative Bellman equation as follows.

$$\hat{Q}(\hat{s}_{i,m}^t, {q_{i,m}}^t) = (1-\varrho)\hat{Q}(\hat{s}_{i,m}^t, q_{i,m}^t) + \rho(u_{i,m}(q_{i,m}^t, g_{i,m}^t) + \rho\Upsilon_{i,m}(q_{i,m}^{t+1})) \quad (46)$$

where $\hat{s}_{i,m}^t$ is the payment state of the content provider at time slot $t+1$ from $\hat{s}_{i,m}^t$ with quality action $q_{i,m}^t$. ρ is the discount factor indicating the myopic views of the edge caching device regarding the future reward, and $\varrho \in (0, 1]$ is the learning rate. $u_{i,m}({q_{i,m}}^t, {g_{i,m}}^t)$ is the utility of the edge caching device with the payment $g_{i,m}^t$ and the quality of secure caching service ${q_{i,m}}^t$, which is also the immediate reward of the quality action ${q_{i,m}}^t$. The $\Upsilon(\cdot)$ denotes the highest value of the Q function, which is calculated by

$$\Upsilon(\hat{s}_{i,m}^{t+1}) = \max_{q_{i,m}} \hat{Q}(\hat{s}_{i,m}^{t+1}, q_{i,m}^{t+1}). \quad (47)$$

The ϵ-greed policy is applied for the edge caching device to choose the quality of secure caching service for the content provider based on payment state, with the consideration of the tradeoff the learning process between the exploitation and exploration. More specifically, the optimal quality strategy is chosen with a high probability $1-\epsilon$, while the other quality strategies are randomly chosen a small probability of ϵ. Thus the quality action of the edge caching device i for content m at time t is given by

$$\Pr\{q_{i,m}^t = q_{i,m}^*\} = \begin{cases} 1-\epsilon, & q_{i,m}^* = \arg\max\ \hat{Q}(\hat{s}_{i,m}^t, q_{i,m}) \\ \epsilon, & \text{otherwiese} \end{cases} \quad (48)$$

VII. Performance Evaluation

In this section, we conduct extensive simulations to verify the performance of the proposed scheme. The simulation setup is introduced, followed by the numerical results and analysis.

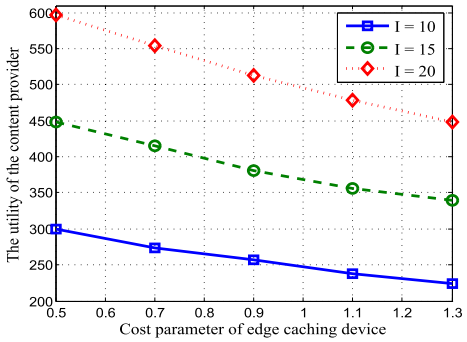


Fig. 2. The utility of the content provider vs. the cost parameter of the edge caching device.

A. Simulation Setup

In our simulation scenario, there is a $500 \times 500\ m^2$ terrain with 10 edge caching devices and one content provider. The cost parameter of each edge caching device is uniformly distributed in [0.5, 1.5]. The number of mobile users within each edge caching device is uniformly distributed in [50, 100]. The ratio of mobile users requesting each content is uniformly distributed in [0, 1]. The adjustment parameter for the payment is 1. The adjustment parameter for the cost of each edge caching device is 1. The satisfaction parameter is 20. For the Q-learning, learning rate of the content provider is $\varpi = 1$ [47]. The discount factor of the content provider is $\vartheta = 0.8$ [47]. The simulation time is $T = 8000$. Other parameters are $\gamma = 1.2$, $\beta = 1$ [24].

To show the superiority of the proposed scheme, following two conventional schemes are employed for comparisons.

- *Random Scheme [48].* In random scheme, the content provider selects a random payment strategy and each edge caching device randomly determines the quality of secure caching service.
- *Auction Scheme [14].* In this scheme, each content is cached on one edge caching device which is selected based on auction game without consideration of quality of secure caching service. Besides, the edge caching device randomly selects the quality of secure caching service.

