IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 August 2024, accepted 8 September 2024, date of publication 11 September 2024,
date of current version 20 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3457853

RESEARCH ARTICLE

Toward Secure and Scalable Vehicular Edge Computing With Zero-Energy RIS Using DRL

ABDUL WAHID[1], MUHAMMAD AYZED MIRZA[2], MANZOOR AHMED[3], MUHAMMAD SHERAZ[4], TEONG CHEE CHUAH[4], IT EE LEE[4], AND WALI ULLAH KHAN[5], (Member, IEEE)

[1]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[2]School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong 250200, China
[3]School of Computer and Information Science, Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan 432000, China
[4]Centre for Wireless Technology, Faculty of Engineering, Multimedia University, Cyberjaya, Selangor 63100, Malaysia
[5]Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg

Corresponding authors: Teong Chee Chuah (tcchuah@mmu.edu.my) and Manzoor Ahmed (manzoor.achakzai@gmail.com)

This work was supported by the Multimedia University Research Fellow under Grant MMUI/240021.

ABSTRACT The escalating demands of advanced automotive technologies exert considerable pressure on modern vehicles' computational capabilities. This scenario underscores the critical need for vehicular edge computing (VEC) networks, which leverage 5G/6G communications to facilitate computational offloading. However, providing seamless access to these services while simultaneously adhering to stringent latency and security requirements presents a formidable challenge. The advent of reconfigurable intelligent surfaces (RIS) heralds a new era of possibilities, which includes enhancing connectivity, boosting data transmission rates, consequently reducing delay, and improving the physical layer security of communication channels. This paper delves into the utilization of zero-energy RIS (ze-RIS) in the context of vehicular computation offloading. Our primary goal is to ensure secure access while optimizing operational efficiency in compliance with various task-related and environmental requirements. The ze-RIS-assisted secure task efficient offloading (DRSTO) scheme is a novel deep reinforcement learning (DRL) framework that cleverly switches communication connections to optimize task offloading efficiency and security thereby resolving this issue. At its core, our assessment strategy revolves around the DRSTO model's secrecy and efficiency factor which serve as both a performance measure and a reward function. Time efficiency and rate of confidentiality are used to evaluate this aspect which provides a thorough evaluation of the scheme's success. Extensive testing and comparison have shown that the DRSTO scheme's efficiency factor can be significantly increased, from 6.05 to 18.10. In addition, the rate of job success has increased dramatically, from 2.12% to 4.63%. When compared to other models that were evaluated, the DRSTO scheme consistently better on several metrics including reward, time frames per step (TFPS) ratio and DRL characteristics.

INDEX TERMS Vehicular communication, security, intelligent surfaces, edge computing, DRL.

I. INTRODUCTION

The merging of Internet of vehicles (IoVs), beyond fifth-generation (B5G) communication, and smart vehicles (SVs) has enabled sophisticated vehicular edge computing networks (VECNs) to emerge. This fusion significantly improves the Quality of Experience (QoE) by enhancing connectivity, facilitating the exchange of data related to road safety, and providing real-time information [1]. However, it also introduces new challenges, such as latency sensitivity and greater resource demands, which are beyond the computational capabilities of modern vehicles. Computation offloading has gained popularity as a solution to these challenges. In order to efficiently dispatch jobs, this method employs vehicular edge computing (VEC) including multi-access edge computing (MEC). Enhanced network performance, decreased latency, and reduced energy consumption can be achieved using proximity-based computing models, which utilize roadside units (RSUs), evolved node-B (eNB), and parking lots [2]. Interactions between

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Khawaja.

vehicles, between vehicles and the edge, and between edges themselves are integral to the proposed collaborative strategy to deal with latency and load imbalance.

Numerous technical alternatives exist for carrying out task outsourcing operations; one example is cellular technologies, which include 4G Long Term Evolution (LTE), 5G New Radio (NR), and 6G [3]. New developments in wireless communication system improvement, such as reconfigurable intelligent surfaces (RIS) have also emerged. With RIS support, VEC network coverage and reliability of vehicle-infrastructure communication are both improved. Coverage in areas with poor connectivity can be improved by putting RIS on buildings or roadways to create signal reflection [4], [5], [6]. The integration of vehicles RISs, and VEC networks results in a win-win situation. Enhanced computing capacity, better communication, lower latency, and greater energy economy are all benefits that vehicles can reap from VEC. On the flip side, RIS simplifies the wireless environment which leads to better communication and resource utilization. In situations involving automotive edge computing, the aforementioned collaboration improves network performance, service quality for vehicles, and task delegation [7].

Utilizing RIS in VEC networks has the potential to enhance the energy efficiency and security of vehicle communication systems. RIS reduces power consumption by reflecting signals to certain vehicles instead of broadcasting them to everyone. The system's overall security is enhanced because directional signal reflection inhibits unlawful communication eavesdropping [8]. RIS technology modifies the electromagnetic environment of transmitted signals for improved signal control. A plethora of sensors such as those for light, temperature, and humidity work in tandem with RIS to improve the precision of environmental sensing and data collection. The RIS controller adjusts the reflection elements based on the data collected from the sensors [9]. The amplitude and phase levels of RIS controllers are optimized using convex optimization and gradient descent. Alternatively, DRL approaches enable RIS controllers to adjust amplitude and phase shift on the fly in reaction to changing channel and network conditions [10], [11]. Research focuses on passive RIS, which are preferred for their cost-effectiveness and simpler deployment. Despite challenges such as the double path fading effect, passive RIS proves advantageous in low-power, low-complexity scenarios. While active RIS with amplifiers can overcome these limitations by boosting signal strength, their higher cost, increased complexity, and greater energy demands may limit their practicality [12]. Findings indicate that although active RIS can enhance performance, passive RIS remains a viable and effective option for vehicular networks due to its applicability in current operational contexts [13]. Offloading duties from devices with limited resources requires energy-efficient data transfer [14]. Sustainable technologies such as zero-energy devices (ZEDs) and wireless power transfer (WPT) are essential for 6G networks due to the increased focus on energy efficiency. These devices transform radio frequency (RF) impulses into usable energy. In addition to enhancing outsourcing processes, ze-RIS enhances data rates, expands coverage, and ensures information privacy, all of which contribute to the sustainability goals of future network generations [15], [16].

A. LITERATURE REVIEW

The issue of offloading tasks from vehicles in VEC has been extensively studied in recent research. Various optimization techniques have been employed to address this problem. In [17], a modified Ant Colony Optimization algorithm was utilized to tackle the task offloading problem. In another study, [18], approached the computation offloading problem in VEC as a graph-based job allocation challenge and used a structure-preserved matching algorithm to solve it. To minimize the overall response time for tasks with dependencies, a modified genetic algorithm-based scheduling approach was introduced in [19]. This approach considered the instability and heterogeneity of VEC. Since executing offloading tasks consumes a vehicle's computing resources and energy, some vehicles might be unwilling to participate. Recent research has proposed incentive mechanisms to encourage vehicles to contribute their idle resources. For instance, a contract-based mechanism was designed in [20], combining resource contribution and utilization incentives. Moreover, in [21] discussed a dynamic resource allocation model using software-defined networking, which utilizes a Stackelberg game theory approach to optimize interactions between network leaders and followers, adapting to fluctuations in network demands and resource availability. This approach is directly applicable to managing the hierarchical and dynamic nature of vehicular networks. Another approach, mentioned in [22], introduced a timeliness-aware incentive mechanism to motivate vehicle participation while accounting for uncertain travel times.

Wireless offloading can be susceptible to attacks due to its broadcasting nature, even despite the benefits of utilizing MEC-enabled IoT networks. This can lead to information leakage, posing a significant security challenge for offloading [23], [24]. Ensuring the security of wireless offloading is crucial for the effectiveness of MEC. Employing physical layer security (PLS) methods can be valuable to maintain the confidentiality of wireless offloading. Therefore, numerous PLS techniques have been suggested to establish secure communication, such as employing friendly jamming signals [25] or cooperative relaying [26]. In [27], PLS were used in multicarrier systems to ensure secure computation offloading an MEC system. Still, traditional PLS methods result in high additional energy consumption and only passively adapt to the wireless environment without influencing it. Moreover, when eavesdroppers possess better channel gains than legitimate users, maintaining satisfactory secrecy remains difficult, even with these PLS methods [28].

To address the limitations of conventional PLS methods, the concept of RIS [29], [30], [31] has emerged as a promising solution for the sixth generation (6G) of wireless communication. RIS can enhance spectrum efficiency,

secrecy performance, and energy efficiency. RIS consists of a uniform planar array that includes many low-cost passive reflecting elements, each capable of adjusting its reflection amplitude and/or phase to improve the characteristics of electromagnetic waves [29]. This enables the manipulation of the reflected signal by the RIS to weaken signals at potential eavesdroppers [30] while strengthening signals for legitimate devices, thereby effectively ensuring the security of wireless communications within RIS-assisted networks. To optimize the secrecy rate in RIS-assisted systems, the authors in [32] conducted joint optimization of the transmit beamforming vector and RIS phase shift using an alternative optimization algorithm. Similarly, [33] employed RIS to strengthen the physical layer security of wireless communication systems and devised an optimization-driven approach to optimize active and passive beamforming techniques simultaneously.

Regarding PLS, many researchers have explored performance optimization challenges within secure MEC networks based on RIS [34], [35]. In [34], scholars delved into RIS-assisted secure MEC networks and introduced an iterative optimization approach. This method simultaneously addresses the design of local computing frequencies, transmit power, RIS phase shifts, and time-slot allocation. The primary goal is maximizing secrecy confidential efficiency (SCE) while adhering to the maximum-min fairness criterion. This optimization process employed successive convex approximation (SCA), Dinkelbach-type algorithms, and penalty function-based techniques. In [35], the authors examined RIS-supported secure MEC systems utilizing NOMA (non-orthogonal multiple access). They jointly optimized RIS phase shifts, local computation rates, and offloading power through an alternative iterative optimization algorithm. The objective was to maximize the secure offloading rate, employing semidefinite relaxation (SDR) and Lagrange dual methods within the optimization process.

The study conducted by [34] and [35] mostly employed SCA and SDR-based traditional optimization techniques to jointly optimize resource allocation for secure MEC systems aided by RIS. However, these methods are impractical for the following reasons: Firstly, precise mathematical models exist as the basis of conventional optimization methods, and accurately specifying RIS-assisted MEC network environments with these models may be difficult. Secondly, as the number of reflecting elements within the RIS grows, the computational complexity of conventional optimization methods increases dramatically. This complexity makes implementation effectively within practical RIS-assisted MEC network scenarios challenging.

The concept of edge intelligence involves the integration of artificial intelligence, particularly DRL, into MEC networks and offers an optimistic approach to achieve efficient and adaptable real-time services [36], [37], [38], [39]. The productivity of DRL within MEC systems supported by RIS has been shown in previous research. For instance, the researcher in [37] investigated RIS-supported MEC networks and introduced a method based on deep deterministic policy gradient (DDPG) for optimizing various parameters such as computation offloading volume, power control, and RIS phase shift. In [38], an offloading algorithm utilizing DRL was proposed to maximize the collective utility of users in RIS-empowered MEC systems. A deep Q-network (DQN) algorithm to design computation offloading and RIS phase shift to minimize long-term total energy consumption presented in [39]. Furthermore, the work in [40] explored the integration of machine learning techniques into future 6G networks to enhance bandwidth management, access capabilities, and reliability. Their findings on machine learning strategies to optimize network operations could be critical for implementing secure and efficient edge computing in vehicular networks.

