CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Journal Paper

DRL-KeyAgree: An Intelligent Combinatorial Deep Reinforcement Learning-Based Vehicular Platooning Secret Key Generation

Harrison Kurunathan*

Kai Li*

Eduardo Tovar*

Alípio Mario Jorge

Wei Ni

Abbas Jamalipour

*CISTER Research Centre

CISTER-TR-240508

2024

DRL-KeyAgree: An Intelligent Combinatorial Deep Reinforcement Learning-Based Vehicular Platooning Secret Key Generation

Harrison Kurunathan*, Kai Li*, Eduardo Tovar*, Alípio Mario Jorge, Wei Ni, Abbas Jamalipour

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: jhk@isep.ipp.pt, kai@isep.ipp.pt, emt@isep.ipp.pt, amjorge@fc.up.pt, Wei.Ni@data61.csiro.au

https://www.cister-labs.pt

Abstract

The exploitation of radio channels' inherent randomness for generating secret keys within a vehicular platoon offers a promising approach to securing communications in dynamic and unpredictable environments. The channel-based key generation leverages the fact that the physical characteristics of the radio channel, such as fading, shadowing, and multipath propagation, vary in a complex manner that makes it difficult for external adversaries to predict or replicate. A challenge lies in accurately assessing the channel's randomness to ensure the generated keys are both secure and consistent across the platooning vehicles, especially in vehicular environments with high mobility and the ever-changing urban landscape. This paper proposes a novel channel-based key generation (DRL-KeyAgree) technique to enhance communication security within vehicular platoons through combinatorial deep reinforcement learning (DRL). DRL-KeyAgree addresses key disagreement among platooning vehicles by training advantage Actor-Critic (A2C), which integrates policy- and value-based strategies to dynamically select optimal quantization intervals adapting to the random wireless channels. Further incorporation of Long Short-Term Memory (LSTM) allows DRL-KeyAgree to capture the characteristics of partially observable radio channels, significantly enhancing the key agreement rate among vehicles. DRL-KeyAgree is rigorously evaluated using the standard National Institute of Standards and Technology (NIST) test suite.

DRL-KeyAgree: An Intelligent Combinatorial Deep Reinforcement Learning-Based Vehicular Platooning Secret Key Generation

Harrison Kurunathan, Kai Li, *Senior Member, IEEE*, Eduardo Tovar, Alípio Jorge, Wei Ni, *Fellow, IEEE*, and Abbas Jamalipour, *Fellow, IEEE*

***Abstract*—The exploitation of radio channels' inherent randomness for generating secret keys within a vehicular platoon offers a promising approach to securing communications in dynamic and unpredictable environments. The channel-based key generation leverages the fact that the physical characteristics of the radio channel, such as fading, shadowing, and multipath propagation, vary in a complex manner that makes it difficult for external adversaries to predict or replicate. A challenge lies in accurately assessing the channel's randomness to ensure the generated keys are both secure and consistent across the platooning vehicles, especially in vehicular environments with high mobility and the ever-changing urban landscape. This paper proposes a novel channel-based key generation (DRL-KeyAgree) technique to enhance communication security within vehicular platoons through combinatorial deep reinforcement learning (DRL). DRL-KeyAgree addresses key disagreement among platooning vehicles by training advantage Actor-Critic (A2C), which integrates policy- and value-based strategies to dynamically select optimal quantization intervals adapting to the random wireless channels. Further incorporation of Long Short-Term Memory (LSTM) allows DRL-KeyAgree to capture the characteristics of partially observable radio channels, significantly enhancing the key agreement rate among vehicles. DRL-KeyAgree is rigorously evaluated using the standard National Institute of Standards and Technology (NIST) test suite.**

***Index Terms*—Vehicular platoons, Secret key agreement, Deep Reinforcement Learning (DRL), Actor-Critic (A2C), Long Short-Term Memory (LSTM).**

I. Introduction

The sixth-generation (6G) communications have the potential to revolutionize data dissemination in vehicular networks, which facilitates fast reaction times, improved vehicle coordination and synchronized movements [1], [2]. 6G is anticipated to substantially increase bandwidth compared to previous generations. Projections indicate a large data rate of 1,000 gigabits/s with a stringent latency of less than 100 microseconds, making it 100 times faster than existing 5G communication models [3]. This advancement is expected to yield highly dependable and low-latency communication, which enables real-time and seamless information exchange between the vehicles, such as positions, speeds, accelerations and intentions [4]. Furthermore, 6G allows the transmission of substantial data volumes between the vehicles, including high-definition video, sensor data, and real-time telemetry. The high-bandwidth communication provided by 6G enables the implementation of applications such as cooperative perception [5], where vehicles share sensor data to enhance situational awareness and collision avoidance.

The capabilities of 6G technology open up possibilities for advanced applications in vehicular platooning that consist of a leader vehicle and several autonomous following vehicles. The leader vehicle can either be an autonomous vehicle or a human-controlled one. The following vehicle in the platoon keeps a small distance from the one ahead of it. Fig. 1 depicts a typical vehicular platoon, where the leader vehicle determines the driving status of the whole platoon, such as speed, heading, and responses to road emergencies [6], [7]. The vehicles in the platoon can take advantage of the reduced air drag, leading to a low fuel consumption [8], which reduces CO2 emissions and promotes environmental sustainability [9].

The leader vehicle periodically broadcasts the vehicle status information to update the following vehicles in the platoon. Each of the following vehicles acts as a data-forwarding node that disseminates the leader´s message to its adjacent next vehicle in the platoon. The following vehicles disseminate data at different times to avoid communication interference. Due to the broadcast nature of wireless channels, data dissemination in the platoon is vulnerable to eavesdropping and replay attacks [10], [11]. An eavesdropper can track the driven route of the platoon to abuse the mobility patterns of the platoon. Data dissemination that includes location coordinates, beacons, and control packets can be used both to improve the platoon service or to target individual vehicles by criminals [12], [13]. Hence, the platooning vehicles must generate secret keys to encrypt the data transmission.

In a channel-based secret key generation, platooning vehicles can leverage the unique properties of the wireless communication channel between them to generate a shared secret key. This process involves measuring the Channel State

H. Kurunathan, and E. Tovar are with the Real-Time and Embedded Computing Systems Research Centre (CISTER), Porto 4249-015, Portugal (E-mail: {jhk, emt}@isep.ipp.pt).

K. Li is with the CISTER Research Centre, Porto 4249-015, Portugal, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (E-mail: kaili@ieee.org).

A.M Jorge is with the Laboratory of Artificial Intelligence and Decision Support (LIAAD), INESC TEC, and also with the University of Porto, Porto 4099-002, Portugal (E-mail: amjorge@fc.up.pt).

W. Ni is with the Digital Productivity and Services Flagship, Commonwealth Scientific and Industrial Research Organization (CSIRO), NSW 2122, Australia, and the School of Computing Science and Engineering, the University of New South Wales (UNSW), Kennington, NSW 2052, Australia (E-mail: wei.ni@data61.csiro.au).

A. Jamalipour is with the School of Electrical and Computer Engineering, University of Sydney, NSW 2006, Australia (E-mail: a.jamalipour@ieee.org).

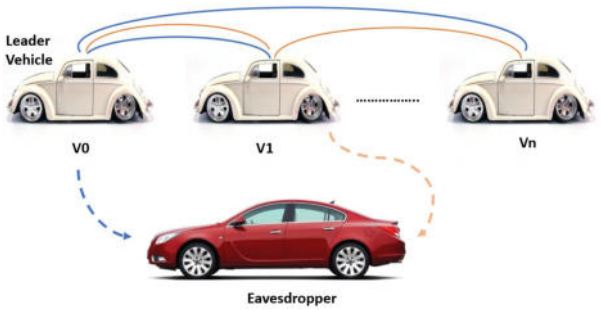

Fig. 1. In the vehicular platoon, the leader vehicle disseminates data to the follower vehicles by using multi-hop data dissemination. Due to the broadcast nature of wireless channels, data transmission is vulnerable to eavesdropping attacks.

Information (CSI), which encompasses various attributes of the wireless channel, such as its phase, amplitude, and multipath effects, reflecting the environment's complexity around the vehicles [14]. The CSI is inherently random due to the dynamic nature of the environment, making it an excellent source for generating keys that are difficult for an eavesdropper to predict. Both vehicles measure the CSI independently and apply quantization to these measurements to discretize the CSI into a sequence of bits. By optimizing the quantization intervals, the vehicles can overcome the differences in their CSI measurements due to noise or other factors, ensuring the generation of the same secret key at the vehicles in a platoon [15]. This key can then be used to encrypt/decrypt vehicular data.

However, this selection of the quantization levels necessitates additional computational resources, thereby augmenting the complexity of the key generation process [16], [17]. In our preliminary work [18], [19], a unanimous key is generated by the cooperative quantization of the channel gains between the vehicles, which is used for encrypting/decrypting the vehicle-to-vehicle communication. The vehicles reconcile upon a shared secret key so that the following vehicle can decode the disseminated data from the preceding.

Optimizing the length of quantization levels in the channel-based key generation is a non-convex problem [22], making it difficult to address by optimization techniques such as regression training [23]. With an increase in the size of the problem, the complexity of resolving the non-convex optimization intensifies at an exponential rate [24], [25].

In [26], a model is developed to balance the number of user vehicles in platoons while ensuring fairness using the reputation score of platoon heads. This reputation score is defined based on parameters such as network bandwidth, latency, and the number of user vehicles within the newly formed platoons. Utilizing reputation scores for head selection in a platoon helps optimize fairness and security. In [27], fairness is improved by determining the platoon formation scheme of vehicles during the control zone. This is done by allocating the passing sequence and time slots of platoons through the merging zone, intending to reduce the makespan and maximum delay to enhance operational efficiency and fairness. By reducing the maximum delay, the variance of travel times is reduced, thereby improving user fairness.

The authors of [28] develop a decentralized speed advisory system (TD-SAS) to improve energy efficiency in autonomous vehicle platooning. This system can identify malicious nodes while enhancing the system's security. By employing a consortium blockchain for managing trust nodes, providing non-repudiation and tamper resistance for reputation data, TD-SAS improves the energy efficiency of the vehicles in the autonomous platoons. The authors of [29] study spectral efficiency and energy efficiency based on deep reinforcement learning, where deep Q networks (DQN) are employed to allocate transmission power in V2X links, thereby enhancing overall energy efficiency. The Actor-Critic algorithm and Proximal Policy Optimization (PPO) [30] have been employed to tackle the energy requirements in eco-driving. These algorithms design cruising speed and power usage profiles based on look-ahead information from connectivity and advanced mapping features. By adjusting the vehicle velocity over the route and implementing energy-efficient approaches at signalized intersections using deep reinforcement learning, a reduction in energy consumption can be observed.

In this paper, we propose a combinatorial deep reinforcement learning (DRL) strategy (DRL-KeyAgree) to generate channel-based confidential key agreements within vehicular platoons, which minimizes the key disagreement rate (KDR) amongst the platooning vehicles. Specifically, DRL-KeyAgree employs A2C, comprising two principal elements: an actor-network responsible for determining the quantization interval and a critic network designated to assess the action selected by the actor by learning the state of the channel quality. Periodic feedback is introduced to calibrate the process by a cost function that stipulates the homogeneity of the keys. DRL-KeyAgree trains the platooning vehicles to learn to make sequential decisions in changing environments. DRL-KeyAgree enables the vehicles to interact with the environment and receive feedback in the form of a reward that minimizes the KDR, which maximizes a long-term cumulative reward.