B. Numerical Results

Fig. 2 shows the average utility of the content provider with the cost parameter of each edge caching device. In this simulation, the experiment is repeated with 10,000 times. All edge caching devices have the same cost parameter. In addition, the number of edge caching devices in the network increases from 10 to 20. The cost parameter increases from 0.5 to 1.3. From Fig. 2, it can be observed that the utility of the content provider decreases with the cost parameter of the edge caching device. The reason is that the payment of the content provider for the secure caching service is large if the cost parameter of each edge caching device is high. In addition, with the different numbers of edge caching devices, the utility of the content provider becomes equal when the cost parameter of the edge caching device increases. The reason is that the content provider has no willingness to stimulate edge caching device to provide the highest-quality of secure caching service

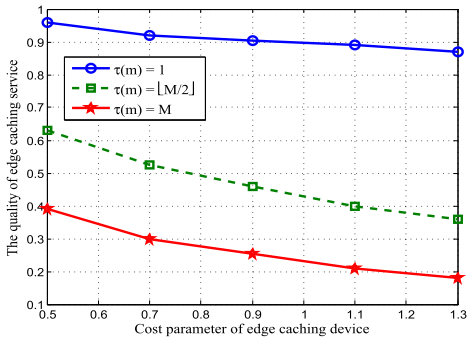


Fig. 3. The quality of secure caching service vs. the cost parameter of the edge caching device.

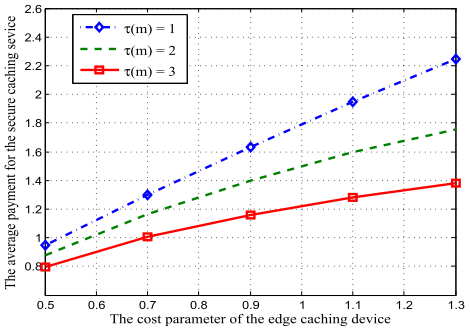


Fig. 4. The average payment of the content provider vs. the cost parameter of the edge caching device.

when the cost parameter increases, whereby the utility of the edge caching device decreases.

Fig. 3 shows the quality of secure caching service with the cost parameter of the edge caching device. To show the quality of secure caching service, three contents are compared. With the decreasing order of requesting times among all contents, the indexes of three contents are 1, $\lfloor M/2 \rfloor$, and M, respectively. Here $\lfloor \cdot \rfloor$ represents the floor function. From Fig. 3, it can be observed that the quality of secure caching service for each content decreases with the cost parameter of the edge caching device. The reason is that each edge caching device can decrease the quality of secure caching service to reduce cost since the high cost parameter increases the power consumption. In addition, as the popularity of the content with $\tau(m) = 1$ is the largest, the content provider has high willingness to offer a high payment for stimulating high-quality secure caching services from edge caching devices. Therefore, the decrease rate of content with $\tau(m) = 1$ is lower than that of other contents with larger indexes.

Fig. 4 is the payment of the content provider for secure caching services with the cost parameter of each edge caching device. In the simulation, three types of contents are used to show results, which are $\tau(m) = 1, 2, 3$, respectively. Other settings are unchanged. From Fig. 4, it can be seen that the payment for the secure caching service increases with the cost parameter of the secure caching service. The reason is that the content provider has to pay more for obtaining stable quality of secure caching service with the increase of the cost parameter. In addition, it can be observed that the payment for securely caching the most popular content with $\tau(m) = 1$ increases fastest compared with those for contents with lower popularity.

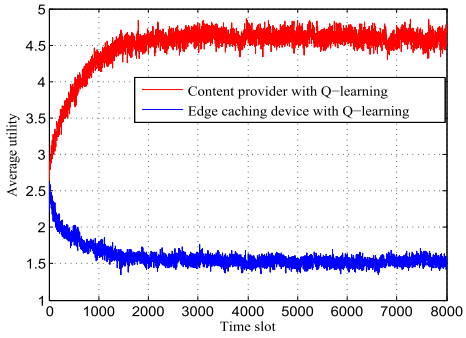


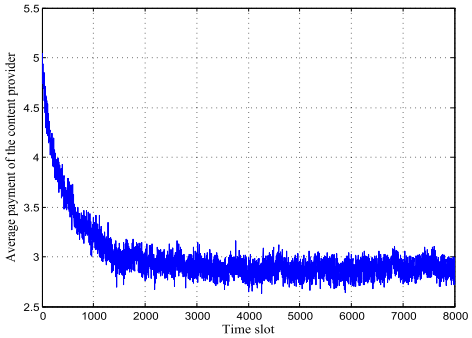
Fig. 5. Q-learning process for the content provider and edge caching devices.

