Reinforcement Learning Based PHY Authentication for VANETs

Xiaozhen Lu, Liang Xiao, *Senior Member*, *IEEE*, Tangwei Xu, Yifeng Zhao, *Member*, *IEEE*, Yuliang Tang, *Member*, *IEEE*, Weihua Zhuang, *Fellow, IEEE*

***Abstract*—Mobile edge computing in vehicular ad hoc networks (VANETs) suffers from rogue edge attacks due to the vehicle mobility and the network scale. In this paper, we present a physical authentication scheme to resist rogue edge attackers whose goal is to send spoofing signals to attack VANETs. This authentication scheme exploits the channel states of the shared ambient radio signals of the mobile device and its serving edge such as the onboard unit during the same moving trace and applies reinforcement learning to select the authentication modes and parameters. By applying transfer learning to save the learning time and applies deep learning to further improve the authentication performance, this scheme enables mobile devices in VANETs to optimize their authentication modes and parameters without being aware of the VANET channel model, the packet generation model, and the spoofing model. We provide the convergence bound such as the mobile device utility, evaluate the computational complexity of the physical authentication scheme, and verify the analysis results via simulations. Simulation and experimental results show that this scheme improves the authentication accuracy with reduced energy consumption against rogue edge attacks.**

***Index Terms*—VANETs, rogue edge attacks, physical authentication, ambient radio signals, reinforcement learning.**

I. Introduction

Mobile devices such as smartphones or smartwatches can offload some computation-intensive tasks to its serving edge nodes such as the onboard units (OBUs) in the same vehicle [1], [2]. Mobile edge computing that enables mobile devices in vehicular ad hoc networks (VANETs) to use the computing, storage and power resources of the serving edge nodes significantly reduces the latency, saves energy, and protects the user privacy as compared with cloud computing [3]–[5]. However, edge computing has to detect spoofing attacks, especially rogue edge attacks, as most communications between mobile devices and OBUs are not well protected with authentications, especially with the OBUs high mobility and the large-scale network topology of VANETs [6]–[8].

The rogue edge attackers can apply the universal software radio peripherals (USRPs) to eavesdrop the signals sent from nearby radio sources such as vehicles [9], and send spoofing signals to fabricate incorrect warnings to attack VANETs [10], [11].

Existing VANET authentication schemes depend on cryptography, trust, certificate, and physical (PHY) authentication. As a lightweight security protocol, the PHY authentication usually exploits the PHY properties under the spatial distribution of wireless transmissions, such as the received signal strength indicators (RSSIs) [12], [13], the received signal strength (RSS) of the packets, and the channel responses [14]–[16]. In particular, ambient radio signals help the mobile device improve the authentication accuracy [17]–[19]. Such PHY authentication uses the ambient signals from multiple ambient radio sources observed by both the mobile device and the OBU in the same vehicle and extracts the PHY features of the ambient radio signals, such as the packet arrival interval and RSSIs. Mobile devices in VANETs generate packets of different priority for different types of applications. The packet priority depends on the arrival interval [20] or the communication channel [21]. For example, the priority of a packet transmitted in the control channel is higher than a packet delivered in the service channel of VANETs.

The optimal authentication mode and parameters are challenging to determine because of the environmental changes and terminal mobility [22]. The authentication parameters can be optimized via reinforcement learning (RL) techniques [12], [18] without depending on the VANET channel model, the packet generation model, and the spoofing model. For example, the RSSI based authentication (RSSI-Q) algorithm in [12] applies Q-learning to choose the test threshold based on the signal RSSI of the message under test. However, its authentication performance in VANETs, such as the authentication accuracy, suffers from degradation. The authentication scheme (ambient-Q) proposed in [18] that uses the ambient radio signals channel states improves the authentication performance as compared with the RSSI based authentication schemes [12], [13]. The mixed-Q based VANET authentication scheme in [23] applies Q-learning to choose the authentication policy based on the authentication accuracy of the previous authentication, the estimated attack rate, and the packet priority. However, these authentication schemes optimize the authentication policy with a lower learning speed in the high dimensional state space that cannot deal with networking environment changes in VANETs [24].

In this paper, we present a PHY authentication framework that uses the spatial decorrelation PHY features of the ambient radio signals received along the vehicular driving traces

This work was supported in part by the Natural Science Foundation of China under Grant 61971366, 61671396, 61731012 and 91638204. *(Corresponding Author: Yifeng Zhao).*

X. Lu, L. Xiao, T. Xu, Y. Zhao and Y. Tang are with the Department of Information and Communication Engineering, Xiamen University, Xiamen 361005, China. Email: {lxiao, zhaoyf}@xmu.edu.cn.

W. Zhuang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Email: wzhuang@uwaterloo.ca.

to detect the spoofing packets sent by rogue edge nodes. This framework aims to improve the authentication accuracy with less security overhead and provide better protection for the packets with higher priority due to the limited computation and communication resource of VANETs. More specifically, this authentication scheme uses the packet arrival interval and RSSIs sent from the base stations (BSs), access points (APs), and roadside units (RSUs) along the road to detect rogue edge nodes outside the vehicle that cannot receive those ambient radio signals, e. g., the radio signal sent from a BS that the rogue edge node has never been close to, but the serving edge does. This scheme uses the ambient radio signals during the edge node transmission intervals instead of the noise information.

Compared with the PHY authentication schemes such as in [12], [18] that depend on the channel state between the receiver and the transmitter under test, our proposed PHY authentication scheme incorporates the PHY features of the ambient radio signals besides the RSSI and the arrival intervals of the signal under test in the authentication. This authentication scheme exploits the Dyna architecture to generate simulated authentication experiences to estimate the Q-values that are the expected long-term discounted reward for each authentication policy. The previous authentication experiences in similar vehicular networks are used to initialize the Q-values (instead of setting as an all-zero matrix) to improve the initial authentication.

We present a deep RL based authentication that uses neural episodic control (NEC) [25], a type of deep reinforcement learning, to reduce the optimization time in the authentication process, and uses transfer learning to initialize learning parameters and thus improve the authentication performance in large-scale VANETs. In this scheme, the deep convolutional neural network (CNN) and the experience replay technique are used to generate the key to look up the memory module named differentiate neural dictionary (DND) that outputs the estimated long-term expected utility of each authentication policy including the authentication mode and parameters such as the test threshold. Our proposed authentication scheme enables the mobile device with more computation resources to optimize the authentication policy and use offloading to reduce the overall computational overhead for the computation-intensive tasks [26], [27].

This authentication scheme that is robust against the dynamic VANET environment and the unknown spoofing model optimizes the authentication policy via error-and-trial. The performance bound regarding the mobile device utility is provided and verified via simulation results, which show that the proposed scheme improves the authentication performance in comparison with mixed-Q in [23], RSSI-Q in [12] and ambient-Q in [18]. The computational complexity of the proposed scheme is evaluated theoretically.

This paper is structured as follows. First, the related work of the authentication in VANET is reviewed in Section II and the VANET model is presented in Section III. A PHY authentication framework is presented in Section IV, RL based authentication schemes are proposed in Section V and Section VI, and the corresponding performance is analyzed in Section VII. Simulation and experimental results are provided in Section VIII and the conclusions are drawn for this work in Section IX.

II. Related Work

Most VANET authentication schemes depend on cryptography, trust, and certificate [11], [28]–[31]. For example, in Sybil detection protocol presented in [11], the motor vehicle as a certificate authority calculates the hash values of the overheard pseudonyms. The privacy preservation scheme proposed in [28] replaces multiple pseudonym certificates with a public key infrastructure to protect privacy and reduce the computational overhead. The VANET authentication strategy proposed in [29] applies the key-insulated pseudonym model to protect location privacy. The VANET authentication scheme in [30] applies cryptology techniques to improve the vehicle-to-grid reliability connections and reduce communication overhead. In the VANET authentication proposed in [31], RSUs perform encryption to reduce computing cost for vehicular crowdsensing systems.

PHY authentication techniques provide lightweight authentication to detect rogue edge attacks [9], [12], [18], [32]–[36]. For instance, the VANET authentication method presented in [33] uses the global positioning system information and the RSS of the signals under test to detect rogue APs with reduced communication overhead. The cyber-physical VANET authentication protocol proposed in [32] uses the channel states against the man-in-the-middle attacks. A detection scheme designed in [9] forms an authentication hypothesis test to defend rogue attackers. The combined spoofing method proposed in [34] uses the spoofer in the wireless networks as a relay and applies the convex optimization method to resist the suspicious links, and thus improve the spoofing rate of the relay. The interference resistance scheme proposed in [35] uses the friendly spoofing nodes with the convex optimization method to reduce the spoofing symbol error rate in the additive white Gaussian noise scenarios.

The ambient radio signal based authentication scheme presented in [23] uses the ambient radio signals and the channel states and applies Q-learning to optimize the authentication parameters in the PHY authentication process. Different from the previous work in [23], in this work, we apply the Dyna architecture and transfer learning to save the convergence time in the authentication process, and use deep reinforcement learning to further decrease the authentication error in VANETs. The performance bound of the PHY authentication is provided and verified via simulations.

III. System Model

A. Network Model

Mobile device Bob and his serving edge Alice that such as an OBU are both located in a vehicle that moves at a speed v_{A} at time slot k, as shown in Fig. 1. Bob offloads computation-intensive tasks such as a computation task in a virtual reality game to Alice. Upon receiving the k-th packet such as the computation results from Alice with her network identity, such as the media access control (MAC) address. Bob has

to determine whether the packet is actually sent by Alice or the rogue edge node Eve. The packet received by Bob has center frequency f_0 and bandwidth W, following the IEEE 802.11 protocol [37]. Each packet consists of the preambles for channel estimation, packet header that contains the frame control and packet content such as the edge computation results. For simplicity, Bob is assumed to receive the k-th packet with Alice's MAC address at time slot k.