B. OBJECTIVE AND CONTRIBUTION

The main focus of this work is to enhance task offloading and execution efficiency while maintaining a higher level of secrecy. However, achieving both improved task rates and security presents challenges. Although MEC/VEC can support task offloading in such scenarios, accessing them promptly, efficiently, and securely remains debatable, especially in dynamic VECN environments. To tackle these challenges, we suggest an innovative method referred to as DRL under the scheme of secure and task-efficient offloading assisted by ze-RIS, known as DRSTO. ze-RIS and conventional passive RIS differ significantly in terms of power needs. Unlike conventional passive RIS, which depends on a constant power supply, ze-RIS functions without external power, utilizing energy harvesting or wireless power transfer. Despite their power differences, both types improve wireless communication by reflecting and adjusting radio waves [15], [41]. The target of DRSTO is to enhance the optimization of task offloading decisions by prioritizing task processing efficiency and secrecy as primary objectives. This scheme leverages RIS to enhance task offloading efficiency and ensure secure communication within the VECN. Following are the main characteristics of DRSTO and contributions.

- This paper introduces an innovative task offloading scheme called VECN dynamics-aware DRSTO for task offloading. It utilizes the NR-V2X framework, which is improved with RIS to enhance security and efficiency. The DRSTO scheme combines different wireless technologies and transmission methods in the sub-6GHz and 28GHz frequency bands. Because of this, vehicles can submit queries to VEC servers for tasks that they normally would handle locally or send elsewhere. There are two ways for vehicles to connect to the VEC server when they want to discharge: directly, through a vehicle to RSU to VEC link, or indirectly, through a vehicle to RIS to RSU to VEC link. The location and connectivity of the vehicle determine the exact route it will take.
- Achieving optimal computation outsourcing is the main goal of this methodology. This will ensure that computational and communication resources are distributed

efficiently while maintaining anonymity. For this purpose, the optimization objective is defined as a markov decision process (MDP), and the VECNs' dynamic configuration is well-suited for it. The decision-making process of the DRL agent is guided by a compensation function that emphasizes both task efficiency and confidentiality. The function takes into account time savings and the rate of secrecy. In an effort to optimize latency secrecy rate and offloading decisions the DRSTO scheme which is supported by DRL uses the advantage actor critic (A2C) algorithm to boost the system's overall efficiency and security.

- It is required to continuously evaluate the state of the VEC network to maintain and improve the system's effectiveness. Central to this assessment should be the degree to which communication components and local and peripheral resources are interconnected. To guarantee an accurate and fair evaluation of computational offloading a new factor called task offloading efficiency and secrecy (TES) has been introduced. Time savings and secrecy rate are both included in this metric which provides a thorough evaluation of the offloading processes performance.
- The simulation results show that the DRSTO scheme outperforms competing RIS-enhanced DRL-driven task offloading methods. It achieves higher rewards faster with lower variation. Furthermore, the DRSTO framework demonstrates shorter learning times, a more favorable ratio of successful to unsuccessful tasks, and an increased TES factor.

C. ARTICLE ORGANIZATION

The structure of this article unfolds as follows: Section II elucidates the system model, encompassing the communication framework, task and processing decision model, and the local and edge computing models. Section III and Section IV delve into the problem formulation and the proposed methodology, respectively. Section V describes the simulation setup, presents the results, and provides an analysis. Lastly, the article concludes in Section VI.

II. SYSTEM MODEL

A collaborative architecture is introduced in this section, which involves vehicles and edge components, and utilizes ze-RIS to enhance NR-V2X communications within the VECN.

A. COMMUNICATION FRAMEWORK

Our communication framework considers next generation evolved node B (gNB) type RSUs along the roadsides. Additionally, RISs are installed on the roadsides or buildings alongside the road, as depicted in Fig. 1. In this setup, resource-constrained vehicles, denoted as ve_i, request processing of resource-intensive tasks from VEC servers at RSU rs_j.

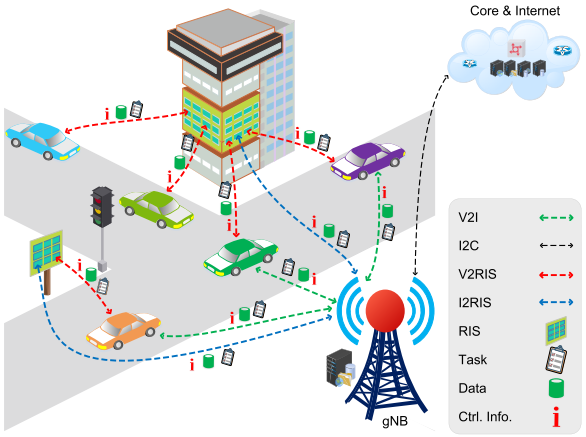


FIGURE 1. Illustration of the system model showcasing the framework's mechanism for task offloading in RIS-assisted symbiotic VECN.

The communication architecture we use in this study is similar to our previous work [42], [43], and it operates in both the mmWave and sub-6GHz bands. Legitimate vehicles are represented as ve_j, while non-legitimate vehicles or users are denoted as ve_k. Both ve_j, ve_k are equipped with N antennas, and the RISs (ri_l) are equipped with M reflection elements. Vehicles ve_j can communicate with each other and with the VEC server attached to rs_i through Uu links or mmWave links, facilitated by the RSUs (rs_i) and RISs (ri_l). Here, the variables $i, j, k, l \in \mathcal{N}$ and $\mathcal{N} = \{1, 2, \ldots, \mathcal{N}\}$.

Incorporating RIS in the system enables vehicles (ve_j) to establish direct and reflected links for communication with RSUs (rs_i). Vectors have been boosted by channel representing the direct link between rs_i and ve_j and ve_k are represented by $\mathbf{h}_{d,ve_j} \in \mathbb{C}^{N\times 1}$ and $\mathbf{h}_{d,ve_k} \in \mathbb{C}^{N\times 1}$, respectively. The channel gain matrix from rs_i to ri_l is denoted by $\mathbf{H}_{rs_i,ri_l} \in \mathbb{C}^{N\times M}$. Furthermore, the channel gain between ri_l and ve_j and ve_k is defined as $\mathbf{g}_{ri_l,ve_j} \in \mathbb{C}^{M\times 1}$ and $\mathbf{g}_{ri_l,ve_k} \in \mathbb{C}^{M\times 1}$, respectively.

The RIS phase shifts are represented by the matrix Θ, which is a diagonal matrix where each diagonal element corresponds to the phase shift introduced by an RIS element. The matrix Θ can be denoted as:

$$\Theta = \text{diag}(\theta_1, \theta_2, \ldots, \theta_M) \tag{1}$$

Here, M represents the number of reflection elements in the RIS. Each element θ_m in the matrix Θ is defined as:

$$\theta_m = \beta_m \exp(z\phi_m) \tag{2}$$

where β_m is the reflection coefficient of the m^{th} reflection element, which determines the amplitude of the reflected signal. It takes values within the range [0, 1], representing the proportion of the incident signal power that is reflected. ϕ_m is the phase angle introduced by the m^{th} reflection element to the incident signal, and its value is restricted within the range $[0, 2\pi)$. This phase angle is crucial for controlling the direction and constructive interference of the reflected signals.

The expression $\exp(z\phi_m)$ represents the complex exponential form of the phase angle ϕ_m, where $z = \sqrt{-1}$. This form is used to describe the phase shift mathematically.

To simplify the analysis, we assume that the reflection coefficient β_m equals 1 for all elements. This assumption does not compromise generality because it simplifies the phase shift design by focusing solely on the phase angle ϕ_m. Thus, the phase shift element θ_m becomes:

$$\theta_m = \exp(z\phi_m) \tag{3}$$

Remark 1: In RIS-based communication, the synchronization technique mirrors the approach used in traditional backscatter communication [44]. The vehicle or RSU utilizes a wireless control link while going through the transmission phase to send synchronization details, for example an activation signal for the RIS controller [45]. Afterwards, synchronization information is interpreted by the controller based on RIS, at the same time it sets up the reflection state mode of the RIS. The synchronization stage allows for the harmonization of the vehicle or RSU with the RIS, hence promoting effective communication.

Remark 2: The primary emphasis of this study is task offloading, with no investigation into the potential integration of RIS with ZEDs. For those interested in comprehending how ZEDs work in tandem with RISs, we suggest consulting the studies presented in [15], [16], [46], and [47] for further insight.

Under the assumption that ve_j communicates with RSU rs_i through the direct channel, the received signal at RSU r_i can be expressed using the formulation presented in [48] as follows:

$$y_{d,ve_j} = \mathbf{h}^{\mathbf{H}}_{d,ve_j} \mathbf{w}_j \, s_j + n. \tag{4}$$

In the direct communication scenario, the expression involves various parameters. The term $\mathbf{h}_{d,ve_j}$ represents the vector denoting the channel gain of the direct link connecting RSU rs_i and vehicle ve_j. The vector $\mathbf{w}_j$ denotes the beamforming vector specifically allocated to the vehicle ve_j. Additionally, s_j corresponds to the information signal intended for vehicle ve_j, while n signifies the existence of noise at the receiver that is designed to be uncorrelated with s_j. Modeled with an average value of 0 and a standard deviation of σ^2, n is characterized, i.e., $n \sim \mathcal{CN}(0, \sigma^2)$. In a similar manner, when ve_j communicates with RSU rs_i using the reflected channel, and the RIS ri_l is employed for signal reflection, the received signal at RSU rs_i can be derived according to the formulation in [49] as follows:

$$y_{r,ve_j} = \left(\mathbf{h}^{\mathbf{H}}_{d,ve_j} \mathbf{H}_{rs_i,ri_l} \Theta \, \mathbf{g}_{ri_l,ve_j}\right)^H \mathbf{w}_j \, s_j + n. \tag{5}$$

In this context, matrix $\mathbf{H}_{rs_i,ri_l}$ represents the channels between RSU rs_i and RIS ri_l. Θ denotes the phase shift matrix of RIS ri_l, which controls the phase shifts of its reflection elements. Moreover, the vector $\mathbf{g}_{ri_l,ve_j}$ represents the channels from RIS ri_l to vehicle ve_j. To compute the total received signal at RSU rs_i, the maximum ratio combining (MRC) approach is used to combine the direct and reflected signals. Suppose there are J vehicles communicating with RSU rs_i via direct and reflected channels. The combined received signal at the RSU can be expressed as follows:

$$y_{i,j} = \sum_{j=1}^{j} \alpha_j \, y_{d,ve_j} + \sum_{j=1}^{j} \beta_j \, y_{r,ve_j} + n. \tag{6}$$

Here, α_j and β_j are the weights assigned to the direct and reflected channels. Both α_j and β_j are typically between 0 and 1, where $\alpha_j + \beta_j = 1$. Similarly, the direct y_{d,ve_k} and indirect y_{r,ve_k} received signals from ve_k at RSU rs_i can be calculated by following Eqs.(4) and (5), respectively. However, the total received signal from ve_k at the RSU rs_i can be expressed as:

$$y_{i,k} = \sum_{k=1}^{k} \alpha_k \, y_{d,ve_k} + \sum_{k=1}^{k} \beta_k \, y_{r,ve_k} + n. \tag{7}$$

Next we adopt the Shannon capacity formula to describe the rate of direct and indirect links. This theorem specifies that the highest data rate (in bit/s) achievable on a communication channel, given its bandwidth B (in hertz) and its signal-to-noise ratio (SNR) γ is determined by this theorem as follows:

$$R = B \log_2(1 + \gamma) \tag{8}$$

$$\text{where } \gamma = \begin{cases} \gamma_{d,j} & \text{for direct link,} \\ \gamma_{i,j} & \text{for indirect link.} \end{cases} \tag{9}$$