Fig. 5 shows the Q-learning based strategies decision processes for the content provider and edge caching devices. In the simulation, we use one edge caching device to show the result, whose cost parameter is 0.5 and the number of mobile users is 50 within the edge caching device. Every content is requested by all mobile users under an edge caching device, i.e., $r_{i,m} = 1, \forall i \in \mathcal{I}, \forall m \in \mathcal{M}$. The importance of the content is uniformly distributed. The experiment is repeated with 1000 times. Other settings are unchanged. From Fig. 5, we can observe that both the utilities of the content provider and the edge caching device reach to be stable with some time slots, where the SE of the dynamic game is obtained. In addition, the average utility of the content provider increases, while the average utility of the edge caching device decreases over time. The reason is that the content provider controls the strategy decision of the edge caching device by the payment. If the payment is high, the edge caching device which wants to gain more profits will provide higher quality secure caching and reduce the utility.

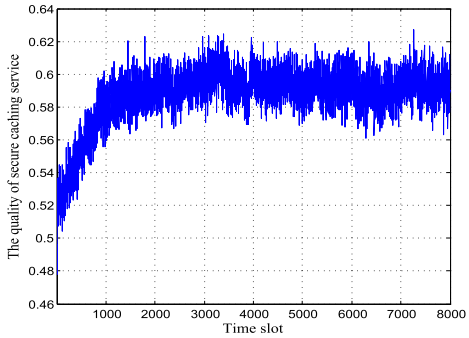
Fig. 6 is the evolution on payment and quality of secure caching service over time. From Fig. 6, it can be observed that the proposed scheme can efficiently stimulate edge caching device to provide high-quality secure caching service with low payment. Specifically, in Fig. 6 (a), the payment of the content provider decreases over the time. The reason is that the initial high payment has stimulated the secure caching service with high quality, where the content provider can reduce payment to improve utility. In Fig. 6 (b), according to the payment strategy of content provider, each edge caching device continuously seeks the optimal quality of secure caching service towards the maximum utility of edge caching device. The quality of secure caching service for each edge caching device first increases and then reaches to be stable.

Fig. 7 is the secure caching service ratio over time. In the simulation, three types of qualities on secure caching services are used, which are 0, 0.5, 1, respectively. From Fig. 7, we can observe that the ratio of high-quality secure caching ratio increases over time. Namely, the high-quality caching service is more provided to the content provider, while the low quality caching service is suppressed. For example, from time slot 0 to 1000, the ratio of secure caching service with $q = 1$ increases 50%, while that with $q = 0$ decreases 50%.

Fig. 8 is the comparison between the proposed scheme with other two schemes about the average quality of secure caching service. Here, the cost parameter changes from 0.5 to 1.3.



(a) Evolution on payment for secure caching service.



(b) Evolution on quality of secure caching service.

Fig. 6. Evolutions of payment and quality on secure caching service over time.

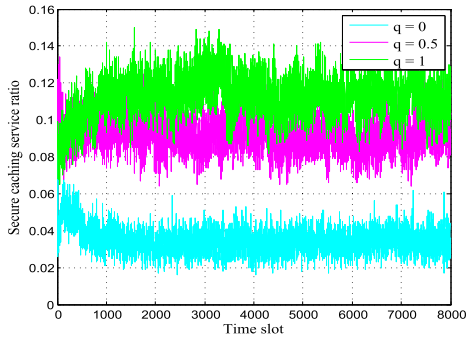


Fig. 7. Secure caching service ratio vs. time slot.

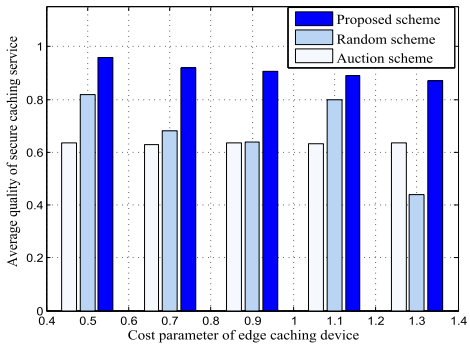


Fig. 8. The comparison of the proposed scheme with conventional schemes on the quality of secure caching service, when the cost parameter of edge caching device changes.

