Jameel, Furqan; Javed, Muhammad Awais; Zeadally, Sherali; Jantti, Riku

Secure Transmission in Cellular V2X Communications Using Deep Q-Learning

Published in:
IEEE Transactions on Intelligent Transportation Systems

DOI:
10.1109/TITS.2022.3165791

Published: 01/10/2022

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Jameel, F., Javed, M. A., Zeadally, S., & Jantti, R. (2022). Secure Transmission in Cellular V2X Communications Using Deep Q-Learning. *IEEE Transactions on Intelligent Transportation Systems*, *23*(10), 17167-17176. https://doi.org/10.1109/TITS.2022.3165791

Secure Transmission in Cellular V2X Communications Using Deep Q-Learning

Furqan Jameel, Muhammad Awais Javed, Sherali Zeadally, and Riku Jäntti

***Abstract*—Cellular vehicle-to-everything (V2X) communication is emerging as a feasible and cost-effective solution to support applications such as vehicle platooning, blind spot detection, parking assistance, and traffic management. To support these features, an increasing number of sensors are being deployed along the road in the form of roadside objects. However, despite the hype surrounding cellular V2X networks, the practical realization of such networks is still hampered by under-developed physical security solutions. To solve the issue of wireless link security, we propose a deep Q-learning-based strategy to secure V2X links. Since one of the main responsibilities of the base station (BS) is to manage interference in the network, the link security is ensured without compromising the interference level in the network. The formulated problem considers both the power and interference constraints while maximizing the secrecy rate of the vehicles. Subsequently, we develop the reward function of the deep Q-learning network for performing efficient power allocation. The simulation results obtained demonstrate the effectiveness of our proposed learning approach. The results provided here will provide a strong basis for future research efforts in the domain of vehicular communications.**

***Index Terms*—Deep Q-learning, Interference management, Physical layer security, V2X communications**

I. Introduction

The development of vehicular networks and the change in public transport through the use of autonomous vehicles is one of the emerging aspects of modern cities [1]. Future vehicular networks promise more efficiency, a reduction in congestion, and better environmental performance. In the traditional traffic system, individual situations such as infrastructure damages or network congestion can lead to systemic effects and failure of the underlying technology [2]–[4]. In this context, it is important to provide a robust communication solution not only between a base station (BS) (infrastructure) and a vehicle (I2V) but also between vehicles themselves, i.e., vehicle-to-Vehicle (V2V). To enable the above connectivity, the C-V2X standard developed by the Third Generation Partnership Project (3GPP) provides two types of communication solutions, a Uu interface connecting the vehicle's transceiver with the enodeB on the downlink/uplink channel, and a PC5 interface using the sidelink channel for V2V communications [5]. I2V communication is thus enabled over the downlink channel over the Uu interface.

The I2V aspect is one of the most explored areas of vehicular communications with efficient protocols and a set of services. However, V2X communication of vehicular networks has not been fully explored. In addition, very few studies take into account the collaboration and optimization between I2V links and V2X links [6].

From the perspective of V2X communications, there are two different regimes, i.e., the cellular-V2X (C-V2X) and dedicated short-range communications (DSRC) [5]. DSRC and C-V2X have been very popular in the past decade due to the increasing number of devices connected to wireless networks. C-V2X can connect to more V2X devices by taking advantage of cellular networks and by establishing wireless links among the vehicles, smart infrastructures, and pedestrians. C-V2X can work in two modes namely the direct communications (DC) mode in which V2X devices are able to directly communicate with each other and the network-based communications (NC) mode in which the cellular BS has a key role, and the V2X devices communicate with the cellular communications. Unlike the static or non-mobile wireless communications, moving vehicles cause the Doppler effect which leads to some challenges for long term evolution (LTE)-based V2X systems [7]. In fact, the carrier frequency offset resulting from the Doppler effect can cause an inter-carrier interference (ICI) in wireless communications. In the remainder of this text, our focus will be on the security of C-V2X communications and vehicular networks in general.

Security is a key aspect of future vehicular networks because several critical applications such as safety awareness and autonomous driving rely on data sharing among vehicles. Recent trends in vehicular communications have also attracted a lot of attention in the physical layer security of vehicular networks. But these research efforts have focused mostly on performance evaluation studies under different fading conditions. For instance, the authors of [8] investigated the secrecy performance (i.e., the information leaked to eavesdroppers to decode the message) of vehicular networks when they experience double Rayleigh fading. Derivations of closed-form expressions of secrecy outage probabilities have remained a key motivation for some recent studies. The secrecy outage probability is defined as the probability of an event when the secrecy rate (i.e., the difference between the rates of the main link and the eavesdropping links) drops below a pre-specified threshold. For instance, the authors of [9] derived the probability of strictly positive secrecy capacity while the authors of [10] provided outage probability expressions for log-normal fading.

Furqan Jameel and Riku Jäntti are with the Department of Communications and Networking, Aalto University, 02150 Espoo, Finland. (email: furqanjameel01@gmail.com and riku.jantti@aalto.fi).

Muhammad Awais Javed is with Electrical and Computer Engineering Department, COMSATS University Islamabad, Islamabad, Pakistan (email: awais.javed@comsats.edu.pk)

Sherali Zeadally is with the College of Communication and Information, University of Kentucky, Lexington, KY 40506-0224, USA (email: szeadally@uky.edu).

Lately, a few studies [11]–[13] have been focusing on the secure transmission of messages in vehicular networks. To mitigate the effect of interferences, a secure and integrated network architecture was proposed for group communications in vehicular networks [14]. Using physical layer security, the authors of [15] proposed a cooperative authentication method for vehicular communications. In [16], the authors proposed a novel channel access scheme that increased the secrecy rate under imperfect channel state estimation conditions. The work in [17] derived an analytical model to use artificial noise for improving physical layer security of C-V2X communications. However, aside from these few works, little attention has been paid to the optimization of link security in vehicular networks. This work aims to address this gap in the literature by focusing on secure transmission in C-V2X communications using deep reinforcement learning. This is particularly useful for the scenarios where a simple reinforcement learning technique can result in high complexity due to large action and state space.