Employing A2C for training the actions of the vehicles might encounter difficulties due to obscured states stemming from partial environmental observation. This can precipitate suboptimal actions; in other words, the quantization intervals that vehicles implement for their key generation may differ, thereby exacerbating the KDR. Vehicular platoons are dynamic, with rapid changes in channel conditions. Thanks to their gated architecture, LSTMs are adept at adapting to such changes, which allows them to learn when to forget previous information and update their internal state based on new inputs. This adaptability is crucial for maintaining the security of the time-varying keys. LSTMs are designed to recognize and exploit long-term dependencies in sequential data, which is crucial when the input has time-varying characteristics of wireless communication channels [31]. The radio channel's state can exhibit dependencies over time due to factors like vehicle movement, environmental changes, and channel fading. LSTMs can capture these temporal patterns effectively, making them more suited for modelling the dynamics of radio channels than CNNs, which excel in spatial data processing. Key agreement in vehicular platoons undergoes partially observable

training environments, where the full state of the system (i.e., the complete characterization of the radio channel) cannot be directly observed. LSTMs can manage such situations better than CNNs due to the ability to maintain an internal state that can infer unobserved aspects of the environment from past observations, thus providing a more comprehensive understanding of the channel characteristics over time. Key agreement in vehicular platoons requires predicting or estimating the future state of the channel to ensure that the key generation process remains synchronized. LSTMs, with their recurrent architecture, are naturally suited for sequence prediction tasks, enabling them to forecast future channel states based on historical data, which is less intuitive for CNNs that are primarily structured for recognizing patterns in spatial data.

The contributions of this paper are as follows:

- A new algorithm, designated as DRL-KeyAgree, is developed predicated on the A2C model to maximize the key agreement within vehicular platoons. By instructing the actions related to the configuration of quantization intervals, the dynamic nature of the channel is quantized to generate the secret key. DRL-KeyAgree empowers vehicles within the platoon to independently apprehend network states, while generating unambiguous secret key bits without sharing their channel information.
- The integration of LSTM into the A2C model enables the learning of acquired hidden states in DRL-KeyAgree and enhances the training environment within A2C. In particular, the hidden states associated with channel dynamics are harnessed to enrich the A2C training environment, facilitating a faster convergence than traditional A2C and Deep Q-Network (DQN) techniques.
- The randomness of the keys generated through DRL-KeyAgree are tested using the standard NIST test suite, showing a confidence of 99% in the randomness. An in-depth performance pertaining to quantization intervals, key lengths, and LSTM sizes that affect the KDR and a comparison of our DRL-KeyAgree against standard DQN and A2C deep learning models have been performed. The results show that our proposed technique reduces the KDR by 60% compared to its counterparts.

The rest of the paper is organized as follows. Section II presents the related works on secret key generation and machine learning (ML)-assisted communication security. Section III presents the security model for data dissemination in platoons. Section IV elaborates on the DRL-KeyAgree, where we delineate the observable states and the cost of learning. Section V presents our performance analysis, followed by concluding remarks and future works in Section VI.

II. Related work

This section provides a background to secret key generation that relies on channel randomness in wireless networks. We also outline contemporary research focusing on communication security based on machine learning.

A. Secret Key Generation Based on Channel Randomness

The exploitation of randomness based on wireless fading channels is advantageous in enhancing security within wireless sensor networks, particularly in the Internet of Things (IoT). The generation of secret keys premised on wireless channel coefficients, such as power measurements, has been suggested as a viable strategy for implementing security on a vast scale [32]. The authors of this study have employed a filter bank-based secret key generation methodology, using Karhunen-Lo‘eve transforms to minimize dependencies and correlations among frequencies.

Reliability pertaining to channel-based key generation hinges on the relationship between the noise impacting the wireless channels and the considered channel samples. In [33], the authors establish the key-gen sensitivity based on a secret key rate. This rate defines the maximum extremity of secret bits that can be derived from observing the channel. The secrecy key rate is computed with a line-of-sight wireless channel by factoring the correlation levels between legitimate users and malicious eavesdroppers.

High channel coherence time is a common phenomenon in IoT devices deployed in a static environment, which can be beneficial in accelerating key generation rates and enhancing randomness. A mapping table-based key distribution approach is proposed in [34] for IoT environments with random channel characteristics, such as scattering materials and distributions, which are integrated to augment the key generation and key agreement rates.

The authors of [35] present a GPS-based key establishment that utilizes the randomness of vehicle movement. Vehicle movement is dependent on factors such as road ground conditions, traffic situations, and pedal operations. Secret bits are generated based on this randomness to establish secure intra-vehicle communication. The authors of [36] study a pairwise shared secret key in full-duplex communication using Received Signal Strength (RSS) in mobile and static IoT networks. They employ cascade reconciliation and the Merkle-Damgard hash function for this purpose, generating a group key disseminated to all users. In addition, the ultrawideband channel pulse response between two transceivers has been employed as a source of common randomness to generate keys in the field of communication [37]. The dimension of the resultant key is determined based on the mutual information between the observations at two radio units. The authors also discern an approximation and upper bound on mutual information for a general multipath channel.

In [38], mapping-varied spatial modulation is introduced to generate secret keys in a multi-hop sensor network. The mapping patterns of the spatial modulation are driven by random Channel Quality Indicator (CQI) patterns over the authorized link to produce the secret key. Given that the mapping patterns of radiated information fluctuate by the CQI pattern, an eavesdropper would be oblivious to the channel where data transmission occurs. Distinct from RSS-based key generation, secret keys can also be derived by exploiting the channel response from multiple orthogonal frequency-division multiplexing subcarriers [39]. The systems possess the ability

to offer detailed channel information for improving the bit generation rate.

B. ML-Assisted Communication Security

Several surveys [40], [41] focus on incorporating ML and deep learning (DL) to realize secure, intelligent systems in the domains of IoT communications. The survey in [42] provides comprehensive insights on machine learning for security in vehicular networks. In this survey, the authors stress building resilient networks dependent on ML algorithms.

Reinforcement Learning-based spoofing attack detection mechanism is studied in [43] with Q-learning authentication to detect rogue nodes. RSSI is utilized for detecting spoofing data by providing an alarm when there is a disparity between the information received from the malicious node and the previously recorded data. Repeated spoofing is formulated through the Markov decision process, and this aids in the decision-making of the Q-learning. Q-learning uses prior knowledge of the attack model for choosing the actions, i.e., detection of the rogue nodes. However, when considering scenarios like platooning, where the channel between vehicles at times becomes partially observable, there is a need for neural networks like LSTM that can aid in enriching the training environment periodically.

ML has been used in securing communications in high-mobility vehicular networks such as the Internet of Vehicles (IoV) and 5G-based vehicular networks (5GVN) [44]. For instance, ML is used for intrusion detection in these connected vehicles. The dynamics of high-mobility networks, such as channel estimation, traffic and trajectory prediction, and congestion control, are presented. Authors of [45] develop an ML-based technique that utilizes the implicit features of wireless channels to generate keys with a high agreement rate. In this work, neural networks are concurrently trained to project their channel estimates into a hidden space, which remains out of reach to potential adversaries. By minimizing the distance between the latent spaces generated by two trusted pairs of wireless nodes the authors were able to reduce the KDR efficiently.

ML has also been used to generate secret keys. The authors of [46] develop an ML-based secret key generation approach. This approach refreshes the weights and data of the input-hidden layer during each dissemination cycle. The confidential key for each iteration is generated by employing the output layer weights.

A temporal difference-based DQN [53], employing a greedy strategy for exploration-exploitation balance, has been utilized to define the longitudinal trajectory on a typical urban roadway with signal-controlled intersections. Discretization of time, space, and speed is employed in this method to simplify the learning process and enable the networks to learn more effectively. By optimizing the trajectory, specifically focusing on fuel-efficient acceleration and deceleration, it is possible to reduce the fuel consumption of the vehicle. DQN-based reinforcement Learning has also been used to train discrete action spaces such as intercept probability, aiming to select the optimal vehicular relay from a set of multiple relays from a security standpoint [48]. Variants of DQN, such as the duelling double deep Q-network (D3QN) [47] have been used to develop a secrecy-aware energy-efficient scheme where it modelled cryptographic key generation process as a Markov Decision Process (MDP). This work formulated time-series parameters such as power control, channel allocation, and beamforming using minimal signal-to-interference-plus-noise ratio constraints. While reinforcement learning techniques like the DQN are much suited for discrete action spaces, it is inefficient while covering long-term temporal dependencies involving sequential data.

LSTM is one of the key technologies used in the literature to capture temporal variations and enrich the environment. They have the capability to capture long-term dependencies in sequential data. For instance, it is used in studying non-linear temporal data like congestion in peak hours to predict the traffic flow of vehicles [54]. Anomalies are deviations from the normal patterns in temporal data. Therefore, LSTM has also been used to detect anomalies and attacks on a Connected-autonomous vehicle using temporal correlation of message contents [55]. It can also learn hierarchical representations of temporal features at different time scales. A Temporal Pattern Attention (TPA) integrated with LSTM is used for evaluating the risk of the candidate trajectory, a risk assessment based on the velocity obstacle method which considers influencing factors such as time to collision [56].

ML also aids in delivering computation-efficient solutions for vehicular security. Authors of [57] use a low-resource deep learning model for enabling control using prediction in self-driving cars. A pre-trained, supervised LSTM model is used to learn the spatial-temporal information in the captured input images and fit the nonlinear relationship between the images and the corresponding steering angles.

Recent ML techniques like transfer learning (TL) have also been used in intrusion detection systems (IDS). In TL, machine learning exploits the knowledge gained from a previous task to improve generalization about another. The authors of [58] use tree-based TL model update schemes to detect intrusions in the system. Their system model is built on two assumptions, contingent on whether or not the cloud can promptly deliver labelled data. TL leverages data sourced either from the cloud or from local accessibility. With each local update, vehicles acquire a pseudo-label for the unlabeled data, and the TL model undergoes multiple updates to counter a new attack. These updates are prioritized over the cloud's completion of the attack data labelling.

Singular learning approaches do not possess the ability to track partially observable network observation transitions resulting from time-varying data such as channel quality. This increases the learning uncertainty in event-driven models like platooning and eventually deteriorates the overall learning accuracy. ML techniques like KNN [59], [60] have been used in the literature to learn the communication data and signal strength in vehicular area networks to take appropriate actions. Our preliminary works [15], [16], [19] were based on secret key agreement algorithms to optimize the channel quantization intervals recursively, thereby maximizing the key agreement probability. With the capability of DRL to address non-convex

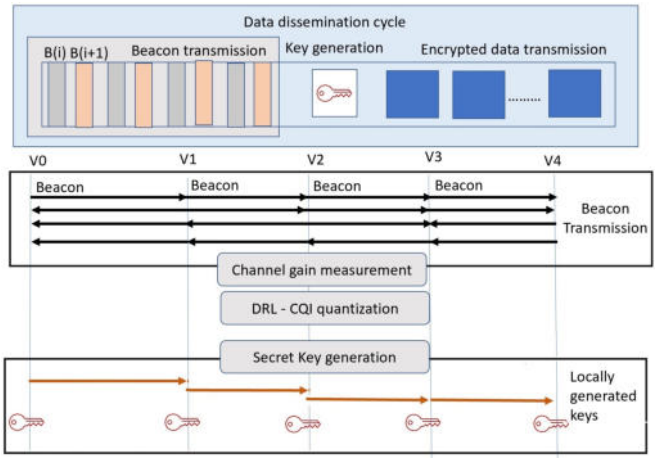

Fig. 2. The encrypted data dissemination, where the platooning vehicles individually generate the secret keys based on the chosen quantization intervals.

optimization problems like secret key generation, in this paper, we employ DRL to aid in the secret key generating process.

III. System Model

The platoon consists of one leader vehicle V_0 which is the first vehicle in the platoon and a number of following autonomous vehicles V_i, where $i \in [1, n]$. The leader vehicle determines crucial driving information of the whole platoon, such as speed adjustments, braking, and lane changes, according to the road conditions. Secret key bits generated at V_i and V_{i+1} are based on the point-to-point estimation of the quantization levels. The beacon packet carries the vehicle's identification (ID) number. The beacon transmission is initialized by the leader vehicle (i.e., V_0), and the adjacent follower vehicle (i.e., V_1) broadcasts the beacon once the leader's beacon is received. V_0 and V_1 individually measure the RSS values on the receipt of the beacon packets.

In particular, the beacon packet from V_1 also acknowledges the reception of V_0 beacon at V_1. The follower vehicles V_i, $i \in [2, n]$ estimate the RSS value of their respective precedent vehicle. Each vehicle in the platoon locally generates its secret key to encrypt or decrypt the vehicular data for dissemination. Moreover, the estimated RSS values are quantized using the proposed combinatorial DRL scheme to achieve the secret key agreement, where details will be introduced in the next section.