From Fig. 8, it can be observed that the proposed scheme outperforms other two schemes. The reason is that the quality of caching service is randomly selected in the random scheme. In the auction scheme, the serving edge caching device is selected based on the price, without consideration of the security quality. In addition, since there is no punishment, the selfish edge caching devices may cheat the content provider without providing caching services. In the proposed scheme, with the Q-learning for the dynamic game, the optimal quality of caching can be obtained.

Fig. 9 is the comparison between the proposed scheme and other two schemes about the utility of the content provider. Here, the number of edge caching devices changes from 10 to 18. From Fig. 9, it can be seen that the proposed scheme obtains the higher average utility of the content provider than that of the other two schemes. The reason is that the payment policy and quality strategy are randomly determined in random scheme, where the utility is not maximum. In the

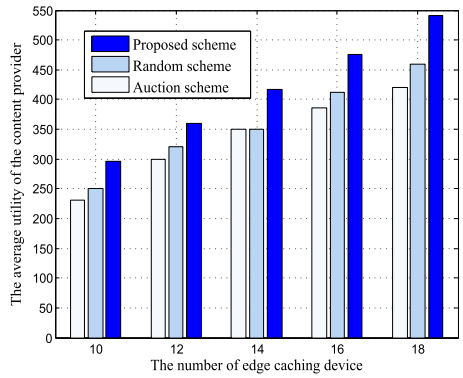


Fig. 9. The comparison of the proposed scheme with conventional schemes on the utility of the content provider, when the number of edge caching devices changes.

auction scheme, the selected edge caching device may be selfish to cheat the content provider, where the utility of the content provider may be negative. In the proposed scheme, with zero payment punishment, the selfishness of edge caching device can be avoided and the optimal strategies of the content provider and edge caching devices are determined by Q-learning with analyzing the dynamic Stackelberg game.

Fig. 10 shows the comparison of the proposed scheme with other two conventional schemes on the average utility of the content provider, where the number of mobile users within the coverage of each edge caching device changes from 50 to 90. From Fig. 10, we can see that the proposed scheme outperforms the other two schemes. In the random scheme, as the payment and qualities of caching services are randomly determined, the content provider cannot obtain the optimal utility. In the auction scheme, though the payment is small, the qualities of caching services are low, which makes a low utility of the content provider. In the proposed scheme, considering the number of mobile users that requires popular content, the optimal payment and qualities of secure caching services are achieved based on the game analysis.

Fig. 11 shows the comparison of the proposed scheme with the other two conventional schemes on the secure ratio of the cached content, where the cost parameter of each edge caching device changes from 0.5 to 1.3. The secure ratio of the cached content is the ratio of attacks that are successfully defended by edge caching devices to all attacks conducted during the simulation time. From Fig. 11, it can be observed

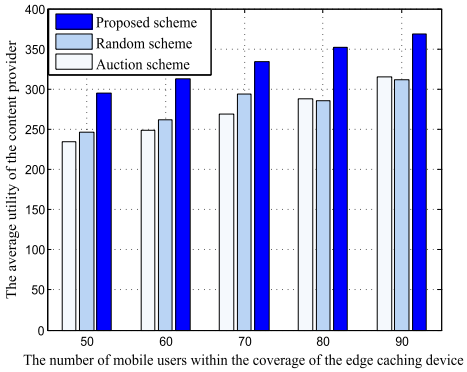


Fig. 10. The comparison of the proposed scheme with conventional schemes on the utility of the content provider, when the number of mobile users within the coverage of the edge caching devices changes.

