How to Mitigate DDoS Intelligently in SD-IoV: A Moving Target Defense Approach

Tao Zhang, *Graduate Student Member, IEEE*, Changqiao Xu, *Senior Member, IEEE*, Ping Zou, Haijiang Tian, Xiaohui Kuang, Shujie Yang, Lujie Zhong, and Dusit Niyato, *Fellow, IEEE*

***Abstract*—Software defined Internet of Vehicles (SD-IoV) is an emerging paradigm for accomplishing Industrial Internet of Things (IIoT). Unfortunately, SD-IoV still faces security challenges. Traditional solutions respond after attacks happening, which is low-effective. To cope with this problem, moving target defense (MTD) was proposed to modify network configurations dynamically. However, current MTD for IIoT has several drawbacks: 1) it cannot handle highly dynamic environments; 2) MTD strategy lacks intelligence because it needs attack–defense models; 3) they are difficult to trace sources. In this article, we propose an intelligent MTD scheme to defend against distributed denial-of-service in SD-IoV. Firstly, we model the configuration mutation of roadside units as a Markov decision process (MDP), and adopt deep reinforcement learning to solve the optimal configuration. Next, we evaluate the trust of vehicles after shuffling, which can distinguish spy vehicles. Finally, extensive simulation results confirm the effectiveness of our solution compared with representative methods.**

***Index Terms*—Deep reinforcement learning (DRL), distributed denial-of-service (DDoS), moving target defense (MTD), software defined Internet of Vehicles (SD-IoV), trust assessment.**

Manuscript received 2 January 2022; revised 20 March 2022 and 23 May 2022; accepted 5 July 2022. Date of publication 13 July 2022; date of current version 8 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 61871048 and 61872253, in part by the National Research Foundation (NRF) and Infocomm Media Development Authority under Grant FCP-NTU-RG-2021-014, in part by the AI Singapore Programme (AISG) under Grant AISG2-RP-2020-019, in part the by the Singapore Ministry of Education (MOE) Tier 1 under Grant RG16/20, and in part by the BUPT Excellent Ph.D. Students Foundation under Grant CX2020123. Paper no. TII-22-0018. (*Corresponding author: Changqiao Xu.*)

Tao Zhang, Changqiao Xu, Ping Zou, Haijiang Tian, and Shujie Yang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: zhangtao17@bupt.edu.cn; cqxu@bupt.edu.cn; pingzoubupt@163.com; hjtian2000@163.com; sjyang@bupt.edu.cn).

Xiaohui Kuang is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the National Key Laboratory of Science and Technology on Information System Security, Beijing 100101, China (e-mail: xiaohui_kuang@163.com).

Lujie Zhong is with the Information Engineering College, Capital Normal University, Beijing 100048, China (e-mail: zhonglj@cnu.edu.cn).

Dusit Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dniyato@ntu.edu.sg).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3190556.

Digital Object Identifier 10.1109/TII.2022.3190556

I. Introduction

With the continuous and rapid development of vehicular ad hoc networks (VANETs), an emerging network paradigm called Internet of Vehicle (IoV) [1] has been proposed recently. As a critical component to realize Industrial Internet of Things (IIoT), IoV will equip vehicles with various sensors and communication modules [2]. Therefore, IoV can enable a large number of concurrent communication connections between vehicles and roadside units (RSUs) as well as among vehicles, referred as vehicle-to-infrastructure and vehicle-to-vehicle. At this circumstance, it is necessary for IoV to adopt a flexible network management. Fortunately, software defined networking (SDN) has gained lots of research interests due to its flexibility and programmability by decoupling the control plane and data plane. Through the seamless integration of SDN and IoV, software defined IoV (SD-IoV) fully inherits the advantages of SDN [3], [4]. In SD-IoV, SDN controllers, which are located at base stations (BSs), will make decisions for vehicular communications by aggregating network information. Actually, whatever in the scenarios of traditional VANETs or emerging SD-IoV, how to protect the security of vehicular system is still an enormous challenge [5]–[7]. For example, if RSUs are compromised by distributed denial-of-service (DDoS) attacks, the entire vehicular network will be possible to collapse and generate incorrect computing results. Many works [8]–[16] have been proposed to mitigate DDoS attacks in VANETs or SD-IoV. However, most of existing solutions are static, i.e., response after attacks happen. Therefore, the adversary can exploit vulnerabilities of vehicular system through enough reconnaissance efforts, which makes current countermeasures low efficiency. What is worse, sophisticated DDoS attacks have become more and more stealthy and persistent in recent years, which causes the difficulty for attack detection.

To cope with aforementioned challenges, moving target defense (MTD) [17]–[21] has emerged as a promising solution, which dynamically modifies network configurations. Compared with traditional security solutions, MTD introduces uncertainty and unpredictability to invalidate the adversary's prior knowledge, thus reducing the success probability of cyber attacks significantly. Existing MTD works have been validated as a suitable security paradigm for IoT [22]–[24] and defending against DDoS [19], [20], [25], [26]. Unfortunately, there are several drawbacks in these MTD approaches. Firstly, current MTD solutions cannot handle highly dynamic environments because MTD for IoT considers the fixed network topology and MTD for DDoS is designed for online web applications. Therefore, highly dynamic wireless environment in SD-IoV produces great challenges for MTD design. Secondly, MTD strategy lacks intelligence because they depend on attack–defense models. Lastly, current methods are difficult to trace the sources of DDoS early, and identify spies accurately. Inspired by MTD, how to combine the inherent dynamics of mobile network (e.g., SD-IoV) with a novel security management needs in-depth study [27].

In this article, we propose an intelligent MTD scheme to defend against DDoS in SD-IoV. The key insight is to mutate the network configurations of RSUs periodically based on deep reinforcement learning (DRL), and evaluate the trust of vehicles after shuffling RSU-vehicle associations dynamically. Recently, DRL has been successfully used to tackle decision-making problems in highly dynamic vehicular network [28]. Benefitted from DRL, our intelligent MTD scheme can reduce the number of innocent vehicles affected by DDoS while separating spy vehicles from innocent vehicles so as to prevent DDoS from the source. To the best of our knowledge, this is the first work that designs an intelligent MTD scheme in dynamic and heterogeneous SD-IoV.

To sum up, the main contributions of this article are summarized as follows.

1) We design an intelligent configuration mutation mechanism. The mutation of network configurations, including the communication ranges and capacities of RSUs, are modeled as a Markov decision process (MDP). Then, how to generate the optimal configurations of RSUs will be transformed into an optimization problem. Based on the thorough analysis, we adopt DRL to solve optimal configuration mutation.
2) To identify spy vehicles and innocent vehicles accurately, we propose to evaluate the trust of vehicles periodically after shuffling RSU-vehicle associations. Considering the high dynamics in SD-IoV, we formalize multiple network constraints for shuffling based on satisfiability modulo theory (SMT) [29], which consists of reachability, accessibility, unpredictability, and capacity constraints. Then, we design a trust assessment algorithm for RSU-vehicle associations.
3) To confirm the effectiveness of our intelligent MTD scheme, we conduct extensive simulations based on the network simulator NS-3. Simulation results confirm that our proposed intelligent MTD scheme outperforms the representative solutions that can be adapted and applied to the scenario of SD-IoV.

The rest of this article is organized as follows. Section II explains the related work. In Section III, network model and threat model are explained. In Section IV, we introduce the system overview. Our proposed intelligent mutation mechanism is introduced in Section V. In Section VI, trust assessment mechanism is proposed. Section VII shows the simulation results. Finally, Section VIII concludes this article.