Our formulation, which focuses on the key agreement between two adjacent platooning vehicles, intentionally decouples the broader multi-agent key agreement problem into manageable subproblems. This methodology is not only pragmatic but also deeply rooted in the intrinsic structure and operational dynamics of vehicular platoons. By concentrating on the pairwise key agreement between adjacent vehicles, the proposed DRL-KeyAgree addresses the most critical and immediate communication security needs within a platoon. This is because secure communication links between neighbouring vehicles are foundational to the platoon's overall security architecture [20]. Adjacent vehicles need to rapidly and reliably share information to coordinate actions, such as speed adjustments and lane changes, making their communication links primary targets for potential adversaries. Optimizing the security of these pairwise links, therefore, has a cascading positive effect on the platoon's collective security [21]. Moreover, the proposed DRL-KeyAgree simplifies the complexity inherent in multi-agent reinforcement learning. Multi-agent systems introduce challenges such as non-stationarity, where the environment's dynamics change as each agent learns and adapts its policy, and the exponential growth of the action space with the addition of each agent. By breaking down the multi-agent problem into dyadic interactions, we reduce the problem space, allowing for more focused learning and optimization processes. Each vehicle pair independently optimizes its key agreement process using the A2C-LSTM framework, making the learning process more manageable and scalable.

The key generation process is carried out through three steps: channel gain measurement, channel quality indicator (CQI) quantization, and platooning secret key generation, as shown in Fig. 2. The vehicles in the platoon transmit their respective beacons, and quantization intervals are estimated. The channel gain $(\Omega_{0,1})$ is measured at the first two vehicles $(V_0$,$V_1)$ according to the beacon packets.

A. Channel Gain Measurement

Channel gain measurement has a crucial role in secret key generation. This refers to measuring the wireless communication channel's characteristics between adjacent vehicles in the platoon. The safety distance between the vehicles depends on several factors, such as speed and deterioration of the braking capacity [61]. A secret key is generated between the leader and the following vehicle, in which V_1 measures the channel gain based on its received signal from V_0. This measurement takes into account the channel impairments such as fading and noise. Assuming that the vehicles with antennas mounted on them travel in the same road segment, a Line of Sight (LOS) communication between the adjacent connected vehicles. For modelling the inter-vehicle communication channel, we take the large-scale path loss, which takes the average effect of multipath is taken into consideration.

For a transmitting power of P_t at vehicle V_i, which is at a distance of $d_{i,j}$ from vehicle V_j, and considering a fixed channel-dependent constant γ and a path loss exponent P_l, the receiving signal power P_r at V_j can be formulated as follows:

$$P_r = P_t + \gamma - P_l \log_{10}(d_{i,j}) + \sigma_{i,j} \tag{1}$$

where $\sigma_{i,j}$ represents the independent shadow fading between the vehicles V_i and V_j over different time epochs. The shadow fading is modelled as a random variable with a log-normal distribution [62]. The received power is not uniform when measured at different locations while maintaining the same distance separation between the vehicle's transmitter and receiver [63]. In this V2V scenario, we incorporate a statistical model for shadow fading that assumes independence between vehicles which means the signal strength varies between the two concerned vehicles. Specifically, the model allows for power attenuation at a rate proportional to $(1/d)^n$ (where d denotes the distance or range), incorporating a random variable

to adequately capture shadowing effects. Considering these dynamics, our channel model offers a realistic evaluation. Moreover, the generation of keys in our framework is contingent upon the inherent randomness of the channel, which is influenced by the vehicular distance. This aspect further underscores the realism and applicability of our evaluation methodology. Now, the channel gain $\Omega_{i,j}$ of the link between the vehicles V_i and V_j can be formulated as:

$$\Omega_{i,j} = P_t - P_r \tag{2}$$

The inter-vehicle distance $d_{i,j}$ can be further written as:

$$d_{i,j} = 10^{\frac{\Omega_{i,j}+\gamma+\sigma_{i,j}}{10P_l}}. \tag{3}$$

We utilized the SUMO simulator [73], to simulate an urban traffic scenario and model traffic flow dynamics to understand the channel gain randomness due to the distance between vehicles. In one of our studies [64], we focused on analyzing inter-vehicle distances within an urban environment. We gathered a comprehensive dataset [65] comprising over 2 million time sequences capturing vehicle movements at velocities ranging from 5 km to 15 km per hour across four lanes in both directions. It's important to note that in our simulation, vehicles were programmed to maintain a minimum following distance of 2.5 m between each other. The variability in vehicle velocities observed in our dataset reflects real-world conditions, such as traffic congestion and signal delays, which influence the speed of vehicles navigating through urban streets. This data includes the timestamps of the vehicles, the types of the vehicles, their speed, and their respective positions. The randomness in the distance between the vehicles and the velocities directly impacts the randomness of channel gain measured between two vehicles. The simulation for key generation using the DRL-KeyAgree architecture is done using TensorFlow. TensorFlow offers a flexible framework for designing and implementing complex neural network architectures, like DRL-KeyAgree. The SUMO dataset provides a comprehensive and realistic representation of urban traffic patterns, incorporating various factors such as vehicle movements, and traffic densities. By integrating this dataset into the DRL-KeyAgree environment, we aim to create a simulation that closely mirrors real-world conditions. The randomness in the channels is learned from our simulations in the SUMO simulator and is used as the dataset in our environment fed to the A2C-LSTM networks.

B. CQI Quantization

The leader and the first follower vehicle quantize the channel gain $(\Omega_{0,1})$, and their respective fading channel randomness is converted into bit vectors. Let l and L denote the l-th quantization interval and higher extremity of quantization intervals, respectively. $l \in [1, L]$. The quantization intervals are obtained by conducting channel gain measurement and CQI quantization. The channel gains are quantized by each vehicle into one of the intervals, i.e., $[\delta_l^-, \delta_l^+]$, where "-" and "+" denote the extremities of the quantization intervals. The underlying quantization intervals can be optimized to their lower and upper extremities in the possibility of significant random noise and estimation errors amongst the inter-vehicle channels.

TABLE I
LIST OF SYMBOLS AND REMARKS

Symbols	Remarks
$\Omega_{(i-1,i)}$	Channel Gain between adjacent vehicles
$l, l+1, ..., L$	Number of Quantization intervals
δ_l	Quantization interval
$\delta_l^\star$	Optimized Quantization interval
V_i	The i-th vehicle
K_i	Key extracted for the ith vehicle
G	Gray code word
K_i^{prv}	Private key
K_i^{pub}	Public key
S_α	states
A_α	Actions
S_α^o	partially observable states

Given an n-vehicle platoon, optimization of the CQI quantization can be formulated as

$$\max_{\delta_i} \sum_{i=1}^{N} \Pr\{\delta_1, \delta_2, \cdots, \delta_i\}, \tag{4}$$

where $\delta_i = [\delta_l^-, \delta_l^+]^*$, $l \in [1, L]$. Pr represents the probability of the uniformity among the secret bits that are quantized to the same optimal l-th CQI quantization interval in $[\delta_l^-, \delta_l^+]^*$. The upper bound of the l-th interval is the lower bound of the adjacent $(l+1)$-th interval, namely, $\delta_l^+ = \delta_{l+1}^-$. However, the problem (4) is non-convex, making it difficult to address using techniques such as regression training. With the increasing size of the problem, the complexity of resolving the non-convex optimization intensifies at an exponential rate. Therefore, DRL-KeyAgree (which will be investigated in Section IV) is proposed to determine the optimal $[\delta_l^-, \delta_l^+]^*$, taking advantage of a combinatorial design of A2C and LSTM to minimize the KDR.

C. Secret Key Generation and Agreement

Classical encoding techniques are used to fix a binary code word to every quantization bin in $[\delta^-, \delta^+]$ to extract secret bits. Based on the binary code word assigned to every quantization bin, a public key cryptosystem generates secret bits to encrypt and secure data transmission in the platooning system [16].

As the value of $\Omega_{i,i+1}$ is within $[\delta^-, \delta^+]$, $F_{enc}^{v_i}(l, l+1)$, an encoding method is used to allocate a binary code to every quantization bin and extract its respective secret key K_i. Gray coding is implemented as an example of $F_{enc}^{v_i}(l, l+1)$ as follows:

Denote $K_i(l),\ l \in [1, L-1]$ as the complement bit of the codeword. We have

$$K_i(l) = \begin{cases} 1, & \text{if } l \mod 3 \geq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

A gray codeword is generated among two neighboring codewords with only a one-bit difference. For a codeword length of G, this respective list consists of 2^G possible codewords. $K_i^+(l) \in \{0,1\}^G$ is defined as the $f_i^+(l)$-th Gray codeword.

$$f_i^+(l) = \lfloor (l-1)/4 \rfloor; \tag{6}$$

$$f_i^-(l) = \lfloor ((l+1) \mod L)/4 \rfloor. \quad (7)$$

Thus, $K_i^-(l) \in \{0,1\}^G$ can be defined as the $f_i^-(l)$-th Gray codeword. The value of $K_i^-(l)$ is responsible for circularly shifting $K_i^+(l)$ by two elements. The value of $F_{enc}^{v_i}(l, l+1)$ can be used by other encoding methods like the Gillham coding [66] and Lucal coding [67]. Following the codeword of $F_{enc}^{v_i}(l, l+1)$, generation of symmetric or asymmetric secret keys is possible for encrypting the data transmission at every hop. In classic Elliptic-curve cryptography (ECC), the vehicles V_i and V_{i+1} concur on an elliptic curve and a defined base point. The private keys K_i^{prv} and K_{i+1}^{prv}, and their corresponding public keys $K_i^{\text{prv}} * G$, which is K_i^{pub}, and $K_{i+1}^{\text{prv}} * G$, which is K_{i+1}^{pub}, are generated.

Similar to this technique, the adjacent vehicles in our system model V_i and V_{i+1} accept on the same private key after our algorithm is executed, i.e., $K_i^{\text{prv}} = K_{i+1}^{\text{prv}}$. Therefore, we have $K_i^{\text{pub}} = K_{i+1}^{\text{pub}}$. V_i computes $K_i^{\text{prv}} * K_{i+1}^{\text{pub}}$, and V_{i+1} computes $K_{i+1}^{\text{prv}} * K_i^{\text{pub}}$, and both vehicles can concur on shared secret key with same bits, $K_S = K_i^{\text{prv}} * K_{i+1}^{\text{pub}} = K_{i+1}^{\text{prv}} * K_i^{\text{pub}} = K_i^{\text{prv}} * K_{i+1}^{\text{prv}} * G = K_{i+1}^{\text{prv}} * K_i^{\text{prv}} * G$.

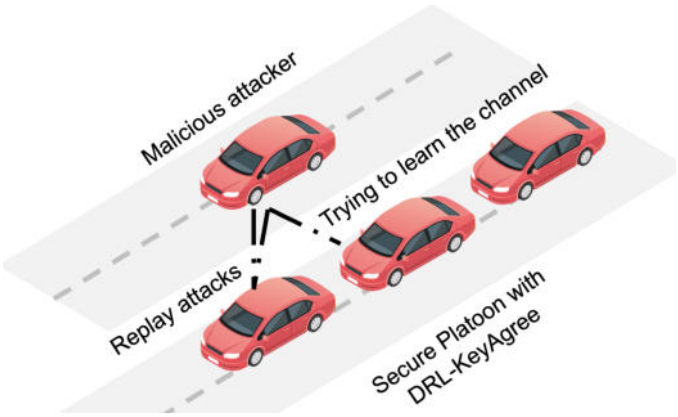

Fig. 3. The attacker is trying to intentionally diminish the quality of the channel or generate the same key to affect the integrity of the platoon.

D. Threat Model

Apart from the noise and interference due to the unpredictability of the environmental conditions, the wireless communication channel in a platooning scenario can be affected by the attacker's intentional jamming. Such attacks have a dire impact on the reliability and quality of channel randomness and eventually the safety of the platoon itself.

Common replay attacks where a malicious adversary captures and replays previously recorded messages to impersonate a legitimate user can be cumbered by analyzing the time stamps from the legitimate and the attacker. Another viable technique is to use message authentication codes to verify the integrity of the message.