$\gamma_{d,j}$, and $\gamma_{i,j}$ represent the SNR for the direct and the indirect links, respectively. To compute the SNR for the direct link, refer to the formula in (10) as per [50]. Similarly, the SNR for the indirect link is derived using the formula in (11), in accordance with [51].

$$\gamma_{d,j} = \frac{\left|\mathbf{h}^H_{d,ve_j} \mathbf{w}_j\right|^2 P_j}{\sigma^2}, \tag{10}$$

$$\gamma_{i,j} = \frac{\left|\mathbf{g}^H_{ri_l,ve_j} \Theta^H \mathbf{H}^H_{rs_i,ri_l} \mathbf{w}_j\right|^2 P_j}{\sigma^2}. \tag{11}$$

$\mathbf{h}^H_{d,ve_j}$ and $\mathbf{g}^H_{ri_l,v_j}$ are the Hermitian transpose (conjugate transpose) of the channel gain vector associated with the direct link and the indirect link. $\mathbf{w}_j$ represents the beamforming vector designated for vehicle ve_j. P_j denotes the transmit power for vehicle ve_j and σ^2 is noise power. The $\mathbf{h}_{d,ve_j}$ between rs_i, and ve_j is impacted by the value produced by a vehicle, such as distance and speed. It can be formulated as given in [52]:

$$\mathbf{h}_{d,ve_j} = \sqrt{\frac{\beta_0 \, d^{\alpha}_{0,d}}{d^{\alpha}_{rs_i,ve_j}}} \, n. \tag{12}$$

Here β_0 is the path loss at a reference distance of $d_{0,d} = 1$ meter, α is the path loss exponent, and d_{rs_i,ve_j} is the distance between rs_i and ve_j. Additionally, $\mathbf{H}rs_i, ri_l$ corresponds to the channel gain matrix from rs_i to ri_l. The indirect link channel

gain is also affected by distances between the vehicle, RSU, and RIS, as well as vehicle speed. Moreover, Θ^H signifies the conjugate transpose of the RIS-associated phase shift matrix. By following [29], the indirect link channel gain among ve_j and ri_l is formulated as:

$$h_{i,j} = f(d_{ri_l,ve_j}, \Theta, \mathbf{H}_{rs_i,ri_l}) \cdot n. \tag{13}$$

Here d_{ri_l,ve_j} denotes the distance of a vehicle from the IRS system. The function $f(d_{ri_l,ve_j}, \Theta, \mathbf{H}_{rs_i,\mathfrak{ri}_l})$ incorporates the channel gain's characteristics, which are influenced by both the distance and phase shifts from the RIS and the vehicle, along with the channel gain matrix $\mathbf{H}_{rs_i,ri_l}$. The rates R_{ve_j} and R_{ve_k} for vehicles ve_j and ve_k can be calculated for both the direct and indirect links by following the (8), (10), and (11). If we assume that any of the ve_k vehicles, may attempt to intercept the communication, then the physical layer security of ve_j can be computed using following expression:

$$\mathfrak{R}_j^s = \max\Big[0, \Big(R_{ve_j} - \max_{\forall j \in j_{max}} \big(R_{ve_k}\big)\Big)\Big]. \tag{14}$$

B. TASK AND PROCESSING DECISION MODEL

Each offloading vehicle ve_j produces τ_j tasks at the start of t-time slot. Each offloading task τ_j consists of a 6-tuple $\tau_j = \{cy_j, sz_j, t_j, t_j^{max}, d_j^{rs}, d_j^{ri}\}$, where $cy_j = \hbar \cdot sz_j$. Here, cy_j and sz_j denote the required CPU cycles and task input size in bits, respectively. In the computing model, the relationship between cy_j and sz_j is represented by the service coefficient $\hbar$. t_j represents the arrival time of τ_j, whereas the maximal permissible execution time available for τ_j is defined by t_j^{max}. Every offloading vehicle is limited to generating a single task per t-time slot. The determination of generating tasks' development follows a Bernoulli distribution, where the probability is $\mathcal{P}$; thus, the average task arrival rate is $\lambda_\tau = \mathcal{P}/t$ [53].

Every task τ_j requires a decision to be made regarding its processing method, either locally or through offloading. This decision is represented by d_j, which is a binary variable with values of either 0 (local processing) or 1 (offloaded processing).

C. LOCAL AND EDGE COMPUTATION MODEL

We assume that all offloading vehicles ve_j have a similar CPU frequency, indicated by fc_j. In implementing a partial task offloading mechanism, we utilize the coefficient $\tilde{\alpha}$ to indicate the portion of task τ_j that will be processed locally. The $\tilde{\alpha}$ is subject to the constraint $0 \leq \tilde{\alpha} \leq 1$. τ_j is not the sole task awaiting processing; there are other tasks in the queue of ve_j as well, and their respective sizes are cumulatively denoted by sz_j^v. Calculation of the average local processing delay T_{l_j} with the help of the equation might be used for sz_j-sized task τ_j.

$$T_{l_j} = (1 - d_j)\Big\{\Big(\hbar(\tilde{\alpha} sz_j + sz_j^v)\Big)/fc_j\Big\}. \tag{15}$$

Similarly, concerning the edge computation model, we assume that each RSU has a multi-CPU VEC server capable of parallel computing. In each time period, the number of available CPUs in the VEC network fluctuates in a dynamic manner. If vehicle ve_j is unable to finish task τ_j within the allotted time t_j^{max}, it is advisable to offload the task to the VEC server, the decision value d_j is set to 1. The CPU resources association on the server is influenced by the incoming tasks τ_j in each time interval. Due to the fact that these tasks appear during distinct time slots and have distinct CPU requirements, different time slots provide disparate amount of available VEC CPU resources. For CPU resources distribution, the VEC server prioritizes tasks τ_j based on the longest given CPU time within its associated time slot. If multiple CPUs have the same available time, one is randomly selected. The aim of this research is to posit the supposition that the VEC server has ample computational capability, and each unit in face of vehicles engaged in offloading can be allocated fc_e computation resources. Transmission delays are an essential factor to consider when processing a task τ_j on the VEC server. The processing time $\mathcal{T}_{e_j}$ for task τ_j at the VEC server can be calculated as follows:

$$T_{e_j} = d_j\Big(\frac{\hbar(\tilde{\beta} sz_j^o)}{fc_e} + \frac{\tilde{\beta} sz_j^o}{\mathfrak{R}_j^s}\Big). \tag{16}$$

The term $\mathfrak{R}_j^s$ is a general reference to the transmission rate, which applies to either direct or indirect links accordingly. The variable $\tilde{\beta}$ denotes the portion of the task to be executed at the VEC, and its value lies between 0 and 1. It is important to note that both $\tilde{\alpha}$ and $\tilde{\beta}$ must satisfy the constraint $\tilde{\alpha} + \tilde{\beta} = 1$.

III. FORMULATION OF OPTIMIZATION PROBLEM

Here we study the optimization of computational efficiency and the maximization of secrecy rates in the VEC network. The main objective is to minimize task processing time while efficiently using computational resources, thereby increasing the ratio of total executed tasks. Additionally, a secondary goal is to enhance the secrecy rate for legitimate task-offloading vehicles to ensure secure communication. The primary aim is to balance efficient computation and a high secrecy rate.

In the dynamic VECN system, where execution time varies between time slots, fairness and robustness are critical concerns. The total task processing time is calculated using (17) to ensure a fair comparison among different scenarios. To address this variability, the TES factor is introduced for rational evaluation. The local processing time, $\mathfrak{P}$, of a specific task, τ_j, is used as a benchmark for comparison. In accordance with our intended objective, the total time consumed by an offloading vehicle can be denoted as follows:

$$\Gamma_j = \big(1 - d_j\big)\ \mathcal{T}_{l_j} + d_j \cdot \mathcal{T}_{e_j}. \tag{17}$$

A. TASK OFFLOADING EFFICIENCY AND SECRECY FACTOR

The evaluation of the offloading decision's effectiveness relies on the TES factor. The TES factor (Ψ) is a

comprehensive metric used to evaluate the effectiveness of task offloading decisions. It integrates two critical aspects, the time efficiency and the secrecy rate. The time efficiency component measures the cumulative time saved from offloading and processing tasks, represented by the difference between the total potential processing time without offloading and the actual time used with offloading, adjusted by specific task-related parameters. The secrecy rate component adds a security dimension, reflecting the level of data protection achieved during the offloading process. The formula for estimating the time saved involves multiplying by $\mathfrak{P}$ and Γ with the difference of t_j^{max} and Γ, providing a measure of relative efficiency as

$$\Psi = \left((\mathfrak{P} - \Gamma_j \cdot t_j^{max} - \Gamma_j) + \mathfrak{R}_j^s\right). \tag{18}$$

The metric Ψ, measures the amount of time that is saved by processing and offloading a single task, and also considers the secrecy rate. By computing the average of Ψ across all J tasks, denoted as Ψ_{avg}, we obtain an overall assessment of the time saved by task processing and offloading, as well as the corresponding secrecy rates. A higher value of Ψ_{avg} signifies a more efficient offloading decision, leading to substantial time savings and enhanced secrecy rates. Eq (19) represents the expression for the average relative TES factor.

By averaging the TES factor across all tasks (Ψ_{avg}), an overall assessment of the time saved and the corresponding security levels is obtained. A higher average TES factor signifies a more efficient offloading decision, leading to substantial time savings and enhanced data security. This metric provides a balanced evaluation, ensuring that offloading strategies optimize both performance and security in VEC networks.

$$\Psi_{avg} = \frac{1}{J}\sum_{j=1}^{J}\left((\mathfrak{P} - \Gamma_j \cdot t_j^{max} - \Gamma_j) + \mathfrak{R}_j^s\right). \tag{19}$$

Hence, the primary goal of optimization is to enhance the average time savings and achieve an elevated level of secrecy rate during the execution and offloading of tasks. The formulation of this objective is outlined as follows:

$$\begin{aligned}
&\underset{\Psi,\,\mathbf{w},\Theta}{\text{maximize}} \ \frac{1}{J}\sum_{j=1}^{J} \Psi_{avg},\\
&s.t. \quad C1: \min\left\{T_{l_j},\ T_{e_j}\right\} \le t_j^{max},\\
&\qquad C2: \mathfrak{P} = \text{mean}(fc_j^{min}, fc_j^{max}),\\
&\qquad C3: P_j \le P_j^{max},\\
&\qquad C4: \mathfrak{R}_j^s \ge \frac{\tilde{\beta} sz_j^o \cdot fc_e \cdot d_j}{d_j \cdot \tilde{\beta} sz_j^o \cdot Te_j - \hbar(\tilde{\beta} sz_j^o)},\\
&\qquad C5: \tilde{\alpha} + \tilde{\beta} = 1,\\
&\qquad C6: \forall\, d_j \in \{0,1\} \text{ and } \forall\, j \in \{1,2,3,\ldots,J\}.
\end{aligned} \tag{20}$$

Constraint C1 ensures that the minimum value between the local task processing time T_{l_j} and the VEC processing time T_{e_j} does not exceed the task threshold time t_j^{max}. Constraint C2 sets the benchmark time $\mathfrak{P}$ as the average of the minimum local CPU frequency and the maximum local CPU frequency. Constraint C3 pertains to the vehicle's transmission power, which stipulates that the maximum transmission power must not surpass the power allocated to the vehicle's transmission unit during the beamforming and transmission of task raw data to the VEC server. Constraint C4 ensures that the minimum secrecy rate $\mathfrak{R}_j^s$ must be greater than or equal to the time required to transfer the task's raw data sz_j^o to the VEC server during offloading. Constraint C5 corresponds to the notion that the local portion of the task and the offloaded part of data together form a complete task. Lastly, constraint C6 dictates that the decision parameters d_j is set to 0 for local processing. However, if there are no available free CPUs locally in the given time slot, d_j is set to 1.