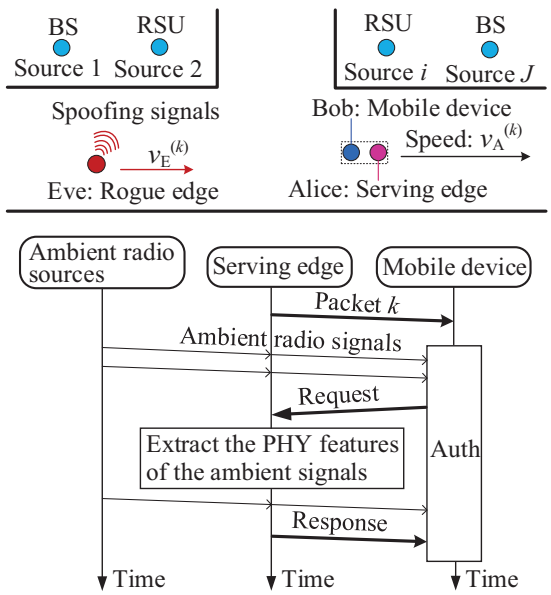


Fig. 1. PHY authentication with the physical layer features of ambient radio signals, where Bob compares the received ambient physical features such as the RSSI and the packet arrival interval sent by the edge node under test with his copy to detect rogue edge attackers.

The vehicle that hosts Bob and Alice receives signals from the J ambient radio sources such as nearby BS, RSU, and other radio devices, with transmit power P_i from the i-th ambient radio sources. As Alice stays in the same vehicle with Bob, she is highly likely to receive most of the ambient signals that Bob receives. Upon receiving the k-th packet, Bob uses the packet preambles to estimate the channel states at M tones over bandwidth W at center frequency f_0 and applies the maximum-likelihood criterion as presented in [38] to formulate channel vector $\mathbf{H}$ for the packet under test during the given time interval, and saves the channel record $\hat{\mathbf{H}}_{\mathrm{A}}$ of Alice after authentication.

B. Attack Model

The attacker Eve monitors the channel to send spoofing signals when the edge node Alice does not transmit her own signals. More specifically, Eve can use the MAC protocols such as the carrier sense multiple access to monitor the transmission channel and send spoofing signals without collisions with Alice [39]. Eve can choose the attack rate denoted by $y^{(k)}$ that corresponds to the number of spoofing packets sent by Eve between the $(k-1)$-th and the k-th packet sent by Alice. For simplicity, the attack rate is normalized with $0 \leq y^{(k)} < Y_{\mathrm{MAX}}$, where $Y_{\mathrm{MAX}} < 1$ is the maximum attack rate. For example, if $y = 0$, Eve sends 0 spoofing packet before Alice sends the k-th packet and after her $(k-1)$-th packet.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

J	Number of the ambient radio sources
N_{A}	Number of the shared ambient radio sources
X	Number of the feasible test threshold levels
h_m	Channel response at tone m for packet k
$\hat{\mathbf{H}}_{\mathrm{A}}/\mathbf{H}$	Channel record/channel response vector
$\bar{R}_j/R_j$	RSSI of the previous $(N_{\mathrm{A}}-j)$-th packet of Bob/the node under test
$\bar{T}_j/T_j$	Arrival interval of the previous $(N_{\mathrm{A}}-j)$-th packet of Bob/the node under test
$\bar{\mathbf{F}}/\mathbf{F}$	Ambient features of Bob/the node under test
$\hat{p}_{\mathrm{F/M}}$	False alarm rate/miss detection rate
θ	Priority of the k-th packet
x_0	Authentication mode
x_1	Test threshold
y	Attack rate
c/C_{R}	Authentication mode cost/spoofing cost

Eve is assumed to be outside Bob's vehicle. Without loss of generality, Eve moves with speed v_{E} and aims to spoof Bob with Alice's identity. The attacker can use a random attack policy with a random attack rate $y^{(k)}$ that uniformly distributed between 0 and Y_{MAX}. As a smart attacker, Eve can also observe the authentication accuracy of the previous authentication and the priority of current packet as her state, and use Q-learning to choose the attack rate to increase its utility denoted by u_{E}. Note that u_{E} is based on Bob's utility and the spoofing cost C_{R}. The goal of the attacker is to cheat Bob in his authentication process with the ambient radio signals. The important symbols are summarized in TABLE I and the time index k is omitted if no ambiguity.

IV. PHY Authentication Framework

We propose a PHY authentication framework that depends on both the features of the signals previously sent from the same edge device as well as the features of the ambient signals that both mobile device Bob and serving edge Alice received during a given time period. This framework chooses the authentication mode and parameters with the authentication experiences and current attack strength. Specifically, a fast authentication mode compares the channel vector under test with Alice's channel record. The safe authentication mode in this framework depends on the ambient signals from multiple ambient radio sources shared by both Alice and Bob and achieves more accurate authentication at the cost of more communication and computational overhead.

Upon receiving the k-th packet under test, the mobile device chooses the authentication mode denoted by $x_0 \in \{1, 2\}$ and test threshold $x_1 \in \{\epsilon/X\}_{0 \leq \epsilon \leq X}$, where X is the number of the parameter quantization levels. Similar to [38], the mobile device applies the maximum-likelihood criterion to estimate the channel response h_m at tone m for the k-th

packet, and formulates channel vector $\mathbf{H} = [h_m]_{1\leq m\leq M}$. If $x_0 = 1$, the mobile device compares $\mathbf{H}$ with Alice's channel record $\hat{\mathbf{H}}_{\mathrm{A}}$.

The safe authentication mode with $x_0 = 2$ uses the ambient signal PHY features to improve the authentication accuracy with more computation and communication complexity compared with the fast authentication mode with $x_0 = 1$. As mobile devices in well-designed VANETs rarely receive both strong ambient signals and the edge signal at the same time on the same frequency channel, this scheme evaluates the signals on another frequency channel or at another time to measure the PHY features of the ambient radio signals. More specifically, the mobile device requests the edge node to extract the PHY features of the previous N_{A} ambient packets from a given set of ambient radio sources in a specific time duration of τ. Let $\bar{R}_j$ and $\bar{T}_j$ be the RSSI and the arrival interval of the previous $(N_{\mathrm{A}}-j)$-th packet sent from the given ambient radio sources received by Bob during the monitoring period, respectively. The ambient feature vector of the mobile device, $\bar{\mathbf{F}} = [\bar{R}_j; \bar{T}_j]_{1\leq j\leq N_{\mathrm{A}}}$, during τ consists of the RSSIs and arrival intervals of previous N_{A} ambient packets.

Upon receiving a request from the mobile device, the node under test extracts the features of the previous N_{A} ambient packets from ambient radio sources as specified in the request. Let R_j and T_j be the RSSI and the arrival interval of the $(N_{\mathrm{A}} - j)$-th packet during the monitoring period τ. The node under test sends ambient feature vector $\mathbf{F} = [R_j; T_j]_{1\leq j\leq N_{\mathrm{A}}}$ in the authentication response to the mobile device.

The PHY authentication framework builds a hypothesis test using the Frobenius norm with a test statistic Δ given by

$$\Delta = \frac{\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{A}}\|^2}{\|\hat{\mathbf{H}}_{\mathrm{A}}\|^2} + \frac{(x_0 - 1)\|\mathbf{F} - \bar{\mathbf{F}}\|^2}{\|\bar{\mathbf{F}}\|^2}. \tag{1}$$

If $\Delta < x_1$, the mobile device accepts the packet and uses the higher-layer authentication protocol to further authenticate it. Otherwise, it sends a spoofing alarm. If the packet also passes the higher-layer authentication, Bob updates Alice's channel record with $\hat{\mathbf{H}}_{\mathrm{A}} = \mathbf{H}$.

V. RL BASED PHY AUTHENTICATION

We propose a reinforcement learning based PHY authentication scheme (RLPA) that optimizes the authentication mode and parameter in the dynamic environment via-trial-and-error. The authentication policy to authenticate the k-th packet with Alice's identity $\boldsymbol{x}^{(k)}$ is selected with current state $\boldsymbol{s}^{(k)}$ and the Q-value denoted by $Q(\boldsymbol{s}^{(k)}, :)$ for each authentication policy. The state includes the estimated authentication accuracy and the packet priority. This algorithm uses the Dyna architecture to generate simulated authentication experiences. It also applies transfer learning to reduce the unnecessary exploration in the initial authentication based on the previous authentication experiences in similar VANETs.

Upon receiving the k-th packet under test, mobile device Bob applies the PHY authentication framework as shown in Fig. 1. More specifically, Bob uses the method in [20] to analyze the arrival interval of the packets with Alice's identity and the channel type, and to evaluate the priority of the packet, $\theta \in [0, 1]$, which is uniformly quantized into M_4 levels. Similar to [16], the false alarm rate $\hat{p}_{\mathrm{F}}^{(k-1)}$ is the ratio of the Bob's received packets that are falsely treated as Eve's packets, and the ratio of Eve's packets accepted by Bob during the previous N_{B} packets by mistake is defined as the miss detection rate $\hat{p}_{\mathrm{M}}^{(k-1)}$ [16]. The attack rate $\hat{y}^{(k-1)}$ represents the ratio of Eve's packets among the previous N_{B} packets. For simplicity, the false alarm rate, the miss detection rate and the attack rate are quantized into M_1, M_2 and M_3 levels, respectively.