The probability of the attacker trying to observe the channel between vehicles V_0 and V_1 can be given as:

$$P_{Malicious} = \Pr\{\delta^{Malicious} \in [\delta^-, \delta^+]\}. \quad (8)$$

As shown in Fig. 3, the channel quality between the two vehicles in the platoon differs from the attacker. Thus, it is difficult for the attacker to generate the same key as the platooning vehicles.

IV. Combinatorial DRL Architecture – DRL-KeyAgree

Reinforcement learning uses feedback from the environment to learn and adapt by choosing a better action by analyzing the environment before and after the previous action. The critical components of our combinatorial DRL architecture are the LSTM layers and the A2C layer, as presented in Fig. 4. The LSTM layer periodically feeds the partially observable states to enrich the environment. The A2C reinforcement learning architecture consists of an actor, state, cost, and underlying policy. An actor chooses the action strategy corresponding to the state output from the environment. Based on the choice of action, a cost is defined. Aiming at reducing the overall cost, the actions are adjusted. A2C network is a combination of the policy gradient, which is the actor and the function approximation, which is the critic. The actor will choose actions, and the critic calculates feedback based on actions taken. The actor then modifies the respective action based on the critic's feedback every time.

A. POMDP for planning under uncertainty

A Partially Observable Markov Decision Process (POMDP) must be formalized to define this scenario. This process be condensed as a tuple (S, A, O, C), where S, A and O are the state, action, and correspondent observation space, respectively. C is the associated cost. We consider the channel quality $\Psi_{t,i,j}$ and the quantization intervals between the observable vehicles i and j as the state. The cycle time of dissemination is mentioned as t. The network state S_α is defined as

$$S_{\alpha,t} = \{\Psi_{\alpha,i,j}, \delta_{\alpha,v_{i,j}}\}. \quad (9)$$

The action in POMDP that encompasses the choosing quantization intervals ($\boldsymbol{\delta_l}$) is unanimously generated over the cycle time t. Thus, we have

$$A_{\alpha,t} = \{\boldsymbol{\delta_{l(i,j)}}\}. \quad (10)$$

Here, S_α^o only covers the observable state of the vehicle under observation. For instance, if the second follower is not visible to the leader vehicle at the state $S_{\alpha,t}$, it cannot know to capture state information. The observation state pertaining to two adjacent vehicles, $S_\alpha^o \in S_\alpha$ is given by

$$S_\alpha^o = \{(\Psi_{\alpha,i,j}^o, \delta_{\alpha,V_{i,j}}^o)\}. \quad (11)$$

The cost of our platooning system is in coherence with the uniformity of secret key bits generated across the vehicles under observation. In a platoon with $(i - N)$ vehicles, the difference in the generated secret bits must be minimal. Thus, the cost of the POMDP is defined as

$$C\{\beta|\alpha, A_\alpha\} = \sum_1^N |\varphi_{(\mathbf{i},\mathbf{i+1})}|. \quad (12)$$

The cost function defines the immediate cost when the action takes place. To change from state α to state β the network

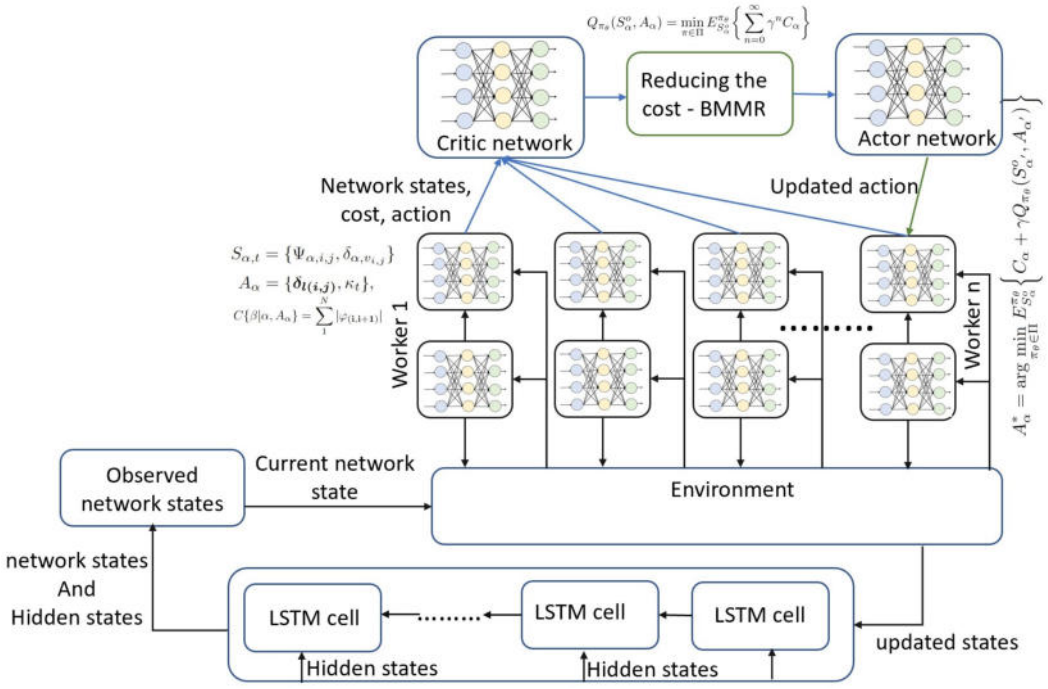

Fig. 4. DRL-KeyAgree with a combinatorial DRL architecture where the LSTM layer periodically feeds the partially observable states to enrich the environment and the A2C reinforcement learning architecture choosing the actions based on the critic feedback.

cost is denoted by $C\{\beta|\alpha, A_\alpha\}$. The cost function affirms the uniformity of the key generation over the platoon. We aim to reduce the overall cost denoted as an action-value function. $\varphi_{(i,i+1)}$ is the difference in the key bits generated between two adjacent vehicles. By reducing the bit difference, we can ensure that the maximum secret bits generated between the vehicles are equal, thus creating a unanimity of key bits among the vehicles. The critic network approximates the action value function, which can be defined as follows.

$$Q_{\pi_\theta}(S^o_\alpha, A_\alpha) = \min_{\pi\in\Pi} E^{\pi_\theta}_{S^o_\alpha}\left\{\sum_{n=0}^{\infty}\gamma^n C\{\beta|\alpha, A_\alpha\}\right\} \tag{13}$$

where $\gamma \in [0,1]$ is a discount factor used as a factor in upcoming state observations. $E^{\pi_\theta}_{S^o_\alpha}\{\cdot\}$ takes the expectation in accordance with the policy π_θ and state observation S^o_α. The optimal action-value function (13) of a state-action pair (S^o_α, A_α) and the value of the concurrent state-action pair $(S_{\alpha'}, A_{\alpha'})$ can be then given as

$$Q_{\pi_\theta}(S^o_\alpha, A_\alpha) = \min_{\pi_\theta\in\Pi} E^{\pi_\theta}_{S_\alpha}\left\{C\{\beta|\alpha, A_\alpha\} + \gamma Q_{\pi_\theta}(S_{\alpha'}, A_{\alpha'})\right\}. \tag{14}$$

The optimal action, A^*_α, which is in line with (14), can be given by

$$A^*_\alpha = \arg\min_{\pi_\theta\in\Pi} E^{\pi_\theta}_{S_\alpha}\left\{C\{\beta|\alpha, A_\alpha\} + \gamma Q_{\pi_\theta}(S_{\alpha'}, A_{\alpha'})\right\}. \tag{15}$$

B. LSTM layer

The partially observable network transitions result from the time-varying data in the platooning scenario, such as the channel quality and quantization levels. This increases learning uncertainty in the underlying model, which results in the eventual deterioration of the overall learning accuracy. The partially observable states can be due to poor channel quality or a long distance between the vehicles. The secret key generation can be compromised, corresponding to partial knowledge about the states of the vehicles. To mitigate this, it is necessary that a state characterization layer is designed to predict the hidden states. Therefore, we extend the LSTM to predict the respective hidden states and enrich the training environment of the A2C.

At state S_α, the hidden states $h^{hid}_{S_\alpha}$ are calculated according to the following composite functions:

$$h^{hid}_{S_\alpha} = o_{S_\alpha}\tanh(C_{S_\alpha}); \tag{16}$$

$$p_{S_\alpha} = \sigma(W_p \cdot [h^{hid}_{S_\alpha-1}, C_{S_\alpha-1}, A_{S_\alpha}] + e_p), \tag{17}$$

$$F_{S_\alpha} = \sigma(W_f \cdot [h^{hid}_{S_\alpha-1}, C_{S_\alpha-1}, A_{S_\alpha}] + e_f); \tag{18}$$

$$o_{S_\alpha} = \sigma(W_0 \cdot [C_{S_\alpha}, h^{hid}_{S_\alpha-1}, A_{S_\alpha}] + e_0); \tag{19}$$

$$C_{S_\alpha} = F_{S_\alpha} C_{S_\alpha-1} + p_{S_\alpha}\tanh(W_c \cdot [h^{hid}_{S_\alpha-1}, A_{S_\alpha}] + e_c). \tag{20}$$

For an input gate of the LSTM layer p_{S_α} in (17), there is a forget gate F_{S_α} given by (18), an output gate o_{S_α} in (19) and corresponding cell activation vectors C_{S_α} in (20). The input gate provides the information on the newer quantization interval information to be stored in the cell state. Forget gate has the previous information that is to be discarded. The output gate provides the corresponding activation to the final output of the LSTM block. The logistic sigmoid and hyperbolic tangent functions can be represented by σ and $\tanh$. The weight matrix and the bias matrix can be given as $\{W_0, W_c, W_f, W_p\}$ and $\{e_0, e_c, e_f, e_p\}$ respectively.

The state processing network of this architecture captures the temporal correlation over the entire training period where any hidden quantization interval is chosen. The LSTM network captures any variation regularity from the observation vectors in line with the quantization intervals. Thus the action space gets improved by the hidden information at every turn. After the entire $T_{learning}$ is completed, the hidden state at the last time step represents the fully updated environment that will be used for optimal action generation and value estimation.

Algorithm 1 encompasses the combinatorial learning part used for learning the environment and helping to take appropriate actions to reduce the overall cost. With the utilization of LSTM, the states enriched are with the hidden inputs and added to the environment. The beacons that are used to initiate the transmissions and the quantization intervals between the vehicles are also defined. These inputs act as the states S_α used to define the action A_α through learning.

C. A2C layer

The A2C layer that we employ utilizes an advantage function in our combinatorial structure. This function delineates the predicted outcome for all future actions. The advantage function determines the anticipated optimal value for the selected quantization interval, which is updated in the action space. When the feedback value calculated by the selected action from the critic network is greater than the expected value, the action is considered optimal, and the gradient follows in that choice of action. The hidden states are fed into the A2C network for obtaining enriched actions based on the hidden states of the network.

An environment of the current state, next state, action on the state and its respective cost $\left(\alpha, \beta, A_\alpha, C\{\beta|\alpha, A_\alpha\}\right)$ are presented to the input states from the environment for the training of the A2C onboard. The training is done for a time of t_{learning}, and the selected actions A_α are carried out in the training environment in every t_{learning}.

Fig. 4 shows that A2C trains the policy gradients to minimize the estimated loss between the actor-critic and target neural networks. It takes advantage of the experiences gleaned from the current state, stored in a memory unit. The critic neural network provides an estimation of the optimal $Q_{\pi_\theta}(S^o_\alpha, A_\alpha)$, which projects the expected cumulative network cost. This cost can be described as the total loss experienced once the state α has been observed and an action A_α has been executed.

The critic network learns the optimal actions for minimizing the cost function that affirms the uniformity of key generation over the platoon as defined in Equation (12). The actor neural network in A2C generates the action of choosing the quantization intervals. The training episodes are carried out in the onboard A2C till the data is disseminated securely between the platoon. However, there is a possibility of multiple token packets being transmitted by v_0 and v_1 during one cycle of data dissemination. In such a case, the aforementioned algorithm can be realized in Δ iterations resulting in ($\Delta \geq 1$) and Δ secret keys at every vehicle. As the magnitude of Δ increases, the estimation errors decrease, leading to a higher probability of consistent keys. Therefore, obtaining a larger Δ is crucial to increase the likelihood of key agreement. Pertaining to this statement, we assume $\Delta = 1$ in this article to analyze the key agreement performance in the worst case.