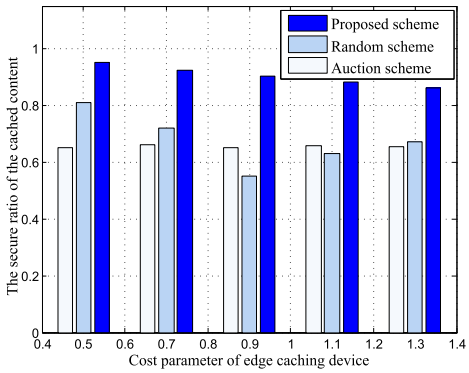


Fig. 11. The comparison of the proposed scheme with conventional schemes on the secure ratio of the cached content, when the cost parameter of edge caching device changes.

that the proposed scheme obtains the highest secure ratio among the three schemes. The reason can be explained as follows. In the random scheme, the quality of secure caching service is randomly determined by edge caching nodes, where the cached contents are easily compromised when the quality of the secure caching service is low. In the auction scheme, the edge caching device with the low caching price is selected as a winner, where the quality of secure caching service is also low. In the proposed scheme, the content provider jointly considers its demand and payment to motivate edge caching devices to provide high-quality caching services.

VIII. Conclusion

This paper has presented a game theory and reinforcement learning based secure edge caching scheme in MSNs. Specifically, to motivate the participation of edge caching devices, the interactions between the content provider and edge caching devices are formulated as a Stackelberg game to stimulate edge caching devices to provide high-quality caching services. The static Stackelberg game is first analyzed to achieve SE for the optimal strategies of both the content provider and edge caching devices, where all players in the game can obtain the maximum utilities. Especially, the zero payment mechanism is adopted to suppress the selfish behaviors of edge caching devices. Furthermore, the Q-learning, a typical type of reinforcement learning mode, is exploited to obtain the optimal strategy for each party in the dynamic network scenarios, without the awareness of the edge caching devices' secure caching model parameters and the content provider's payment model parameters. At last, the simulation results show that the proposed scheme can efficiently motivate edge caching devices to provide high-quality secure caching services and improve the content provider's utility. For the future work, the quality of secure caching service evaluation model and deep reinforcement learning method for accelerating learning process will be investigated.

References

[1] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, "An edge caching scheme to distribute content in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5346–5356, Jun. 2018, doi: 10.1109/TVT.2018.2824345.

[2] O. A. Wahab, J. Bentahar, H. Otrok, and A. Mourad, "Towards trustworthy multi-cloud services communities: A trust-based hedonic coalitional game," *IEEE Trans. Services Comput.*, vol. 11, no. 1, pp. 184–201, Jan. 2018.

[3] M. Chen, L. Wang, J. Chen, X. Wei, and L. Lei, "A computing and content delivery network in the smart city: Scenario, framework, and analysis," *IEEE Netw.*, vol. 33, no. 2, pp. 89–95, Mar. 2019.

[4] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecastupdate 2016–2021*, Mar. 2017.

[5] Z. Su, Y. Hui, and T. H. Luan, "Distributed task allocation to enable collaborative autonomous driving with network softwarization," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 10, pp. 2175–2189, Oct. 2018.

[6] Y. Yu, R. Chen, H. Li, Y. Li, and A. Tian, "Toward data security in edge intelligent IIoT," *IEEE Netw.*, vol. 33, no. 5, pp. 20–26, Sep. 2019.

[7] Z. Su, Q. Xu, and Q. Qi, "Big data in mobile social networks: A QoE-oriented framework," *IEEE Netw.*, vol. 30, no. 1, pp. 52–57, Jan. 2016.

[8] Z. Zhou, H. Yu, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 953–964, Mar. 2018.

[9] Z. Li, Y. Liu, A. Liu, S. Wang, and H. Liu, "Minimizing convergecast time and energy consumption in green Internet of Things," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[10] Y. Hui, Z. Su, and T. H. Luan, "Collaborative content delivery in software-defined heterogeneous vehicular networks," *IEEE/ACM Trans. Netw.*, to be published.

[11] S. Josilo and G. Dan, "Selfish decentralized computation offloading for mobile cloud computing in dense wireless networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 1, pp. 207–220, Jan. 2019.

[12] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, "Edge computing security: State of the art and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1608–1631, Aug. 2019.

[13] M. Dai, J. Li, Z. Su, W. Chen, Q. Xu, and S. Fu, "A privacy preservation based scheme for task assignment in Internet of Things," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[14] Z. Su, Q. Xu, J. Luo, H. Pu, Y. Peng, and R. Lu, "A secure content caching scheme for disaster backup in fog computing enabled mobile social networks," *IEEE Trans Ind. Informat.*, vol. 14, no. 10, pp. 4579–4589, Oct. 2018, doi: 10.1109/TII.2018.2849984.