II. Related Work

Whatever are the scenarios of traditional VANETs or emerging SD-IoV, the security of vehicular network is still an important issue that must be considered [5], [6]. Many methods have been proposed to solve the aforementioned security issue. For example, Biron *et al.* The authors in reference[8] proposed a real-time DDoS detection scheme that estimates the effect of cyber attacks. The work by Mejri *et al.* [9] proposed to detect greedy behaviors in highly mobile network. This approach consists of suspicion phase and decision phase based on linear regression and fuzzy logic. These works are just considered in VANETs, thus may not be appropriate for SD-IoV. Specially in SD-IoV, Biasi *et al.* [10] proposed to detect DDoS by time series analysis of packets and mitigate DDoS by searching for the source of spoofed packets. On the other hand, some works focus on trust evaluation to improve the security of vehicular network [11]. Xia *et al.* [12] proposed a trust inference model to quantify the trust levels of vehicles, which combine subjective and recommendation trust. The work by Najafi *et al.* [13] proposed a prediction and reputation method in decentralized manner. To defend against cyber attacks from anomalous nodes, Nigam *et al.* [14] proposed an intelligent trust-based routing protocol based on multiobjective optimization. However, all these approaches are static and only respond after attacks happen, so that the adversary can eventually exploit vulnerabilities with enough reconnaissance efforts.

Fortunately, MTD has emerged as a promising solution and gained ever-growing attention [17], [18]. The core idea of MTD is to dynamically modify the components or configurations of network systems, which will introduce uncertainty and unpredictability to confuse the adversary. To ensure security in IoT environments, Nizzi *et al.* [22] proposed an address shuffling mechanism with limited network overhead. Duan *et al.* [23] proposed a random range mutation that allows for changing the coverage ranges of access points (APs) periodically and randomly. Based on SDN, the work by Ge *et al.* [24] proposed two proactive defense mechanisms that reconfigure the network topology. Besides considering security in IoT, other MTD works focused on specific cyber attacks, e.g., DDoS. Lin *et al.* [25] proposed a cost-effective approximation algorithm to mitigate application layer DDoS attacks with guaranteed performance. The work by Zhou *et al.* [26] proposed to dynamically control the admission of devices and migrate service replicas to mitigate DDoS attacks early near sources. Our previous work [19]–[21] proposed an intelligent route mutation scheme that optimizes the mutation selections, which defends against DDoS attacks effectively. Because SD-IoV is highly dynamic and depends on wireless communication paradigm, whatever MTD methods for IoT or DDoS cannot be adopted directly.

SD-IoV is regarded as a promising paradigm for implementing future IoV-based IIoT. It has been proved that security will still be the main challenge in SD-IoV. Fortunately, MTD has emerged as a game-changing way to defend against various cyber attacks, e.g., DDoS. Nonetheless, existing MTD approaches are not suitable in SD-IoV because of highly dynamic and complex wireless environments. Therefore, it is meaningful to design a novel MTD scheme in SD-IoV. To the best of our knowledge, our work is the first contribution that designs an intelligent MTD to mitigate DDoS in SD-IoV, and further evaluates the trust of vehicles to distinguish spy vehicles.

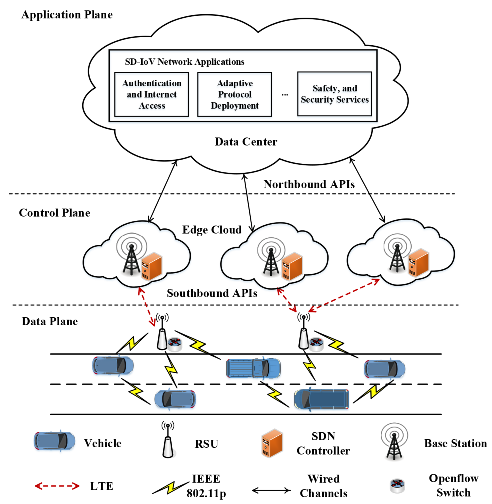


Fig. 1. System architecture of SD-IoV.

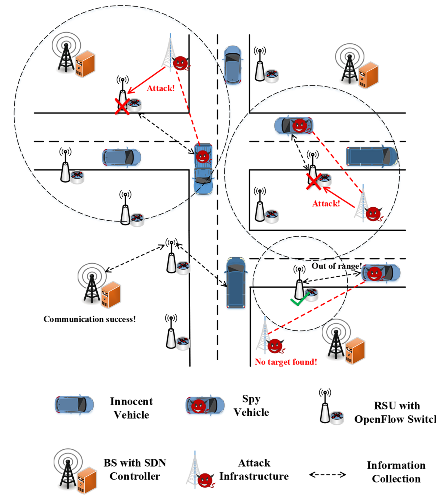


Fig. 2. Example of attack scenario.

III. Model Formulation

In this section, we introduce network model and threat model.

A. Network Model

Fig. 1 illustrates the system architecture of SD-IoV. The wireless data plane consists of RSUs and vehicles, which correspond to clients. RSUs, where OpenFlow switches are located at, also act as access points for vehicles. The control plane consists of BSs, where SDN controllers are located at. Based on southbound application programming interface (API), control plane has ability to communicate with data plane. Apart from control plane and data plane, there also exists the application plane, where network functions (e.g., Internet access, security services, and so on) are implemented in the data center. Northbound API is used to decompose network functions into lower-level controller functions.

We consider an SD-IoV scenario with n RSUs, m vehicles, and k BSs. The SD-IoV is modeled as a undirected graph $\mathbb{G} = (\mathcal{N}, \mathcal{M}, \mathcal{K}, \mathcal{E})$ as follows.

1) $\mathcal{N}$ is the set of RSU nodes n_i $(1 \leq i \leq n)$.
2) $\mathcal{M}$ is the set of vehicle nodes v_j $(1 \leq j \leq m)$.
3) $\mathcal{K}$ is the set of BS nodes b_x $(1 \leq x \leq k)$.
4) $\mathcal{E}$ is the set of wireless links e connecting different nodes.

Within the area of interest, vehicles move according to Manhattan mobility model [30]. In actual, our proposed scheme can work well in other mobility models because its working mechanism does not depend on specific mobility model. BSs and RSUs are located uniformly in the selected area.

B. Threat Model

In this article, we consider a complex attack scenario in SD-IoV, where the adversary has two kinds of malicious identities: 1) the spy vehicle that connects with RSUs, and further collects information (e.g., IP addresses, geographical locations, and so on) without any malicious behaviors; 2) the attacking infrastructure that launches bandwidth-based or volumetric DDoS to targeted RSUs based on information collected by spy vehicles. Finally, all vehicles that connect with targeted RSUs will be

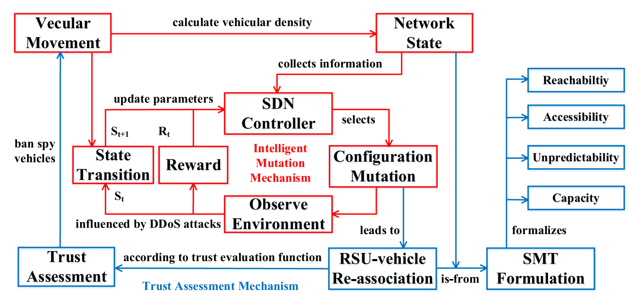


Fig. 3. Framework of our proposed MTD scheme.

affected, which leads to their network services interrupted. This attack scenario is named spy-enabled DDoS.

An adversary can deploy a number of attacking infrastructures in the selected area, which have more transmission power than that of vehicles and RSUs. Vehicles will be spies if they belong to the adversary or are compromised by the adversary. Spy vehicles communicate with attacking infrastructures periodically to upload collected network information. RSUs are considered as attacked targets because they are more vulnerable to hacking than expensive and large BSs [23]. Fig. 2 depicts an example of DDoS attacks launched by spy vehicles and attacking infrastructures. In this article, we do not consider security problems about control plane because some works [31]–[33] have focused on them.