Algorithm 1 Combinatorial LSTM-A2C for key agreement

Initialize:
  LSTM states: $h^{hid}_{S_\alpha}$, critic Parameters: $\Psi_{t,i,j,id}$, S_α, actor parameters: A_α
Initialize:
  an empty buffer of length $T_{Learning}$ and the time-step t;
for Event time: $t = n$ **do**
  LSTM: enriches observable network states S_α by adding $h^{hid}_{S_\alpha}$ to the environment
  Action: corresponding action A_α choosing of δ_l based on the network state S_α at t
  Key Generation: Secret key K_t is generated based on the chosen δ_l
  at the end of t the agent gets the observation S_α of the environment
  Append the observations S_α to the end of the buffer
end for
for Event time: $t = n + 1$ **do**
  The agent samples the next action $A_{\alpha'}$ based on the reduction of Cost $C\{\beta|\alpha, A_\alpha\}$
  New action $A_{\alpha'}$ is updated to the environment $\left(\alpha, \beta, A_{\alpha'}, C\{\beta|\alpha, A^{'}_{\alpha}\}\right)$
  Buffer Clearance: the first observation in the buffer is deleted and S_α is appended to the end of the buffer
  The critic approximates the optimal action-value function $Q_{\pi_\theta}(S^o_\alpha, A_{\alpha'})$
  The optimal action A^*_α is defined by satisfying the optimal action-value function in (11) and updated to the environment
  Update the timestep $t \rightarrow t + n$
  Until The maximum number of iterations $T_{Learning}$ is completed
end for
Output: δ^*_L, K_t

To analyze the *space complexity* of an algorithm, we need to consider the additional memory requirements as a function of the input size. The algorithm initializes LSTM states ($h^{hid}_{S_\alpha}$), critic parameters ($\Psi_{t,i,j,id}$), and actor parameters (A_α). The space complexity for these parameters is constant, denoted as O(1), as they do not depend on the input size. The algorithm maintains a buffer of observations with a length of $T_{Learning}$. The space complexity of the buffer is O($T_{Learning}$) since it grows with the number of time steps.

The *time complexity* of the provided algorithm is primarily determined by the two main loops: the first loop for event time $t = n$ and the second loop for event time $t = n + 1$, with each iteration involving constant-time operations of the LSTM enrichment, action selection, key generation, and buffer updates. The second loop focuses towards action sampling, environment updates, buffer clearance, critic approximation, and updating the optimal action. The algorithm's runtime

grows linearly with the value of $T_{Learning}$ till the maximum number of iterations.

V. Performance Analysis

This section thoroughly delves into the performance metrics and the evaluation of the test outcomes. The performance indicators have been designed to measure the secret bits' randomness and KDR of the proposed DRL-KeyAgree system.

- **Secret bits randomness:** This performance metric shows the randomness amongst the secret bits generated between every vehicle. A higher level of randomness helps in reducing the possibility of cracking the generated keys.
- **KDR:** This performance metric shows the uniformity in the number of secret bits generated throughout the training period. This is a ratio of the number of secret bits to the overall secret bits generated. A smaller value of the KDR signifies the uniformity of secret bits generated by the vehicles.

A. Secret bits randomness

We used the standard randomness test suite from NIST [68] to verify the randomness of the keys generated by the DRL algorithm. The NIST provides a set of 16 tests to verify the randomness. Among them, seven are selected to verify the randomness of the key generated by DRL-KeyAgree.

The P-value represents the probability that an efficiently tuned random number generator might yield a sequence with lesser randomness than the evaluated input sequence. For the test to be considered successful, all P-values must exceed 0.01 [69]. Specific tests within this suite required an input bit stream greater than 106 bits. As a result, we selected tests that were suitable for our dataset.

As shown in Table II, the keys generated by DRL-Keyagree between every pair of vehicles in the platoon have P-values larger than the threshold of 0.01 to pass the test. This specific threshold indicates that the generated secret bit streams are completely random with a confidence of 99%. This large randomness increases the complexity of cracking the keys by any malicious eavesdropper.

B. Impact of the quantization intervals in platooning

The action defined by the DRL-KeyAgree is the selection of quantization levels between the vehicles. As discussed in Section III, the quantization intervals are optimized to the lower and upper extremities, and they can have a finite set of quantization intervals to choose from. The experiments were run with a learning rate of 0.001 which determines the step size at each iteration while moving toward a minimum of a loss function. It controls the size of the updates to the model's weights. Higher values of the discount factor (0-1) make the agent focus more on long-term rewards, hence we took a discount factor of 0.99. We take varying amounts of LSTM cells for the enrichment of the model and learning its impact.

In Fig. 5, we present the comparison of the KDR taken for 100 and 500 quantization levels across four vehicles in the platoon. When the number of quantization intervals is lesser, reaching convergence is much easier. Thus, the convergence is achieved in around 200 episodes. When the choice of action is made from a larger pool of values, it results in more time to reach convergence, this can be seen that the KDR values do not converge early, and the training patterns are still visible around 200 episodes. This shows the importance of channel quality in a platooning scenario. Erratically varying channel quality can result in higher extremities in the quantization intervals, which can delay the KDR's convergence. A good channel with its quality varying only because of fading will result in a smaller pool of quantization intervals.

C. Comparison with state-of-the-art techniques

1) DQN-based secret key generation: In a DQN-based channel randomness key generation, the action is the quantization levels chosen to generate secret key bits. As the training progresses, the agent adjusts its actions, exploring different quantization levels to maximize the reward which is formulated as the unanimity of the secret bits. The experience replay allows the agent to store and revisit past interactions with the environment, including the selection of quantization levels and the resulting secret key generation. The agent is incentivized to select quantization levels that lead to the generation of secret keys with high levels of agreement, thereby maximizing the cumulative reward over time.

2) DDQN-based secret key generation: Double DQN-based secret key generation [74] employs a separate target Q-network for generating target Q-values based on the quantization levels at every state. DDQN decouples the quantization level selection and its respective value estimation steps instead of using the Q-values directly, similar to the standard DQN networks. It also uses an experience replay similar to the DQN networks to learn from past interactions with the environment.

3) Dueling DQN-based secret key generation: Duelling DQN [75] explicitly separates the estimation of the state value function and the advantage of choosing the action of the formulation of the quantization intervals. This enables the network to understand the advantages of specific states and their respective actions. The target Q-values, representing the quantization levels resulting in unanimous keys, serve as the reference during training to stabilize and prevent overestimation or oscillation of Q-values.

4) HRKF-based secret key generation: Higher Rate Secret Key Formation (HRKF) [71] is a channel randomness-based secret key generation technique. This traditional method undergoes four stages for key generation such as channel characteristics measurement, reciprocal enhancement, quantization and encoding, randomness extraction and key verification.

On one hand, DDQN decouples the action selection of key generation and its respective value estimation, thus making it more stable than the original DQN. By using two separate Q-networks, DDQN provides a more accurate estimation of Q-values and leads to slightly better convergence during training.

TABLE II
P-VALUES FROM NIST STATISTICAL TEST SUITE, TO PASS THE TEST, ALL P-VALUES MUST BE GREATER THAN 0.01. THESE ARE THE P VALUES OF THE KEYS GENERATED BETWEEN THE VEHICLES

NIST tests	V0 - V1	V1 - V2	V2 - V3	V3 - V4	V4 - V5
Frequency test	0.7	0.69	0.71	0.67	0.73
Block frequency test	0.9	0.89	0.87	0.88	0.9
Cumulative sums (FWD)	0.6	0.71	0.73	0.68	0.649
Cumulative sums (REV)	0.9	0.83	0.87	0.88	0.9
Runs	0.56	0.61	0.58	0.55	0.59
FFT	0.33	0.39	0.32	0.33	0.3
Approx. Entropy	0.3	0.39	0.3	0.34	0.35
Serial	0,48, 0.49	0,48, 0.47	0,47, 0.48	0,46, 0.49	0,48, 0.47

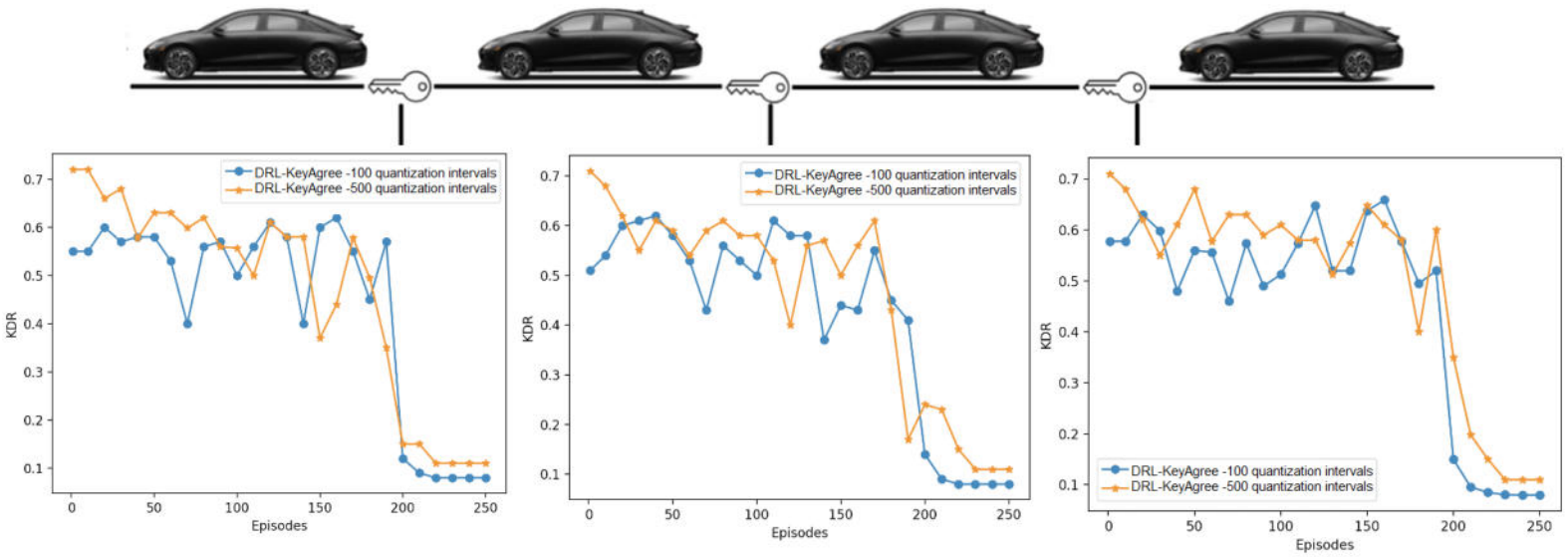

Fig. 5. KDR values for a set of four vehicles across the platoon, and convergence with the number of quantization intervals compared against each other. The lower number of quantization levels results in faster convergence.