The state space of RLPA denoted by $\mathbf{S}$ consists of all the quantized false alarm rates and miss detection rates of the PHY authentication in the previous N_{B} packets, the quantized estimated attack rates in the previous N_{B} packets, and the quantized priority levels of the current packet under test, i. e., the state space dimension is $M_1 \times M_2 \times M_3 \times M_4$. The state is formed as $\boldsymbol{s}^{(k)} = [\hat{p}_{\mathrm{F}}^{(k-1)}, \hat{p}_{\mathrm{M}}^{(k-1)}, \hat{y}^{(k-1)}, \theta^{(k)}] \in \mathbf{S}$. The authentication policy $\boldsymbol{x}^{(k)}$ is selected based on the Q-function $Q(\boldsymbol{s}^{(k)}, :)$ and the ε-greedy strategy. The authentication policy set denoted by $\mathbf{A}$ is a $2(X + 1)$-dimensional vector that consists of 2 feasible authentication modes and $X+1$ authentication parameters. Bob then uses the resulting authentication policy $\boldsymbol{x}^{(k)}$ and follows the PHY authentication framework in Fig. 1 to authenticate the k-th packet.

Bob estimates $\hat{p}_{\mathrm{F}}^{(k)}$, $\hat{p}_{\mathrm{M}}^{(k)}$ and $\hat{y}^{(k)}$ of the previous $(N_{\mathrm{B}} - 1)$ packets and packet k with Alice's identity. Bob then evaluates the cost c that indicates the computation and communication overheads of the chosen authentication mode to authenticate the k-th packet and the utility, u, that relies on the packet priority, the authentication cost and authentication accuracy that indicate the authentication accuracy given by

$$u(\boldsymbol{x}, y) = -\,\hat{y}^{(k)}\theta\left(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}}^{(k)} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}^{(k)}\right) - \beta c \tag{2}$$

where ϖ_{F} (ϖ_{M}) is the coefficient of false alarm rate (miss detection rate) in the authentication process, and $\beta \in [0, 1]$ denotes the authentication cost coefficient.

In the authentication process, the learning rate $\alpha \in (0, 1]$ and the discount factor $\delta \in [0, 1]$ are used to update the Q-function with the Bellman iterative equation,

$$\begin{aligned} Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) \leftarrow (1-\alpha)Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) \\ + \alpha\left(u + \delta \max_{\boldsymbol{x}^*\in\mathbf{A}} Q\left(\boldsymbol{s}^{(k+1)}, \boldsymbol{x}^*\right)\right). \end{aligned} \tag{3}$$

The Dyna architecture uses authentication experience including the current state, authentication policy, utility and the next state, i.e., $\{\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, u, \boldsymbol{s}^{(k+1)}\}$, to generate D simulated authentication experiences. The learning model depends on the next occurrence count vector Φ' given by

$$\Phi'\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right) \leftarrow \Phi'\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right) + 1. \tag{4}$$

Bob updates the occurrence count vector Φ in the simulated experience in each real authentication experience using Φ' as

Algorithm 1: RL based PHY authentication scheme

Initialize α, δ, D, N_{A}, N_{B}, $\boldsymbol{s}^{(0)}$, $\Phi\left(\boldsymbol{s}^{(0)}, \boldsymbol{x}^{(0)}\right) = \mathbf{0}$, and $\mathbf{Q} = \mathbf{Q}^*$
for $k = 1, 2, \cdots$ **do**
  Receive packet k
  Extract $\mathbf{H}$ under test
  Evaluate the packet priority θ
  $\boldsymbol{s}^{(k)} = \left[\hat{p}_{\mathrm{F}}^{(k-1)}, \hat{p}_{\mathrm{M}}^{(k-1)}, \hat{y}^{(k-1)}, \theta^{(k)}\right]$
  Choose $\boldsymbol{x}^{(k)}$ with the ε-greedy strategy
  if $x_0^{(k)} > 1$ **then**
    Extract $\bar{\mathbf{F}} = \left[\bar{R}_j; \bar{T}_j\right]_{1 \le j \le N_{\mathrm{A}}}$ of the previous N_{A} ambient packets
    Send request including τ and the ambient radio source set to the node under test
    Obtain $\mathbf{F} = [R_j; T_j]_{1 \le j \le N_{\mathrm{A}}}$ from the node under test
  end if
  Calculate Δ via (1)
  if $\Delta < x_1^{(k)}$ **and** pass the authentication in higher-layer **then**
    Accept packet k
    $\hat{\mathbf{H}}_{\mathrm{A}} = \mathbf{H}$
  else
    Send a spoofing alarm
    Keep the RSSI record $\hat{\mathbf{H}}_{\mathrm{A}}$
  end if
  Estimate $\hat{p}_{\mathrm{F}}^{(k)}$, $\hat{p}_{\mathrm{M}}^{(k)}$, and $\hat{y}^{(k)}$ of the previous $N_{\mathrm{B}} - 1$ packets and current packet k with Alice's identity
  Evaluate utility u via (2)
  Update Q-value $Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)$ via (3)
  Update $\Phi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)$ via (4) and (5)
  Calculate $\Psi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right)$ via (6)
  Calculate $\omega\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)$ via (7)
  for $\eta = 1, 2, \cdots, D$ **do**
    Choose $\left(\hat{\boldsymbol{s}}^{(\eta)}, \hat{\boldsymbol{x}}^{(\eta)}\right)$ at random
    Obtain $\hat{\boldsymbol{s}}^{(\eta+1)}$ via (6)
    Compute $\omega\left(\hat{\boldsymbol{s}}^{(\eta)}, \hat{\boldsymbol{x}}^{(\eta)}\right)$ via (7)
    Update $Q\left(\hat{\boldsymbol{s}}^{(\eta)}, \hat{\boldsymbol{x}}^{(\eta)}\right)$ via (3)
  end for
end for

follows:

$$\Phi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) \leftarrow \sum_{\boldsymbol{s}^{(k+1)} \in \mathbf{S}} \Phi'\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right). \quad (5)$$

According to Φ' in (4) and Φ in (5), the transition probability Ψ from $\boldsymbol{s}^{(k)}$ to $\boldsymbol{s}^{(k+1)}$ in the simulated experience is calculated by

$$\Psi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right) \leftarrow \frac{\Phi'\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}, \boldsymbol{s}^{(k+1)}\right)}{\Phi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)}. \quad (6)$$

By definition, let ω be the average over all the previous real authentication experiences, calculated by

$$\omega\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) \leftarrow \frac{1}{\Phi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)} \sum_{\varsigma=1}^{\Phi\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)} u. \quad (7)$$

The Q-value is updated D more times, each based on a chosen $(\hat{\boldsymbol{s}}^{(\eta)}, \hat{\boldsymbol{x}}^{(\eta)})$ whose Q-value is updated via the iterated Bellman equation in (3), and $\hat{\boldsymbol{s}}^{(\eta+1)}$ is selected with transition probability Ψ in (6). Bob exploits the authentication experiences from Γ similar VANET scenarios to initialize the Q-value as $\mathbf{Q}^*$, where each scenario trains K packets, as shown Algorithm 1.

VI. Deep RL based PHY Authentication

We design a deep reinforcement learning based PHY authentication named DRLPA to save the optimization time and improve the authentication performance for a mobile device with more computing resources. More specifically, reinforcement learning, deep learning, and transfer learning are used to choose the authentication policy for the k-th packet. This scheme uses a CNN and a DND similar to the NEC algorithm in [25] to reduce the dependence on the authentication experiences, also, it applies transfer learning to initialize the CNN weights.

Upon receiving the k-th packet with the Alice's identity, mobile device Bob evaluates packet priority $\theta^{(k)}$ and the authentication accuracy of the previous N_{B} packets to formulate the state $\boldsymbol{s}^{(k)} = [\hat{p}_{\mathrm{F}}^{(k-1)}, \hat{p}_{\mathrm{M}}^{(k-1)}, \hat{y}^{(k-1)}, \theta^{(k)}]$. Bob chooses his authentication policy based on current state sequence $\boldsymbol{\varphi}^{(k)}$ including the previous E state and authentication policies, i.e., $\boldsymbol{\varphi}^{(k)} = \{\boldsymbol{s}^{(k-E)}, \boldsymbol{x}^{(k-E)}, \boldsymbol{s}^{(k-E-1)}, \boldsymbol{x}^{(k-E-1)}, \cdots, \boldsymbol{x}^{(k-1)}, \boldsymbol{s}^{(k)}\}$. The previous state sequence $\boldsymbol{\varphi}^{(k-1)}$, authentication policy $\boldsymbol{x}^{(k-1)}$, utility $u^{(k-1)}$, and current state sequence $\boldsymbol{\varphi}^{(k)}$ are used to form the k-th authentication experience $\boldsymbol{e}^{(k)} = \{\boldsymbol{\varphi}^{(k-1)}, \boldsymbol{x}^{(k-1)}, u^{(k-1)}, \boldsymbol{\varphi}^{(k)}\}$ that is saved in an experience pool $\mathcal{D}$.