A. Related Work

Reinforcement learning is one of the emerging techniques for obtaining optimal solutions through hit and trial methods in a dynamic environment [18]. In order to improve the efficiency and enable fast convergence, several algorithms have been proposed in the past. Some of these algorithms include the SARSA algorithm, temporal difference algorithm (TD), Q-learning algorithm, and function approximation algorithm [19]. If the environment encountered is Markov, the interaction is the Markov decision-making process (MDP). Each action is performed on the Q value in the Q-learning algorithm. Reinforcement learning techniques have been widely used in the literature. For instance, in [20], the Bayesian reinforcement learning provides an optimized tradeoff in uncertainty learning. The proposed scheme proved to be effective in large scale Bayesian reinforcement learning. The optimization of the online reinforcement learning algorithm is presented in [21] where the predefined learning cycles are replaced by the adaptively learning cycles. Moreover, unnecessary calculations are minimized based on the state transition. The authors also showed that the adaptive approach which minimizes the duration of learning cycles is suitable for reinforcement learning. In [22] the authors discussed the pre-training framework (PRELSA) for improving the learning speed in autonomous vehicles and showed a 20% increase as compared to existing learning algorithms.

However, all complex problems cannot be solved by reinforcement learning alone. Due to this reason, we need to combine reinforcement learning with deep neural networks a technique referred to as deep reinforcement learning. Recently, there has been a lot of interests in deep reinforcement learning techniques. The reinforcement learning was applied in computer vision applications in [23] and an active agent is used to identify the location of a target object, to refine the geometry of the object, and to optimize the representation and show the effectiveness of the proposed method. In [24], the authors applied a deep reinforcement learning algorithm on Atari 2600 games and improved the data efficiency via the utilization of unrewarded experiences and hierarchical reinforcement learning (HRL). To improve the performance, the intrinsic reward and the representations learned during the training are crucial. From the perspective of intelligent transportation systems, a deep Q-learning algorithm was implemented in [25] for automated driving cars. The authors evaluated the algorithm's performance based on the knowledge of the surrounding environment but the packet loss ratio obtained was high. In [26], the authors considered video games with image output and they used Q-learning and deep neural networks. Similarly, in [27], the authors proposed an adaptive resource allocation using reinforcement learning and the conducted experiments on the CloudSim platform. The reinforcement learning-based schemes proved to be superior compared with other allocation schemes.

Deep reinforcement learning techniques are also well suited for wireless resource allocation. In [28], the authors proposed a novel deep deterministic policy gradient (DDPG) algorithm and a normalization algorithm using a neural network. The proposed algorithm outperforms the existing ones in terms of the average bit rate. Similarly, in [29], the authors proposed a deep reinforcement learning-based energy management technique. The simulation results showed that the proposed approach performs better than the dynamic programming technique in terms of saving energy of the vehicle. In [30], the authors implemented a deep reinforcement learning algorithm for vehicular networks. The results showed that the proposed algorithm can maximize the sum throughput and achieve fairness. In [31], the authors proposed an actor-critic reinforcement learning-based radio resource management for vehicular networks to improve the system capacity, throughput, and efficiency. The proposed method achieves the desired throughput and convergence. The concept of reinforcement learning has also been applied in 5G mmWave communication in a massive multiple-input-multiple-output (MIMO) network in [32]. Despite these developments, there are few studies that have focused on providing deep Q-learning solutions to improve link security of vehicular networks. In this context, the authors of [33] addressed the resource allocation problem for multi-platooning vehicular networks. They proposed a security-aware power allocation strategy to improve the security of the network. Similarly, the authors of [34] considered stringent latency requirements and addressed the efficiency of the vehicular network using deep reinforcement learning techniques. In [35], the authors investigated the problem of vehicle to vehicle transmission of wireless messages. Platooning is a key technology in smart cities for efficient V2V communications. The authors proposed a Cooperative Reinforcement Learning (CRL) approach that uses LTE technology. The proposed scheme outperforms the distributed reinforcement learning-based resource selection scheme in terms of delay in cooperative awareness messages. Moreover, resource utilization was also efficient in the proposed scheme. The work in [36] uses Reconfigurable Intelligent Surfaces (RIS) to ensure secure vehicular communications in the presence of an eavesdropper. The analytical results in terms of secrecy rate and outage probability were derived for RIS-enabled vehicular communications.

Figure 1. Illustration of system model for C-V2X communications.

B. Motivations and Research Contributions

1) Motivations: Most recent works on vehicular communications have conducted performance preliminary studies on V2X communications. Furthermore, optimization studies have mostly restricted their scope to spectrum or energy efficiency [29]–[31] with very few efforts focused on the secrecy of information [33], [34]. Although spectrum and energy efficiency of V2X communications are important aspects and need further investigation, the significance of link security cannot be overlooked. Vehicles today supply a lot of data and leave electronic traces by using a navigation aid or by transmitting data to the car manufacturer. However, future vehicles will be permanently networked thereby requiring efficient link security solutions.

2) Research Contributions: Motivated by the recent results of the related works discussed above, this work makes the following key research contributions:

1) We propose a novel vehicular communication architecture that considers vehicular infrastructure, roadside objects, and eavesdroppers in one single study. This consideration helps us in evaluating the impact of interferences (between I2V and V2X links) on the secrecy performance of a vehicular network.
2) To ensure optimal secrecy performance while guaranteeing the required quality of service for I2V and V2X communications, we describe a secrecy optimization problem for the network considered. We also take the signal power and interference constraints into consideration.
3) To effectively solve the optimization problem, we develop an efficient deep Q-learning-based power allocation strategy. The performance results obtained show the feasibility of our proposed solution in ensuring link security for V2X communications.

C. Organization

The remainder of the paper is organized as follows. Section II presents the system model and problem formulation. In Section III, we describe in detail the proposed optimization framework. Section IV presents the simulation results and relevant discussions. Finally, Section V makes some concluding remarks and presents our future work.

II. System Model and Problem Formulation

In this section, we present the details of the system model along with a discussion on the problem formulation.

A. Network Setup

In Fig. 1, we consider a downlink cellular V2X network operating over L orthogonal subbands in the presence of E eavesdroppers, such that $\mathcal{L} = \{1, 2, 3 \dots L\}$ and $\mathcal{E} = \{1, 2, 3, \dots E\}$. In order to guarantee a minimum average signal-to-interference-and-noise ratio (SINR), i.e., Ω_0, the BS only serves one vehicle per subband for I2V communications. However, the BS also enables V2X communication between the vehicles and the roadside objects. To do so, the BS assigns a random subband to the other vehicles (different from the one in I2V) for downlink communication with roadside objects. We assume that each subband $l \in \mathcal{L}$ can accommodate a total of K V2X links, such that $\mathcal{K} = \{1, 2, 3, \dots K\}$. Similar to the I2V link, the minimum guaranteed SINR for each V2X link is given as Ω_k. In parallel, the eavesdroppers overhear the V2X transmission and try to decode the message on I2V and V2X links. Thus, the objective of this work is to efficiently allocate

power to reduce the interference in the network and maximize the secrecy rate (i.e., the difference between the rates of main link and eavesdropping links).