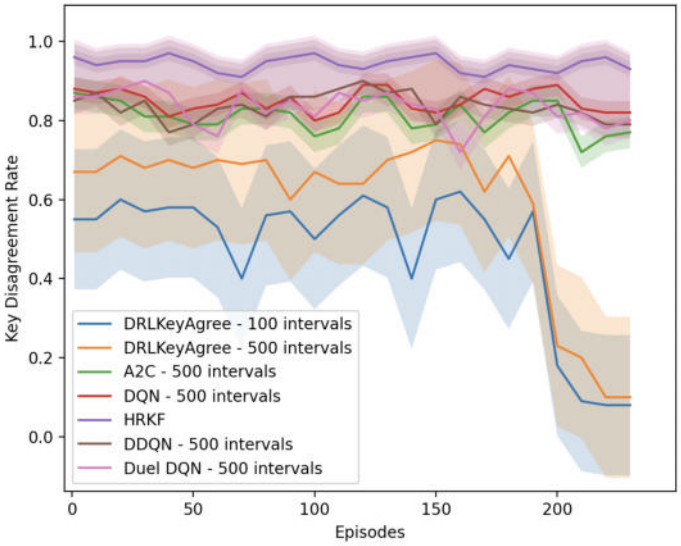

Fig. 6. Given 250 training episodes, we examine the KDR of the proposed DRL-KeyAgree in contrast to A2C, DQN, DDQN, Dueling DQN, and HRKF methods

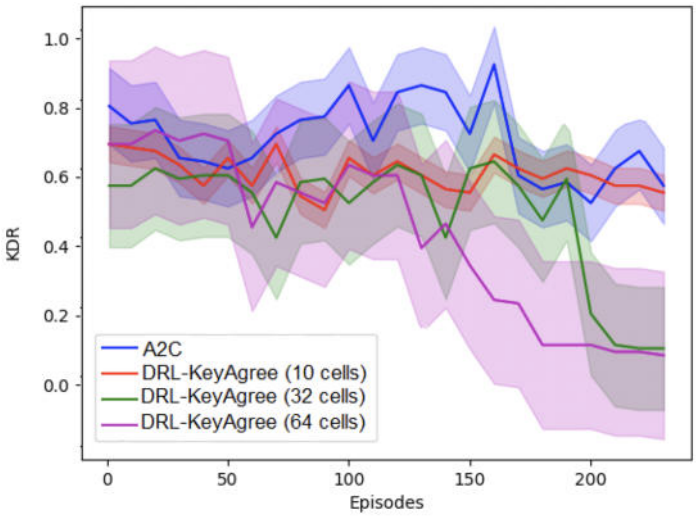

Fig. 7. When the number of LSTM cells in the DRL-KeyAgree increases, its exploration of the state is significantly higher than its counterparts, thus leading to the least KDR and an early convergence.

On the other hand, Dueling DQN splits the Q values into two streams, the state value stream to estimate the value of being in a particular state and the advantage stream to estimate the

advantage of choosing a set of secret bits compared to the target Q value. The output of both these streams is combined to get the final Q-values. This explicit separation of functions will enable Dueling DQN to be efficient where different actions are expected in the same state space. However, due to the impact of dimensionality, Q-learning is not practical for use in learning problems that contain a large state or action space [70].

A2C networks are typically more data and sample-efficient than the DQN and its variants operating in smaller action spaces [72]. In our case, the action was to choose the respective quantization level to generate the secret bits. HRKF does not focus on accurately assessing the channel randomness for unanimous secure key agreement and does not use any machine learning strategies to improve the key agreement. Rather, it just generates keys based on the channel randomness between the vehicles, resulting in high KDR values. Considering the temporal nature of the data trained, in collaboration with LSTM, DRL-KeyAgree outperforms DQN and its variants. Another prominent reason for this performance is A2C is an on-policy algorithm that learns from the current policy, whereas DQN and its variants are off-policy algorithms that learn from a target policy. On-policy algorithms tend to require fewer samples to converge to an optimal policy. DRL-KeyAgree optimizes the policy directly using policy gradients, allowing for more fine-grained control over the learning process compared to the value-based methods like DQN, DDQN, and Dueling DQN.

As shown in Fig. 6, the DRL-KeyAgree that conducts the combination of A2C and LSTM achieved faster convergence than the counterparts. The proposed DRL-KeyAgree captures the temporal variations and adjusts the respective actions more efficiently, resulting in lower KDR and faster convergence than its counterparts. Because of its on-policy methods, the average of the KDR since DRL-KeyAgree achieves convergence (around 220 episodes), is almost 40% better when compared to DQN, Duel DQN, DDQN, HRKF, and A2C.

Another important factor that affects the KDR is the number of quantization intervals the DRL has to choose from to generate the secret keys. For this evaluation, we used a set of 500 and 100 quantization intervals for the DRL-KeyAgree to choose from and plotted them against A2C, DQN, DDQN, and Duel DQN with 500 intervals. Fig. 6 shows a large KDR at the initial episodes, from 1 to 150. This is because the DRL has a large number of values to select from when the sample size is large. However, as the training proceeds, the difference becomes smaller as the proposed DRL-KeyAgree that uses A2C LSTM trains the network efficiently to reduce the overall KDR. Towards convergence, the difference between the KDR amongst the techniques with the same quantization intervals is around 70%.

D. Impact of the LSTM size on KDR

The size of an LSTM defines the number of memory cells in the network. This size determines the capacity to learn and represent complex patterns in selecting the channels between vehicles. When the size of the LSTM is larger, more intricate dependencies and relationships within the input sequence can be captured. With larger sizes of LSTM, the predictive accuracy improves and results in an early convergence and lesser key generation rates.

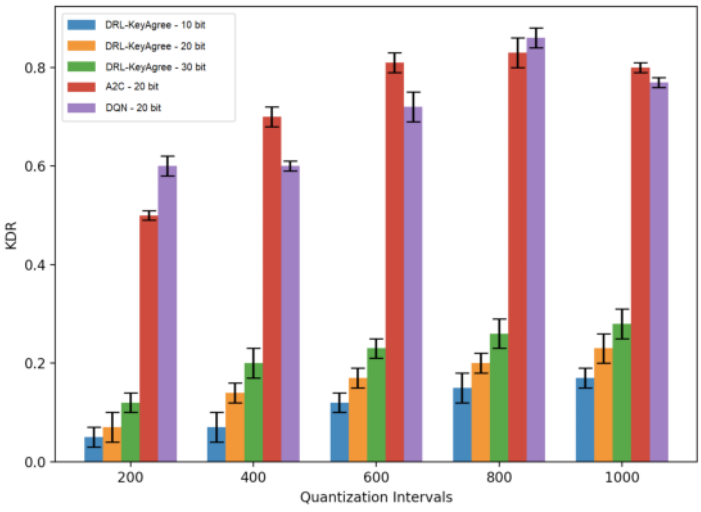

Fig. 8. KDR values of the DQN and A2C rise acutely high when we consider a state pool with more than 500 quantization intervals.

It must be noted that larger LSTM sizes have more parameters, resulting in overfitting and increasing the overall computational complexity. Hence, it is vital to choose the right LSTM size based on the size of the input. As shown in Fig. 7, When considering a larger LSTM size of around 64 cells, the convergence occurs around 170 episodes with a lower KDR. However, without LSTM or minimal cells of just ten cells, they do not reach convergence compared to 64 cells. We also have plotted the KDR gain in Fig. 7, which shows the exploration of the LSTM to define an optimal action. When the value of the LSTM is more significant (LSTM 64), its exploration of the state is significantly higher than its counterparts in both extremities, thus increasing the standard deviation visible in this plot, resulting in the least KDR value and early convergence.

E. Impact of the number of quantization intervals on DRL

The state space has a crucial impact on the KDR. When the amount of samples to choose from is large, it takes more time to reach convergence. For this experiment, we consider three key bit variants of the DRL-KeyAgree and pit them against DQN and A2C, which have a 20-bit fixed key length for various quantization intervals.

From the results of the experiment in Fig. 8, it is evident that for larger quantization intervals, the KDR rises high. We see the same behavior here. But interestingly, the KDR values of the DQN and A2C rise acutely high when we consider a state pool with more than 500 quantization intervals. We can witness almost a 50% - 60% increase in the values of KDR in the case of A2C and DQN. In this experiment, we also present an interesting result that even with a larger key bit length of 30 bits, the proposed DRL-KeyAgree could outperform the standard A2C and DQN methods by 20%- 30%.

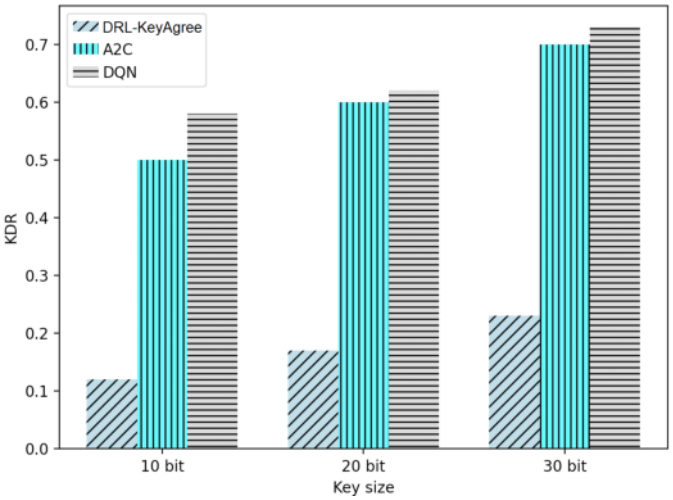

Fig. 9. Impact of quantization intervals on the KDR: DRL-KeyAgree constantly provides a lower KDR than A2C and DQN by around 40%.

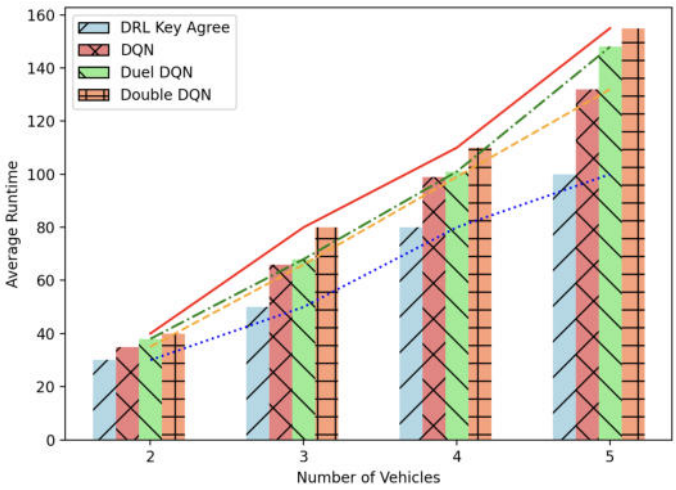

Fig. 10. Impact of the increasing number of vehicles on the overall runtime of secret key generation

F. Impact of the key length on the KDR

The secret key's length is one of the crucial factors determining the security level provided by an underlying cryptographic system. A longer key length generally makes it more tedious for an attacker to break the encryption and decipher the message. However, there can be trade-offs in line with security and delay when we have a longer key.

We compare DQN, A2C, and DRL-KeyAgree, which are trained to provide a key of the same length. From Fig. 10, it is evident that KDR increases with the length of the keys. When the key size is longer, achieving a unanimous key is difficult as it is tougher to attain unanimity among a large number of key bits. The results in the graph represent the average KDR that has been taken since our DRL-Key agree method reached convergence. Our method constantly provides a lower KDR than A2C and DQN by around 40%.

G. Impact on scalability

Runtime plays a crucial role in learning by influencing the system's scalability. In this experiment, we compared the runtime (in milliseconds) for secret key generation using our DRL-KeyAgree method against state-of-the-art deep learning techniques, such as Double DQN [74], Duelling DQN [75], and the classic DQN. Analyzing the running average (illustrated by the dotted lines in Figure 10), we observe that the average runtime for secret key generation using DRL-KeyAgree is consistently 10-20% faster than DQN, Dueling DQN, and Double DQN, especially evident when the number of vehicles in the platoon increases from four to five. The increased runtime in dueling DQN architectures is primarily due to the separate estimation of the value and advantage streams within the neural network architecture. In the case of the double DQN architecture, it updates and maintains two separate Q-value estimators for higher stability, but this is a trade-off for reducing the overall runtime.

VI. Conclusion

In this paper, we proposed DRL-KeyAgree, a combined DRL approach that employs A2C and LSTM to curtail the secret key disagreement between the platooning vehicles and bolster the vehicular platoon's communication security in 6G. DRL-KeyAgree learns fluctuating network states, including vehicles' channel quality, and constructs a secret key to secure transmission. We tested the proposed DRL-KeyAgree against state-of-the-art methods like DDQN, Dueling DQN, DQN, A2C and HRKF, demonstrating its superiority in KDR across diverse key lengths, runtime and quantization stages. The keys produced by our method also successfully passed the NIST's standard randomness test suite, indicating a heightened difficulty in breaking the generated keys.