Upon close examination of the objective function in optimization problem (20), we notice that it takes the form of a fraction and incorporates multiple random variables. Additionally, the constraints in (C1), (C3), and (C4) are non-convex. Moreover, the variables are interrelated, making the optimization problem (20) non-convex and inherently stochastic in nature. Due to these complexities, it is not feasible to directly solve this problem using traditional optimization techniques.

IV. DRL METHODOLOGY FOR SOLVING TES MAXIMIZATION PROBLEM

Improving the propriety across tasks is the optimization access's main challenge. This approach uses the way of reducing the medium processing delay for every J tasks/subtasks coming at each time slot t while simultaneously maximizing the secrecy rate. However, conventional optimization methods encounter challenges in fulfilling this goal because of influences from the dynamic nature of tasks and intricacies inherent in the VECN system. To overcome these challenges, we propose a novel strategy known as the DRL-based ze-RIS-assisted secure task efficient offloading (DRSTO) approach. This approach leverages the capabilities of DRL to make effective, time-delay, and secrecy-efficient computation offloading-based choices. In this section, we formulate Problem (20) as an MDP, rendering it suitable for the DRL method to learn the optimal policy within the training environment.

A. MDP MODEL FORMULATION

An MDP is usually composed of five elements, namely $\mathcal{M} = \{S, A, P, R, \gamma\}$. Each component plays a distinct role in tackling the issue of task offloading. The following subsections describe in detail the operations carried out by elements of MDP.

1) STATE SPACE

The state space is directly associated with the environmental states in which a DRL learning agent interacts. Our scenario defines the state space S, where an environmental state at time slot t is represented as $s_t \in S$. In line with the task offloading

environment, the state s_t for each task τ_j is described as follows:

$$s_t = \left\{(cy_j, sz_j, t_j, st_{loc}, st_{off}, st_{vec}, d_j^{rs}, d_j^{ri}, \Re_j^s, \Psi_{avg})\right\} \quad (21)$$

The DRL agent gathers information on various aspects at time slot t, including the CPU cycles (cy_j), task size (sz_j), and arrival time (t_j) of τ_j. Additionally, the DRL agent observes the states related to processing and offloading, such as the local processing state (st_{loc}), the data offloading state (st_{off}), and the VEC processing state (st_{vec}). Furthermore, the DRL agent also observes the inter-vehicle-RSU distances (d_j^{rs}), vehicle-RIS distances (d_j^{ri}), secrecy rate ($\Re_j^s$), and the average TES factor (Ψ_{avg}) associated with the task τ_j.

2) ACTION SPACE

The DRL learning agent engages with the environment, observing the state elements mentioned in (21). After observing these state elements, the agent takes actions A. At time slot t, the learning agent performs action a_t based on state s_t. The action a_t for each task τ_j is described as follows:

$$a_t = \{loc^e, \tau^{off}, vec^e\} \quad (22)$$

The learning agent has three potential actions to take based on the observed state at time slot t. If there is sufficient CPU capacity available locally, it can choose the local execution action (loc^e). However, if there is a shortage of CPU cycles locally or no free CPU is available at the same time slot, the agent selects the task for offloading to the VEC server. In this scenario, it first takes the offloading action (τ^{off}) and then the VEC execution action (vec^e) in an interleaved manner. Additionally, the agent examines the communication environment before making the offloading and VEC execution actions to ensure secure communication. Consequently, while observing the communication environment at time slot t, the agent selects suitable values for ($\mathbf{w}$) and (Θ), accordingly.

3) REWARDS

The reward function is of utmost importance as it determines the state transition quality level. Suppose the action a_t is taken by agent at time slot t, leading to a transition from state s_t to s_{t+1} and receiving a reward denoted as r_t.

$$R(s_t, a_t, s_{t+1}) = \left(\Psi_{avg}(s_t, a_t, s_{t+1})\right) \cdot w. \quad (23)$$

The parameter w in (23) serves the purpose of fine-tuning the magnitude of the rewards within the system. Its role is to guarantee that the range of rewards is accurately adjusted to match the intended goals and linked limitations within the issue.

B. THE DRSTO SCHEME

The proposed DRSTO scheme integrates a VECN environment with the A2C algorithm. This scheme comprises the A2C algorithm, which serves as the central framework for reinforcement learning, and the DRSTO environment, which emulates VECN theory. Under the proposed DRSTO scheme, the components interact with each other, consistently enhancing secure offloading decision functionality.

The A2C algorithm is a model-free, synchronous method in reinforcement learning, combining the advantages of both policy-based and value-based methods. It consists of two primary components, the Actor and the Critic. The Actor is responsible for selecting actions based on the policy, while the Critic evaluates the chosen actions by computing the value function, which estimates the expected future rewards.

In the DRSTO scheme, The Actor component outputs a probability distribution over actions, enabling the agent to make stochastic decisions that are not solely deterministic. This allows for exploration of the action space, which is crucial for learning in dynamic VECN environment. The Critic component calculates the value function, providing feedback on the quality of the actions taken by the Actor. The advantage function, which is the difference between the expected reward of the chosen action and the average expected reward, helps in reducing variance and stabilizing the learning process. The agent is equipped with the ability to optimize its actions using the A2C algorithm, which considers both the observable environmental conditions and the rewards obtained. By continuously interacting with the DRSTO environment, the agent learns to make optimal offloading decisions based on different VECN scenarios through training. The synchronous nature of A2C ensures efficient utilization of computational resources and facilitates faster convergence compared to asynchronous methods.

1) DRSTO TRAINING ALGORITHM

The DRSTO training procedure initiates by setting up the actor and critic networks with randomly assigned weights w_θ and w_v, respectively. Additionally, critical parameters such as the learning rate and the target return discount factor are defined. The initial state is set and the experience buffer is created. In its iterative process, the algorithm selects actions, updates the VECN environment, logs transitions in the experience buffer, and computes target returns as well as value function targets, utilizing rewards and the discount factor. The weights w_θ of the actor network are then updated according to the policy gradient method, i.e,

$$\nabla_\theta J \approx \frac{1}{|D|}\sum_{t=0}^{|D|} \nabla_\theta \log(\pi(a_t|s_t; w_\theta)) \cdot A(s_t, a_t; w_\theta, w_v). \quad (24)$$

The process involves determining the advantage function and policy function of logarithm, which are derived by averaging the gradient across transitions stored in the experience buffer D. The weights of the critic network are updated by using the mean squared error loss, reflecting the difference between predicted and actual outcomes, i.e,

$$L_{\text{critic}} = \frac{1}{|D|}\sum_{t=0}^{|D|}(z_t - V(s_t; w_v))^2. \quad (25)$$

This update process is based on comparisons between the estimated value functions and their respective targets. Through given learning rate and calculated gradients, the parameters of actor and critic are updated. After completing this step, the experienced buffer is cleared, and process continues until convergence, signifying that the algorithm has reached its optimal performance.

Algorithm 1 A2C Based DRSTO Training Algorithm

Input: Ψ, Ψ_{avg},, st_{loc}, st_{ofld}, st_{vec}
Output: secure offloading decision policy d_j
Initialize $\Psi = \Psi_{avg} = 0$ Initialize the actor and critic network with w_θ, w_v, and random weights Initialize discount factor γ, learning rate α, empty buffer D, initial state s
while (*notconverged*) **do**
  Interact with VECN environment, initialize state s Select and execute action a on current state s, and observe next state s' and reward r Store the transition (s, a, r, s') in buffer D Update the current state s to the next state s'
  if (*episode* $\neq$ *over*) **then**
    Initialize empty lists G, and z Set G_t for the last time step as final reward r Set z_t for the last time step as the target return G_t **foreach** *t time step in episode* **do**
      $G_t = r_t + \gamma \cdot G_{t+1}$
      $z_t = G_t$
    end
    Update weights w_θ using (24)
    Update weights w_v using (25)
    $w_\theta \leftarrow w_\theta + \alpha \cdot \nabla_\theta J$
    $w_v \leftarrow w_v + \alpha \cdot \nabla_v L_{\text{critic}}$
    Reset the experience buffer D
  end
end

a: TRAINING OPTIMIZATION

In the phase in training that involves exploration, the agent randomly selects actions, but not all of these actions are valid due to constraints related to the system model. To address this, we introduce a penalty strategy during sampling to discourage non-legitimate actions and prioritize valid ones. These penalties are applied to actions like local processing, offloading, and edge processing if they are not feasible. The penalty term Υ is included in the reward function to guide the agent towards making optimal decisions considering constraints, as

$$R(s_t, a_t, s_{t+1}) = \left(\Psi_{avg}(s_t, a_t, s_{t+1}) - \Upsilon\right) \cdot w. \tag{26}$$

In this context, Υ can take on one of three values: Υ_l, Υ_o, or Υ_e. Specifically, Υ_l stands for penalties related to local actions, Υ_o represents penalties associated with offloading actions, and Υ_e pertains to penalties linked to edge processing actions. The DRL agent in DRSTO scheme is to train as a primary objective to make efficient offloading decisions while adhering to system constraints. The agent operates within the VECN environment as outlined in Section IV-A.

2) DRSTO ALGORITHM

The environment of task offloading in VECN and DRL agent is integrated in DRSTO algorithm. The DRL agent operations is described by algorithm 2 while training environment is described by Algorithm 1 while algorithm 1 provides a comprehensive detail of the training method for the DRL agent. Collectively, these algorithms collaborate coherently to achieve their shared goal of secure and efficient task offloading.

The core of the Algorithm 2 operates within a loop, persisting until the task τ_j is marked as "done," signaling the completion of all tasks. Within this iterative loop, several key steps are executed. First, the algorithm retrieves the task offloading decisions from Algorithm 1). It subsequently computes and returns both the TES factor and the average TES factor, regardless of local or VEC computation decision. When the decision dictates local processing $d_j == 0$, the local state st_{loc} is updated. If the local state reaches zero, it signifies task completion, prompting the algorithm to proceed to the next task in the queue.

The algorithm also accounts for offloading decisions. When the decision points offloading $d_j == 1$, the algorithm evaluates the vertical distance of the vehicle ve_j concerning both the RSU and the RIS. Depending on this comparison, it calculates values for the vehicle's stay time $t^s_{ve_j}$ for both direct and indirect links as

$$t^s_{ve_j} = \begin{cases} \dfrac{\left(2\sqrt{\left(rs^r_i\right)^2 - \mathfrak{e}_{rs_i}}\right)}{\|\vec{\mu_{ve_j}}\|} & \text{under RSU,} \\ \dfrac{\left(2\sqrt{\left(ri^r_l\right)^2 - \mathfrak{e}_{ri_l}}\right)}{\|\vec{\mu_{ve_j}}\|} & \text{under RIS.} \end{cases} \tag{27}$$

In this context, $\vec{\mu_{ve_j}}$ represents the velocity vector of vehicle ve_j. The notation $\mathfrak{e}_{rs_i}$ refers to vertical distance within RSU rs_i and vehicle ve_j. Similarly, $\mathfrak{e}_{ri_l}$ indicates the vertical distance within RIS ri_l and vehicle ve_j. Additionally, rs^r_i and $\mathfrak{r}i^r_l$ are used to specify the communication range radius of rs_i and ri_l, correspondingly.