We note that, in a dense deployment, distribution of resources can lead to a significant amount of interference. This not only affects the I2V link but also deteriorates the V2X links thereby making it difficult to decode the information. For the sake of simplicity, we focus on a single subband such that the 0 index refers to the I2V link between the BS and the vehicle, and index k refers to the V2X link. When the BS sends a message to the vehicle over l band, the received SINR at the vehicle can be written as:

$$\gamma_0 = \frac{p_0|h_{0,0}|^2}{N_0 + \sum_{k\in\mathcal{K}} p_k|h_{k,0}|^2}, \tag{1}$$

where p_0 is the transmit power by the BS, $h_{0,0}$ represents the channel gain between the BS and the vehicle, p_k denotes the transmit power from the vehicle to the roadside object over k-th link, and $h_{k,0}$ is the channel between the vehicle in V2X and the vehicle in I2V. Furthermore, N_0 represents the variance of additive white Gaussian noise with zero mean. The SINR at the k-th V2X link over the l subband is given as:

$$\gamma_k = \frac{p_k|h_{k,k}|^2}{N_k + p_0|h_{0,k}|^2 + \sum_{j\in\mathcal{K}\{k\}} p_j|h_{j,k}|^2} \tag{2}$$

where $h_{0,k}$ represents the channel between the BS and the k-th vehicle, $h_{k,k}$ is the channel gain between the k-th vehicle and the k-th roadside object, $h_{j,k}$ denotes the channel gain between the j-th vehicle and the k-th roadside object, and p_j represents the transmit power of the j-th vehicle. In addition, N_k is the variance of additive white Gaussian noise with zero mean.

During the V2X transmission, the non-cooperative eavesdroppers also receive the signal and attempt to decode the message. In general, the V2X links are considered more vulnerable to eavesdropping attacks due to their ad hoc and low-powered nature. Thus, our aim in this study is to ensure the maximum secrecy rate for V2X links while guaranteeing the required SINR for I2V communications. The received signal SINR at $e \in \mathcal{E}$, as a result of V2X communication, can be written as:

$$\gamma_{e,k} = \frac{p_k|h_{k,e}|^2}{N_e + p_0|h_{0,e}|^2 + \sum_{j\in\mathcal{K}\{k\}} p_j|h_{j,e}|^2}, \tag{3}$$

where $h_{k,e}$ is the channel gain between the k-th vehicle and the e-th eavesdropper, $h_{0,e}$ represents the channel between the BS and the e-th vehicle, $h_{j,e}$ denotes the channel gain between the j-th vehicle and the e-th eavesdropper, and p_j represents the transmit power of the j-th vehicle. Furthermore, N_e is the variance of additive white Gaussian noise with zero mean. The achievable secrecy rate for k-th V2X link can be expressed as:

$$R_k^{\text{sec}} = \log_2(1+\gamma_k) - \log_2(1 + \max_{e\in\mathcal{E}} \gamma_{e,k}). \tag{4}$$

B. Problem Formulation

As stated earlier, the main objective of this work is to ensure the secrecy of V2X links without compromising the quality of service requirements of the I2V link. We intend to accomplish this task by efficiently allocating the power for V2X links. This is because the transmit power of V2X links has a twofold effect on the performance of the network. Firstly, since I2V links are given priority over V2X links, it is important to minimize the interference for I2V which is caused by V2X communications. Secondly, the efficient power allocation for V2X communication can be considerably helpful in improving the secrecy rate. Let us now represent transmit power vector of K vehicles communicating over V2X links as $p = \{p_1, p_2, \dots p_K\}$, then, the power allocation problem can be expressed as

$$\begin{aligned} &\max_{p} \sum_{k\in\mathcal{K}} R_k^{\text{sec}} \\ \text{s.t.}\quad &\mathbf{C1}: \gamma_0 \geq \Omega_0 \\ &\mathbf{C2}: 0 \leq p_k \leq p_{\max}, \quad k \in \mathcal{K} \\ &\mathbf{C3}: \gamma_k \geq \Omega_k, \quad k \in \mathcal{K}, \end{aligned} \tag{5}$$

where $p_{\max}$ denotes the maximum transmit power of the vehicles communicating over the V2X link. The constraints are defined as follows: the constraint **C1** ensures that the maximum SINR requirements for I2V communications are met; the constraint **C2** refers to the maximum power limit of the vehicles communicating over V2X links; the constraint **C3** guarantees that the SINR requirements of the V2X communications are met.

The formulated optimization problem is non-trivial to solve due to the presence of interference terms and constraints [37]. In other words, it is a non-convex optimization problem due to the objective function and the SINRs expressions. The interference term from the neighboring cell in the denominator makes the optimization non-convex. A less complex solution to this problem has been provided in [38] where the authors ignore the interference term to simplify the analysis. Yet, there is a need to address the interference factor in the optimization of practical communication systems. Furthermore, as the network scales up and the number of vehicles, roadside objects, and eavesdroppers increases, it becomes difficult to use conventional iterative algorithms to find the solution [39]. The extensive computations often require bisection technique or matrix inversion to efficiently solve the problem which is challenging in practical systems.

It is also worth pointing out that a single vehicle can only be aware of its own transmit power. It does not need to know the transmit power of the other vehicles in the network. Thus, the optimization framework needs to consider the interaction between the vehicles without interacting with the network. Thus, we aim to individually improve the performance of each vehicle in the network such that a vehicle could adapt continuously according to the network conditions. In this way, the vehicle can efficiently allocate power to communicate with the roadside object without compromising the secrecy or degrading the performance of I2V links. We describe our

proposed secrecy optimization framework in the following section.

III. Proposed Secure Power Allocation for V2X Communications

In this section, we describe our proposed deep Q-learning framework for solving the optimization problem described earlier.

A. Learning Framework

Since it is well-established that vehicles communicate with the roadside objects independent of the behavior of other vehicles, it is important to formulate a network-wide policy for optimization. For this reason, we adopt a multi-agent approach, whereby, each vehicle can be considered as a single agent. An agent in a typical Q-learning scenario interacts with the environment based on a state set and an action set. The agent transitions from one state to another after performing an action which consequently determines the overall policy of the agent based on the reward it receives. The major goal of an agent is to maximize the cumulative reward in the long term.