Similar to [25], a target CNN is used to reduce the correlation between the target Q-value and the current Q-value and thus reduce oscillations or divergence of the authentication policies. The data mining results from the Γ similar VANET authentication scenarios are used to initialize the CNN weights denoted by $\boldsymbol{\phi}^{(k)}$. Bob updates CNN weights $\boldsymbol{\phi}^{(k)}$ each time slot by using the experience reply technique and the stochastic gradient descent (SGD) algorithm to avoid local minima and reduce the computational complexity in the PHY authentication process. Specifically, Z authentication experiences $\{\boldsymbol{a}^{(z)}\}_{1 \le z \le Z}$ are randomly selected from the authentication experience pool $\mathcal{D}$ to formulate a minibatch $\mathcal{B} = \{\boldsymbol{a}^{(z)}\}_{1 \le z \le Z}$. Bob minimizes the loss function representing the mean squared error of the target Q-value to update $\boldsymbol{\phi}$ given by,

$$\boldsymbol{\phi} = \arg\min_{\boldsymbol{\phi}^*} \mathbb{E}_{\mathcal{B}}\Bigg[\Bigg(u + \delta \max_{\boldsymbol{x}^* \in \mathbf{A}} Q\left(\boldsymbol{\varphi}^{(k+1)}, \boldsymbol{x}^*; \boldsymbol{\phi}\right) - Q(\boldsymbol{\varphi}^{(k)}, \boldsymbol{x}; \boldsymbol{\phi}^*)\Bigg)^2\Bigg] \quad (8)$$

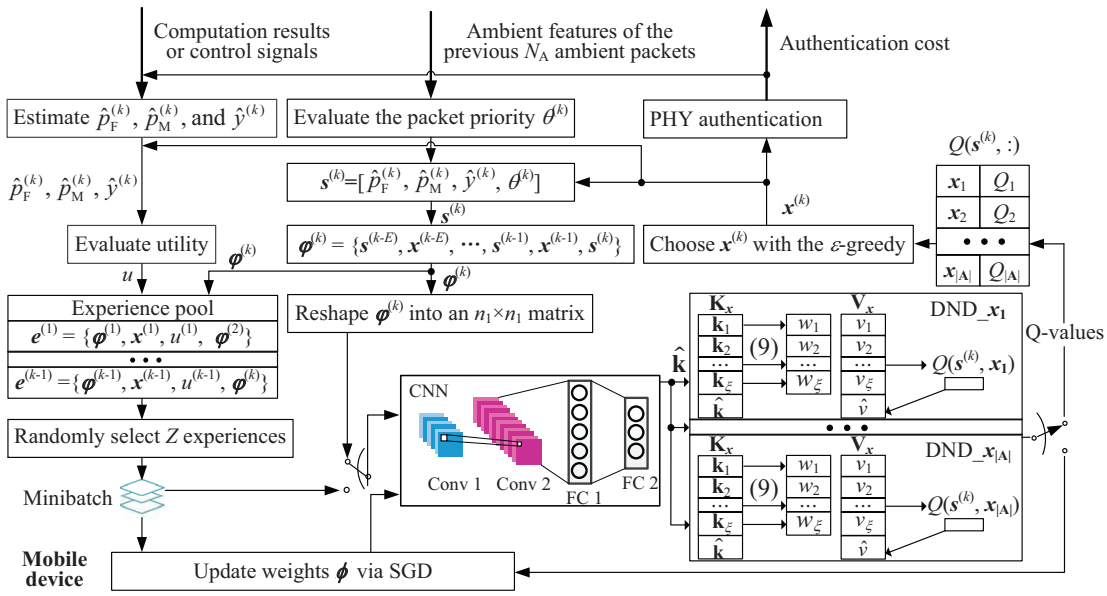


Fig. 2. Deep RL based PHY authentication for VANETs.

Algorithm 2: Deep RL based PHY authentication

1: Initialize E, Z, N_A, N_B, ξ, $\boldsymbol{s}^{(0)}$, $\boldsymbol{\phi}^{(0)} = \boldsymbol{\phi}^*$, and $\mathcal{D} = \emptyset$
2: **for** $k = 1, 2, \cdots$ **do**
3: Receive packet k
4: Extract $\mathbf{H}$ under test
5: Evaluate the packet priority θ
6: $\boldsymbol{s}^{(k)} = \left[\hat{p}_F^{(k-1)}, \hat{p}_M^{(k-1)}, \hat{y}^{(k-1)}, \theta^{(k)}\right]$
7: **if** $k \leq E$ **then**
8: Select $\boldsymbol{x}^{(k)}$ at random
9: **else**
10: $\varphi^{(k)} = \left\{\boldsymbol{s}^{(k-E)}, \boldsymbol{x}^{(k-E)}, \cdots, \boldsymbol{x}^{(k-1)}, \boldsymbol{s}^{(k)}\right\}$
11: $\mathcal{D} \leftarrow \left\{\boldsymbol{\varphi}^{(k-1)}, \boldsymbol{x}^{(k-1)}, u^{(k-1)}, \boldsymbol{\varphi}^{(k)}\right\} \cup \mathcal{D}$
12: Select Z experiences $\left\{\boldsymbol{a}^{(z)}\right\}_{1 \leq z \leq Z}$ randomly
13: Update $\boldsymbol{\phi}^{(k)}$ via (8)
14: CNN input, $\boldsymbol{\varphi}^{(k)}$ and $\boldsymbol{\phi}^{(k)}$
15: CNN output, $\hat{\mathbf{k}}$
16: Calculate w_χ via (9)
17: $Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) = \sum_{\chi=1}^{\xi} w_\chi v_\chi$
18: **if** $\hat{\mathbf{k}} \in \mathbf{K}_{\boldsymbol{x}}$ **then**
19: Update the Q-value with $Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)$ in $\mathbf{V}_{\boldsymbol{x}}$
20: **else**
21: $\mathbf{K}_{\boldsymbol{x}} = \mathbf{K}_{\boldsymbol{x}} \cup \hat{\mathbf{k}}$
22: $\mathbf{V}_{\boldsymbol{x}} = \mathbf{V}_{\boldsymbol{x}} \cup Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right)$
23: **end if**
24: Choose $\boldsymbol{x}^{(k)} = \left\{x_0^{(k)}, x_1^{(k)}\right\}$ with the ε-greedy strategy
25: **end if**
26: PHY authentication as shown in steps 8-22 in Algorithm 1
27: **end for**

Bob reshapes state sequence $\boldsymbol{\varphi}^{(k)}$ into an $n_1 \times n_1$ matrix and inputs the matrix to the CNN whose weights are updated with $\boldsymbol{\phi}^{(k)}$.

The CNN includes 2 convolutional (Conv) layers whose activation function is the rectified linear units (ReLUs) and 2 fully connected (FC) layers, as shown in Fig. 2. More specifically, Conv 1 has f_1 filters each with $n_2 \times n_2$ sizes and stride s_1, and its output is sent to Conv 2 that has f_2 filters each with size $n_3 \times n_3$ and stride s_2. The output of Conv 2 that has $f_2(n_1 - n_2 - n_3 + 2)^2$ feature maps is reshaped into an $f_2((n_1 - n_2)/s_1 s_2 - (n_3 - 1)/s_2 + 1)^2$-dimensional vector and then input to the first FC layer that has r_1 ReLUs. The output of FC 1 is sent to the second FC layer that outputs a key denoted by $\hat{\mathbf{k}}$ with $2(X+1)$ dimensions for the feasible authentication policies. The key, $\hat{\mathbf{k}}$, is input to all the $2(X+1)$ DNDs that depend on the 2 feasible authentication modes and the $X+1$ feasible authentication parameters.

Let $\mathbf{K}_{\boldsymbol{x}}$ be an array that saves the keys to lookup the estimated Q-values that are saved in a database vector denoted by $\mathbf{V}_{\boldsymbol{x}}$. Let $\mathbf{k}_\chi$ denote the χ-th vector in $\mathbf{K}_{\boldsymbol{x}}$ of the DND and v_χ be the χ-th element in the vector $\mathbf{V}_{\boldsymbol{x}}$, with $1 \leq \chi \leq \xi$, and $\xi \leq k$. Both key vector $\mathbf{K}_{\boldsymbol{x}}$ and database $\mathbf{V}_{\boldsymbol{x}}$ are saved in the DND memory module, i.e., $\text{DND} \leftarrow \text{DND} \cup (\mathbf{K}_{\boldsymbol{x}}, \mathbf{V}_{\boldsymbol{x}})$. Similar to [25], the previous L keys in $\mathbf{K}_{\boldsymbol{x}}$ is chosen to calculate the weight w_χ for the χ-th Q-value in $\mathbf{V}_{\boldsymbol{x}}$ with a Gaussian kernel as follows:

$$w_\chi = \frac{\exp\left(\frac{-\|\hat{\mathbf{k}} - \mathbf{k}_\chi\|_2^2}{2}\right)}{\sum_{l=1}^{L} \exp\left(\frac{-\|\hat{\mathbf{k}} - \mathbf{k}_l\|_2^2}{2}\right)}. \tag{9}$$

The Q-value in $\mathbf{V}_{\boldsymbol{x}}$ is updated with $Q\left(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)}\right) = \sum_{\chi=1}^{\xi} w_\chi v_\chi$ if the key $\hat{\mathbf{k}}$ has already existed in $\mathbf{K}_{\boldsymbol{x}}$; Otherwise, the DND updates the array with $\mathbf{K}_{\boldsymbol{x}} = \mathbf{K}_{\boldsymbol{x}} \cup \hat{\mathbf{k}}$

and the database with $\mathbf{V}_{\boldsymbol{x}} = \mathbf{V}_{\boldsymbol{x}} \cup Q(\boldsymbol{s}^{(k)}, \boldsymbol{x}^{(k)})$. Similar to Algorithm 1, the mobile device chooses authentication policy $\boldsymbol{x}^{(k)}$ similar to Algorithm 1, and evaluates the utility via (2) after performing the PHY authentication process.