Then the DRSTO algorithm selects the secrecy rate $\Re^s_j$ either for the direct link between the vehicle ve_j and the RSU rs_j or the indirect link via the RIS ri_l. To increase the secrecy rate, a secure beamforming mechanism is adopted while fine-tuning the beamforming vectors to direct the main signal precisely toward the intended legitimate receiver, while simultaneously nullifying signals in the directions of unauthorized eavesdroppers [54]. We employed the maximal ratio transmission (MRT) technique to maximize the γ at ve_j while minimizing the γ at the ve_k. Considering the MIMO configurations the MRT beamforming vector $\mathbf{w}$ is calculated as follows [55]:

$$\mathbf{w} = \frac{\mathbf{h}^H \mathbf{e}_{rx}}{\|\mathbf{h}\mathbf{e}_{rx}\|} \tag{28}$$

$\mathbf{h}$ represents the channel vector, which contains the complex channel gains from each transmitter antenna to the receiver

Algorithm 2 The DRSTO Algorithm

Input: Task τ_j's tuple, rs_j, ri_l, $\mathfrak{e}_{rs_i}$, $\mathfrak{e}_{ri_l}$, $\mathfrak{R}_j^s$, $t_{ve_j}^s$
Output: Ψ, Ψ_{avg},, st_{loc}, st_{ofld}, st_{vec}
Initialize: $\tau_j = !done$, $\Psi = \Psi_{avg} = st_{loc} = st_{ofld} = st_{vec} = 0$
while $(\tau_j \neq done)$ **do**
  fetch d_j from *Algorithm 1*, calculate and return Ψ, Ψ_{avg}
  if $(d_j == 0)$ **then**
    update the state st_{loc}, and return **if** $(st_{loc} == 0)$ **then**
      set task completion flag of $\tau_j = done$ and move to next τ_j
    end
  else if $(d_j == 1)$ **then**
    if $(\mathfrak{e}_{rs_i} < \mathfrak{e}_{ri_l})$ **then**
      select $t_{ve_j}^s$, and $\mathfrak{R}_j^s$ of direct ve_j–rs_j link
    else
      select $t_{ve_j}^s$, and $\mathfrak{R}_j^s$ of indirect ve_j–ri_l–rs_j link
    end
    if $(t_{ve_j}^s > (\tilde{\beta} sz_j^o / \mathfrak{R}_j^s))$ **then**
      update states st_{ofld}, st_{vec}, and return **if** $(st_{ofld} == st_{vec} == 0)$ **then**
        set flag of $\tau_j = done$ and move to next τ_j
      end
    end
  end
  if $\left(t_{ve_j}^s > (\tilde{\beta} sz_j^o / \mathfrak{R}_j^s)\ \&\ d_j == 1\ \&\ \tau_j == done\right)$ **then**
    transfer output result to ve_j **else**
      transfer output result to rs_{j+1} or ri_{l+1} in headway of ve_j
    end
  end
end

antennas. $\mathbf{h}^H$ is the conjugate transpose of the channel vector $\mathbf{h}$. The $\|\mathbf{h}\|$ is the Euclidean norm of the channel vector $\mathbf{h}$, and $\mathbf{e}_{rx}$ is the unit vector representing the receiver's antenna. The transmitted signal $\mathbf{x}$ using MRT is formed by scaling the information symbol s with the MRT beamforming vector $\mathbf{w}$ [55]:

$$\mathbf{x} = \sqrt{P} \cdot \mathbf{w} \cdot s + n. \tag{29}$$

Subsequently, the DRSTO algorithm verifies if the vehicle can remain for an adequate duration to facilitate the task upload while considering the secrecy rate. When the conditions align, it updates both the offloading state st_{ofld} and the edge processing state st_{vec}. Task completion is recognized when both the offloading and VEC states reach zero, allowing the algorithm to transition to the subsequent task.

Disseminating the task results, if the vehicle's stay time proves sufficient, and the offloading decision d_j prescribes VEC execution, in addition to the task completion flag being set to "done," the output result is directly conveyed to the vehicle ve_j. Alternatively, when any of these conditions are unmet, the output is forwarded to next RSU rs_{j+1} or RIS ri_{l+1} within the vehicle's headway.

In an VECN environment the DRL agent progresses, navigating and interacting, until the learning process reaches a point of convergence. At this juncture, where the training process reaches its peak performance, the DRSTO scheme is deemed to be thoroughly trained and optimized.

TABLE 1. Parameters for simulation.

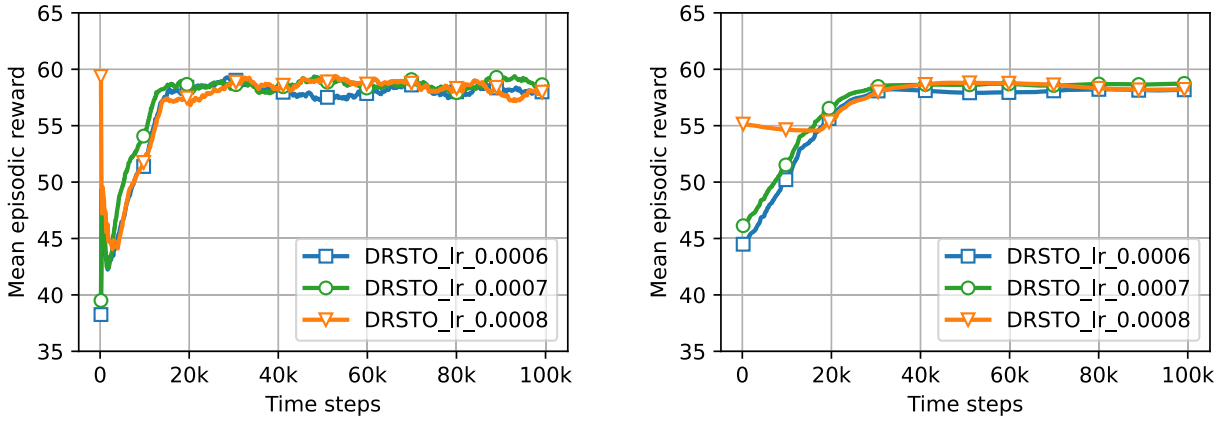
Parameter	Value
Task generation probability	0.7
CPU cycles needed for processing a single bit $\hbar$	30 CPU cycles
Frequency of v_v & VEC per CPU	1MHz–1GHz, 1.5GHz–2.5GHz
Energy coefficient κ	$2 * 10^{-9} - 2 * 10^{-6}$
Number of RSU antennas N	4
Number of RIS elements	50
Max P_j, and σ	30dBm, & -100dBm
Rician fading factor in RIS to RSU	10dB
Rician fading factor in RSU to Vehicle	3dB
PL exponent for RSU to RIS	2dB
PL exponent for RIS to Vehicle	2.8dB
PL exponent for RSU to Vehicle	3.5dB
Cellular-Bandwidth	20MHz at 800MHz
Millimeter-wave bandwidth	200MHz at 28GHz
Parameters for DRL	
Number of update steps per iteration.	100
LR, γ	0.0005, 0.99
Time steps used for training	100k
Value-function coefficient & GAE λ	0.5, 1.0

V. SIMULATIONS

In this section, we have provided comprehensive information about the simulation setup, alternative approaches, the criteria used for evaluation, results and discussion.

A. SIMULATION SETUP

The communication range of the RSUs and ze-RISs in our vehicular network scenario is 200 meters. Similarly, the communication range of the vehicles comprising the network is set at 100 meters. These vehicles are considered to move randomly along the road at speeds ranging from 10 meters per second to 20 meters per second. In the described channel configurations, the ze-RIS utilizes a UPA comprising an $M_x \times M_y$ matrix of reflection elements. A Rician fading model is adopted to characterize our channel, accounting for both LoS and Non-LoS components across the connections involving RSU-RIS, RIS-Vehicle, and RSU-Vehicle links. Our path loss model is $PL(\alpha, d) = \beta_0 (d/d_0)^{-\alpha}$, where β_0 is the path loss at $d_0 = 1$ meter, d is the transmission distance, and the path loss exponent is α. Our configuration defines the non-LoS component of the RSU-RIS connection as $H_R^{LoS} \sim \mathcal{CN}(0, 1)$ and the LoS component as $H_R^{LoS} = a_l a_t^H$. In this context, a_l and a_t represent the steering vectors, which are calculated using the angles of arrival and departure, respectively, adhering to the methodology described in [56]. Other communication links also use similar channel models, and we have set the path loss parameter β_0 to -40dB, as per the 3GPP UMi model. We choose the task size and processing threshold time randomly from a predefined set of values. The task size can vary between 1 and 50 megabytes, while the processing threshold time falls within the range of 200 to 500 milliseconds. Moreover, we have employed the widely



(a) Reward with variable LR

(b) Smoothed reward with variable LR

FIGURE 2. Examining the impact of different learning rate (LR) on the reward obtained per episode.

accepted *M/M/1-FCFS* queue processing model for task scheduling.

B. ALTERNATE METHODS AND EVALUATION

The DRSTO scheme is compared against other DRL algorithms, including DQN, PPO, QRDQN, and DDPG. Similar to our approach, the DQN-based scheme proposed in [57] learns from the environment to optimize tasks like computation offloading, resource allocation, and IRS phase shift policies. The QRDQN scheme, a modified version of DQN using quantile regression, provides a distribution of Q-values, making it suitable for tasks where risk and uncertainty are significant factors. Meanwhile, the DDPG scheme, introduced in [58], is another MEC-based secure task offloading scheme utilizing RIS assistance. Additionally, the PPO-based task offloading scheme, a variant of DDPG, is included for comparison due to its stability and reliable learning process achieved through techniques like advantage clipping and the use of surrogate objective functions.

We comprehensively evaluated the DRSTO and other competing approaches, considering two key aspects: the efficiency of task offloading and learning agents within DRL effectiveness. In terms of DRL performance, we assessed learning agents based on their rewards and the ratio of TFPS, indicating the time taken per environmental step. Regarding task offloading efficiency, we utilized the average TES factor to evaluate the joint efficiency of task offloading, considering factors such as communication and computation delays and secrecy. Additionally, we employed task success and drop ratios as quantitative measures to determine the exact number of tasks that were dropped or completed effectively in the VECN environment.

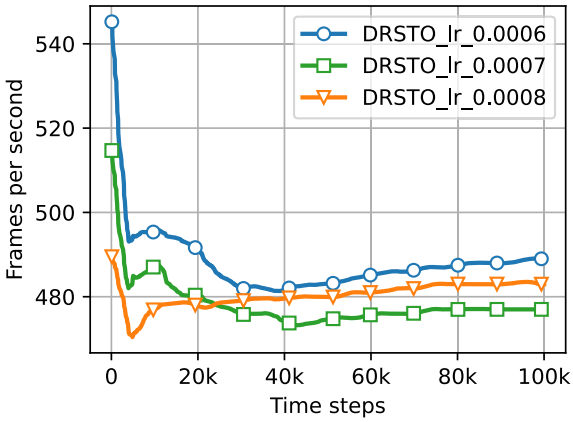
C. SIMULATION RESULTS AND ANALYSIS

Within this subsection, we have presented the results and conducted an analysis concerning both DRL performance and task offloading efficiency.