IV. System Overview

To defend against spy-enabled DDoS effectively, we propose an intelligent MTD scheme. Unlike other security solutions in SD-IoV [10], our method can reduce the DDoS damage efficiently, and further distinguish spy vehicles based on trust assessment. Fig. 3 depicts the framework of our proposed MTD scheme, where the red part is the workflow of intelligent mutation mechanism and blue part is the workflow of trust assessment mechanism. Intelligent mutation mechanisms adjust the network configurations of RSUs based on DRL, and trust assessment mechanism evaluates the trust of vehicles after association shuffling. The detailed working mechanisms of them are explained in Sections IV-A and IV-B, respectively.

A. Intelligent Mutation Mechanism

Intelligent mutation mechanism is to deploy the optimal configuration of RSUs by DRL, which aims to reduce the number of vehicles affected by DDoS attacks. As shown in Fig. 3, SDN controller collects information about vehicular traffic density, which is regarded as network state. Then, it selects the mutated configurations of RSUs, and observes the state transition and reward. Lastly, the parameters of neural networks are updated. With enough number of iterations, SDN controller will obtain the optimal configurations of RSUs. We named the aforementioned intelligent mutation mechanism as proximal policy optimization algorithm for optimal configuration mutation (PPO-OCM). This part is illustrated in Section V in detail.

B. Trust Assessment Mechanism

Trust assessment mechanism is to distinguish spies by evaluating the trust of vehicles after shuffling. It should be noted that spy vehicles connect more frequently to RSUs that were attacked than innocent vehicles because of their own nature. Over time, the trust of innocent and spy vehicles will diverge, thus revealing spy vehicles. When configurations are mutated, vehicles will reassociated with RSUs by considering multiple constraints, which is modeled as a constrained satisfaction problem based on SMT. By using the Z3 theorem prover [34], feasible RSU-vehicle reassociations can be generated. After selecting an association between vehicles and RSUs, the trust of vehicles will be evaluated. Lastly, spy vehicles will be banned from connecting to any RSUs. The aforementioned workflow is called trust assessment algorithm for RSU-vehicle associations (TA-RVA). More details are described in Section VI.

V. Intelligent Mutation Mechanism

In this section, we firstly formulate the process of configuration mutation as an MDP. Then, how to find the optimal configuration of RSUs will be transformed to an optimization problem. Finally, we proposed a PPO-OCM to solve this optimization problem.

A. MDP Model

Time is divided into equal slots, whose basic length is ΔT, then time is slotted with the index $t \in \{0, 1, 2, \cdots\}$. In this subsection, to capture the dynamics caused by vehicular movements, we model the configuration mutation as a MDP. The key features of MDP are shown as follows.

1) State Space: A geographical area is divided into uniform-size squares called grids, and each grid has different vehicular densities. Assuming that the selected area has h grids in total, the vector of vehicular densities is regarded as network state $\mathcal{S}_t = \{s_1, s_2, \ldots, s_h\}$, where s_z $(1 \leq z \leq h)$ denotes the number of vehicles in gird z. The selected area is supposed to have m vehicles, the size of state space is $\binom{m+h-1}{h-1}$, which grows with the increasing number of vehicles. To address the scalability issue, we utilize the DRL, which is presented in Section III-B.

2) Action Space: Network configurations including the communication ranges and access capacities of RSUs are mutated periodically. Selecting a joint network configuration of all RSUs is considered as an action denoted as $\mathcal{A}_t = [\tilde{R}_t^{\text{rsu}}, \tilde{Q}_t^{\text{rsu}}]$, where $\tilde{R}_t^{\text{rsu}} = \{R_{t,1}^{\text{rsu}}, R_{t,2}^{\text{rsu}}, \ldots, R_{t,n}^{\text{rsu}}\}$ denotes the communication ranges of RSUs, and $\tilde{Q}_t^{\text{rsu}} = \{Q_{t,1}^{\text{rsu}}, Q_{t,2}^{\text{rsu}}, \ldots, Q_{t,n}^{\text{rsu}}\}$ denotes the access capacities of RSUs (the maximum numbers of vehicles that can connect to them). The range is $\alpha_1, \ldots, \alpha_k$ and the capacity is $\beta_1, \ldots, \beta_l$, which can be represented by natural numbers. When the communication range and access capacity of an RSU becomes larger, more vehicles can connect to it. However, on this condition, RSU will have more energy consumption, and if it is attacked by DDoS successfully, more vehicles will be affected. When the communication range and access capacity of an RSU become smaller, contrary conclusion will be done. The size of action space is $k^n l^n$, which will grow exponentially with the increasing number of RSUs. The detail that how to obtain the scalability is also presented in Section III-B.

3) State Transitions: Vehicles travel through multiple grids when they move. For a vehicle, switching from one grid to another grid at each time slot is considered as a state transition.

4) Reward Function: When a mutated configuration is selected at time slot t, we define the reward $\mathcal{R}_t$ with three factors: 1) service quantity; 2) energy consumption; 3) security situation; as $\mathcal{R}_t = \alpha\mathcal{R}_t^n - \beta\mathcal{R}_t^e - \zeta\mathcal{R}_t^s$, where α, β, and ζ are all coefficients. Thereinto, the reward of service quantity is defined by $\mathcal{R}_t^n = \sum_{i\in\mathcal{N}} \mathbb{N}_{t,i}$, where $\mathbb{N}_{t,i}$ denotes the number of vehicles that connect to RSU n_i at time slot t, and satisfies $\mathbb{N}_{t,i} \leq Q_{t,i}^{\text{rsu}}$. The reward of energy consumption is defined by $\mathcal{R}_t^e = \sum_{i\in\mathcal{N}} \mathbb{E}(R_{t,i})$, where $\mathbb{E}(R_{t,i})$ is the energy consumption when communication range is $R_{t,i}$ at time slot t. According to [35], $\mathbb{E}(R_{t,i})$ is computed as follows:

$$\mathbb{E}(R_{t,i}) = \begin{cases} \sum_{\mathbb{N}_{t,i}} \Phi_t * \left(E_{\text{elec}} + \epsilon_f R_{t,i}^2\right), & R_{t,i} \leq R_d \\ \sum_{\mathbb{N}_{t,i}} \Phi_t * \left(E_{\text{elec}} + \epsilon_t R_{t,i}^4\right), & R_{t,i} \geq R_d \end{cases}$$

where Φ_t denotes the amount of transmitted data from RSU to each connected vehicle at time slot t, E_{elec}, ϵ_f, and ϵ_t are all constants, R_d is the threshold distance usually set as 75 m. The reward of security situation is defined by $\mathcal{R}_t^s = \sum_{i\in\mathcal{N}} \mathbf{Y}_{t,i}\mathbb{N}_{t,i}$, where $\mathbf{Y}_{t,i}$ is an indicator function that denotes whether RSU n_i is compromised successfully by DDoS, if so, its value is assigned as 1, otherwise its value is assigned as 0.

The objective of configuration mutation is to choose the optimal configuration of RSUs, which equals to how to maximize the cumulative reward obtained from the environment. Therefore, we formulate the optimization problem of configuration mutation as follows:

$$\mathbf{P1}: \max_{\pi} \mathbf{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R}_{t+k}\right] \tag{1}$$

where π is an policy that chooses the action of mutated configurations, $\mathbf{E}$ is the expectation operator, and γ is a discount factor between 0 and 1.

B. DRL for Optimal Configuration Mutation

Theoretically, optimization problem P1 can be solved by dynamic programming or exhaustive search. However, because of frequent vehicular movements, it is impossible to trace state transition probabilities mathematically, which must be used for computation. In recent years, reinforcement learning (RL) has shown its advantage of obtaining the optimal policy in MDP, thus becoming an promising method for optimization problem. Next, we analyze the rationality of using RL in SD-IoV. Because vehicles are driven by humans nowadays, they likely frequent several places, e.g., home, office, and so on. It can be seen that vehicle mobility exhibits temporal locality, which is similar with human mobility [37]. Based on the conclusion in [38], the number of vehicles inside each grid, i.e., vehicular density, is relatively stable in different days. The historical trajectories of vehicles will have very strong regularity. Usually, the behaviors of spy vehicles including mobility are the same as innocent vehicles because they want to hide themselves as much as possible. Therefore, in the training process of RL, environment will not influence the convergence negatively. According to the observation and analysis, RL can learn the optimal configuration of RSUs by interacting with environment in a certain period of time, then optimal configurations will be deployed later in practice.