The single-agent framework works well in point-to-point communication scenarios. However, it cannot be used for dense cellular V2X communications. Thus, as clarified earlier, we chose a multi-agent approach. Specifically, we consider K agents in the environment selecting different actions to transition between the states. There are four major components of the considered multi-agent learning model, i.e., state, action, reward, and transition probability. The following subsections describe each component:

1) State: The state of the learning model is a set of random variables, wherein, the random variable is denoted as χ_i, where i is the index of the state set which is expressed as:

$$\chi = \{\chi_1, \chi_2, \chi_3, \dots \chi_n\}, \tag{6}$$

where each state indicates a unique characteristic of the network. In the setup considered, the state of the system directly impacts the performance of I2V and V2X communications. We define the different state variables based on the different constraints of the formulated optimization problem. More specifically, $\chi_1 \in \{0, 1\}$ denotes whether the required SINR is guaranteed for V2X communications such that $\chi_1 = 1_{\{\gamma_k \geq \Omega_k\}}$, and $\chi_2 \in \{0, 1\}$ indicates the guaranteed SINR at the I2V link such that $\chi_2 = 1_{\{\gamma_0 \geq \Omega_0\}}$. Furthermore, χ_3 and χ_4 define the communication range of the V2X vehicles for I2V vehicle and V2X vehicles around BS respectively, such that $\chi_3 \in \{0, 1, \dots, N_1\}$ and $\chi_4 \in \{0, 1, \dots, N_2\}$. Here, N_1 and N_2 indicate the number of coverage regions with radius $d_1, \dots d_{N_1}$ and $d_1^{'}, \dots d_{N_2}^{'}$, respectively.

2) Action: We represent the joint action of all agents as a single set Λ. A k-th agent can choose different actions (i.e., Λ_k) from this set and the joint action of all agents can be written as:

$$\Lambda = \Lambda_1, \Lambda_2, \Lambda_3, \dots \Lambda_K = \prod_{k=1}^{K} \Lambda_k. \tag{7}$$

In our study, the transmit power of the vehicle in V2X communication is considered as the action. The k-th vehicle selects a transmit power from Λ_k which is less than the maximum transmit power of the vehicle. Since a vehicle is unaware of the other entities in the environment, it selects the action with the same probability during the training. Hence, it uses a fixed step size of Δp to select different values of transmit power within the minimum and maximum range.

3) Reward: A reward can be considered as a positive or a negative response the agent receives upon performing an action. For the model we are considering, we define reward $\Upsilon(s, a)$ as a function of state $s \in \chi$ and actions $a \in \Lambda$. We provide more details on the reward function in the next section.

4) Transition Probability: The transition probability is one of the key components of the learning model and describes the interacting environment. As the name suggests, it is the probability of transitioning from one state to another as a result of an action. In other words, it can be defined as the probability of ending up in $s^{(t+1)}$ from $s^{(t)}$ as a result of action $a^{(t)}$.

The objective of the agent is to find the best policy function against a state denoted as $\pi(s)$. To evaluate this policy, it is necessary to find the action-value function $Q_\pi(s, a)$ which is expressed as [40]:

$$Q_\pi(s^{(t)}, a^{(t)}) = \beta \sum_{s^{(t+1)} \in \chi} V_\pi(s^{(t+1)}) \Pr(s^{(t+1)} | s^{(t)}, a^{(t)}) + \Upsilon(s^{(t)}, a^{(t)}), \tag{8}$$

where β denotes the discount factor and $V_\pi(.)$ is the value function. According to [40], the value function is expressed as:

$$V_\pi(s^{(t+1)}) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \Upsilon^{(t+1)} \beta^t | s^{(0)} = s^{(t+1)} \right], \tag{9}$$

where $s^{(0)}$ is the initial state and $\Upsilon^{(t+1)}$ represents the reward received at $t + 1$ time. Based on the above expression, the optimal policy must satisfy the Bellman optimality expression for an optimal value function given as:

$$V(s^{(t)}) = \max_a Q(s^{(t)}, a^{(t)}), \tag{10}$$

where $Q(s, a)$ denotes the optimal Q-function. Next, we use the deep Q-learning approach to solve this expression.

B. Secure Power Allocation using Deep Q-Learning

Next, we solve the Bellman equation using the deep Q-learning approach to ensure secure power allocation. Deep Q-learning models generally consist of two main components, i.e., a deep neural network and a Q-learning framework. A deep Q-learning approach is more suitable for large-scale dense networks where it is difficult to provide a solution using the conventional tabular Q-learning approach [41]. This is because power allocation in wireless networks is challenging and requires high dimensionality and Q-learning may take a substantial amount of time to converge for these problems.

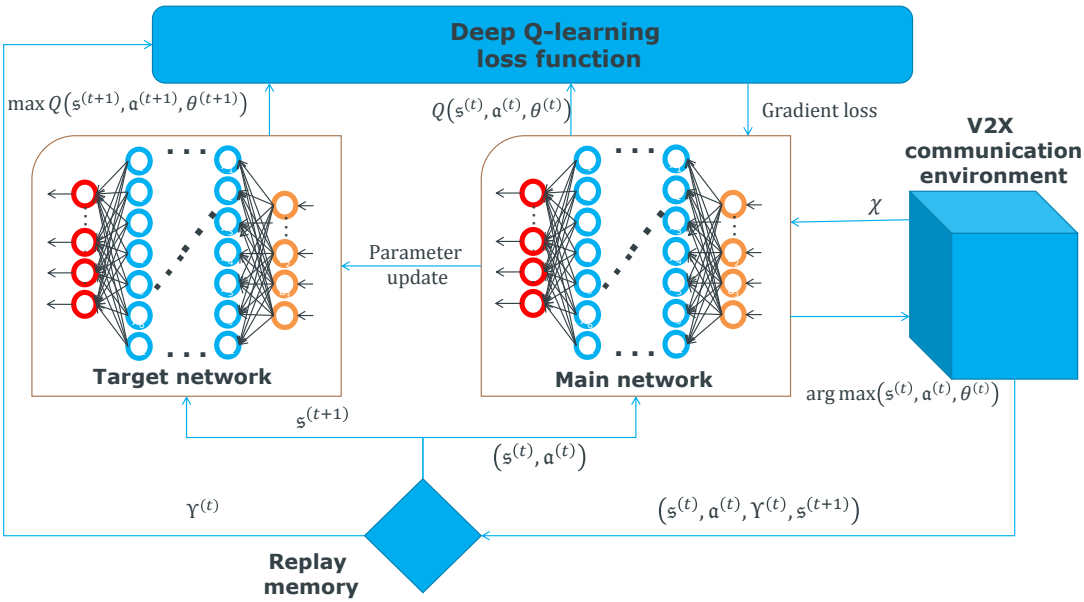

Figure 2. Deep Q-learning algorithm for secure power allocation.