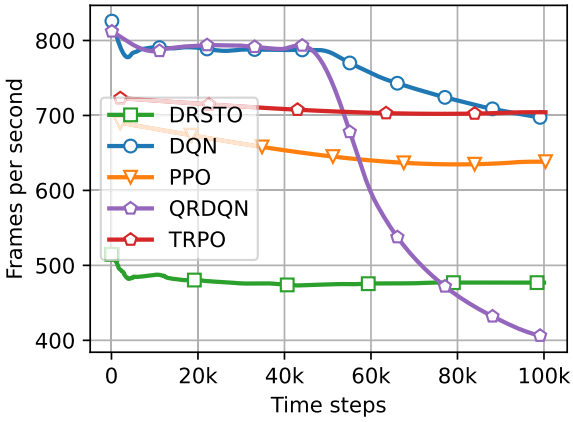
1) DRL PERFORMANCE ASSESSMENT

The performance of the DRSTO scheme is analyzed with different learning rates, lr-0.0006, lr-0.0007, and lr-0.0008, focusing on both the reward obtained and the TFPS rate achieved as shown in Fig. 2 At lr-0.0006, the DRSTO scheme demonstrated a reward of 56.83, indicating its effectiveness in optimizing the task offloading process. Additionally, it achieved a TFPS rate of 487.97, showcasing its computational efficiency in processing 487.97 TFPS. A lower TFPS rate often signifies superior performance. The algorithm achieved an impressive value of 58.56 at lr-0.0007, showcasing its outstanding performance in incentive. At the given learning rate, this discovery implied that the algorithm maximized the combined savings in time and secrecy rate, leading to a higher reward. It was able to achieve this reward with a lower TFPS of 478.06 in contrast when comparing the TFPS rates suggesting a faster implementation than lr-0.0006. similarly, the algorithm was rewarded with a notable 57.92 at lr-0.0008. This learning rates TFPS rate of 480.14 was very closed to the performance achieved at LR-0.0007. Even though the TFPS rate was slightly higher the results show that the DRSTO scheme nevertheless maintained better performance.

Finally, over a range of learning rates the DRSTO scheme showed a broad variety of performance characteristics. At the lowest possible TFPS of 478.06 the lr-0.0007 model showed the highest reward of 58.56 indicating that it performed tasks more efficiently and made the best possible decisions. On the other hand the lowest reward was 56.83 and the highest TFPS rate was 487.97 with lr-0.0006. The goals of maximizing rewards and decreasing TFPS must be considered while selecting the ideal learning rate. The findings highlight the significance of fine-tuning the learning rate for achieving the best possible balance between learning efficiency and task processing speed. Therefore, after thorough evaluation, we identified lr-0.0007 as the most appropriate choice for all subsequent experiments for the DRSTO scheme.



(a) TFPS with variable LR



(b) TFPS of contender schemes

FIGURE 3. Investigating how different learning rate (LR) impact the TFPS for different schemes based DRL.

To further explore the DRL attributes of VECN task offloading, a comparative analysis is performed that included the DRSTO scheme and other existing approaches, as shown in Figs. 3-(b), 4-(a), and 4-(b).

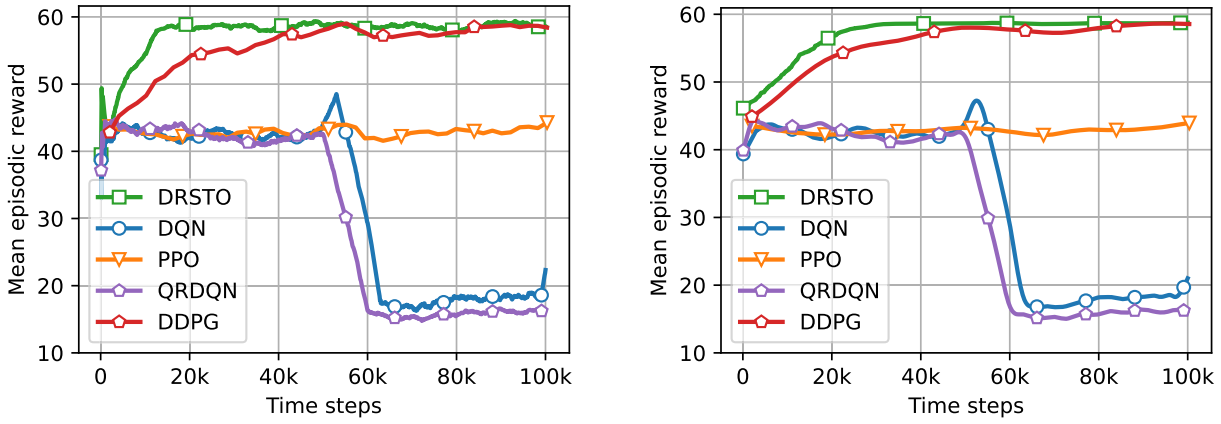
Figure 3-(b) shows a detailed analysis of the TFPS for each algorithm. At first, all schemes have somewhat larger over time decreasing TFPS values as learning progress runs. The DRSTO approach begins with a higher TFPS but adjusted quickly, eventually achieving the lowest TFPS rates, averaging at 478.06, which indicates efficient learning. In contrast, the DQN method starts at high TFPS levels and, despite shifting to lower rates over time, it consistently records higher averages than DRSTO, with an average TFPS of 759.89. On the other hand, the QRDQN scheme, a variant of DQN, performs better than DQN, achieving an average TFPS of 652.15. However, DDPG and PPO learning behaviors are very similar. DDPG has a horizontal trajectory at the beginning, but it starts learning at TFPS levels that gradually go down. Despite this, with an estimated mean TFPS of 706.26, it falls short of PPO, QRDQN, and DRSTO. The average TFPS of 647.97 shows that the PPO scheme outperforms the others, with the exception of DRSTO.

In addition to its usage in TFPS evaluation, we consider episodic reward as a crucial performance metric for making fair comparisons in task-offloading situations. The rate of secrecy, processing and communication latency, and other components make up the reward measure. Section IV-A explains how to calculate the episodic reward by integrating the time saved and the rate of secrecy. Fig. 4-(a) and 4-(b) display the median rewards per episode for each method. The comparison involves the DRSTO scheme, built on A2C, against alternatives employing DQN, QRDQN, DDPG, and PPO in particular. The DRSTO program achieved an average reward of 57.76, demonstrating remarkable performance. The average payout for the DDPG scheme was in close proximity at 55.61. However, DQN earned an average of 32.68 points and QRDQN earned 30.53 average points. The PPO-based scheme obtained an average reward of 42.82. These results indicate that both DQN and QRDQN faced challenges in their learning processes, experiencing a noticeable drop in reward ratios midway through training. In contrast, DRSTO and the DDPG schemes displayed consistent reward trends and demonstrated convergence, with DRSTO ultimately obtaining an average reward that was slightly greater. Notably, the PPO-based scheme struggled to progress in reward accumulation, maintaining a consistent reward level throughout the learning process, suggesting limited learning.

Summary: In the VECN environment, the results unmistakably demonstrate that the DRSTO scheme surpasses alternative DRL algorithms, excelling in terms of lower TFPS and higher rewards. The DRSTO scheme efficacy attributed to the utilization of A2C as its fundamental algorithm. It provides multiple benefits, such as its straightforwardness, the integration of policy gradients with value-based methods, and decreased variability. A2C is especially suitable for real-time learning situations due to its capabilities, which allow for faster convergence and the flexibility to adapt to changing environments. The best DRL algorithm is the one that reduces TFPS ratios while still achieving better rewards. Thanks to the DRSTO plan, this is accomplished. As a result of its constant reward patterns throughout the entire process, the PPO-based scheme showed the least gain in reward acquisition and learning, even though it achieved a lower TFPS. Although it had better TFPS ratios, the DDPG plan was far less effective at achieving rewards than the DRSTO program. QRDQN and DQN schemes exhibited poor outcomes in terms of TFPS ratios and reward mechanisms.

2) TASK OFFLOADING PERFORMANCE ASSESSMENT

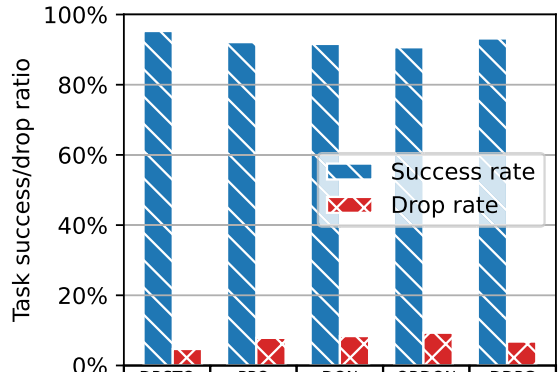
During this stage, we assess the efficiency of task processing and transmission, ensuring that the level of confidentiality conforms to the limitations defined by our DRSTO system and satisfies the specified criteria. Job completion rates and the TES factor are significant indicators of performance. Both the level of confidentiality and the reduction in hiring expenses are essential factors in evaluating the TES factor.



(a) Reward of contender schemes

(b) Smoothed reward of contender schemes

FIGURE 4. Examination of the reward received per episode across different DRL based strategies.



	DRSTO	PPO	DQN	QRDQN	DDPG
Success rate	95.30 %	92.13 %	91.62 %	90.67 %	93.18 %
Drop rate	04.70 %	07.87 %	08.38 %	09.33 %	06.82 %

FIGURE 5. Analysis of the task execution efficiency in terms of task success and drop ratio.

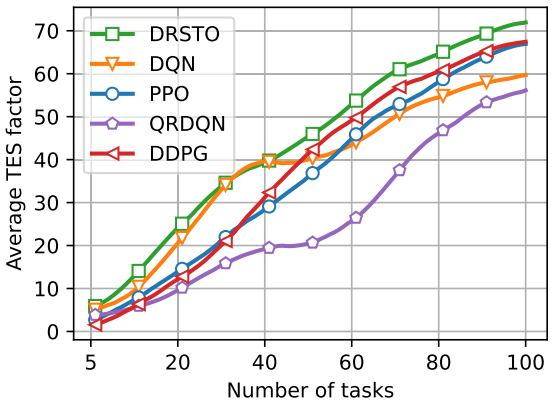


FIGURE 6. Aaverage TES factor of all contender schemes Vs. increasing the number of tasks.

Figure 5 depicts an exhaustive analysis of the pass/fail rates of several task outsourcing techniques. This research highlights the differences in performance among the schemes by providing a thorough understanding of how they compare with regard to the management of task success and decline rates. When choosing between running an operation on the vehicle's own Local Processing Unit (LPU) and entrusting it to the VEC server, the evaluation is crucial. To do a task successfully, one must take into account the work at hand, any available resources, and make judgments based on accurate information. Poor decision-making in this setting could lead to the project's demise and potential breaches of confidentiality.

The DRSTO scheme outperforms all other task offloading schemes in terms of both task success and failure rates. Demonstrating remarkable efficacy, the DRSTO scheme attains an impressive success task rate of 95.30% with a minimal failure task rate of only 4.70%. In contrast, DQN, QRDQN, DDPG, and PPO obtain task success rates of 91.62%, 90.67%, 93.18%, and 92.13%, respectively. Likewise, the performance degradation (drop rate of the task) for DQN, QRDQN, DDPG, and PPO are 8.38%, 9.33%, 6.82%, and 7.87%, respectively. While DDPG exhibits relatively commendable performance compared to other schemes, it falls behind in comparison to our proposed DRSTO scheme. The PPO-based scheme ranks third, followed by DQN, while the QRDQN-based scheme exhibits the least favorable performance among all the schemes.

In any task offloading scheme, it is critical to emphasize reducing delays in task processing and lowering the risk of secrecy breaches, since these elements significantly impact the overall effectiveness of task execution. This emphasis on minimizing processing times and addressing security concerns is integral for optimizing the effectiveness of task processing systems. Fig. 6 offers a visual representation of the average task offloading efficiency and secrecy statistics with an increasing number of tasks, providing insights into the scheme's performance considering these critical factors. In a comprehensive assessment of the TES factor, our study engaged in the evaluation of more than 100 tasks, with each task encompassing a variable range of 1 to 100 randomly generated sub-tasks. Adherence to the specifications detailed in Section V was maintained for all parameters, guaranteeing

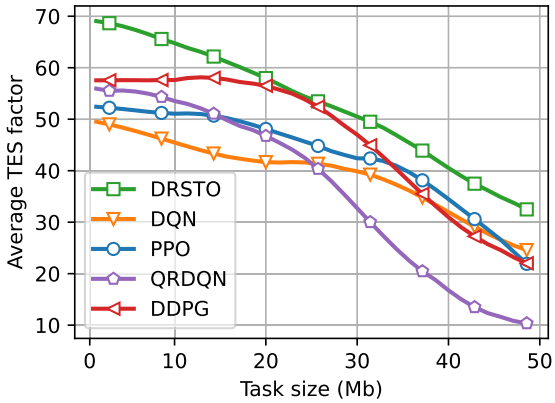


FIGURE 7. Average TES factor behavior against increasing task size.