Based on the definitions of MDP, the size of state and action spaces will be very large with the increasing number of vehicles and/or RSUs. Therefore, it is impractical to use traditional RL algorithm such as Q-learning. Fortunately, DRL [36] approximates policy and/or value function by deep neural networks (DNNs). With DNNs, large state and/or action spaces will be represented powerfully. Existing DRL algorithms are usually classified into value-based and policy-based algorithms. Value-based DRL algorithms approximate value function by DNNs, which aim to handle the large state space. For example, at each time slot, deep Q-network (DQN) minimizes the loss function, which is defined as $\mathcal{L} = \boldsymbol{E}[(y_t - \mathcal{Q}_t(\mathcal{S}_t, \mathcal{A}_t, \theta))^2]$, where $\boldsymbol{E}$ is the expectation operator, $\mathcal{Q}_t$ is the state-action value at time slot t, and θ is a parameter. Target value y_t is expressed as $y_t = \mathcal{R}_t + \max_{\mathcal{A}'} \mathcal{Q}_t(\mathcal{S}_{t+1}, \mathcal{A}', \theta_{\text{old}})$, where $\mathcal{A}'$ is an arbitrary action and θ_{old} is the value of parameter θ before N time slots. However, value-based DRL algorithms cannot handle the large action space. To address this problem, policy-based DRL algorithms approximate the parameterized policy by DNNs. In policy-based DRL, policy gradient is computed as $\nabla\mathcal{L} = \boldsymbol{E}_t[\nabla_\theta \log \pi(\mathcal{A}_t \mid \mathcal{S}_t; \theta)\hat{A}_t]$, where π is a policy and $\hat{A}_t$ is an estimator function at time slot t. Recently, the state-of-the-art DRL algorithm called PPO [39] was proposed, whose objective function is $\mathcal{L}_{\text{clip}} = \boldsymbol{E}_t[\min(r_t, \text{clip}(r_t, 1-\epsilon, 1+\epsilon))\hat{A}_t]$, where ϵ is a hyperparameter to control the clip range and r_t is the policy probability ratio defined as $r_t = \frac{\pi(\mathcal{A}_t|\mathcal{S}_t;\theta)}{\pi(\mathcal{A}_t|\mathcal{S}_t;\theta_{\text{old}})}$. The clip function $clip(r_t, 1-\epsilon, 1+\epsilon)$ aims to constrain the value of r_t, which avoids moving outside the interval $(1-\epsilon, 1+\epsilon)$.

In this article, we adopt proximal policy optimization (PPO) to solve the optimization problem P1. The pseudo-code of PPO-OCM is shown in Algorithm 1. Parameters, replay buffer, and DNNs are initialized firstly (lines 1–5). Line 6 starts the main loop of our algorithm, which is divided into two main parts: 1) generate samples by interacting with the environment (lines 7–18). The iteration starts from an initial state until finishing T time slots, which is called an episode (lines 7 and 8). On each episode, SDN controllers will run policy $\pi_{\theta_{\text{old}}}$ to select an action

Algorithm 1: PPO-OCM.

1: Set parameters ξ, ϵ, and γ.
2: Set batch size T and minibatch size K.
3: Initialize the experience replay buffer $\mathcal{B} = \varnothing$.
4: Randomly initialize Critic network $V(\mathcal{S}, \phi)$.
5: Randomly initialize Actor network π_θ with weight θ.
6: **for** $iteration = 1, 2, \cdots$ **do**
7:   **for** $episode = 1, 2, \ldots, K$ **do**
8:     **for** $t = 0, 1, \ldots, T-1$ **do**
9:       Obtain the current network state $\mathcal{S}_t$.
10:       Run policy $\pi_{\theta_{old}}$ to select action $\mathcal{A}_t$.
11:       Execute the configuration mutation for RSUs.
12:       Observe the outcome reward $\mathcal{R}_t$.
13:       Obtain the next network state $\mathcal{S}_{t+1}$.
14:       Collect $\mathcal{U}_t = (\mathcal{S}_t, \mathcal{A}_t, \mathcal{R}_t, \mathcal{S}_{t+1})$, and $\mathcal{B} \cup \mathcal{U}_t$.
15:       Calculate δ_t and $\hat{A}_t = \sum_{q \geq t}^{K} (\gamma\xi)^{q-t}\delta_q$.
16:       Estimate $\hat{V}_t = \hat{A}_i + V(\mathcal{S}_t, \phi)$.
17:     **end for**
18:   **end for**
19:   **for** $epoch = 1, 2, \ldots, U$ **do**
20:     $\mathcal{J}_a = \dfrac{1}{T}\sum_{i=1}^{T} \min(r_i, clip(r_i, 1-\epsilon, 1+\epsilon))\hat{A}_i$.
21:     Update θ by $\nabla_\theta \mathcal{J}_a$.
22:     $\mathcal{J}_c = -\frac{1}{T}\sum_{i=1}^{T}(\hat{V}_i - V(\mathcal{S}_i, \phi))^2$.
23:     Update ϕ by $\nabla_\phi \mathcal{J}_c$.
24:   **end for**
25:   $\pi_{\theta_{old}} \leftarrow \pi_\theta$.
26: **end for**

that mutates configurations of RSUs, and then observing the reward (lines 9–12). The sample of state transition is stored in the replay buffer, and then advantage function and value functions are estimated by generalized advantage estimation method (lines 13–16); 2) the second part is to learn from samples in the replay buffer. Policy gradient and value function gradient are calculated respectively, and corresponding parameters are updated (lines 19–24). Finally, the selection policy will also be updated (line 25).

VI. Trust Assessment Mechanism

PPO-OCM reduces the effects from DDoS attacks by mutating the configurations of RSUs intelligently, which will cause reassociations between vehicles and RSUs. By utilizing this characteristic, spy vehicles can be distinguished from innocent vehicles by trust assessment in the shuffling process, so as to prevent DDoS from the sources. In this section, we introduce how to reassign RSU-vehicle associations, and formulate the trust of vehicles in each shuffle. Since SD-IoV is dynamic wireless environment, we firstly formalize multiple network constraints based on SMT. Then, we design a trust assessment algorithm.

A. SMT Formalizations for Shuffling Constraints

In this subsection, shuffling RSU-vehicle associations are formalized as a constrained satisfaction problem. Let boolean variable $f_{t,j}^i$ denote whether vehicle v_j connects to RSU n_i at time slot t. If so, $f_{t,j}^i$ equals 1. Otherwise, $f_{t,j}^i$ equals 0. Based on practical network conditions, RSU-vehicle associations should satisfy multiple constraints based on SMT.