The general solution for finding the optimal policy in (10) is to start off with an arbitrary policy and iteratively find the optimal solution. However, the deep Q-learning technique uses temporal differences to provide an efficient and practical solution for a generalized policy iteration. According to [42], the Q-learning update rule for evaluating the global Q-function is expressed as:

$$Q_k(s^{(t)}, a^{(t)}) \leftarrow Q_k(s^{(t)}, a^{(t)}) + \alpha^{(t)}(s,a)\Big[\beta \max_{a} Q(s^{(t+1)}, a^{(t+1)}) + \Upsilon^{(t+1)}(s^{(t)}, a^{(t)}) - Q(s^{(t)}, a^{(t)})\Big], \tag{11}$$

where $\alpha^{(t)}(s,a) = (t(s,a)+1)^{-1}$ is the decaying learning rate and $t(s,a)$ represents the number of times a (s,a) pair is visited before time step t. For new vehicles joining the network, they can train independently and try to maximize their own Q-function. Thus, the local update rule can be defined for the reward function $\Upsilon^{(t+1)}(s_k^{(t)}, a_k^{(t)})$ as follows:

$$Q_k(s_k^{(t)}, a_k^{(t)}) \leftarrow Q_k(s_k^{(t)}, a_k^{(t)}) + \alpha^{(t)}\Big[\beta Q_k(s_k^{(t+1)}, a_k) + \Upsilon^{(t+1)}(s_k^{(t)}, a_k^{(t)}) - Q_k(s_k^{(t)}, a_k^{(t)})\Big], \tag{12}$$

where a_k is expressed as $\arg\max_{a_k} Q_k(s_k^{(t+1)}, a_k^{(t+1)})$.

We now focus our attention on the design of the reward function because it is one of the important aspects in the optimization of V2X communications. In this context, it is worth pointing out that there are no fixed guidelines for the formulation of a reward function in deep Q-learning. The deep Q-learning problems are generally treated as a black box and the agent tries to learn the relationship between different aspects of the environment. Due to this reason, any systematic approach aimed at designing a reward function has not been discussed in the literature. In our study, the reward function Υ_k for any k-th vehicle is a function of four different variables, i.e., the secrecy rate of V2X links, the data rate of the V2X link, the rate of the I2V link, the required SINR threshold of the V2X link and the required SINR threshold of the I2V link.

The reward function developed is expected to make progress toward finding the optimized solution for the problem considered. It is worth highlighting that our network setup does not consider any friendly jammers. In their absence, the BS can do little to maximize the secrecy rate other than improving the achievable secrecy rate of the V2X links while meeting the SINR requirements. Thus, to control the reward function, we define it as follows:

$$\Upsilon_k = \Xi_1 + \Xi_2 + \Xi_3 + \Xi_4 \tag{13}$$

where

$$\Xi_1 = [\log_2(1+\gamma_k) - \log_2(1 + \max_{e\in\mathcal{E}} \gamma_{e,k})]^{2q-1}, \tag{14}$$

$$\Xi_2 = [\log_2(1+\gamma_0) - \log_2(1+\Omega_0)]^{2q-1}, \tag{15}$$

$$\Xi_3 = [\log_2(1+\gamma_k) - \log_2(1+\Omega_k)]^{2q-1}, \tag{16}$$

and Ξ_4 is the bias of the reward function which has a constant real value while q is an integer of the polynomial function. To achieve the objective of maximum secrecy rate and guaranteed SINR, the reward function is designed so as to motivate an increase in the desired rates as much as possible. This also means that higher values of γ_k and γ_0 yield larger rewards thereby ensuring that the SINR requirements for V2X and I2V communications are met.

It is worth noting that it is a challenge to find the optimal strategy in the V2X scenario considered by looking up the Q-values in the table. To address this challenge, we use deep neural networks (as Fig. 2 shows) to fit the optimal value function and the optimal strategy such that we get:

$$Q_k(s_k^{(t)}, a_k^{(t)}; \theta^{(t)}) \approx Q_k(s_k^{(t)}, a_k^{(t)}), \tag{17}$$

where θ represents the weights of the neural network. The deep neural network aims to learn the complex interdependence of inputs and outputs. Specifically, the deep neural network consists of a number of hidden layers along with a single input layer and a single output layer. The main neural network observes a state from the V2X environment and outputs $Q(s_k^{(t)}, a_k^{(t)}; \theta^{(t)})$ [43]. The target neural network outputs the $Q_k(s_k^{(t+1)}, a_k^{(t+1)}; \theta^{(t+1)})$ and the objective of the neural network is to minimize the loss function $L(\theta)$ expressed as follows:

$$\begin{aligned} L(\theta) = \mathbb{E}\{(\Upsilon + \beta \max Q_k(s_k^{(t+1)}, a_k^{(t+1)}; \theta^{(t+1)}) \\ - Q(s_k^{(t)}, a_k^{(t)}; \theta^{(t)}))^2\}. \end{aligned} \tag{18}$$

In the framework considered, the entire communication network is the environment that feeds into the main network. To be more specific, each vehicle sends the state information to the main network. The weights of the deep neural network are updated using the sample from the replay memory which collects the experiences from the other vehicles in parallel and inputs the information into the loss function. This replay memory data is randomly sampled for training the network. The process of obtaining experience from the agent and, subsequently, randomly sampling these experiences is known as experience replay. After a pre-specified time duration, the weights from the main network are copied into the target neural network. During this process, the loss function is also updated iteratively to eventually minimize it to the desired level.

C. Complexity

In general, the complexity of deep learning techniques varies significantly and heavily depends on the application. For instance, a larger amount of data may require more training time for the convergence of the learning model. Additionally, an increase in the number of hidden layers may consume more time in order to reliably train the models. Thus, due to the hierarchical structure and the inherent "black box" nature of these learning models, it is difficult to compute the precise computational complexity. However, due to the latest advances in deep learning approaches, the training can be performed independently. For instance, the training of these models can be performed in the cloud [44]. Then, these trained models can be used for testing in real-time. After a pre-specified period of time, these models can be completely or partially retrained in the cloud.

IV. Performance Evaluation

In this section, we describe the simulation parameters and we discuss the simulation results obtained. The transmission power of vehicles is selected between a range of 1dBm and 15 dBm [45]. We have considered 4 V2X links and 4 eavesdroppers in the network. We used a noise power of -174dBm and Rayleigh fading channels. Although the agents are trained in a simulation environment, there are different ways the vehicles can be trained in practical conditions. For instance, the training can occur in different frames and the vehicle can select a transmit power at the beginning of a frame. The vehicle can then send the frame to the roadside object which sends the channel quality indicator as feedback for the vehicle to estimate the SINR and calculate the reward.