VII. Performance Analyses

We analyze the computational complexity of the RL based PHY authentication scheme and the deep RL based PHY authentication scheme according to the RL complexity analyzed in [40] and the CNN complexity evaluated in [41]. The performance bound of the utility of the mobile device is provided based on the game theoretic study of the authentication process.

The computational complexity of the deep RL based PHY authentication scheme depends on the number of the received packets k, the sampled length of the experiences Z and the complexity of the CNN. For simplicity, let f_0 be the number of the CNN input channels, and $\mathcal{M}_\iota$ is the output feature map of Conv ι, with $\iota = 1$ and 2.

Theorem 1. *The computational complexity of the RL based PHY authentication scheme and the deep RL based PHY authentication scheme at time k are given by $O(kXD)$ and $O\left(kZf_1f_2n_1^2n_3^2\right)$.*

Proof: See Appendix A. ■

Remark: The RL based PHY authentication scheme has a computational complexity that increases with the number of the learning samples k, the number of the feasible test threshold levels X, the length of simulated experiences D. The computational complexity of the deep RL based PHY authentication scheme in Algorithm 2 relies on the number of the learning samples k, the sampled length of the experiences Z, the reshaped input size n_1, the filter numbers of Conv 1 and Conv 2 f_1 and f_2, and the filter size of Conv 2 n_3.

The performance bound of the RL based PHY authentication algorithm can be provided based on the Stackelberg equilibrium (SE) of the authentication game under various network scenarios. In this game, the defender Bob as the leader first chooses both the authentication mode x_0 and test threshold x_1 to maximize its utility given by (2), while attacker Eve as the follower selects her attack rate y to maximize its utility, chosen as $-u - yC_{\mathrm{R}}$. In this case, we have the following result.

Theorem 2. *The performance bound of the RL based PHY authentication scheme is given by*

$$u = \begin{cases} -\beta, & \text{if} \quad C_{\mathrm{R}} \geq \theta(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}) \quad (10a) \\ -Y_{\mathrm{MAX}}\theta(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}) - \beta, & \text{o.w.} \quad (10b) \end{cases}$$

Proof: The attack rate, y, is chosen to maximize $u_E = -u - yC_{\mathrm{R}}$, if $C_{\mathrm{R}} \geq \theta(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}})$, we have

$$\frac{\partial u_{\mathrm{E}}}{\partial y} = \theta\left(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}\right) - C_{\mathrm{R}} \leq 0 \tag{11}$$

and $\forall y \in [0, Y_{\mathrm{MAX}}]$ and $\boldsymbol{x} \in \mathbf{A}$, we have

$$\begin{aligned} u_{\mathrm{E}}(\boldsymbol{x}, 0) &= \beta x_0^2 \\ &\geq y\theta\left(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}\right) + \beta x_0^2 - yC_{\mathrm{R}} = u_{\mathrm{E}}(\boldsymbol{x}, y). \end{aligned} \tag{12}$$

Thus,

$$0 = \arg \max_{y \in [0, Y_{\mathrm{MAX}}]} u_{\mathrm{E}}(\boldsymbol{x}, y) \tag{13}$$

holds for $y^* = 0$.

By (2), we have $u(\boldsymbol{x}, 0) = -\beta x_0^2$ and $\partial u(\boldsymbol{x}, 0)/\partial x_0 = -2\beta x_0 \leq 0$. Thus, we have

$$u\left([1, x_1^*], 0\right) = -\beta \geq -\beta x_0^2 = u(\boldsymbol{x}, 0). \tag{14}$$

Hence,

$$[1, x_1^*] = \arg \max_{\boldsymbol{x} \in \mathbf{A}} u(\boldsymbol{x}, 0). \tag{15}$$

By (12) and (15), we have

$$u_{\mathrm{E}}\left([1, x_1^*], 0\right) \geq u_{\mathrm{E}}(\boldsymbol{x}, y). \tag{16}$$

Thus, $([1, x_1^*], 0)$ is a SE of the game.

Therefore, by (2), we have $u = -\beta$. Similarly, we have the utility given by $-Y_{\mathrm{MAX}}\theta(\varpi_{\mathrm{F}}\hat{p}_{\mathrm{F}} + \varpi_{\mathrm{M}}\hat{p}_{\mathrm{M}}) - \beta$ in (10b). ■

Remark: In (10a), the mobile device chooses the light-weight authentication mode with lower security cost, and the rogue edge does not send spoofing signals due to the high spoofing cost and/or authentication accuracy. Otherwise, the utility, u, decreases with the authentication cost and the packet priority as shown in (10b). In this case, the rogue edge sends the maximum number of spoofing packets due to the low spoofing cost.

VIII. Simulation and Experimental Results

A. Simulation Results

Simulations were performed to evaluate the authentication performance of the physical authentication framework with an initial network topology illustrated in Fig. 3. The channel data used as the authentication basis in the simulations were generated based on the channel model [13] for the Alice-Bob link, and the two-ray VANET model [42] for the Eve-Bob channel. The channels between ambient radio sources and Bob/Alice/Eve follow the log-normal shadowing model [23]. Each authentication accuracy in Fig. 4 was an averaged result of 200 authentications over 2.44 s.

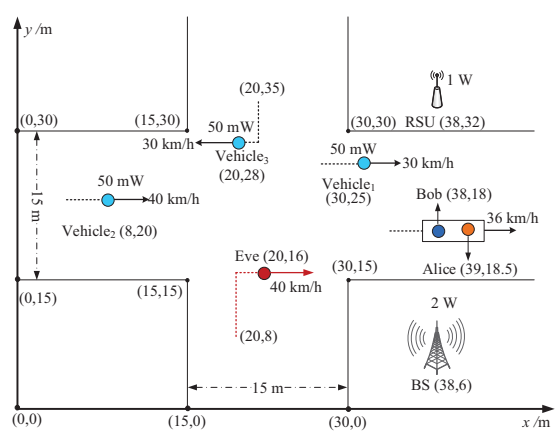


Fig. 3. Initial network topology in the simulations setting in meter, and the initial ambient radio sources of the mobile device including a BS, an RSU, and three vehicles.

Both Bob and Alice were located in the same vehicle that initially moved at speed 36 km/h from location (38, 18) m along a road. Bob offloads the computation-intensive tasks such as virtual reality games to the serving edge Alice.

TABLE II
CNN PARAMETERS OF THE PHY AUTHENTICATION IN ALGORITHM 2

Layer size	Conv 1	Conv 2	FC 1	FC 2
Input size	$1 \times 8 \times 8$	$20 \times 5 \times 5$	360	180
Filter size	4×4	3×3	/	/
Stride	1	1	/	/
No. filters	20	40	180	16
Output	$20 \times 5 \times 5$	$40 \times 3 \times 3$	180	16

Alice processes the computation-intensive tasks from Bob and sends the computation results to Bob with transmit power 100 mW. Eve that is initially located 20.6 m away from that vehicle moves at speed 40 km/h. Bob uses USRP to estimate the channel state of received packets and determines whether the packet with Alice's identity was indeed sent by Alice or Eve.

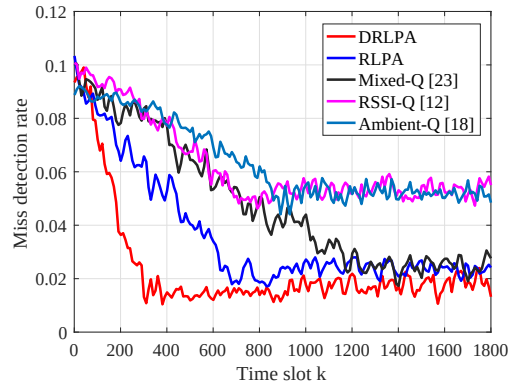
Bob receives signals from 5 ambient radio sources initially, including 3 vehicles each with transmit power 50 mW, a BS that is located 12 m away from Bob with transmit power 2 W and an RSU that is located 14 m away from Bob with transmit power 1 W. The 3 ambient vehicles are initially located 10.6, 24.8 and 30.1 m away from Bob and move at speeds 30, 40 and 30 km/h, respectively.

The attacker Eve sends spoofing signals at power 100 mW with Alice's identity to spoof Bob. With the goal to minimize Bob's utility u in (2) with less spoofing cost, Eve applies Q-learning to choose her attack rate to send the spoofing packets in a time slot. Similar to [43], the additive white Gaussian noise with variance 0.1 is assumed in the simulations.

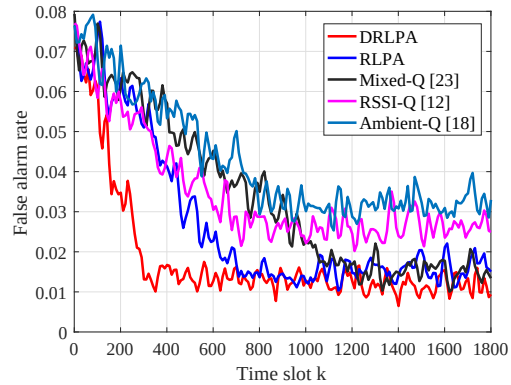
Unless specified otherwise, the authentication process uses 0.5 (or 1) false alarm (or miss detection) weight factor, with 0.01 authentication cost coefficient, 0.1 random exploration probability, 0.7 learning rate and 0.5 discount factor based on the simulation results not shown here for high average authentication accuracy. The test threshold of RLPA is chosen from 0 to 0.35 with 8 levels, yielding 16 feasible authentication policies. The Q-values in the learning algorithm are set to be the convergence values of reinforcement learning after 5 authentication experiences, each lasting 200 time slots in similar scenarios. According to Algorithm 1, the Q-values are updated with 3 simulated experiences and one real experience every time slot afterwards.