1) Reachability Constraint: Vehicles connect to RSUs within N-hop communications. There are two categories for vehicles to connect with RSUs: 1) if the vehicle is covered by an RSU, it will establish a connection with RSU directly; 2) if the vehicle is not covered by any RSUs, it will need other neighbour vehicles to relay request until connecting to an RSU successfully. Therefore, reachability constraint is formalized as follows:

$$f_{t,j}^i \cdot \mathcal{D}(n_i, v_j) \leq \mathcal{W}_{t,i}, \forall n_i \in \mathcal{N}, \forall v_j \in \mathcal{M} \tag{2}$$

where $\mathcal{D}(n_i, v_j)$ denotes the geographical distance between RSU n_i and vehicle v_j. Variable $\mathcal{W}_{t,i}$ denotes the maximum communication range of RSU n_i, and is defined as follows:

$$\mathcal{W}_{t,i} = \begin{cases} \sum_{j \in N-1} R_{t,j}^v + R_{t,i}^{\text{rsu}}, & R_{t,i}^{\text{rsu}} \leq R_{t,j}^v \\ \sum_{j \in N} R_{t,j}^v, & R_{t,i}^{\text{rsu}} > R_{t,j}^v \end{cases} \tag{3}$$

where $R_{t,i}^{\text{rsu}}$ is the communication range of RSU n_i from the output of PPO-OCM at time slot t, $R_{t,j}^v$ is the communication range of vehicle v_j at time slot t, and N is the largest number of hops.

2) Accessibility Constraint: Vehicles must establish connections with RSUs, which is formalized as following:

$$\sum_{n_i \in \mathcal{N}} \sum_{v_j \in \mathcal{M}} f_{t,j}^i = m, \sum_{n_i \in \mathcal{N}} f_{t,j}^i = 1, \forall v_j \in \mathcal{M} \tag{4}$$

where m denotes the number of vehicles. The former equation guarantees that all vehicles can connect with RSUs after shuffling. The latter equation guarantees that a vehicle can only connect to one RSU at each time slot.

3) Unpredictability Constraint: Unpredictability can be improved by minimizing the similarity of RSU-vehicle associations in consecutive time slots, which is formalized as follows:

$$\mathbb{D}_{n \times m}^{t,t+1} \triangleq \begin{bmatrix} d_{1,1}^{t,t+1} & \cdots & d_{1,m}^{t,t+1} \\ \vdots & \ddots & \vdots \\ d_{n,1}^{t,t+1} & \cdots & d_{n,m}^{t,t+1} \end{bmatrix} \tag{5}$$

$$\sum_{n_i \in \mathcal{N}} \sum_{v_j \in \mathcal{M}} (d_{i,j}^{t,t+1})^2 \geq \Psi \tag{6}$$

where $\mathbb{D}_{n \times m}^{t,t+1}$ is called similarity matrix, and $d_{i,j}^{t,t+1} = f_{t+1,j}^i - f_{t,j}^i$ denotes the difference value of association decision variable between two consecutive time slots. Inequation (6) indicates that the sum of $d_{i,j}^{t,t+1}$ must exceed threshold Ψ, which guarantees the unpredictability.

4) Capacity Constraint: The access capacity of RSU $Q_{t,i}^{\text{rsu}}$ is also from the output of PPO-OCM at time slot t. Then, capacity constraint is formalized as follows:

$$\sum_{v_j \in \mathcal{M}} f_{t,j}^i \leq Q_{t,i}^{\text{rsu}}, \forall n_i \in \mathcal{N}. \tag{7}$$

Above inequation guarantees that the number of vehicles that connect to RSU n_i will not exceed its access capacity, which prevents the degradation of service quality.

Algorithm 2: TA-RVA.

1: Initialize the trust scores of all vehicles as 0 s.
2: Initialize the set of feasible associations as null.
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Execute the selected action $\mathcal{A}_t$ based on Algorithm 1.
5: **for** $i = 1, 2, \ldots, n$ **do**
6: Acquire $\mathcal{W}_{t,i}$ and $Q_{t,i}^{rsu}$ from PPO-COM.
7: **end for**
8: Solve feasible RSU-vehicle associations by Z3 solver.
9: Modify the IP addresses of all RSUs.
10: Shuffle with random RSU-vehicle association.
11: **for** $j = 1, 2, \ldots, m$ **do**
12: Update malicious scores of all vehicles by (8).
13: **if** $\mathcal{T}_{t,j} \geq \Upsilon$ **then**
14: Vehicle v_j is considered as a spy and banned.
15: **end if**
16: **end for**
17: **end for**

B. Trust Assessment for RSU-Vehicle Associations

Since association decision variable $f_{t,j}^i$ takes only a value from $\{0, 1\}$, finding solutions that satisfy above constraints is typically a satisfiability problem, which has been proved to be non-deterministic polynomial (NP)-complete [40]. Because of dynamic environment caused by vehicular frequent movements, SDN controllers should solve the satisfiability problem in real-time. Considering that solving NP-complete problem is time-consuming, we calculate the feasible RSU-vehicle associations by Z3 solver [34]. Usually, there exists more than one feasible RSU-vehicle association, which will be selected randomly. To separate spy and innocent vehicles effectively, the trust of all vehicles will be evaluated after each shuffle. With multiple factors, we define the malicious score $\mathcal{T}_{t,j}$ for vehicle v_j when arriving at time slot t as follows:

$$\mathcal{T}_{t,j} = \mu \sum_{\tilde{t}=1}^{t} \sum_{i=1}^{n} f_{\tilde{t},j}^i \mathbf{Y}_{\tilde{t},i} Q_{\tilde{t},i} + \nu \sum_{\tilde{t}=1}^{t} G_{\tilde{t},j} - \xi \sum_{\tilde{t}=1}^{t} \tilde{t} \tag{8}$$

where μ, ν, and ξ are all coefficients, and $G_{\tilde{t},j}$ denotes the number of DDoS experienced by vehicle v_j at time slot $\tilde{t}$. Thereinto, the first term denotes whether RSUs are attacked successfully. If so, the malicious scores of vehicles that connect to these RSUs will increase linearly with the capacities of RSUs. The second term denotes the total number of DDoS experienced by vehicle v_j. The third term denotes that the malicious scores of vehicles will decrease over time slots. When $\mathcal{T}_{t,j}$ is more than the banning threshold Υ, vehicle v_j will be considered as spy, and be banned from connecting to all RSUs.

The pseudo-code of TA-RVA is shown in Algorithm 2. The malicious scores of vehicles and the set of feasible associations are both initialized (lines 1 and 2). On each time slot, the selected configuration mutation action $\mathcal{A}_t$ is executed (lines 3 and 4). Then, parameters $\mathcal{W}_{t,i}$ and $Q_{t,i}^{\text{rsu}}$ are both acquired for each RSU from PPO-COM (lines 5–7). A satisfiability problem is formulated based on SMT, and feasible RSU-vehicle associations are solved by Z3 solver (line 8). SDN controller modifies the

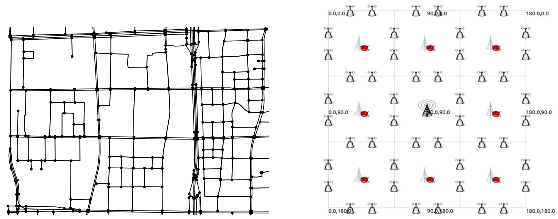


Fig. 4. Simulation scenarios. (a) Extracted road map. (b) Topologies of RSUs, BS, and attacking infrastructures.

IP addresses of RSUs [41] to avoid all IP addresses collected by spy vehicles (line 9). Then, SDN controller will shuffle the RSU-vehicle association (line 10). For all vehicles, malicious scores are updated according to (8). If the malicious score of vehicle v_j is greater than banning threshold, it will be banned from connecting to any RSUs (lines 11–16).