Table I
Simulation parameters and their values.

Simulation parameters	Values
Minimum power	1 dBm
Fading Channel	Rayeligh
Maximum power	15 dBm
Eavesdroppers	4
α	0.2
Noise	-174 dBm
V2X links	4
Realizations	10^4
Hidden layers	3
β	0.9

A. Simulation Setup

For the simulation setup, we assume that the vehicle chooses an action using the greedy exploration approach. This allows an agent to balance the tradeoff between exploration and exploitation of the environment. In simple terms, the agent would have to decide whether it needs to focus on the current immediate reward or further explore the environment for future rewards. The simulation starts with one I2V link and one V2X link. Once the first vehicle in the V2X link is trained, a new V2X link is introduced in the network for training. In the presence of another V2X link, the first one acts greedily and goes on to exploit the environment. After the convergence of second V2X, another vehicle is introduced and the process continues as more vehicles and links are introduced. The main performance metrics are average secrecy rate which is the average difference between the capacity of main and eavesdropping links and the obtained reward for the agent.

Table I presents the simulation parameters and their values we have used in all our tests. In this context, it is worth pointing out that the selection of hyperparameters (i.e., learning rate and discount factor) of the learning network directly affects the performance of the agent. Hyperparameters act as a knob for tweaking the learning agent during training and require multiple training runs. The values of parameters are selected to minimize the loss function as much as possible [46]. This process is repeated multiple times for different configurations of hyperparameters and the best model is used. Besides the exhaustive hyperparameter tuning, the number of hidden layers is carefully selected to maintain the desired degree of freedom in the learning network.

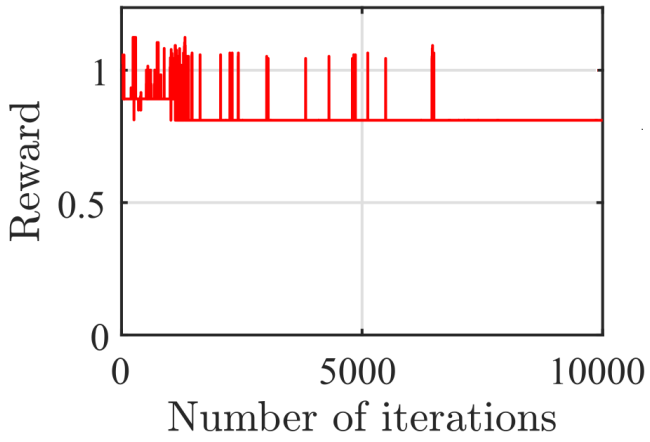

Figure 3. Reward as a function of number of iterations.

B. Numerical Results

Fig. 3 shows the network reward against the number of iterations. We note that an increase in the number of iterations results in stabilizing the reward of the network. Specifically, the total reward varies between 0.7 and 1.4 during the early iterations. Although there seem to be some fluctuations initially in the reward, it gradually converges as the number of iterations increases beyond 5000.

Fig. 4 shows the impact on the reward results when the number of V2X links increases in the network. It is worth noting that our simulation setup starts with a single V2X link and gradually adds the V2X links in the network. This apparently has an impact on the reward of the different links. In other words, we note that the convergence of the reward function becomes more difficult as the number of V2X links in the network increases. It can also be seen that the 3rd and 4th V2X links converge slowly after many iterations when compared to the 1st and 2nd links. This shows that the complexity of the network has an impact on the convergence of the solution.

Fig. 5 further demonstrates the impact of the proposed deep Q-learning solution. Here, we compare our solution with the two baseline learning techniques. For the case of "Baseline Learning 1" the reward is selected as a random number between 0 and 1 while for the case of "Baseline Learning 2" the reward is selected as either 1 or 0. In general, the average secrecy rate increases when the number of iterations increases. In the short term, all the secrecy rates behave similarly. However, when the number of iterations exceeds 5000, the difference in the rewards between the proposed and baseline learning techniques increases. Our proposed technique yields a $18-20\%$ improvement in average secrecy rate over the long term as compared to the other baseline methods.

Fig. 6 shows the average secrecy rate against increasing values of learning rate α and discount factor β. We observe an increase in the average secrecy rate for higher values of α. However, the same is not true for β when the secrecy rate generally increases (when β increases). Fig. 6 (a), (b), and (c) show similar trends where the change in the maximum power

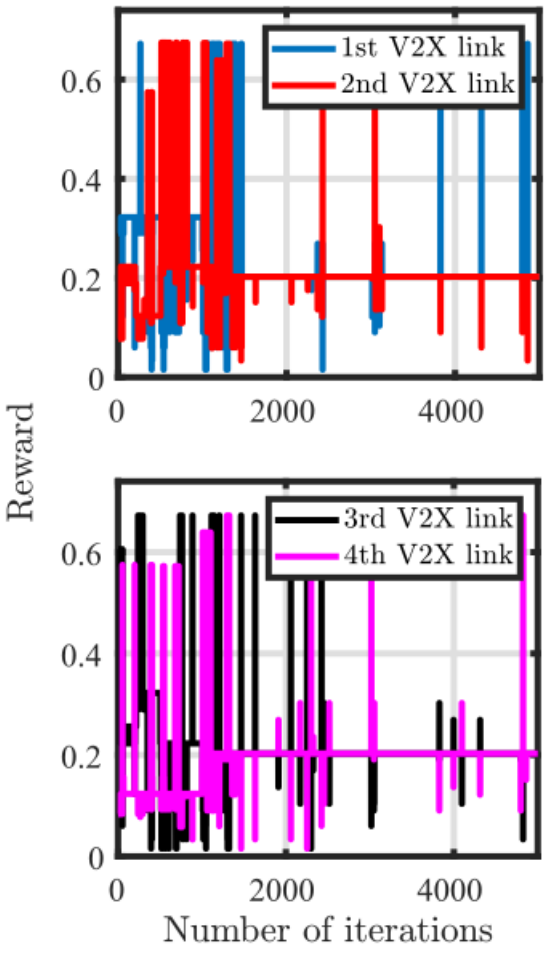

Figure 4. Reward for different V2X links.