[15] L. Lu *et al.*, "Pseudo trust: Zero-knowledge authentication in anonymous p2ps," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 10, pp. 1325–1337, Oct. 2008.

[16] Q. Xu, Z. Su, M. Dai, and S. Yu, "APIS: Privacy-preserving incentive for sensing task allocation in cloud and edge-cooperation mobile Internet of Things with SDN," *IEEE Internet Things J.*, to be published.

[17] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging SDN and NFV security mechanisms for IoT systems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 812–837, 1st Quart., 2019.

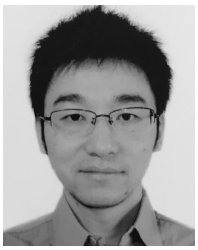
[18] G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and M. Rajarajan, "Scale inside-out: Rapid mitigation of cloud DDoS attacks," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 6, pp. 959–973, Nov. 2018.

[19] Q. Xu, Z. Su, Q. Zheng, M. Luo, and B. Dong, "Secure content delivery with edge nodes to save caching resources for mobile users in green cities," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2550–2559, Jun. 2018, doi: 10.1109/TII.2017.2787201.

[20] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Unravelling the impact of temporal and geographical locality in content caching systems," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1839–1854, Oct. 2015.

[21] M. N. Soorki, W. Saad, M. H. Manshaei, and H. Saidi, "Social community-aware content placement in wireless Device-to-Device communication networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 8, pp. 1938–1950, Aug. 2019.
[22] Q. Xu, Z. Su, and Q. Yang, "Blockchain-based trustworthy edge caching scheme for mobile cyber-physical system," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1098–1110, Feb. 2020.
[23] J. Dai, F. Liu, B. Li, B. Li, and J. Liu, "Collaborative caching in wireless video streaming through resource auctions," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 458–466, Feb. 2012.
[24] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2115–2129, Aug. 2016.
[25] L. Ma, X. Liu, Q. Pei, and Y. Xiang, "Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing," *IEEE Trans. Services Comput.*, vol. 12, no. 5, pp. 786–799, Sep. 2019.
[26] Z. Wang, Z. Ma, S. Luo, and H. Gao, "Enhanced instant message security and privacy protection scheme for mobile social network systems," *IEEE Access*, vol. 6, pp. 13706–13715, 2018.
[27] E. K. Wang, Y. Li, Y. Ye, S. M. Yiu, and L. C. K. Hui, "A dynamic trust framework for opportunistic mobile social networks," *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 1, pp. 319–329, Mar. 2018.
[28] Z. Li, C. Wang, S. Yang, C. Jiang, and I. Stojmenovic, "Space-crossing: Community-based data forwarding in mobile social networks under the hybrid communication architecture," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4720–4727, Sep. 2015.
[29] F. M. Awuor, C.-Y. Wang, and T.-C. Tsai, "Motivating content sharing and trustworthiness in mobile social networks," *IEEE Access*, vol. 6, pp. 28339–28355, 2018.
[30] Y. Meng, C. Jiang, T. Q. S. Quek, Z. Han, and Y. Ren, "Social learning based inference for crowdsensing in mobile social networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 8, pp. 1966–1979, Aug. 2018.
[31] W. Jiang, G. Feng, S. Qin, and T. S. P. Yum, "Efficient D2D content caching using multi-agent reinforcement learning," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 511–516.
[32] L. Chhangte, A. Garg, D. Manjunath, and N. Karamchandani, "Wicache: Towards an SDN based distributed content caching system in WLAN," in *Proc. 10th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2018, pp. 503–506.
[33] K. Xu, X. Li, S. K. Bose, and G. Shen, "Joint replica server placement, content caching, and request load assignment in content delivery networks," *IEEE Access*, vol. 6, pp. 17968–17981, 2018.
[34] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, and M. Song, "Joint computation offloading and content caching for wireless blockchain networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 517–522.
[35] Q. Yang and D. Gunduz, "Coded caching and content delivery with heterogeneous distortion requirements," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4347–4364, Jun. 2018.
[36] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani, "Security in mobile edge caching with reinforcement learning," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 116–122, Jun. 2018.
[37] X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, "Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach," *IEEE Access*, vol. 6, pp. 25463–25473, 2018.
[38] C. Shen, Y. Gonzalez, L. Chen, S. B. Jiang, and X. Jia, "Intelligent parameter tuning in optimization-based iterative CT reconstruction via deep reinforcement learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1430–1439, Jun. 2018.
[39] H. Liu, S. Liu, and K. Zheng, "A reinforcement learning-based resource allocation scheme for cloud robotics," *IEEE Access*, vol. 6, pp. 17215–17222, 2018.
[40] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018.
[41] A. Araldo, G. Dan, and D. Rossi, "Caching encrypted content via stochastic cache partitioning," *IEEE/ACM Trans. Netw.*, vol. 26, no. 1, pp. 548–561, Feb. 2018.
[42] R. T. Tiburski *et al.*, "Lightweight security architecture based on embedded virtualization and trust mechanisms for IoT edge devices," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 67–73, Feb. 2019.
[43] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin, and W. Zhen, "A blockchain-based trusted data management scheme in edge computing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2013–2021, Mar. 2020.
[44] T. Dbouk, A. Mourad, H. Otrok, H. Tout, and C. Talhi, "A novel ad-hoc mobile edge cloud offering security services through intelligent resource-aware offloading," *IEEE Trans. Netw. Service Manage.*, vol. 16, no. 4, pp. 1665–1680, Dec. 2019.
[45] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "FCSS: Fog-Computing-based content-aware filtering for security services in information-centric social networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 7, no. 4, pp. 553–564, Oct. 2019.
[46] H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, and Z. Han, "Computing resource allocation in three-tier IoT fog networks: A joint optimization approach combining Stackelberg game and matching," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1204–1215, Oct. 2017.
[47] Y. Li, L. Xiao, H. Dai, and H. V. Poor, "Game theoretic study of protecting MIMO transmissions against smart attacks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
[48] Z. Su, Q. Qi, Q. Xu, S. Guo, and X. Wang, "Incentive scheme for cyber physical social systems based on user behaviors," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 1, pp. 92–103, Jan. 2020.