VII. Performance Evaluation

To simulate a realistic street scenario, we select a partial map of Beijing as an urban topology, whose extracted road map is shown in Fig. 4(a). The experimental scenario has the size of $1.8 \times 1.8\ km^2$, which is divided into 3×3 grids. Two simulation platforms SUMO [43] and NS-3 [44] are applied together to conduct following experiments. The traces of vehicles are generated by SUMO based on the Manhattan grid mobility model, and are imported into NS-3. Each vehicle is equipped with two types of wireless interfaces: 1) IEEE 802.11p; 2) 3GPP long term evolution (LTE). RSUs are deployed every 300 m along streets, and BS is located at the center of the selected area. We also deploy attacking infrastructures in the center of each grid. The topologies of RSUs, BS, and attacking infrastructures is shown in Fig. 4(b), which is drawn by NetAnim 3.108 [42]. Besides, BS hosts the SDN controller, which executes our proposed PPO-OCM and TA-RVA. At each episode, SDN controller decides current mutated configurations and RSU-vehicle associations, then distributes these decisions to all RSUs in coverage. Finally, RSUs will modify their own configurations and be reassociated with corresponding vehicles. To interface NS-3 simulator to DRL, we use the NS3gym interface [45], which allows for seamless integration between OpenAI Gym and NS-3 simulator. Table I summarizes the main parameters of SD-IoV. The main parameters of DNNs are shown in Table II.

We observe that there are no MTD methods similarly to ours in SD-IoV. Therefore, we choose following baselines that can be adapted and applied to the scenario of SD-IoV.

1) Random network mutation (RNM) [23]: A network agility scheme that randomly changes the communication range of RSUs. This approach forces vehicles to switch their associated RSUs, which aims to confuse the adversary because RSUs appear and disappear randomly and frequently.
2) Trust inference model (TIM) [12]: A trust evaluation scheme based on trust inference model that integrates the subjective trust and recommendation trust.

TABLE I
Simulation Parameters

Parameter	Value or Range
Time slot ΔT, sending interval	$0.1s$, $0.1s$
Simulation area	$1.8 \times 1.8\ km^2$
Size of request packets, data packets	$64byte$, $1000byte$
Communication range of vehicles, BSs	$50m$, $1000m$
Communication range of RSUs	$[50, 60, \cdots, 100]m$
Capacity of RSUs	$[15, 20, 25, 30]$
Grid length	$600m$
Number of vehicles m	$[50, 60, \cdots, 100]$
Number of spy vehicles	10
Number of RSUs n	$[40, 48]$
Number of attacking infrastructures	9
Number of BS k	1
Largest number of hops N	3
Velocity of vehicles	$10 - 60km/h$
Coefficient α, β, ζ	2, 1×10^{-3}, 0.5
Coefficient E_{elec}, ϵ_f, ϵ_t	50, 10, 1.3×10^{-3}
Coefficient μ, ν, ξ	0.4, 1, 1
Banning threshold Υ	10000

TABLE II
Hyperparameters for PPO-OCM

Parameter	Value or Range
Maximum episodes	4000
Discount factor	0.9
GAE discount	0.95
Policy learning rate (Adam)	0.005
Timesteps per episode	10
Policy epochs	4
PPO Clipping	0.2
Actor network hidden layers	[256]
Critic network hidden layers	[16]
Activation function	$tanh$
Output function	$softmax$

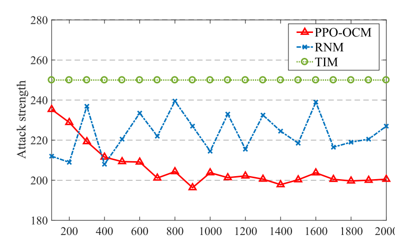


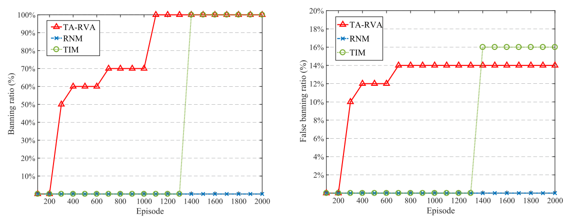
Fig. 5. Defense performance comparison.

3) Costconstrained association control algorithm (CACA) [46]: A handoff scheme that is expected to maximize the minimum throughout of mobile nodes.

A. Defense Performance

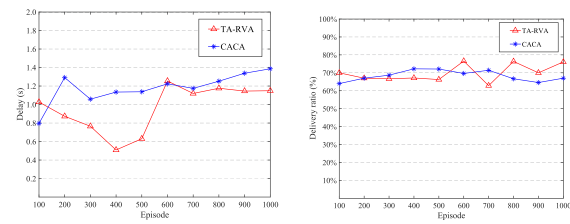
Attack strength (AS) is an important metric to measure the defense performance, which is defined as the sum of attacked RSUs' capacities. We carry out 2000 episodes of simulations, and calculate the attack strength on each episode while PPO-OCM, RNM, and TIM are deployed.

As shown in Fig. 5, AS under TIM does not change over episodes, reaching about 250. The reason is that RSUs in TIM have constant communication ranges and capacities, which will result in constant AS. Under RNM, AS shows fluctuations



(a) (b)

Fig. 6. Trustworthiness performance comparison. (a) BR. (b) FBR.



(a) (b)

Fig. 7. Network performance for different episodes. (a) Average delay. (b) Average delivery ratio.

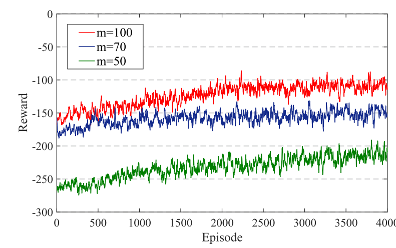


Fig. 8. Convergence performance of PPO-OCM.

continuously over episodes, and average AS is about 225. This is because that RNM changes the communication ranges and access capacities of RSUs periodically, which can decrease AS to some extent. AS under PPO-OCM converges from 235 to 200 because DRL can learn the optimal configurations of RSUs to decrease AS gradually. The convergence time is about 1000 episodes. Simulation results strongly confirm that PPO-OCM can decrease the AS of DDoS, which is better than that under RNM and TIM.

B. Trustworthiness Performance

The trustworthiness of our proposed TA-RVA is considered as the probabilities that spy vehicles are banned and innocent vehicles are normal. That is, we observe trustworthiness of TA-RVA by the reliability of trust assessment [47]. Therefore, we use banning ratio (BR) and false banning ratio (FBR) as reliability metrics. BR is defined as the portion of spy vehicles that are banned out of total spy vehicles. FBR is defined as the portion of innocent vehicles that are banned out of total innocent vehicles. We carry out 2000 episodes of simulations, and calculate the BR and FBR on each episode while TA-RVA, RNM, and TIM are deployed. Because RNM has no trust assessment, the BR and FBR of RNM will always be zero.

As shown in Fig. 6(a), the BR of TA-RVA increases over episodes, which will reach 100% after 1100 episodes. The reason is that spy vehicles always tell attacking infrastructures to launch DDoS attacks, which will lead to the rapid increasing of malicious scores. The BR of TIM reaches 100% after 1300 episodes because TIM needs more episodes to observe spy vehicles. It means that TA-RVA spends less episodes to ban all spy vehicles compared to TIM. Evaluation results in Fig. 6(b) show that the FBR of TA-RVA grows slowly, which will be almost 14%. This is because that some innocent vehicles have similar mobility trajectories with spy vehicles, which will cause that malicious scores are higher than banning threshold. In addition, the FBR of TIM is about 16% because of similar reasons. Simulation results fully confirm that TA-RVA has better reliability of trust assessment compared to TIM. As a conclusion, TA-RVA can work with high trustworthiness.

C. Network Performance

To evaluate the network performance, we consider two metrics: 1) delay; 2) delivery ratio. Delay denotes the average interval of time required for successfully delivered packets. Delivery ratio denotes the portion of packets that are received by the destinations out of the total packets generated. We carry out 1000 episodes of simulations, and calculate average delay and delivery ratio over episodes compared with CACA.

Fig. 7(a) and (b) indicate that the average delay and delivery ratio of TA-RVA are similar with that of CACA because TA-RVA considers multiple network constraints to guarantee quality of service (QoS). Therefore, it can make a conclusion that our proposed MTD scheme only affects the QoS in an acceptable range.

D. Convergence Performance

To evaluate the convergence performance adequately, we utilize the reward in the train process of DRL. With the number of vehicles as 50, 70, and 100, we evaluate reward with 4000 episodes, whose results are shown in Fig. 8.