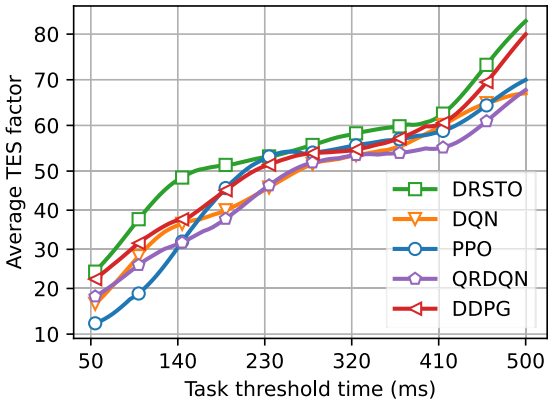


FIGURE 8. Average TES factor behavior against increasing execution threshold time.

a fair and impartial comparison across different task offloading schemes. Analyzing Fig. 6, reveals the competence of all task offloading schemes in achieving task offloading while maintaining secrecy, albeit with varying performance levels. Notably, for increasing number of tasks in one second, all schemes tend towards convergence. Significantly, the proposed DRSTO scheme consistently outperforms the other schemes, achieving an average TES factor of 44.16. It is evident that with fewer tasks, the TES factor tends to be lower, but it increases as the number of tasks grows. The DDPG and PPO schemes exhibit a similar performance pattern to that of the DRSTO scheme concerning the TES factor as the number of tasks increases. However, DDPG attains an average TES factor of 37.55, while PPO achieves an average TES factor of 36.07. In contrast, the DQN and QRDQN schemes achieve average TES factors of 38.11 and 26.06, respectively. Nevertheless, in this scenario, QRDQN remains the least effective among the schemes.

The average behavior of TES with respect to different task sizes is examined in Fig. 7. The task management patterns of all rival offloading techniques are very similar, which is interesting because the amount of time saved usually decreases as the task sizes get higher. At first, when handling smaller task sizes, all systems get greater average TES factors. Nevertheless, as the number of the tasks grows, their efficiency factors gradually decrease. The DRSTO scheme demonstrates a consistently smooth pattern throughout this scenario, reaching the highest average TES factor of 52.67. In this context, the PPO scheme closely follows a TES factor pattern similar to that of the DRSTO scheme, getting an average TES factor of 43.82. Conversely, the DQN and QRDQN schemes attain average TES factors of 41.09 and 36.11, respectively. Meanwhile, the DDPG scheme remains competitive with the DRSTO scheme but falls slightly behind, achieving an average TES factor of 50.43. Besides, the QRDQN scheme once again exhibits the least favorable performance among all the schemes.

The assessment of task offloading schemes is significantly influenced by the task threshold time. In our research, we scrutinized diverse offloading strategies, where the effectiveness of the schemes is assessed over various task threshold, as illustrated in Fig. 8. Each scheme exhibited competence in handling tasks, even when confronted with the most minimal threshold time.However, their performance, in terms of the TES factor, showed variation. The DRSTO scheme significantly improved when the task threshold time ranged from 50ms to 140ms. However, all schemes, including DRSTO, did not exhibit substantial progress when the threshold time was between 230ms and 410ms. Beyond the 410ms threshold time, all schemes experienced a notable increase in the TES factor. The DRSTO scheme achieved the highest average TES factor of 54.83 on average. DQN and QRDQN schemes achieved nearly identical average TES factors of 48.96 and 48.33, respectively. The PPO and DDPG schemes attained average TES factors of 50.78 and 51.39, respectively. Once again, the QRDQN scheme exhibited the poorest performance in this scenario. However, the DRSTO scheme performed well across lower, average, and higher task threshold times. It consistently achieved the maximum average TES factor compared to all other schemes.

To evaluate the effectiveness of our proposed DRSTO scheme in terms of secrecy during offloading, we conducted an analysis on the impact of eavesdroppers on the average achievable secrecy rate. The results of this evaluation, including scenarios with different numbers of eavesdroppers and a scenario with no eavesdroppers, are presented in Fig. 9. The analysis reveals that the average achievable secrecy rate is 1.99 bits/sec/Hz when no eavesdroppers are present. As we introduce a single eve, the average achievable secrecy rate slightly decreases to 1.96 bits/sec/Hz. Similarly, including 2, 3, 4, and 5 eavesdroppers, the average achievable secrecy rates are 1.93, 1.89, 1.86, and 1.83 bits/sec/Hz, respectively. Analyzing these results, it becomes evident that our DRSTO scheme performs consistently well across all cases, demonstrating a relatively stable performance even in the presence of eavesdroppers. The variation in secrecy rate is minimal as the number of eavesdroppers increases. The discrepancy in achievable secrecy rates across the scenarios with no eavesdroppers and 1-5 eavesdroppers in our system is only a degradation of 1.42%, 2.93%, 4.72%, 6.22%, and 7.92%, respectively. Considering the importance of maintaining the secrecy of on-road vehicle data during task

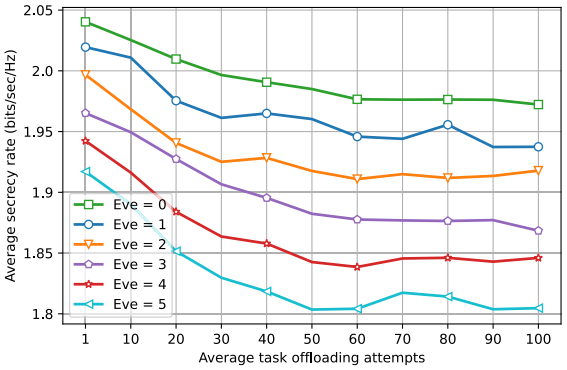


FIGURE 9. Average secrecy rate against increasing number of eavesdroppers in the system.

offloading, these variations in secrecy rate are manageable and do not significantly compromise the overall effectiveness of our scheme.

Summary: The evaluation of task offloading performance, as seen in Figs. 5 to 9, highlights the unique characteristics of the DRSTO scheme that enhance its effectiveness in minimizing delays while ensuring the secrecy rate is maintained. This scheme considers the dynamic nature of vehicle headway, allowing it to adapt to variations in inter-vehicle distances and movement patterns. Furthermore, the architectural preference for the advanced 5G-NR-V2X RAT gives the DRSTO scheme a competitive advantage in optimizing delays.

A topic of interest to investigate is the integration of RIS as the primary tool in all schemes. This introduces a dynamic selection procedure that relies on the vehicle's proximity to either a RIS or an RSU to choose the communication route. Direct connection is established between the vehicle and the RSU while they are in close proximity but the path towards the RIS is decided when the vehicle is closer to it. This technique plays a vital role in reducing delays and ensuring the necessary level of secrecy. Using the RIS to communicate with the edge server is significantly more efficient in terms of decreasing delays.

VI. CONCLUSION

This paper presents an in-depth analysis of enhancing VECNs using the DRL-based DRSTO scheme. DRSTO tailored for 5G-NR-V2X heterogeneous RATs and supported by zero-energy RIS (ze-RIS) offers vehicles the choice between local processing and offloading tasks to VEC servers. The system facilitates two offloading pathways: a route between vehcile and VEC using RSU which the direct link or between vehicle to VEC through RSU assisted by RIS which the indirect link, influenced by the vehicles' duration of stay. The TES factor stands out as a significant innovation for DRL agent acting as a reward mechanism. It merges time savings with secrecy rates to optimize offloading actions. The DRL is utilized to reduce delay and simultaneously enhance secrecy rates during offloading. Extensive evaluations show that DRSTO significantly improves task success rates, from 2.12% to 4.63%, and elevates the TES factor from 6.05 to 18.10. Compared to other schemes, DRSTO stands out in reward generation and TFPS ratios, demonstrating its effectiveness in VECNs.

In light of the promising results of the ze-RIS-assisted secure task efficient offloading (DRSTO) scheme, we propose several future research directions. Exploring alternative DRL algorithms such as Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) could uncover new optimization techniques and performance enhancements. Adopting more realistic vehicular mobility models, such as the Intelligent Driver Model (IDM) or Simulation of Urban Mobility (SUMO), can better reflect actual vehicular movement, improving the robustness and validity of simulations. Introducing multi-agent DRL frameworks for collaborative optimization among vehicles could enhance scalability and resource management in densely populated areas. Additionally, incorporating advanced technologies like non-orthogonal multiple access (NOMA) and full-duplex communication could improve spectral efficiency and reduce latency. These directions will further advance the DRSTO scheme and enhance the effectiveness of vehicular edge computing networks.

REFERENCES

[1] F. Tang, B. Mao, N. Kato, and G. Gui, "Comprehensive survey on machine learning in vehicular network: Technology, applications and challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 2027–2057, 3rd Quart., 2021.

[2] M. Ahmed, S. Raza, M. A. Mirza, A. Aziz, M. A. Khan, W. U. Khan, J. Li, and Z. Han, "A survey on vehicular task offloading: Classification, issues, and challenges," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4135–4162, Jul. 2022.

[3] G. Naik, B. Choudhury, and J.-M. Park, "IEEE 802.11bd & 5G NR V2X: Evolution of radio access technologies for V2X communications," *IEEE Access*, vol. 7, pp. 70169–70184, 2019.

[4] M. W. Shabir, T. N. Nguyen, J. Mirza, B. Ali, and M. A. Javed, "Transmit and reflect beamforming for max-min SINR in IRS-aided MIMO vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1099–1105, Jan. 2023.

[5] L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. D. Renzo, C.-B. Chae, and L. Hanzo, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45913–45923, 2020.

[6] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3187–3196, May 2020.

[7] X. Lei, M. Wu, F. Zhou, X. Tang, R. Q. Hu, and P. Fan, "Reconfigurable intelligent surface-based symbiotic radio for 6G: Design, challenges, and opportunities," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 210–216, Oct. 2021.

[8] S. Arzykulov, A. Celik, G. Nauryzbayev, and A. M. Eltawil, "Artificial noise and RIS-aided physical layer security: Optimal RIS partitioning and power control," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 992–996, Mar. 2023.

[9] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on STAR-RIS: Use cases, recent advances, and future research challenges," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 1–19, Jul. 2023.

[10] T. Zhang, P. Ren, D. Xu, and Z. Ren, "RIS subarray optimization with reinforcement learning for green symbiotic communications in Internet of Things (IoT)," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19454–19465, Apr. 2023.

[11] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[12] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[13] M. Ahmed, S. Raza, A. A. Soofi, F. Khan, W. U. Khan, S. Z. U. Abideen, F. Xu, and Z. Han, "Active reconfigurable intelligent surfaces: Expanding the frontiers of wireless communication—A survey," *IEEE Commun. Surveys Tuts.*, vol. 1, no. 1, pp. 1–11, 1st Quart., 2024.

[14] N. Rahmatov and H. Baek, "RIS-carried UAV communication: Current research, challenges, and future trends," *ICT Exp.*, vol. 9, no. 5, pp. 961–973, Oct. 2023.