Qichao Xu received the Ph.D. degree from the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China, in 2019. He is currently an Assistant Professor with the School of Mechatronic Engineering and Automation, Shanghai University. His research interests are in trust and security, general area of wireless network architecture, the Internet of Things, vehicular networks, and resource allocation.



Zhou Su received the Ph.D. degree from Waseda University, Tokyo, Japan, in 2003. His research interests include multimedia communications, wireless communications, and network traffic. He is a TPC Member of some flagship conferences, including the IEEE INFOCOM, the IEEE ICC, and the IEEE Globecom. He received the Best Paper Award of the International Conference IEEE BigdataSE2019, the IEEE ComSoc GCCTC2018, the IEEE CyberSciTech2017, and the Funai Information Technology Award for Young Researchers, in 2009. He is the Chair of the Multimedia Services and Applications over Emerging Networks Interest Group (MENIG), the IEEE ComSoc Society, and the Multimedia Communications Technical Committee. He also served as the Co-Chair for several international conferences, including the IEEE Vehicular Technology Conference (VTC) Spring 2016 and the IEEE Consumer Communications and Networking Conference (CCNC) 2011. He is an Associate Editor of the IEEE OPEN JOURNAL OF COMPUTER SOCIETY and *IET Communications*.

Rongxing Lu received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Canada, in 2012. From May 2012 to April 2013, he worked as a Post-Doctoral Fellow with the University of Waterloo. From April 2013 to August 2016, he worked as an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. He is currently an Associate Professor with the Faculty of Computer Science (FCS), University of New Brunswick (UNB), Canada. He has published extensively in his areas of expertise. His research interests include applied cryptography, privacy enhancing technologies, and the IoT-big data security and privacy. He is currently a Senior Member of the IEEE Communications Society. He was a recipient of the eight best (student) paper awards from some reputable journals and conferences. He received the most prestigious Governor General's Gold Medal, in 2012, the 8th IEEE Communications Society (ComSoc) Asia–Pacific (AP) Outstanding Young Researcher Award, in 2013, and the 2016-17 Excellence in Teaching Award, FCS, UNB. He currently serves as the Vice-Chair (Publication) for the IEEE ComSoc Communications and Information Security Technical Committee (CIS-TC).