The experience pool can store at most 1800 authentication experiences, each DND stores 10 previous keys, the state sequence consists of 10 state action pairs, and the minibatch has 32 sample experiences. FC1 has 360 ReLUs as a tradeoff between the computational complexity and the authentication accuracy, similar to [26]. FC 2 outputs a 16-dimensional vector corresponding to the Q-values of each authentication policy. The CNN architecture parameters in Algorithm 2 are summarized in Table II.

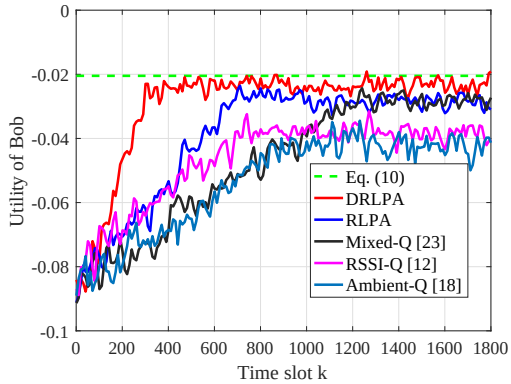
As shown in Fig. 4, the proposed RLPA improves the authentication performance as compared with mixed-Q in [23], RSSI-Q in [12] and ambient-Q in [18]. For instance, RLPA improves the authentication accuracy by 3.3%, 3.5% and 4.4% at time slot 800 compared with mixed-Q, RSSI-Q, and ambient-Q, respectively. That's because the Q-value in



(a) Miss detection rate



(b) False alarm rate

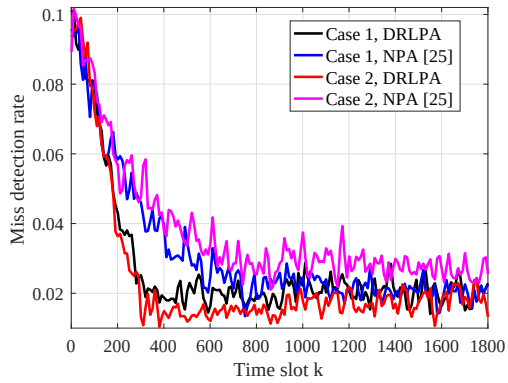


(c) Utility of Bob

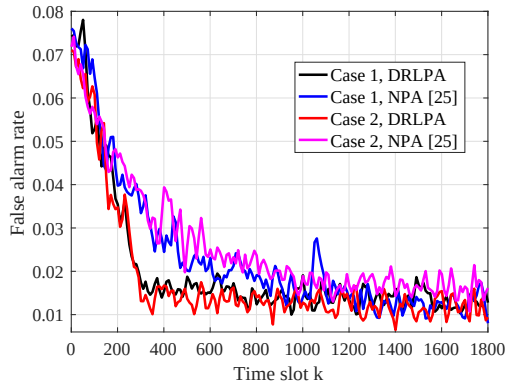
Fig. 4. PHY authentication performance in the VANET with reinforcement learning as shown in Fig. 3, with $\varpi_{\mathrm{F}} = 0.5$, $\varpi_{\mathrm{M}} = 1$, $\beta = 0.01$, $\alpha = 0.7$, and $\delta = 0.5$.

RLPA is updated with more frequently simulated experiences formulated from each practical authentication experience to accelerate the PHY authentication. Due to the flexible control of the authentication mode and parameters, the authentication performance can be further improved by DRLPA, which decreases the miss detection rate by 51.7% and false alarm rate by 38.5%, respectively. The mobile device can converge to the performance bound given by Theorem 2 after 500 time slots, which is 64.3%, 50%, and 58.3% faster than mixed-Q, RSSI-Q, and ambient-Q, respectively.

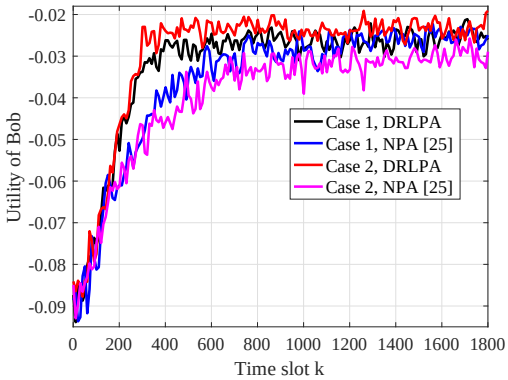
As shown in Fig. 5, DRLPA is robust against the smart



(a) Miss detection rate

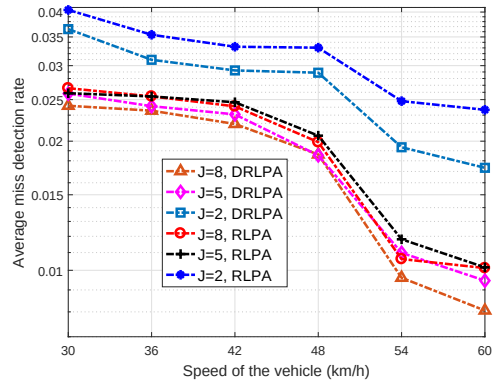


(b) False alarm rate

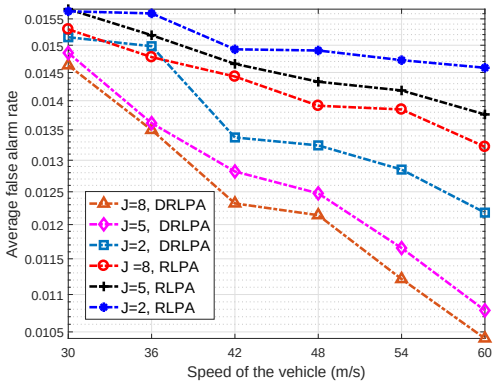


(c) Utility of Bob

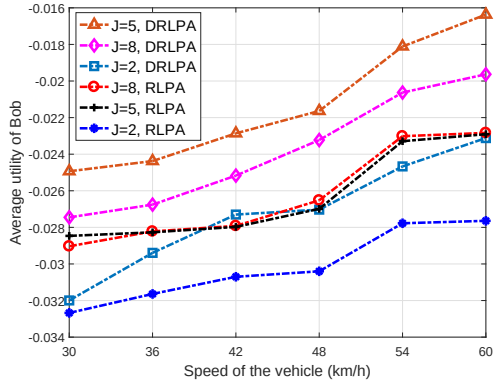
Fig. 5. Performance of the deep RL based authentication with the VANET topology as shown in Fig. 3 to resist a faked edge attacker who uses a fixed attack rate in Case 1 and a smart attacker who applies Q-learning to choose the attack rate in Case 2.



(a) Average miss detection rate



(b) Average false alarm rate

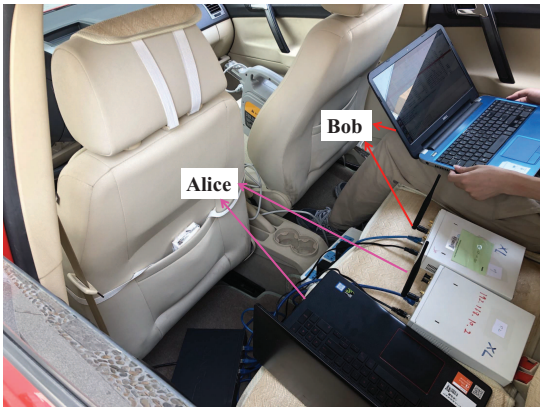


(c) Average utility of Bob

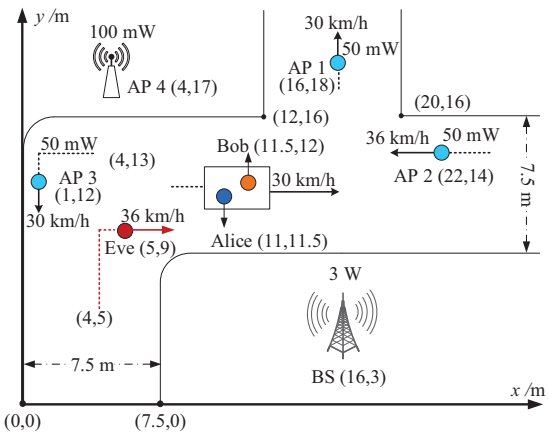
Fig. 6. Average performance of the PHY authentication with speed and the number of the initial ambient radio sources, with $\varpi_{\mathrm{F}} = 0.5$, $\varpi_{\mathrm{M}} = 1$, $\beta = 0.01$, $\alpha = 0.7$, and $\delta = 0.5$.

attackers who apply Q-learning to choose the attack rate. For example, DRLPA decreases the miss detection rate by 64.7% and the false alarm rate by 49.1% at time slot 600, and saves 58.3% convergence time as compared with PHY authentication with standard NEC (NPA) in [25]. The improvement is achieved by applying transfer learning to exploit the previous authentication experiences in similar VANETs to initialize the CNN weights and the learning parameters such as the learning rate, which accelerates the optimization over that of NPA.