Acknowledgement

This Paper was supported by the CISTER Research Unit (UIDP/UIDB/04234/2020), financed by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology), and by project Aero.Next Portugal (ref. C645727867-00000066), funded by the EU/Next Generation, within call n.º 02/C05-i01/2022 of the Recovery and Resilience Plan (RRP)

References

[1] Chen, X., Leng, S., He, J. and Zhou, L., 2020. Deep-learning-based intelligent intervehicle distance control for 6G-enabled cooperative autonomous driving. IEEE Internet of Things Journal, 8(20), pp.15180-15190.

[2] Prathiba, S.B., Raja, G., Anbalagan, S., Gurumoorthy, S., Kumar, N. and Guizani, M., 2021. Cybertwin-driven federated learning based personalized service provision for 6G-V2X. IEEE Transactions on Vehicular Technology, 71(5), pp.4632-4641.

[3] Li, Y., Yu, Y., Susilo, W., Hong, Z. and Guizani, M., 2021. Security and privacy for edge intelligence in 5G and beyond networks: Challenges and solutions. IEEE Wireless Communications, 28(2), pp.63-69.

[4] Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J., Niyato, D., Dobre, O. and Poor, H.V., 2021. 6G Internet of Things: A comprehensive survey. IEEE Internet of Things Journal, 9(1), pp.359-383.

[5] Adhikari, M., Hazra, A., Menon, V.G., Chaurasia, B.K. and Mumtaz, S., 2021. A roadmap of next-generation wireless technology for 6G-enabled vehicular networks. IEEE Internet of Things Magazine, 4(4), pp.79-85.

[6] Li, K., Ni, W., Zheng, J., Tovar, E. and Guizani, M., 2021. Confidentiality and timeliness of data dissemination in platoon-based vehicular cyber-physical systems. IEEE Network, 35(4), pp.248-254.

[7] Kurunathan, H., Santos, J., Moreira, D. and Santos, P.M., 2023, June. Towards Safe Cooperative Autonomous Platoon Systems Using COTS Equipment. In 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) (pp. 464-469). IEEE.

[8] Pereira, J., Kurunathan, H. and Santos, P.M., 2023, June. Edge-Aided V2X Collision Avoidance with Platoons: Towards a Hybrid Evaluation Toolset. In 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W) (pp. 68-71). IEEE.

[9] Maglaras, L.A., Basaras, P. and Katsaros, D., 2013, December. Exploiting vehicular communications for reducing co2 emissions in urban environments. In 2013 International Conference on Connected Vehicles and Expo (ICCVE) (pp. 32-37). IEEE.

[10] Li, K., Ni, W., Tovar, E. and Guizani, M., 2019. Optimal rate-adaptive data dissemination in vehicular platoons. IEEE Transactions on Intelligent Transportation Systems, 21(10), pp.4241-4251.

[11] Rui Rafael, Harrison Kurunathan, and Eduardo Tovar, "An RT-cloud solution towards security in Vehicular platooning systems". in RT-Cloud 2023, 35th Euromicro conference on Realtime Systems.

[12] Li, K., Ni, W., Tovar, E. and Guizani, M., 2018, May. LCD: Low latency command dissemination for a platoon of vehicles. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[13] Taylor, S.J., Ahmad, F., Nguyen, H.N. and Shaikh, S.A., 2022. Vehicular platoon communication: architecture, security threats and open challenges. Sensors, 23(1), p.134.

[14] Zeng, K., 2015. Physical layer key generation in wireless networks: challenges and opportunities. IEEE Communications Magazine, 53(6), pp.33-39.

[15] Li, K., Ni, W., Emami, Y., Shen, Y., Severino, R., Pereira, D. and Tovar, E., 2019. Design and implementation of secret key agreement for platoon-based vehicular cyber-physical systems. ACM Transactions on Cyber-Physical Systems, 4(2), pp.1-20.

[16] Li, K., Lu, L., Ni, W., Tovar, E. and Guizani, M., 2019. Secret key agreement for data dissemination in vehicular platoons. IEEE Transactions on Vehicular Technology, 68(9), pp.9060-9073.

[17] Ye, C., Mathur, S., Reznik, A., Shah, Y., Trappe, W. and Mandayam, N.B., 2010. Information-theoretically secret key generation for fading wireless channels. IEEE Transactions on Information Forensics and Security, 5(2), pp.240-254.

[18] Li, K., Kurunathan, H., Severino, R. and Tovar, E., 2018, April. Cooperative key generation for data dissemination in cyber-physical systems. In 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS) (pp. 331-332). IEEE.

[19] Li, K., Lu, L., Ni, W., Tovar, E. and Guizani, M., 2019, May. Cooperative secret key generation for platoon-based vehicular communications. In ICC 2019-2019 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[20] Al-Shareeda, S. and Özgüner, F., 2014, November. Secure pairwise key establishment in vehicular networks. In 2014 International Conference on Connected Vehicles and Expo (ICCVE) (pp. 105-111). IEEE.

[21] Balador, A., Bazzi, A., Hernandez-Jayo, U., de la Iglesia, I. and Ahmadvand, H., 2022. A survey on vehicular communication for cooperative truck platooning application. Vehicular Communications, 35, p.100460.

[22] Xu, P., Cumanan, K., Ding, Z., Dai, X. and Leung, K.K., 2016. Group secret key generation in wireless networks: Algorithms and rate optimization. IEEE Transactions on Information Forensics and Security, 11(8), pp.1831-1846.

[23] Gascón, A., Schoppmann, P., Balle, B., Raykova, M., Doerner, J., Zahur, S. and Evans, D., 2016. Privacy-preserving distributed linear regression on high-dimensional data. Cryptology ePrint Archive.

[24] Wang, D., Bai, B., Zhao, W. and Han, Z., 2018. A survey of optimization approaches for wireless physical layer security. IEEE Communications Surveys and Tutorials, 21(2), pp.1878-1911

[25] Vavasis, S.A., 1995. Complexity issues in global optimization: a survey. Handbook of global optimization, pp.27-41.

[26] Datta, S., Nikolaou, P. and Michael, M.K., 2023, July. Reputation-based User Vehicle Assignment in Intelligent and Connected Vehicle Platoons. In 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS) (pp. 1-6). IEEE.

[27] Deng, Z., Yang, K., Shen, W. and Shi, Y., 2023. Cooperative platoon formation of connected and autonomous vehicles: toward efficient merging coordination at unsignalized intersections. IEEE Transactions on Intelligent Transportation Systems.

[28] Li, S., Li, J., Liang, Y., Zhang, H., Wu, S., Wang, S. and Cheng, L., 2023. TD-SAS: A Trust-Aware and Decentralized Speed Advisory System for Energy-Efficient Autonomous Vehicle Platoons. IEEE Transactions on Intelligent Vehicles.

[29] Liu, Z., Han, Y., Fan, J., Zhang, L. and Lin, Y., 2020, July. Joint optimization of spectrum and energy efficiency considering the c-v2x security: A deep reinforcement learning approach. In 2020 IEEE 18th International Conference on Industrial Informatics (INDIN) (Vol. 1, pp. 315-320). IEEE.

[30] Zhu, Z., Gupta, S., Gupta, A. and Canova, M., 2023. A deep reinforcement learning framework for eco-driving in connected and automated hybrid electric vehicles. IEEE Transactions on Vehicular Technology.

[31] Bas, C.U. and Ergen, S.C., 2012, April. Spatio-temporal characteristics of link quality in wireless sensor networks. In 2012 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1152-1157). IEEE.

[32] Mitev, M., Barreto, A.N., Pham, T.M., Matthé, M. and Fettweis, G., 2022, December. Filterbank secret key generation rates in multipath channels. In GLOBECOM 2022-2022 IEEE Global Communications Conference (pp. 4057-4062). IEEE.

[33] Del Prete, S., Fuschini, F. and Barbiroli, M., 2022. A Study on Secret Key Rate in Wideband Rice Channel. Electronics, 11(17), p.2772.

[34] Usman, M., Althunibat, S. and Qaraqe, M., 2022. A channel state information-based key generation scheme for the Internet of Things. Security and Communication Networks, 2022.

[35] Yang, E. and Fang, S., 2022, April. GPSKey: GPS-based secret key establishment for intra-vehicle environment. In Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2022.

[36] Xu, P., Cumanan, K., Ding, Z., Dai, X. and Leung, K.K., 2016. Group secret key generation in wireless networks: Algorithms and rate optimization. IEEE Transactions on Information Forensics and Security, 11(8), pp.1831-1846.

[37] Wilson, R., Tse, D. and Scholtz, R.A., 2007. Channel identification: Secret sharing using reciprocity in ultrawideband channels. IEEE Transactions on Information Forensics and Security, 2(3), pp.364-375.

[38] Yang, Y. and Guizani, M., 2018. Mapping-varied spatial modulation for physical layer security: Transmission strategy and secrecy rate. IEEE Journal on Selected Areas in Communications, 36(4), pp.877-889.

[39] Liu, H., Wang, Y., Yang, J. and Chen, Y., 2013, April. Fast and practical secret key extraction by exploiting channel response. In 2013 Proceedings IEEE INFOCOM (pp. 3048-3056). IEEE.

[40] Sharma, P., Jain, S., Gupta, S. and Chamola, V., 2021. Role of machine learning and deep learning in securing 5G-driven industrial IoT applications. Ad Hoc Networks, 123, p.102685.

[41] Sarker, I.H., Khan, A.I., Abushark, Y.B. and Alsolami, F., 2022. Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions. Mobile Networks and Applications, pp.1-17.

[42] Olowononi, F.O., Rawat, D.B. and Liu, C., 2020. Resilient machine learning for networked cyber-physical systems: A survey for machine learning security to securing machine learning for CPS. IEEE Communications Surveys and Tutorials, 23, pp.524-552.

[43] Lu, X., Wan, X., Xiao, L., Tang, Y. and Zhuang, W., 2018, May. Learning-based rogue edge detection in VANETs with ambient radio signals. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[44] Liang, L., Ye, H. and Li, G.Y., 2018. Toward intelligent vehicular networks: A machine learning framework. IEEE Internet of Things Journal, 6(1), pp.124-135.

[45] Wei, X. and Saha, D., 2022, May. Knew: Key generation using neural networks from wireless channels, in Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning (pp. 45-50).

[46] Zhang, J. and Qu, G., 2019. Physical unclonable function-based key sharing via machine learning for IoT security. IEEE Transactions on Industrial Electronics, 67(8), pp.7025-7033.

[47] Sharma, H., Kumar, N. and Tekchandani, R.K., 2023. SecBoost: Secrecy-Aware Deep Reinforcement Learning Based Energy-Efficient Scheme for 5G HetNets. IEEE Transactions on Mobile Computing.

[48] Ghourab, E.M., Bariah, L., Muhaidat, S., Sofotasios, P.C., Al-Qutayri, M. and Damiani, E., 2023. Secure Relay Selection with Outdated CSI in Cooperative Wireless Vehicular Networks: A DQN Approach. IEEE Access.

[49] Li, Q., Ou, B., Liang, Y., Wang, Y., Yang, X. and Li, L., 2023. TCN-SA: A Social Attention Network Based on Temporal Convolutional Network for Vehicle Trajectory Prediction. Journal of Advanced Transportation, 2023.

[50] Chen, Y., Shu, T., Zhou, X., Zheng, X., Kawai, A., Fueda, K., Yan, Z., Liang, W., Kevin, I. and Wang, K., 2022. Graph attention network with spatial-temporal clustering for traffic flow forecasting in an intelligent transportation system. IEEE Transactions on Intelligent Transportation Systems.

[51] de Araujo-Filho, P.F., Naili, M., Kaddoum, G., Fapi, E.T. and Zhu, Z., 2023. Unsupervised GAN-based intrusion detection system using temporal convolutional networks and self-attention. IEEE Transactions on Network and Service Management.

[52] Zhang, Z., Han, S., Wang, J. and Miao, F., 2023, May. Spatial-temporal-aware safe multi-agent reinforcement learning of connected autonomous vehicles in challenging scenarios. In 2023 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5574-5580). IEEE.

[53] Cheng, Y., Hu, X., Chen, K., Yu, X. and Luo, Y., 2023. Online longitudinal trajectory planning for connected and autonomous vehicles in mixed traffic flow with deep reinforcement learning approach. Journal of Intelligent Transportation Systems, 27(3), pp.396-410.