Evaluation results indicate that the convergence speeds are similar when the number of spy vehicles is ten and the number of innocent vehicles is 40, 60, and 90. PPO-OCM will converge within about 2000 episodes. When the total number of vehicles is 100, reward is the largest, which converges from -150 to -100. On the other hand, when the total number of vehicles is 50, reward is the least, which converges from -250 to -210. This is because that more innocent vehicles in SD-IoV will bring positive reward in the training process.

E. Overhead in SMT Formalization

SMT solving time is the main overhead in the process of removal and reassociation between RSUs and vehicles. By utilizing the Z3 solver, we calculate SMT solving time under multiple environments that have different number of vehicles when the number of RSUs is 40 and 48. The number of vehicles is set as 50, 60, 70, 80, 90, and 100 respectively.

As shown in Fig. 9, SMT solving time increases significantly with the increasing number of vehicles, especially when the number of vehicles reaches 100 in SD-IoV. In addition, SMT solving time also increases significantly with the increasing number of RSUs. This is because the size of association matrix,

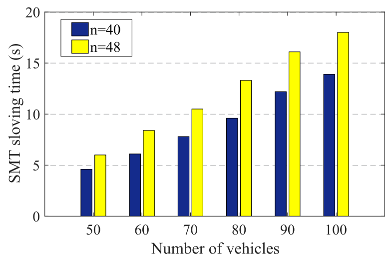


Fig. 9. SMT solving time for different numbers of vehicles.

denoted as $m \times n$, will grow with the increasing number of vehicles and/or RSUs. When the network scale of SD-IoV increases, SMT formalization of feasible RSU-vehicle associations needs more time to be solved.

VIII. Conclusion and Future Work

In this article, we proposed an intelligent MTD scheme that consists of two mechanisms: 1) PPO-OCM; 2) TA-RVA. Firstly, we modeled the configuration mutation of RSUs as an MDP, then formulate an optimization problem, which is solved by DRL. Secondly, trust assessment mechanism is proposed to identify spy vehicles with multiple constraints. Lastly, we conduct simulations to confirm the effectiveness of our method.

In future work, we will consider the storage capacity and transmission delay of flow table in our proposed MTD scheme. In addition, we will investigate how to solve new security problems after introducing SDN into IoV.

References

[1] Q. Zhang et al., "Graph neural networks-driven traffic forecasting for connected Internet of Vehicles," *IEEE Trans. Netw. Sci. Eng.*, to be published, doi: 10.1109/TNSE.2021.3126830.

[2] C. Hu et al., "A digital twin-assisted real-time traffic data prediction method for 5G-enabled Internet of Vehicles," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2811–2819, Apr. 2022.

[3] J. Chen et al., "Service-oriented dynamic connection management for software-defined Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2826–2837, Oct. 2017.

[4] T. Mekki et al., "Software-defined networking in vehicular networks: A survey," *Trans. Emerg. Telecommun. Technol.*, 2021, Art. no. e4265.

[5] W. B. Jaballah et al., "Security and design requirements for software-defined VANETs," *Comput. Netw.*, vol. 169, 2020, Art. no. 107099.

[6] A. Akhunzada and M. K. Khan, "Toward secure software defined vehicular networks: Taxonomy, requirements, and open issues," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 110–118, Jul. 2017.

[7] N. Sharma et al., "Secure authentication and session key management scheme for Internet of Vehicles," *Trans. Emerg. Telecommun. Technol.*, vol. 33, 2022, Art. no. e4451.

[8] Z. A. Biron, S. Dey, and P. Pisu, "Real-time detection and estimation of denial of service attack in connected vehicle systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3893–3902, Dec. 2018.

[9] M. N. Mejri and J. Ben-Othman, "GDVAN: A new greedy behavior attack detection algorithm for VANETs," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 759–771, Mar. 2017.

[10] G. de Biasi, L. F. M. Vieira, and A. A. F. Loureiro, "Sentinel: Defense mechanism against DDoS flooding attack in software defined vehicular network," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.

[11] A. Hbaieb et al., "A survey of trust management in the Internet of Vehicles," *Comput. Netw.*, vol. 203, 2022, Art. no. 108558.

[12] H. Xia, S.-S. Zhang, Y. Li, Z.-K. Pan, X. Peng, and X.-Z. Cheng, "An attack-resistant trust inference model for securing routing in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7108–7120, Jul. 2019.

[13] M. Najafi et al., "Decentralized prediction and reputation approach in vehicular networks," *Trans. Emerg. Telecommun. Technol.*, vol. 13, 2022, Art. no. e4456.

[14] R. Nigam et al., "AI-enabled trust-based routing protocol for social opportunistic IoT networks," *Trans. Emerg. Telecommun. Technol.*, 2021, Art. no. e4330.

[15] J. Li, Z. Xue, C. Li, and M. Liu, "RTED-SD: A real-time edge detection scheme for sybil DDoS in the Internet of Vehicles," *IEEE Access*, vol. 9, pp. 11296–11305, 2021.

[16] J. Zhang et al., "AntiConcealer: Reliable detection of adversary concealed behaviors in EdgeAI assisted IoT," *IEEE Internet of Things J.*, to be published, doi: 10.1109/JIOT.2021.3103138.

[17] J.-H. Cho et al., "Toward proactive, adaptive defense: A survey on moving target defense," *IEEE Commun. Surv. Tut.*, vol. 22, no. 1, pp. 709–745, Jan.–Mar. 2020.

[18] R. E. Navas, F. Cuppens, N. B. Cuppens, L. Toutain, and G. Z. Papadopoulos, "MTD, where art thou? A systematic review of moving target defense techniques for IoT," *IEEE Internet of Things J.*, vol. 8, no. 10, pp. 7818–7832, May 2021.

[19] T. Zhang, X. Kuang, Z. Zhou, H. Gao, and C. Xu, "An intelligent route mutation mechanism against mixed attack based on security awareness," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.

[20] C. Xu, T. Zhang, X. Kuang, Z. Zhou, and S. Yu, "Context-aware adaptive route mutation scheme: A reinforcement learning approach," *IEEE Internet of Things J.*, vol. 8, no. 17, pp. 13528–13541, Sep. 2021.

[21] T. Zhang et al., "DQ-RM: Deep reinforcement learning-based route mutation scheme for multimedia services," in *Proc. Int. Wireless Commun. Mobile Comput.*, 2020, pp. 291–296.

[22] F. Nizzi, T. Pecorella, F. Esposito, L. Pierucci, and R. Fantacci, "IoT security via address shuffling: The easy way," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3764–3774, Apr. 2019.

[23] Q. Duan et al., "Range and topology mutation based wireless agility," in *Proc. 7th ACM Workshop Moving Target Defense*, 2020, pp. 59–67.

[24] M. Ge et al., "Proactive defense mechanisms for the software-defined Internet of Things with non-patchable vulnerabilities," *Future Gener. Comput. Syst.*, vol. 78, pp. 568–582, 2018.

[25] Y.-H. Lin, J.-J. Kuo, D.-N. Yang, and W.-T. Chen, "A cost-effective shuffling-based defense against HTTP DDoS attacks with SDN/NFV," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–7.

[26] Y. Zhou, G. Cheng, Y. Zhao, Z. Chen, and S. Jiang, "Towards proactive and efficient DDoS mitigation in IIoT systems: A moving target defense approach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2734–2744, Apr. 2022.

[27] W. Soussi, M. Christopoulou, G. Xilouris, and G. Gür, "Moving target defense as a proactive defense element for beyond 5G," *IEEE Commun. Standards Mag.*, vol. 5, no. 3, pp. 72–79, Sep. 2021.