As shown in Fig. 6, DRLPA is robust against the interference from the ambient radio signals. For example, DRLPA decreases the miss detection rate by 50.6% and false alarm rate by 12.5% at speed 54 km/h, respectively, as the number of ambient radio resources increases from 2 to 8. The speed of the vehicle holds both Bob and Alice $v_1 \in \{30, 36, 42, 48, 54, 60\}$ km/h improves the authentication performance. For instance, DRLPA decreases the miss detection rate by 63.4% and the false alarm rate 27.6% in the PHY authentication framework, respectively, as the vehicle speed increases from 30 km/h to 60 km/h. That is because the difference between the relative static Alice-Bob link and



(a) VANET edge offloading snapshot



(b) Initial network topology

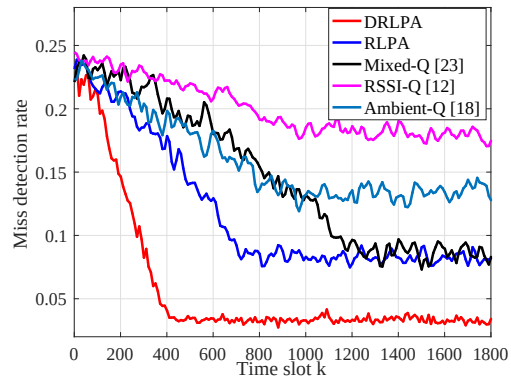
Fig. 7. Experimental settings in which both Bob and Alice a vehicle resist Eve outside the vehicle, and both Alice and Eve sent signals using USRP with Alice's identity.

the Eve-Bob link increases with the moving speed of the vehicle that carries both mobile device Bob and edge node Alice.

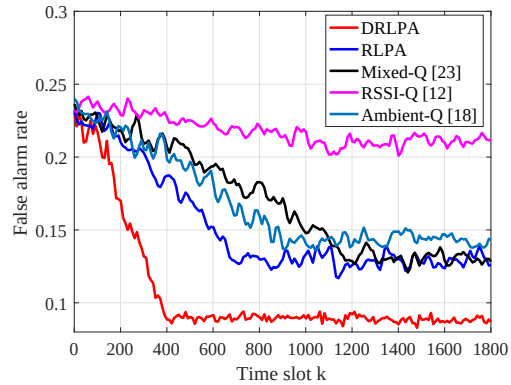
B. Experimental Results

The authentication performance of RLPA and DRLPA was evaluated via experiments, as shown in Fig. 7. A mobile device Bob receives signals from serving edge node Alice (a USRP-aided laptop) in the same vehicle against a rogue edge Eve outside the vehicle. In most cases, both Alice and Bob receive ambient radio signals from 5 sources including 4 APs and a BS. The vehicle that carries both Alice and Bob initially moved at speed 30 km/h from location (11.5,12) m along a road, as shown in Fig. 7. Bob offloaded some computation tasks to Alice that returned the computation results with transmit power 100 mW afterward. Each of the 3 moving ambient APs (wih moving speeds 30, 40 and 30 km/h) transmitted with power 50 mW, a BS located at (16,3) m had transmit power 3 W and a static AP at (4,17) m transmitted with power 100 mW. Each authentication performance in Fig. 8 was an average result of 200 authentications over 5.22 s.

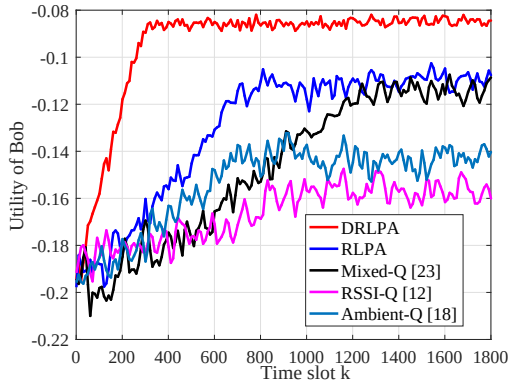
Eve initially located at (5,9) m moves at speed 36 km/h and sends spoofing signals with transmit power 100 mW during



(a) Miss detection rate



(b) False alarm rate



(c) Utility of Bob

Fig. 8. Experimental performance in the VANET topology as shown in Fig. 7, with $\varpi_{\mathrm{F}} = 0.5$, $\varpi_{\mathrm{M}} = 1$, $\beta = 0.01$, $\alpha = 0.7$, and $\delta = 0.5$.

Alice transmission intervals. With the goal to minimize Bob's utility with less spoofing cost, Eve applies Q-learning to choose her attack rate between 0 and 0.9 with 3 levels. According to the chosen attack rate, Eve uses the MAC protocols such as the carrier sense multiple access to monitor the transmission channel, and sends the spoofing packets without collisions with Alice in a time slot.

As shown in Fig. 8, the proposed DRLPA achieves higher authentication accuracy and accelerates the optimization speed in the authentication process compared with the benchmarks such as mixed-Q [23], RSSI-Q [12] and ambient-Q [18]. For instance, our proposed DRLPA decreases the

miss detection rate by 79.4% and false alarm rate by 52.3% compared with mix-Q at time slot 800, because DRLPA uses transfer learning technique and the CNN to accelerate the optimization speed in the authentication process to improve the authentication accuracy. DRLPA improves the authentication accuracy with lower energy consumption compared with ambient-Q, e.g., DRLPA improves 10.8% authentication accuracy compared with ambient-Q at time slot 800.

IX. Conclusion

In this paper, we have proposed a PHY authentication framework that uses the PHY properties of both the received signals and the ambient radio signals against rogue edge attacks for VANETs. This framework applies RL to optimize the authentication policy without requiring to know the packet generation model, the VANET channel model, and the spoofing model and a deep RL based authentication algorithm for the mobile device with more computation resources. The proposed authentication algorithms can reach the provided convergence performance bound. Simulation and experimental results have been provided to demonstrate performance gains over the benchmark schemes such as mixed-Q in [23], RSSI-Q in [12] and ambient-Q in [18]. For example, DRLPA decreases the miss detection rate by 76.08% and false alarm rate by 70.44% after 1000 time slots compared with mixed-Q.

Appendix A
Proof of Theorem 1

Proof: According to [40], the computational complexity of mixed-Q in [23] is given by $O\left(kX\right)$. Compared with mixed-Q, the RL based PHY authentication scheme updates the Q-values with D more times. Thus, the computational complexity of the RL based PHY authentication scheme is given by $O\left(kXD\right)$.

According to the CNN architecture in Algorithm 2, by [25], we have

$$n_2^2(n_1-n_2+1)^2 \ll f_2 n_3^2(n_1-n_2-n_3+2)^2. \tag{17}$$

Thus, similar to [40] and [41], as $s_1=s_2=1$ and $n_1>n_2>n_3$, the computational complexity of the deep RL based PHY authentication scheme is given by

$$\begin{aligned}
&O\left(kZ\sum_{\iota=1}^{2} f_{\iota-1}f_{\iota}n_{\iota+1}^2\mathcal{M}_{\iota}\right)\\
&=O\left(kZf_1\left(n_2^2\left(\frac{n_1-n_2}{s_1}+1\right)^2\right.\right.\\
&\qquad\left.\left.+f_2n_3^2\left(\frac{n_1-n_2}{s_1s_2}-\frac{n_3-1}{s_2}+1\right)^2\right)\right) &(18)\\
&=O\left(kZf_1\left(n_2^2(n_1-n_2+1)^2\right.\right.\\
&\qquad\left.\left.+f_2n_3^2(n_1-n_2-n_3+2)^2\right)\right) &(19)\\
&=O\left(kZf_1f_2n_3^2(n_1-n_2-n_3+2)^2\right) &(20)\\
&=O\left(kZf_1f_2n_3^2(n_1^2-2n_1(n_2+n_3-2)\right.\\
&\qquad\left.+(n_2+n_3-2)^2)\right) &(21)\\
&=O\left(kZf_1f_2n_1^2n_3^2\right). &(22)
\end{aligned}$$

■

References

[1] K. Zhang, S. Leng, X. Peng, P. Li, S. Maharjan, and Y. Zhang, "Artificial intelligence inspired transmission scheduling in cognitive vehicular communications and networks," *IEEE Internet of Things J.*, vol. 6, no. 2, pp. 1987–1997, Apr. 2019.

[2] H. Peng, Q. Ye, and X. Shen, "Spectrum management for multi-access edge computing in autonomous vehicular networks," *IEEE Trans. Intelligent Transportation Systems*, vol. 1, no. 1, pp. 1–12, Jun. 2019.

[3] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, Aug. 2017.

[4] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, "Mobile edge computing for vehicular networks: A promising network paradigm with predictive offloading," *IEEE Veh. Technol. Mag.*, vol. 12, no. 2, pp. 36–44, Jun. 2017.

[5] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV empowered future IoV with enhanced communication, computing, and caching," *Proceedings of the IEEE*, vol. 1, no. 1, pp. 1–18, Nov. 2019.

[6] L. Xiao, W. Zhuang, S. Zhou, and C. Chen, *Learning-based VANET Communication and Security Techniques*. Springer, 2019.

[7] S. Woo, J. Jin, and L. Hoon, "A practical wireless attack on the connected car and security protocol for in-vehicle CAN," *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 2, pp. 993–1006, Apr. 2015.

[8] W. Si, D. Starobinski, and M. Laifenfeld, "A robust load balancing and routing protocol for intra-car hybrid wired/wireless networks," *IEEE Trans. Mobile Computing*, vol. 18, no. 2, pp. 250–263, Feb. 2019.