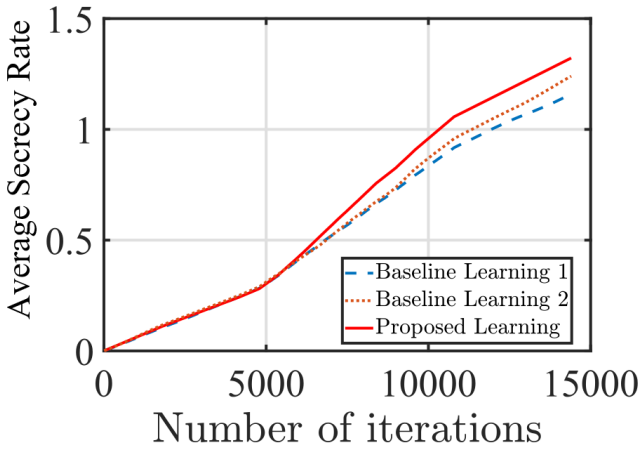

Figure 5. Average secrecy rate versus the number of iterations.

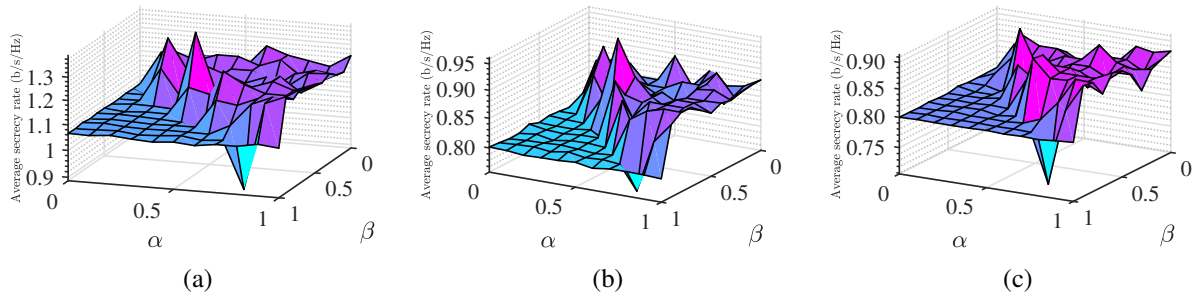

Figure 6. Average secrecy rate versus different values of α and β, where (a) $p_{\max} = 25$ dBm, (b) $p_{\max} = 20$ dBm, and (c) $p_{\max} = 15$ dBm.

affects the average secrecy rate. In particular, we note that when the transmit power decreases, the overall secrecy rate also drops.

V. Conclusion and Future Work

Advances in vehicular communications have been rapidly changing the landscape of future intelligent transportation systems. Due to an ever-increasing number of cyber-attacks on connected networks, it is becoming increasingly important to ensure the security the vehicular networks. In this context, this work has provided a secure power allocation strategy for V2X communications. Specifically, we have considered the high interference scenario where the signal from I2V communication affects the V2X communications and vice versa. To mitigate the decoding capabilities of eavesdroppers while guaranteeing the SINR requirements, we have proposed a deep Q-learning power allocation strategy. The simulation results obtained demonstrate the impact of different parameters of the learning technique on secrecy performance. We have also shown that the proposed reward solution is more efficient when compared to conventional baseline learning strategies.

Although the proposed solution provides considerable performance gains, it can be improved further in several ways. For instance, cooperation among different vehicles can be helpful to further improve the secrecy performance of the vehicular network. Moreover, efficient resource sharing for UL, DL, and V2V communications need further investigation. Additionally, future research can explore how to jointly optimize the spectral efficiency and secrecy rate of cellular V2X networks as it would be insightful to explore this tradeoff for vehicular networks. These challenging yet interesting extensions will be part of our future work.

References

[1] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA Broadcasting/Multicasting for Low-Latency and High-Reliability 5G Cellular V2X Communications," *IEEE Internet of Things Journal*, 2019.

[2] F. Jameel, S. Wyne, M. A. Javed, and S. Zeadally, "Interference-aided vehicular networks: Future research opportunities and challenges," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 36–42, 2018.

[3] H. Gao, J. Zhu, T. Zhang, G. Xie, Z. Kan, Z. Hao, and K. Liu, "Situational assessment for intelligent vehicles based on stochastic model and gaussian distributions in typical traffic scenarios," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–11, 2020.

[4] H. Gao, J. Zhu, X. Li, Y. Kang, J. Li, and H. Su, "Automatic parking control of unmanned vehicle based on switching control algorithm and backstepping," *IEEE/ASME Transactions on Mechatronics*, pp. 1–1, 2020.

[5] S. Zeadally, M. A. Javed, and E. B. Hamida, "Vehicular Communications for ITS: Standardization and Challenges," *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 11–17, 2020.

[6] P. Wang, B. Di, H. Zhang, K. Bian, and L. Song, "Platoon cooperation in cellular V2X networks for 5G and beyond," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3919–3932, 2019.

[7] J. Mei, K. Zheng, L. Zhao, Y. Teng, and X. Wang, "A latency and reliability guaranteed resource allocation scheme for LTE V2V communication systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3850–3860, 2018.

[8] Y. Ai, M. Cheffena, A. Mathur, and H. Lei, "On physical layer security of double Rayleigh fading channels for vehicular communications," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1038–1041, 2018.

[9] X. Liu, "Probability of strictly positive secrecy capacity of the Rician-Rician fading channel," *IEEE Wireless Communications Letters*, vol. 2, no. 1, pp. 50–53, 2012.

[10] X. L. Liu, "Outage probability of secrecy capacity over correlated log-normal fading channels," *IEEE communications letters*, vol. 17, no. 2, pp. 289–292, 2012.

[11] Y. Liu, W. Wang, H.-H. Chen, L. Wang, N. Cheng, W. Meng, and X. Shen, "Secrecy Rate Maximization via Radio Resource Allocation in Cellular Underlaying V2V Communications," *IEEE Transactions on Vehicular Technology*, 2020.

[12] Z. Zhu, Z. Wang, Z. Chu, D. Zhang, and B. Shim, "Robust energy harvest balancing optimization with V2X-SWIPT over MISO secrecy channel," *Computer Networks*, vol. 137, pp. 61–68, 2018.

[13] X. Peng, H. Zhou, B. Qian, K. Yu, F. Lyu, and W. Xu, "Enabling Security-Aware D2D Spectrum Resource Sharing for Connected Autonomous Vehicles," *IEEE Internet of Things Journal*, 2020.

[14] C. Lai, H. Zhou, N. Cheng, and X. S. Shen, "Secure group communications in vehicular networks: A software-defined network-enabled architecture and solution," *IEEE Vehicular Technology Magazine*, vol. 12, no. 4, pp. 40–49, 2017.