[15] S. Naser, L. Bariah, S. Muhaidat, and E. Basar, "Zero-energy devices empowered 6G networks: Opportunities, key technologies, and challenges," *IEEE Internet Things Mag.*, vol. 6, no. 3, pp. 44–50, Sep. 2023.

[16] H. Ren, Z. Chen, G. Hu, Z. Peng, C. Pan, and J. Wang, "Transmission design for active RIS-aided simultaneous wireless information and power transfer," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 600–604, Apr. 2023.

[17] J. Feng, Z. Liu, C. Wu, and Y. Ji, "AVE: Autonomous vehicular edge computing framework with ACO-based scheduling," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10660–10675, Dec. 2017.

[18] Z. Gao, M. Liwang, S. Hosseinalipour, H. Dai, and X. Wang, "A truthful auction for graph job allocation in vehicular cloud-assisted networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 10, pp. 3455–3469, Oct. 2022.

[19] F. Sun, F. Hou, N. Cheng, M. Wang, H. Zhou, L. Gui, and X. Shen, "Cooperative task scheduling for computation offloading in vehicular cloud," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11049–11061, Nov. 2018.

[20] J. Zhao, M. Kong, Q. Li, and X. Sun, "Contract-based computing resource management via deep reinforcement learning in vehicular fog computing," *IEEE Access*, vol. 8, pp. 3319–3329, 2020.

[21] J. Du, C. Jiang, A. Benslimane, S. Guo, and Y. Ren, "SDN-based resource allocation in edge and cloud computing systems: An evolutionary Stackelberg differential game approach," *IEEE/ACM Trans. Netw.*, vol. 30, no. 4, pp. 1613–1628, Aug. 2022.

[22] X. Chen, L. Zhang, Y. Pang, B. Lin, and Y. Fang, "Timeliness-aware incentive mechanism for vehicular crowdsourcing in smart cities," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3373–3387, Sep. 2022.

[23] X. He, R. Jin, and H. Dai, "Physical-layer assisted secure offloading in mobile-edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4054–4066, Jun. 2020.

[24] B. Ai, A. F. Molisch, M. Rupp, and Z.-D. Zhong, "5G key technologies for smart railways," *Proc. IEEE*, vol. 108, no. 6, pp. 856–893, Jun. 2020.

[25] Y. Wu, G. Ji, T. Wang, L. Qian, B. Lin, and X. Shen, "Non-orthogonal multiple access assisted secure computation offloading via cooperative jamming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7751–7768, Jul. 2022.

[26] Q. Li and L. Yang, "Beamforming for cooperative secure transmission in cognitive two-way relay networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 130–143, 2020.

[27] J. Xu and J. Yao, "Exploiting physical-layer security for multiuser multicarrier computation offloading," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 9–12, Feb. 2019.

[28] W. Wang, W. Ni, H. Tian, and L. Song, "Intelligent omni-surface enhanced aerial secure offloading," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5007–5022, May 2022.

[29] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[30] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[31] J. Xu and B. Ai, "When mmWave high-speed railway networks meet reconfigurable intelligent surface: A deep reinforcement learning method," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 533–537, Mar. 2022.

[32] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[33] K. Feng, X. Li, Y. Han, S. Jin, and Y. Chen, "Physical layer security enhancement exploiting intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 734–738, Mar. 2021.

[34] S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. M. Leung, "Reconfigurable intelligent surface-assisted secure mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6647–6660, Jun. 2022.

[35] D. Wang, X. Li, L. Pang, Y. He, F. Zhou, L. Wang, and R. Zhang, "IRS-aided secure mobile edge computing for NOMA networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Foshan, China, Aug. 2022, pp. 25–30.

[36] Y. Chen, Z. Liu, Y. Zhang, Y. Wu, X. Chen, and L. Zhao, "Deep reinforcement learning-based dynamic resource management for mobile edge computing in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4925–4934, Jul. 2021.

[37] J. Xu, B. Ai, L. Chen, and L. Wu, "Deep reinforcement learning for communication and computing resource allocation in RIS aided MEC networks," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 3184–3189.

[38] X. Zhang, Y. Shen, B. Yang, W. Zang, and S. Wang, "DRL based data offloading for intelligent reflecting surface aided mobile edge computing," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–7.

[39] Z. Wang, Y. Wei, Z. Feng, F. R. Yu, and Z. Han, "Resource management and reflection optimization for intelligent reflecting surface assisted multi-access edge computing using deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1175–1186, Feb. 2023.

[40] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 122–134, Dec. 2020.

[41] Y. Zheng, S. A. Tegos, Y. Xiao, P. D. Diamantoulakis, Z. Ma, and G. K. Karagiannidis, "Zero-energy device networks with wireless-powered RISs," *IEEE Trans. Veh. Technol.*, vol. 1, no. 1, pp. 1–5, Apr. 2023.

[42] M. A. Mirza, J. Yu, S. Raza, M. Krichen, M. Ahmed, W. U. Khan, K. Rabie, and T. Shongwe, "DRL-assisted delay optimized task offloading in automotive-Industry 5.0 based VECNs," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 6, Jun. 2023, Art. no. 101512.

[43] M. A. Mirza, Y. Junsheng, S. Raza, M. Ahmed, M. Asif, A. Irshad, and N. Kumar, "MCLA task offloading framework for 5G-NR-V2X-based heterogeneous VECNs," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14329–14346, Dec. 2023.

[44] G. Yang, Y.-C. Liang, R. Zhang, and Y. Pei, "Modulation in the air: Backscatter communication over ambient OFDM carrier," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1219–1233, Mar. 2018.

[45] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[46] N. A. Mitsiou, V. K. Papanikolaou, P. D. Diamantoulakis, and G. K. Karagiannidis, "Energy-aware optimization of zero-energy device networks," *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 858–862, Apr. 2022.

[47] D. Tyrovolas, S. A. Tegos, V. K. Papanikolaou, Y. Xiao, P.-V. Mekikis, P. D. Diamantoulakis, S. Ioannidis, C. K. Liaskos, and G. K. Karagiannidis, "Zero-energy reconfigurable intelligent surfaces (zeRIS)," 2023, *arXiv:2305.07686*.

[48] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[49] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[50] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8627–8640, Dec. 2021.

[51] S. Zhang and R. Zhang, "Intelligent reflecting surface aided multi-user communication: Capacity region and deployment strategy," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5790–5806, Sep. 2021.

[52] S. Abeta, *Further Advancements for E-UTRA Physical Layer Aspects*, Standard TR 36.814, 3GPP, 2017.

[53] Q. Ye, W. Shi, K. Qu, H. He, W. Zhuang, and X. Shen, "Joint RAN slicing and computation offloading for autonomous vehicular networks: A learning-assisted hierarchical approach," *IEEE Open J. Veh. Technol.*, vol. 2, pp. 272–288, 2021.

[54] B. Lyu, C. Zhou, S. Gong, D. T. Hoang, and Y.-C. Liang, "Robust secure transmission for active RIS enabled symbiotic radio multicast communications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 1–18, May 2023.
[55] R. Kaur and B. Bansal, "Secure beamforming for intelligent reflecting surface assisted MISO wireless communications," in *Proc. 3rd Int. Conf. Intell. Commun. Comput. Techn. (ICCT)*, Jan. 2023, pp. 1–5.
[56] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted MIMO symbiotic radio networks," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, Jul. 2021.
[57] D. Wang, B. Song, P. Lin, F. R. Yu, X. Du, and M. Guizani, "Resource management for edge intelligence (EI)-assisted IoV using quantum-inspired reinforcement learning," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 12588–12600, Jul. 2022.
[58] J. Xu, A. Xu, L. Chen, Y. Chen, X. Liang, and B. Ai, "Deep reinforcement learning for RIS-aided secure mobile edge computing in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 2455–2464, Feb. 2024.

ABDUL WAHID received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2017, and the M.S. degree in computer science and technology from the Southwest University of Science and Technology (SWUST), Mianyang, Sichuan, China, in 2020. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China. His current research interests include wireless communications, physical layer security, reconfigurable intelligent surface, machine learning, and 6G technology.

MUHAMMAD AYZED MIRZA received the B.S.C.S. degree from Government College University, Faisalabad, in 2008, the M.S.C.S. degree in computer science from the National Textile University, Faisalabad, in 2016, and the Ph.D. degree from the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2023. He was a Lecturer with Punjab College University, Faisalabad Campus, for four years; and with the National Textile University, from 2016 to 2018. Currently, he is an Associate Professor with the School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong, China. His research interests include B5G/6G based vehicular edge computing networks (VECNs), data/task offloading in VECNs, RIS/ze-RIS/BD-RIS/NOMA assisted vehicular communications, and RL/DRL-based resource optimization in VECNs.

MANZOOR AHMED received the B.E. and M.S. degrees in electrical engineering and computer science from Pakistan universities, in 1996 and 2010, respectively, and the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications, China, in 2015. From 1997 to 2000, he was a Lecturer with Balochistan Engineering University; and a Telecomm Engineer with government-owned telecommunication service provider NTC, Pakistan, from 2000 to 2011. He was a Postdoctoral Researcher with the Electrical Engineering Department, Tsinghua University, China, from 2015 to 2018. He was an Associate Professor with the Department of Computer Science and Technology, Qingdao University. He is currently a Professor with the School of Computer and Information Science and also with the Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan. He has several research publications in IEEE top journals and conferences. His research interests include resource allocation and offloading in vehicular communications and networking, fog and edge computing, socially aware D2D communication, physical layer security, RIS, backscattering, and UAV communication. He received several awards, including the Distinction Award from the President of Pakistan, the Best Employee Award from NTC, and the Best Paper Award from the 2014 GameNets Conference.

MUHAMMAD SHERAZ is a Research Fellow with the Faculty of Engineering, Multimedia University, Malaysia. His research interests include information-theoretic aspects of wireless networks, signal processing, mmWave and THz communications, AI, digital twin networks, resource allocation, non-orthogonal multiple access (NOMA), reconfigurable intelligent surfaces (RIS), ultra-reliable low latency communications (URLLC), edge computing, the IoT, intelligent transportation systems (ITS), and Industry 4.0. He has served as a TPC member for several conferences.

TEONG CHEE CHUAH received the B.Eng. and Ph.D. degrees from Newcastle University, U.K., in 1999 and 2002, respectively. He is currently a Professor with the Faculty of Engineering, Multimedia University, Malaysia. His research interests include resource management for fixed-mobile convergent networks, applications of artificial intelligence in communications, and cellular network planning and optimization.

IT EE LEE received the Bachelor of Engineering degree (Hons.) in electronics and the M.Eng.Sc. degree from Multimedia University, Cyberjaya, Malaysia, in 2003 and 2009, respectively, and the Ph.D. degree specializing in optical wireless communication from the University of Northumbria, Newcastle Upon Tyne, U.K., in 2014. She is currently the Deputy Director of the Research Management Centre (RMC) and an Assistant Professor with the Faculty of Engineering, Multimedia University. Her research interests include wireless optical communications and the Internet of Things.

WALI ULLAH KHAN (Member, IEEE) received the master's degree in electrical engineering from COMSATS University Islamabad, Pakistan, in 2017, and the Ph.D. degree in information and communication engineering from Shandong University, Qingdao, China, in 2020. He is currently with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg. He has authored/co-authored more than 100 publications, including international journals, peer-reviewed conferences, and book chapters. His research interests include convex/nonconvex optimizations, non-orthogonal multiple access, reflecting intelligent surfaces, ambient backscatter communications, the Internet of Things, intelligent transportation systems, satellite communications, unmanned aerial vehicles, physical layer security, and applications of machine learning.