[28] X. Lu, L. Xiao, T. Xu, Y. Zhao, Y. Tang, and W. Zhuang, "Reinforcement learning based PHY authentication for VANETs," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3068–3079, Mar. 2020.

[29] L. D. Moura and N. Bjørner, "Satisfiability modulo theories: Introduction and applications," *Commun. ACM*, vol. 54, no. 9, pp. 69–77, 2011.

[30] F. Bai et al., "The important framework for analyzing the impact of mobility on performance of routing protocols for ad hoc networks," *Ad Hoc Netw.*, vol. 1, no. 4, pp. 383–403, 2003.

[31] A. J. Siddiqui and A. Boukerche, "On the impact of DDoS attacks on software-defined Internet-of-Vehicles control plane," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, 2018, pp. 1284–1289.

[32] K. S. Sahoo et al., "An early detection of low rate DDoS attack to SDN based data center networks using information distance metrics," *Future Gener. Comput. Syst.*, vol. 89, pp. 685–697, 2018.

[33] B. Yuan, D. Zou, S. Yu, H. Jin, W. Qiang, and J. Shen, "Defending against flow table overloading attack in software-defined networks," *IEEE Trans. Serv. Comput.*, vol. 12, no. 2, pp. 231–246, Mar./Apr. 2019.

[34] L. D. Moura et al., "Z3: An efficient SMT solver," in *Proc. Int. Conf. Tools Algorithms Construction Anal. Syst.*, 2008, pp. 337–340.

[35] J. Tao, L. Zhu, X. Wang, J. He, and Y. Liu, "RSU deployment scheme with power control for highway message propagation in VANETs," in *Proc. IEEE Glob. Commun. Conf.*, 2014, pp. 169–174.

[36] J. Zhang, M. Z. A. Bhuiyan, X. Yang, A. K. Singh, D. F. Hsu, and E. Luo, "Trustworthy target tracking with collaborative deep reinforcement learning in EdgeAI-Aided IoT," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1301–1309, Feb. 2022.

[37] T. He et al., "What is the human mobility in a new city: Transfer mobility knowledge across cities," in *Proc. Web Conf.*, 2020, pp. 1355–1365.

[38] F. Li, X. Song, H. Chen, X. Li, and Y. Wang, "Hierarchical routing for vehicular ad hoc networks via reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1852–1865, Feb. 2019.

[39] J. Schulman et al., "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

[40] J. H. Jafarian, E. Al-Shaer, and Q. Duan, "An effective address mutation approach for disrupting reconnaissance attacks," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 12, pp. 2562–2577, Dec. 2015.

[41] Y. He, M. Zhang, X. Yang, Q. T. Sun, J. Luo, and Y. Yu, "The intelligent offense and defense mechanism of Internet of Vehicles based on the differential Game-IP hopping," *IEEE Access*, vol. 8, pp. 115217–115227, 2020.

[42] The NetAnim, 2017, Accessed: Mar. 04, 2022. [Online]. Available: https://www.nsnam.org/wiki/NetAnim_3.108

[43] The Simulation of Urban Mobility, 2001, Accessed: Mar. 04, 2022. [Online]. Available: http://sumo.sourceforge.net

[44] The NS-3 Network Simulator, 2011, Accessed: Mar. 04, 2022. [Online]. Available: https://www.nsnam.org

[45] P. Gawlowicz et al., "NS-3 meets OpenAI Gym: The playground for machine learning in networking research," in *Proc. ACM Int. Conf. Model., Anal. Sim. Wireless Mobile Sys.*, 201, pp. 113–120.

[46] W. Wong, A. Thakur, and S.-H. G. Chan, "An approximation algorithm for AP association under user migration cost constraint," in *Proc. 35th IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.

[47] G. Fortino, F. Messina, D. Rosaci, G. M. L. Sarne, and C. Savaglio, "A trust-based team formation framework for mobile intelligence in smart factories," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6133–6142, Sep. 2020.

Tao Zhang (Graduate Student Member, IEEE) received the B.S. degree in Internet of Things engineering from the Beijing University of Posts and Telecommunications, Beijing, China, and the Queen Mary University of London, London, U.K., in 2018. He is currently working toward the Ph.D. in computer science and technology with the School of Computer Science, Beijing University of Posts and Telecommunications.

His research interests include network security, moving target defense, and reinforcement learning.

Mr. Zhang was a recipient of the Best Paper Award of International Conference on Networking and Network Applications (NaNA) 2018 and International Wireless Communications and Mobile Computing Conference (IWCMC) 2021.

Changqiao Xu (Senior Member, IEEE) received the Ph.D. degree in computer science and technology from the Institute of Software, Chinese Academy of Sciences (ISCAS), Beijing, China, in Jan. 2009.

From 2002 to 2007, he was an Assistant Research Fellow and R&D Project Manager with ISCAS. From 2007 to 2009, he was a Researcher with the Athlone Institute of Technology, Athlone, Ireland, and Joint Training Ph.D. with Dublin City University, Dublin, Ireland. In Dec. 2009, he joined Beijing University of Posts and Telecommunications (BUPT), Beijing, China. He is currently a Professor with the State Key Laboratory of Networking and Switching Technology, and Director of the Network Architecture Research Center, BUPT. His research interests include future internet technology, mobile networking, multimedia communications, and network security.

Dr. Xu has edited two books and authored or coauthored over 200 technical papers in prestigious international journals and conferences, including *IEEE Communication Magazine*, IEEE/ACM TRANSACTIONS ON NETWORKING (ToN), IEEE TRANSACTIONS ON MOBILE COMPUTING (TMC), International Conference on Computer Communications (INFOCOM), ACM Multimedia, etc. He has served a number of international conferences and workshops as a Cochair and TPC member. He is currently serving as the Editor-in-Chief of *Transactions on Emerging Telecommunications Technologies (Wiley)*.

Ping Zou received the B.S. degree in computer science, in 2021, from the Beijing University of Posts and Telecommunications, Beijing, China, where he is currently working toward the master degree in computer science and technology with the School of Computer Science.

His research interests include reinforcement learning and network security.

Haijiang Tian is currently working toward the B.S. degree in telecommunications engineering and management with International School, Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include network security and IoV.

Xiaohui Kuang received the Ph.D. degree in computer science and technology from the School of Computer, National University of Defense Technology, Changsha, China, in 2003.

He is currently a Research Fellow and Professor with the National Key Laboratory of Science and Technology on Information System Security, Beijing, China. He is also a Guest Professor with the Beijing University of Posts and Telecommunications, Beijing, China. His research interest includes network and information security, and wireless network.

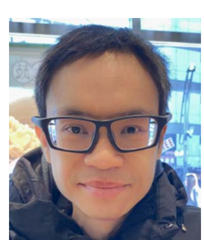
Shujie Yang received the Ph.D. degree in computer science and technology from the Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2017.

He is currently a Lecturer with the State Key Laboratory of Networking and Switching Technology, Beijing, China. His major research interests include the areas of wireless communications and wireless networking.

Lujie Zhong received the Ph.D. degree in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2013.

She is currently an Associate Professor with the Information Engineering College, Capital Normal University, Beijing, China. Her research interests include communication networks, computer system and architecture, and mobile networks.

Dr. Zhong has published papers in prestigious international journals and conferences in the related areas, including *IEEE Communication Magazine*, IEEE TRANSACTIONS ON MOBILE COMPUTING (TMC), IEEE TRANSACTIONS ON MULTIMEDIA (TMM), *IEEE Internet of Things Journal* (IoTJ), IEEE International Conference on Computer Communications (INFOCOM), and ACM Multimedia (MM).



Dusit Niyato (Fellow, IEEE) received the B.Eng. degree in computer engineering from the King Mongkuts Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand, and the Ph.D. in electrical and computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1999 and 2008, respectively.

He is a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include the areas of Internet of Things (IoT), machine learning, and incentive mechanism design.