[15] H. J. Jo, I. S. Kim, and D. H. Lee, "Reliable cooperative authentication for vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1065–1079, 2017.

[16] B. Qiu, H. Xiao, A. T. Chronopoulos, D. Zhou, and S. Ouyang, "Optimal access scheme for security provisioning of c-v2x computation offloading network with imperfect csi," *IEEE Access*, vol. 8, pp. 9680–9691, 2020.

[17] C. Wang, Z. Li, X. G. Xia, J. Shi, J. Si, and Y. Zou, "Physical layer security enhancement using artificial noise in cellular vehicle-to-everything (c-v2x) networks," *IEEE Transactions on Vehicular Technology*, 2020, to appear.

[18] J.-G. Cao, "Research on electronic commerce automated negotiation in multi-agent system based on reinforcement learning," in *2009 International Conference on Machine Learning and Cybernetics*, vol. 3. IEEE, 2009, pp. 1419–1423.

[19] W. Qiang and Z. Zhongli, "Reinforcement learning model, algorithms and its application," in *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*. IEEE, 2011, pp. 1143–1146.

[20] B. Wu and Y. Feng, "Monte-Carlo Bayesian Reinforcement Learning Using a Compact Factored Representation," in *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*. IEEE, 2017, pp. 466–469.

[21] A. Notsu, K. Yasuda, S. Ubukata, and K. Honda, "Optimization of learning cycles in Online reinforcement learning systems," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 3530–3534.

[22] J.-J. Kim, S.-H. Cha, M. Ryu, and M. Jo, "Pre-training Framework for Improving Learning Speed of Reinforcement Learning based Autonomous Vehicles," in *2019 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 2019, pp. 1–2.

[23] Z. Wu, N. M. Khan, L. Gao, and L. Guan, "Deep Reinforcement Learning with Parameterized Action Space for Object Detection," in *2018 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2018, pp. 101–104.

[24] N. Dilokthanakul, C. Kaplanis, N. Pawlowski, and M. Shanahan, "Feature control as intrinsic motivation for hierarchical reinforcement learning," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3409–3418, 2019.

[25] A. Chandramohan, M. Poel, B. Meijerink, and G. Heijenk, "Machine Learning for Cooperative Driving in a Multi-Lane Highway Environment," in *2019 Wireless Days (WD)*. IEEE, 2019, pp. 1–4.

[26] R. Tan, J. Zhou, H. Du, S. Shang, and L. Dai, "An modeling processing method for video games based on deep reinforcement learning," in *2019 IEEE 8th joint international information technology and artificial intelligence conference (ITAIC)*. IEEE, 2019, pp. 939–942.

[27] I. John, A. Sreekantan, and S. Bhatnagar, "Efficient Adaptive Resource Provisioning for Cloud Applications using Reinforcement Learning," in *2019 IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS* W)*. IEEE, 2019, pp. 271–272.

[28] C. Qiu, Y. Hu, Y. Chen, and B. Zeng, "Deep Deterministic Policy Gradient (DDPG)-Based Energy Harvesting Wireless Communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8577–8588, 2019.

[29] R. Liessner, C. Schroer, A. M. Dietermann, and B. Bäker, "Deep Reinforcement Learning for Advanced Energy Management of Hybrid Electric Vehicles," in *ICAART (2)*, 2018, pp. 61–72.

[30] Y. Liu, H. Yu, S. Xie, and Y. Zhang, "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 158–11 168, 2019.

[31] H. Khan, A. Elgabli, S. Samarakoon, M. Bennis, and C. S. Hong, "Reinforcement Learning-Based Vehicle-Cell Association Algorithm for Highly Mobile Millimeter Wave Communication," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 4, pp. 1073–1085, 2019.

[32] T. K. Vu, M. Bennis, M. Debbah, M. Latva-Aho, and C. S. Hong, "Ultra-reliable communication in 5G mmWave networks: A risk-sensitive approach," *IEEE Communications Letters*, vol. 22, no. 4, pp. 708–711, 2018.

[33] X. Peng, H. Zhou, B. Qian, K. Yu, N. Cheng, and X. Shen, "Security-Aware Resource Sharing for D2D Enabled Multiplatooning Vehicular Communications," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–6.

[34] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.

[35] S. Sharma and B. Singh, "Cooperative Reinforcement Learning Based Adaptive Resource Allocation in V2V Communication," in *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2019, pp. 489–494.

[36] Y. Ai, F. A. Pereira de Figueiredo, L. Kong, M. Cheffena, S. Chatzinotas, and B. Ottersten, "Secure vehicular communications through reconfigurable intelligent surfaces," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.

[37] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on selected areas in communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

[38] S. Niknam and B. Natarajan, "On the regimes in millimeter wave networks: Noise-limited or interference-limited?" in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.

[39] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438–5453, 2018.

[40] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[41] K. I. Ahmed and E. Hossain, "A deep q-learning method for downlink power allocation in multi-cell networks," *arXiv preprint arXiv:1904.13032*, 2019.

[42] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[43] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 705–10 717, 2017.

[44] X. Peng, K. Ota, and M. Dong, "Edge Computing Based Traffic Analysis System Using Broad Learning," in *International Conference on Artificial Intelligence for Communications and Networks*. Springer, 2019, pp. 238–251.

[45] R. Aslani, E. Saberinia, and M. Rasti, "Resource allocation for cellular v2x networks mode-3 with underlay approach in lte-v standard," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8601–8612, 2020.

[46] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7331–7376, 2019.

Furqan Jameel received his master's degree in Electrical Engineering (funded by prestigious Higher Education Commission Scholarship) at the Islamabad Campus of COMSATS Institute of Information Technology (CIIT), Pakistan. Currently, he is with the Department of Communications and Networking, Aalto University, Finland, where his research interests include modeling and performance enhancement of vehicular networks, machine/ deep learning, ambient backscatter communications, and wireless power transfer.

Muhammad Awais Javed is currently working as an Assistant Professor at COMSATS University Islamabad, Pakistan. His research interests include intelligent transport systems, vehicular networks, protocol design for emerging wireless technologies and Internet of things.

Sherali Zeadally is a Professor at the University of Kentucky. His research interests include Cybersecurity, privacy, Internet of Things, computer networks, and energy-efficient networking. He is a Fellow of the British Computer Society and the Institution of Engineering Technology, England.

Riku Jäntti is currently a Professor in communications engineering and the Head of the Department of Communications and Networking School of Electrical Engineering, Aalto University (formerly known as TKK), Finland. His research interests include radio resource control and optimization for machine type communications, cloud-based radio access networks, spectrum and co-existence management, and RF inference.