[54] Narmadha, S. and Vijayakumar, V., 2023. Spatio-temporal vehicle traffic flow prediction using multivariate CNN and LSTM model. Materials today: proceedings, 81, pp.826-833.

[55] Mansourian, P., Zhang, N., Jaekel, A. and Kneppers, M., 2023. Deep learning-based anomaly detection for connected autonomous vehicles using spatiotemporal information. IEEE Transactions on Intelligent Transportation Systems.

[56] Wu, Y., Gao, J., Wu, H. and Wei, H., 2023. Driving intention prediction algorithm based on TPA-LSTM for autonomous vehicles. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, p.09544070231214348.

[57] Jiang, H., Chang, L., Li, Q. and Chen, D., 2020, October. Deep transfer learning enable end-to-end steering angles prediction for self-driving car. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 405-412). IEEE.

[58] Otoum, Y., Wan, Y. and Nayak, A., 2022, May. Transfer learning-driven intrusion detection for Internet of Vehicles (IoV). In 2022 International Wireless Communications and Mobile Computing (IWCMC) (pp. 342-347). IEEE.

[59] So, S., Sharma, P. and Petit, J., 2018, December. Integrating plausibility checks and machine learning for misbehavior detection in VANET. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 564-571). IEEE.

[60] Grover, J., Prajapati, N.K., Laxmi, V. and Gaur, M.S., 2011. Machine learning approach for multiple misbehavior detection in VANET. In Advances in Computing and Communications: First International Conference, ACC 2011, Kochi, India, July 22-24, 2011, Proceedings, Part III 1 (pp. 644-653). Springer Berlin Heidelberg.

[61] Tiberti, W., Vieira, B., Kurunathan, H., Severino, R. and Tovar, E., 2020, April. Tightening up security in low power deterministic networks. In 2020 16th IEEE International Conference on Factory Communication Systems (WFCS) (pp. 1-7). IEEE.

[62] ETSI, T., 2017. 138 901 v14. 0.0,"5G; study on channel model for frequencies from 0.5 to 100GHz,". 3GPP TR 38.901 version 14.0. 0 Release 14)," ETSI, Tech. Rep.

[63] Munoz, D., Bouchereau, F., Vargas, C. and Enriquez, R., 2009. Signal parameter estimation for the localization problem. In Position location techniques and applications (pp. 23-29). Elsevier Inc.

[64] R. Reddy, L. Almeida, H. Kurunathan, M. G. Gaitan, P. M. Santos and E. Tovar, "Worst-Case Response Time of Mixed Vehicles at Complex Intersections," in IEEE Open Journal of Intelligent Transportation Systems, doi: 10.1109/OJITS.2024.3368797.

[65] R. Reddy, L. Almeida, H. Kurunathan, M. G. Gaitan, P. M. Santos and E. Tovar, "Worst-Case Response Time of Mixed Vehicles at Complex Intersections," GiThub Repository: https://github.com/radhareddy31/FCD

[66] Aircraft Electrical and Electronic Systems - Principles, Operation and Maintenance (1 ed.). Butterworth-Heinemann (Elsevier Ltd.). p. 69. ISBN 978-0-7506-8695-2.

[67] H. M. Lucal, "Arithmetic Operations for Digital Computers Using a Modified Reflected Binary Code," in IRE Transactions on Electronic Computers, vol. EC-8, no. 4, pp. 449-458, Dec. 1959, doi: 10.1109/TEC.1959.5222057.

[68] Pareschi, Fabio, Riccardo Rovatti, and Gianluca Setti. "On statistical tests for randomness included in the NIST SP800-22 test suite and based on the binomial distribution." IEEE Transactions on Information Forensics and Security 7.2 (2012): 491-505.

[69] Georgescu, C., Simion, E., Nita, A.P. and Toma, A., 2017, June. A view on NIST randomness tests (in) dependence. In 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-4). IEEE.

[70] Wei, F., Feng, G., Sun, Y., Wang, Y., Qin, S. and Liang, Y.C., 2020. Network slice reconfiguration by exploiting deep reinforcement learning with large action space. IEEE Transactions on Network and Service Management, 17(4), pp.2197-2211.

[71] Dewi, I.T., Sudarsono, A., Kristalina, P. and Yuliana, M., 2020. Higher Rate Secret Key Formation (HRKF) based on Physical Layer for Securing Vehicle-to-Vehicle Communication. EMITTER International Journal of Engineering Technology, 8(1), pp.140-160.

[72] Zhong, C., Lu, Z., Gursoy, M.C. and Velipasalar, S., 2019. A deep actor-critic reinforcement learning framework for dynamic multichannel access. IEEE Transactions on Cognitive Communications and Networking, 5(4), pp.1125-1139.

[73] Reddy, R., Almeida, L., Gaitan, M., Kurunathan, H., Santos, P. and Tovar, E., 2021, December. Work-In-Progress: Worst-Case Response Time of Intersection Management Protocols. In 2021 IEEE Real-Time Systems Symposium (RTSS) (pp. 556-559). IEEE.

[74] Van Hasselt, H., Guez, A. and Silver, D., 2016, March. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

[75] Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M. and Freitas, N., 2016, June. Dueling network architectures for deep reinforcement learning. In International conference on machine learning (pp. 1995-2003). PMLR.

Harrison Kurunathan is an integrated researcher at CISTER, Porto, Portugal. He received his PhD in Electrical and Computer Engineering from the University of Porto, Portugal, in 2021. He holds a bachelor's degree in Electronics and Communication Engineering from SRM University, India (2012), and a master's degree in Very Large-Scale Integration from SSN College of Engineering, Anna University, India (2014). Before his current role, he served as an assistant professor in the Department of Electronics and Communication Engineering at SRM University, India. His research interests encompass machine learning, wireless sensor networks, visible light communication (VLC), network protocols, network security, the Internet of Things (IoT), and cooperative cyber-physical systems (CPS).

Kai Li (S'09–M'14–SM'20) received the B.E. degree from Shandong University, China, in 2009, the M.S. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2010, and the Ph.D. degree in computer science from The University of New South Wales, Sydney, NSW, Australia, in 2014. Currently, he is a Visiting Research Scientist with the Division of Electrical Engineering, Department of Engineering, University of Cambridge, U.K., and a Senior Research Scientist with the CISTER Research Centre, Porto, Portugal. He is also a CMU-Portugal Research Fellow, jointly supported by Carnegie Mellon University (CMU), Pittsburgh, PA, USA, and the Foundation for Science and Technology (FCT), Lisbon, Portugal. In 2022, he was a Visiting Research Scholar with the CyLab Security and Privacy Institute, CMU. Prior to this, he was a Post-Doctoral Research Fellow with the SUTD-MIT International Design Centre, Singapore University of Technology and Design, Singapore, from 2014 to 2016. He has also held positions as a Visiting Research Assistant with the ICT Centre, CSIRO, Brisbane, QLD, Australia, from 2012 to 2013, and a full-time Research Assistant with the Mobile Technologies Centre, The Chinese University of Hong Kong, Hong Kong, from 2010 to 2011. He has been an Associate Editor of journals, such as *Internet of Things* (Elsevier) since 2024, *Nature Computer Science* (Springer) since 2023, *Computer Communications* (Elsevier) and *Ad Hoc Networks* (Elsevier) since 2021, and IEEE ACCESS from 2018 to 2024.

Eduardo Tovar (Member, IEEE) received the Licentiate, M.Sc., and Ph.D. degrees in electrical and computer engineering from the University of Porto, Porto, Portugal, in 1990, 1995, and 1999, respectively. Currently, he is a Professor with the Computer Engineering Department, School of Engineering (ISEP), Polytechnic Institute of Porto (P.Porto), where he is also engaged in research on real-time distributed systems, WSN, multiprocessor systems, CPS, and industrial communication systems. He heads the CISTER Research Laboratories. He was the Vice Chair of ACM SIGBED (2015–2019) and a member of the Executive Committee of the IEEE TC-RTS. Since 1991, he has ben authored or coauthored more than 300 scientific papers in the areas of RTES, CPS, and WSN. He has been consistently participating in top-rated scientific events as the Program Chair (e.g., ECRTS—2005, RTCSA—2010 and 2016, RTAS—2013, and ICCPS—2016). He has been the General Chair/Co-Chair of various scientific events, including the IEEE/ACM CPS Week—2018.

Alípio Mário Jorge received the PhD degree in computer science from U. Porto. He is currently an associate professor with the Department of Computer Science of the Faculty of Science of the U. Porto and the coordinator of LIAAD/INESC TEC, the Artificial Intelligence and Decision Support Lab of U. Porto since 2012. His research interests include data mining and machine learning, in particular recommender systems, NLP, and web intelligence. He lectures on information processing and data mining. He leads research projects on data mining and Web intelligence. He has co-chaired international conferences (ECML/PKDD 2005, Discovery Science 2009, ECML/PKDD 2015), workshops and seminars in data mining and artificial intelligence.

Wei Ni (Fellow, IEEE) received the B.E. and Ph.D. degrees in Communication Science and Engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He is a Principal Research Scientist at CSIRO and a Conjoint Professor at the University of New South Wales. He is also an Adjunct Professor at the University of Technology Sydney, and an Honorary Professor at Macquarie University. He was a Postdoctoral Fellow at Shanghai Jiaotong University from 2005 to 2008; Deputy Project Manager at Bell Labs, Alcatel/Alcatel-Lucent from 2005 to 2008; and Senior Researcher at Nokia from 2008 to 2009. He has (co-)authored one book, ten book chapters, more than 300 journal papers, more than 100 conference papers, 26 patents, and ten standard proposals accepted by IEEE. His research interests include machine learning, online learning, stochastic optimization, and their applications to the security, integrity and efficiency of network systems. Dr. Ni has been an Editor for IEEE Transactions on Vehicular Technology since 2022, IEEE Transactions on Wireless Communications since 2018, IEEE Transactions on Information Forensics and Security since 2024, IEEE Communications Surveys and Tutorials since 2024, and Cambridge Press New Research Directions: Cyber-Physical Systems since 2022. He served as the Chair of the IEEE Vehicular Technology Society New South Wales Chapter from 2020 to 2022, the Secretary and then the Vice-Chair of the Chapter from 2015 to 2019, Track Chair for VTC-Spring 2017, Track Co-chair for IEEE VTC-Spring 2016, Publication Chair for BodyNet 2015, and Student Travel Grant Chair for WPMC 2014.

Abbas Jamalipour (Fellow, IEEE) received the Ph.D. degree in Electrical Engineering from Nagoya University, Nagoya, Japan in 1996. He holds the position of Professor of Ubiquitous Mobile Networking with the University of Sydney and, since January 2022, the Editor-in-Chief of the IEEE Transactions on Vehicular Technology. He has authored nine technical books, eleven book chapters, over 550 technical papers, and five patents, all in the area of wireless communications and networking. Prof. Jamalipour is a recipient of a number of prestigious awards, such as the 2019 IEEE ComSoc Distinguished Technical Achievement Award in Green Communications, the 2016 IEEE ComSoc Distinguished Technical Achievement Award in Communications Switching and Routing, the 2010 IEEE ComSoc Harold Sobol Award, the 2006 IEEE ComSoc Best Tutorial Paper Award, as well as over 15 Best Paper Awards. He was the President of the IEEE Vehicular Technology Society (2020-2021). Previously, he was Executive Vice-President and Editor-in-Chief of VTS Mobile World and has been an elected member of the Board of Governors of the IEEE Vehicular Technology Society since 2014. He was the Editor-in-Chief of IEEE WIRELESS COMMUNICATIONS, the Vice President-Conferences, and a member of the Board of Governors of the IEEE Communications Society. He sits on the Editorial Board of the IEEE ACCESS and several other journals and is a member of the Advisory Board of the IEEE Internet of Things Journal. He has been the General Chair or Technical Program Chair for several prestigious conferences, including IEEE ICC, GLOBECOM, WCNC, and PIMRC. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electrical, Information, and Communication Engineers (IEICE), the International Artificial Intelligence Alliance (AIIA), the Institution of Engineers Australia, an ACM Professional Member, and an IEEE Distinguished Speaker.