[9] K. Zaidi, M. Milojevic, V. Rakocevic, A. Nallanathan, and M. Rajarajan, "Host-based intrusion detection for VANETs: A statistical approach to rogue node detection," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6703–6714, Aug. 2016.

[10] B. Yu, C. Xu, and B. Xiao, "Detecting Sybil attacks in VANETs," *J. Parallel and Distributed Comput.*, vol. 73, no. 6, pp. 746–756, Jun. 2013.

[11] T. Zhou, C. Roy, P. Ning, and K. Chakrabarty, "P2DAP-Sybil attacks detection in vehicular ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 582–594, Mar. 2011.

[12] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016.

[13] L. Xiao, X. Wan, and Z. Han, "PHY-layer authentication with multiple landmarks with reduced overhead," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1676–1687, Mar. 2018.

[14] C. Pei, N. Zhang, X. Shen, and J. Mark, "Channel-based physical layer authentication," in *Proc. IEEE Global Commun. Conf.*, pp. 4114–4119, Austin, TX, Dec. 2014.

[15] L. Shi, M. Li, S. Yu, and J. Yuan, "BANA: Body area network authentication exploiting channel characteristics," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1803–1816, Sept. 2013.

[16] L. Xiao, G. Sheng, X. Wan, W. Su, and P. Cheng, "Learning-based PHY-layer authentication for underwater sensor networks," *IEEE Commun. Letters*, vol. 23, no. 1, pp. 60–63, Jan. 2019.

[17] L. Xiao, Q. Yan, W. Lou, G. Chen, and Y. Hou, "Proximity-based security techniques for mobile users in wireless networks," *IEEE Trans. Inf. Forensics & Security*, vol. 8, no. 12, pp. 2089–2100, Oct. 2013.

[18] J. Liu, L. Xiao, G. Liu, and Y. Zhao, "Active authentication with reinforcement learning based on ambient radio signals," *Springer Multimedia Tools and Applications*, vol. 76, no. 3, pp. 3979–3998, Oct. 2015.

[19] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam, "Proximate: Proximity-based secure pairing using ambient wireless signals," in *Proc. ACM Int'l Conf. Mobile Systems, Applications, and Services (MobiSys)*, pp. 211–224, New York, NY, Jun. 2011.

[20] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance modeling of message dissemination in vehicular ad hoc networks with priority," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 61–71, Jan. 2011.

[21] C. Shao, S. Leng, Y. Zhang, and H. Fu, "A multi-priority supported medium access control in vehicular ad hoc networks," *Computer Commun.*, vol. 39, pp. 11–21, Feb. 2014.

[22] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5948–5956, Dec. 2009.

[23] X. Lu, X. Wan, L. Xiao, Y. Tang, and W. Zhuang, "Learning-based rogue edge detection in VANETs with ambient radio signals," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, Kansas City, MO, May 2018.

[24] A. Wasef, Y. Jiang, and X. Shen, "DCS: An efficient distributed-certificate-service scheme for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 533–549, Jul. 2009.

[25] A. Pritzel, B. Uria, S. Srinivasan, *et al.*, "Neural episodic control," in *Proc. Int'l Conf. Machine Learning (ICML)*, pp. 2827–2836, Sydney, Australia, Aug. 2017.

[26] M. Min, L. Xiao, Y. Chen, P. Cheng, D. Wu, and W. Zhuang, "Learning-based computation offloading for IoT devices with energy harvesting," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1930–1941, Feb. 2019.

[27] K. Zhang, Y. Zhu, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Deep learning empowered task offloading for mobile edge computing in urban informatics," *IEEE Internet of Things J.*, vol. 6, no. 5, pp. 7635–7647, Apr. 2019.

[28] D. Huang, S. Misra, M. Verma, and G. Xue, "PACP: An efficient pseudonymous authentication-based conditional privacy protocol for VANETs," *IEEE Trans. Intelligent Transportation Systems*, vol. 12, no. 3, pp. 736–746, Sept. 2011.

[29] R. Lu, X. Lin, T. Luan, X. Liang, and X. Shen, "Pseudonym changing at social spots: An effective strategy for location privacy in VANETs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 86–96, Jan. 2012.

[30] A. Abdallah and X. Shen, "Lightweight authentication and privacy-preserving scheme for V2G connections," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2615–2629, Mar. 2017.

[31] S. Basudan, X. Lin, and K. Sankaranarayanan, "A privacy-preserving vehicular crowdsensing-based road surface condition monitoring system using fog computing," *IEEE Internet of Things J.*, vol. 4, no. 3, pp. 772–782, Jun. 2017.

[32] A. Chan and J. Zhou, "Cyber-physical device authentication for the smart grid electric vehicle ecosystem," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 7, pp. 1509–1517, Jul. 2014.

[33] H. Han, F. Xu, C. Tan, Y. Zhang, and Q. Li, "VR-defender: Self-defense against vehicular rogue APs for drive-thru Internet," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3927–3934, Oct. 2014.

[34] J. Xu, L. Duan, and R. Zhang, "Fundamental rate limits of physical layer spoofing," *IEEE Wireless Commun. Letters*, vol. 6, no. 2, pp. 154–157, Apr. 2017.

[35] J. Xu, L. Duan, and R. Zhang, "Transmit optimization for symbol-level spoofing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 41–55, Jan. 2018.

[36] A. Ferrante, N. Laurenti, C. Masiero, M. Pavon, and S. Tomasin, "On the error region for channel estimation-based physical layer authentication over Rayleigh fading," *IEEE Trans. Inf. Forensics & Security*, vol. 10, no. 5, pp. 941–952, May 2015.

[37] G. R. Hiertz, D. Denteneer, S. Max, *et al.*, "IEEE 802.11s: The WLAN mesh standard," *IEEE Wireless Commun.*, vol. 17, no. 1, pp. 104–111, Feb. 2010.

[38] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[39] Y. Kawamoto, H. Nishiyama, N. Kato, Y. Shimizu, A. Takahara, and T. Jiang, "Effectively collecting data for the location-based authentication in Internet of Things," *IEEE Systems J.*, vol. 11, no. 3, pp. 1403–1411, Sept. 2017.

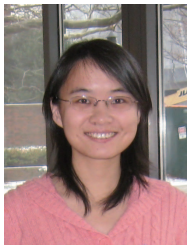
[40] C. Jin, Z. Allen-Zhu, S. Bubeck, and M. I. Jordan, "Is Q-learning provably efficient?" in *Proc. Conf. Advances in Neural Inf. Processing Systems (NIPS)*, pp. 4868–4878, Montreal, Canada, Dec. 2018.

[41] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 5353–5360, Boston, MA, Jun. 2015.

[42] C. Sommer and F. Dressler, "Using the right two-ray model? A measurement based evaluation of PHY models in VANETs," in *Proc. ACM Annual Int'l Conf. Mobile Computing and Networking (MobiCom)*, pp. 1–3, Las Vegas, NV, Sept. 2011.

[43] E. Perez-Cabre and B. Javidi, "Scale and rotation invariant optical ID tags for automatic vehicle identification and authentication," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1295–1303, Jul. 2005.

Xiaozhen Lu received the B.S. degree in communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017. She is currently pursuing the Ph.D. degree with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. Her research interests include network security and wireless communications.



Liang Xiao (M'09-SM'13) is currently a Professor with the Department of Information and Communication Engineering and department head with the Department of Cybersecurity, Xiamen University, Xiamen, China. She is an IEEE Senior member, and member of IEEE Technical Committee on Big Data. She has served in several editorial roles, including an associate editor of IEEE Trans. Information Forensics & Security, IEEE Trans. Dependable & Secure Computing and IEEE Trans. Commun. Her research interests include wireless security, smart grids and wireless communications.

She won the best paper award for 2017 IEEE ICC and 2016 IEEE INFOCOM Bigsecurity WS. She received the B.S. degree in communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, NJ, in 2009. She was a Visiting Professor with Princeton University, NJ, USA, Virginia Tech, VA, USA, and University of Maryland, College Park, MD, USA.

Tangwei Xu received the B.S. degree in electronic information engineering from Nanjing University of Science and Technology, Nanjing, China, in 2018. He is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. His current research interests include network security and wireless communications.

Yifeng Zhao received his B.S. degree in communication engineering in 2002, M.S degree in electronic circuit system in 2005 and Ph.D. degree in communication engineering in 2014 from Xiamen University, Xiamen, China. He is currently an Assistant Professor with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. His current research interests include mmWave communication, massive MIMO and machine learning applied in wireless communications.

Yuliang Tang received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 1987, the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1996, and the Ph.D. degree in communication engineering from Xiamen University, Xiamen, China, in 2009.

He is currently a Professor with the Department of Information and Communication Engineering, Xiamen University. He has authored or coauthored over 90 papers in refereed mainstream journals and reputed international conferences and has been granted over ten patents. His current research interests include wireless communication systems and vehicular ad hoc networks.

Weihua Zhuang (M'93-SM'01-F'08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair in Wireless Communication Networks.

Dr. Zhuang was a recipient of the 2017 Technical Recognition Award from the IEEE Communications Society Ad Hoc and Sensor Networks Technical Committee, and a co-recipient of several Best Paper Awards from IEEE conferences. She was the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2013, the Technical Program Chair/Co-Chair of IEEE VTC Fall 2017 and Fall 2016, and the Technical Program Symposia Chair of IEEE Globecom 2011. She is an elected member of the Board of Governors and Vice President-Publications of the IEEE Vehicular Technology Society. She was an IEEE Communications Society Distinguished Lecturer from 2008 to 2011. Dr. Zhuang is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada.