CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Conference Paper

Exploring LSTM-assisted A2C For Physical Layer Security in Vehicular Cyber-Physical Systems

Harrison Kurunathan*

Kai Li*

Wei Ni

Na Li

Eduardo Tovar*

Mohsen Guizani

*CISTER Research Centre

CISTER-TR-240402

2024/05/27

Exploring LSTM-assisted A2C For Physical Layer Security in Vehicular Cyber-Physical Systems

Harrison Kurunathan*, Kai Li*, Wei Ni, Na Li, Eduardo Tovar*, Mohsen Guizani

*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, Wei.Ni@data61.csiro.au, xiaonanahd163@163.com, emt@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Physical layer security is of paramount importance in vehicular cyber-physical systems, as it safeguards not only the privacy of sensitive data exchanged between vehicles and infrastructure but also ensures the integrity and reliability of the entire transportation network. Key generation plays a crucial role in establishing secure communication channels and facilitating the creation of unique cryptographic keys used for encryption, decryption, and authentication purposes. The secret key generation involves deriving secret bits by harnessing the inherent randomness present within the communication channels. The difficulty lies in precisely evaluating the randomness of the channel to achieve unanimous agreement on secure key generation within an unpredictable environment. In this line, we propose a combinatorial approach involving A2C and LSTM to decrease the key disagreement rate. A2C employs policy and value-based strategies to choose quantization levels predicated on the randomness of wireless channels and LSTM includes the partially observable radio channels and improves the environment. Based on our performance evaluation, the proposed A2C-LSTM method substantially accelerates the convergence rate by 50-60% and reduces the Key Disagreement Rate (KDR) by 40%.

Exploring LSTM-assisted A2C For Physical Layer Security in Vehicular Cyber-Physical Systems

Harrison Kurunathan*, Kai Li*†, Wei Ni‡, Na Li§, Eduardo Tovar*, Mohsen Guizani¶
*CISTER, Portugal, {jhk, emt}@isep.ipp.pt
†University of Cambridge, U.K., kaili@ieee.org
‡CSIRO, Australia, wei.ni@data61.csiro.au
§Beijing University of Posts and Telecommunications, China, lina_lena@bupt.edu.cn
¶MBZUAI, United Arab Emirates, mguizani@ieee.org

***Abstract*—Physical layer security is of paramount importance in vehicular cyber-physical systems, as it safeguards not only the privacy of sensitive data exchanged between vehicles and infrastructure but also ensures the integrity and reliability of the entire transportation network. Key generation plays a crucial role in establishing secure communication channels and facilitating the creation of unique cryptographic keys used for encryption, decryption, and authentication purposes. The secret key generation involves deriving secret bits by harnessing the inherent randomness present within the communication channels. The difficulty lies in precisely evaluating the randomness of the channel to achieve unanimous agreement on secure key generation within an unpredictable environment. In this line, we propose a combinatorial approach involving A2C and LSTM to decrease the key disagreement rate. A2C employs policy and value-based strategies to choose quantization levels predicated on the randomness of wireless channels and LSTM includes the partially observable radio channels and improves the environment. Based on our performance evaluation, the proposed A2C-LSTM method substantially accelerates the convergence rate by 50-60% and reduces the Key Disagreement Rate (KDR) by 40%.**

***Index Terms*—Physical Layer Security, Vehicular Cyber-Physical Systems, Actor-Critic (A2C), Long Short-Term Memory (LSTM).**

I. Introduction

The vulnerability of interconnected vehicular cyber-physical systems to cyber threats necessitates robust encryption protocols, intrusion detection mechanisms, and authentication procedures to thwart potential attacks and maintain users' trust in the safety and efficacy of automotive technologies. Without adequate security measures, the potential consequences of unauthorized access or manipulation of critical vehicular systems could lead to disastrous outcomes, jeopardizing the safety, stability and functionality of modern transportation systems as a whole [1]. Vehicle-to-vehicle (V2V) communication allows vehicles to share real-time data about their speed, position, acceleration, deceleration, and other relevant information with nearby vehicles. Usually, the leader vehicle periodically broadcasts the vehicle status information to adjacent vehicles or the series of following vehicles [2]. This broadcasting technique using wireless channels is very vulnerable to cyber attacks such as eavesdropping and message replay. By tracking the driving route a malicious actor can abuse the mobility patterns of the vehicle. Hence, physical layer security is of paramount importance in vehicular cyber-physical systems.

Traditional methods like authentication protocols are essential to ensure that only authorized vehicles can participate in communications, preventing malicious entities from infiltrating the network [3]. Additionally, in recent days there are also intrusion detection systems that detect and mitigate any attempts at cyberattacks or unauthorized access [4]. One efficient way of securing the communication channels is the channel-based key generation, which generates secret bits based on the channel randomness of the wireless channels. In a channel-based key generation, the quantization levels are the number of discrete levels used to represent the observed characteristics of the communication channel. These levels are used to discretize the continuous measurements obtained from the channel into a finite set of values. The selection of quantization levels significantly influences the granularity of channel characteristics representation [5]. Optimizing the length of quantization levels in the channel-based key generation is a non-convex problem [6], making it difficult to address by optimization techniques such as regression training [7]. With an increase in the size of the problem, the complexity of resolving the non-convex optimization intensifies at an exponential rate [8], [9]. Hence, instead of relying solely on traditional cryptographic algorithms or predefined key generation methods, AI algorithms analyze the properties of the channel, such as signal strength, noise, and propagation characteristics, to derive secure keys. In this work, we employ a combination of A2C and LSTM to generate channel-based confidential key agreements in V2V networks, that minimize the overall key disagreement rate (KDR). The contributions of this paper are as follows:

- Our proposal introduces a combinatorial AI approach aimed at optimizing key agreements within V2V networks. This technique facilitates the generation of secret key bits autonomously by capturing network states without divulging channel information.
- We propose the integration of A2C and LSTM models which improves the training environment, leading to quicker convergence compared to alternative state-of-the-art AI techniques.
- The randomness of the generated keys is assessed using the standard NIST test suite. Findings demonstrate that our proposed technique decreases the key disagreement

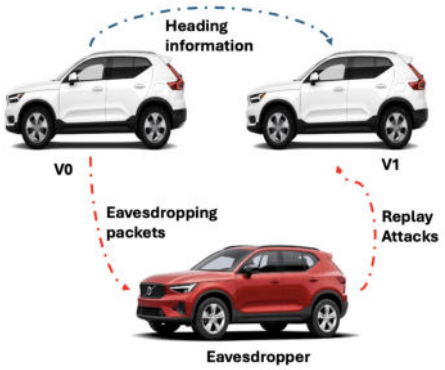

Fig. 1. V2V network where the leader vehicle disseminates data to the following vehicle using single hop data dissemination and an eavesdropper trying to eavesdrop the packets and attack the vehicles in the network.

rate by 60% in comparison to other AI techniques.

The rest of the paper is organized as follows. Section II presents the related works on secret key generation. Section III the system model for data dissemination in pairwise vehicles. Section IV elaborates on our proposed architecture, 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

Physical layer secret key generation provides lightweight encryption for resource-constrained Internet of Things (IoT) devices. Static IoT environments lead to ultra-low/zero secret key rates. Recent works have developed a reconfigurable intelligent surface (RIS) proposed for multiple-input single-output (MISO) communications [10]. In these techniques phase shifts at the RIS are tuned randomly in each frame, which makes time-varying reflection channels superimposed on the original channels. This randomness in frames is used in developing secret bits. However, in the context of vehicles, we confront mobile entities where communication channels exhibit significant randomness. Researchers in [11] introduce a cross-layer scheme for multi-carrier authentication, harnessing spatially/temporally correlated wireless channel characteristics to enable key verification without compromising secrecy. By associating a time-stamped hashed key with channel phase responses, we construct a PHY-layer signature, enabling sender identity verification while utilizing correlated channel responses across subcarriers to validate message integrity.

Similar cross-layer authentication schemes [12] have been employed in vehicular communication, integrating the transient reciprocal attributes of the wireless channel to re-authenticate the associated terminal, thereby mitigating overall complexity and reducing computation and communication overheads. The proponents of this approach design a location-dependent shared key with minimal mismatched bits, which is extracted between both networks for authentication purposes.

A scheme for physical key generation based on LoRa technology has been devised to secure V2V/V2I communications in [13]. Leveraging the Long Range (LoRa) protocol, this approach utilizes Received Signal Strength Indicator (RSSI) measurements over long distances as consensus information for generating secure keys. Adaptive wiretap code rates are proposed with a secrecy outage constraint to align with the time-varying wireless channels. In another study [14], beamforming and artificial noise (AN) schemes are introduced to enhance the ergodic secrecy rates of uplink and downlink channels, even without knowledge of eavesdropper channel state information (CSI). Additionally, researchers in [15] propose a GPS-based key establishment method that capitalizes on the randomness of vehicle movement, which is influenced by factors such as road conditions, traffic scenarios, and pedal operations. Secret bits are derived from this inherent randomness to establish secure intra-vehicle communication. The authors of [16] investigate pair-wise shared secret keys in full-duplex communication using Received Signal Strength (RSS) in both mobile and static IoT networks. They employ cascade reconciliation and the Merkle-Damgard hash function to generate a group key disseminated to all users. Moreover, the ultrawideband channel pulse response between two transceivers is utilized as a source of common randomness for key generation in communication contexts [17]. The dimension of the resulting key is determined based on the mutual information between observations at two radio units.

Surveys addressing artificial intelligence (AI) applications [18] have emphasized its significant role in enhancing security measures. The survey presented in [19] offers extensive insights into machine learning for security within vehicular networks. The authors of this survey emphasize the development of resilient networks reliant on ML algorithms. However, singular learning approaches cannot track partially observable network observation transitions resulting from time-varying data [20], such as channel quality, thereby increasing learning uncertainty in event-driven models and ultimately diminishing overall learning accuracy. ML techniques like K-nearest neighbours (KNN) [21] have been employed in the literature to learn communication data and signal strength in vehicles to execute appropriate actions. Our initial research efforts [5], [22], [23], [24] were centred on secret key agreement algorithms aimed at optimizing channel quantization intervals recursively, thus maximizing key agreement probability. Leveraging the capability of deep reinforcement learning (DRL) to address non-convex optimization problems such as secret key generation, in this study, we utilize a combination of A2C and LSTM to assist in the secret key generation process.

III. System Model

In our study, we focus on pairwise transmissions between 2 wirelessly connected vehicles and the generation of security bits between them based on the respective randomness of channel quality. Our system comprises a leader vehicle, denoted as V_0, and its adjacent follower vehicle, V_1. The leader vehicle plays a critical role in determining essential driving parameters such as speed adjustments, braking, and lane changes, based on prevailing road conditions. Secret key

bits are generated independently at both V_0 and V_1 through point-to-point estimation of quantization levels. As illustrated in the figure, beacons carrying vehicle identifications are transmitted among vehicles. Upon receiving beacon packets, both V_0 and V_1 measure the Received Signal Strength (RSS) values. Notably, the beacon packet from V_1 also confirms the reception of the beacon from V_0 at V_1. Each vehicle then generates its secret key locally for encrypting or decrypting vehicular data for dissemination. Furthermore, the estimated RSS values are quantized using the proposed A2C-LSTM scheme to achieve consensus on secret key generation. The key generation process comprises three steps: channel gain measurement, channel quality indicator (CQI) quantization, and secret key generation, as depicted in Figure 2. In the initial step of channel gain measurement, the channel gain $\Omega_{0,1}$ for the link between vehicles V_0 and V_1, with transmission power P_t and receiving power P_r, is measured as follows:

$$\Omega_{0,1} = P_t - P_r \tag{1}$$

A secret key is established between the leading and following vehicles, where V_1 assesses the channel gain using the signal received from V_0. When the vehicles are positioned at a distance of $d_{0,1}$, with a constant γ dependent on the channel, and a path loss exponent of P_l, the received signal power P_r at V_1 can be expressed as:

$$P_r = P_t + \gamma - P_l \log_{10}(d_{i,j}) + \sigma_{i,j} \tag{2}$$

$\sigma_{i,j}$ denotes the independent shadow fading occurring between vehicles V_0 and V_1 across various time intervals. This shadow fading is characterized as a random variable following a log-normal distribution, assuming independence between the vehicles, thereby indicating variability in signal strength between the two. This modelling choice enables a realistic portrayal of the fluctuations in communication conditions.

V_0 and V_1 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.

Our system model can be extrapolated to N number of vehicles. In our model we have decoupled it to just two vehicles, however, more vehicles can be added to the adjacent set of vehicles and the number of vehicles will not affect any key generation rate or hinder the QoS as the secret key generation will be always done for a pair of vehicles. Given an n-vehicle scenario, optimization of the CQI quantization can be formulated as

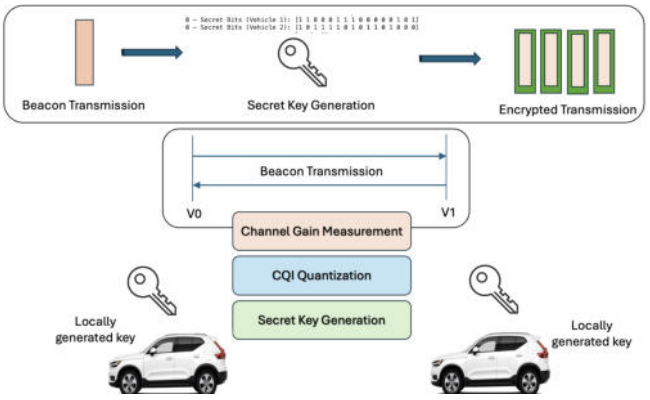

Fig. 2. Dissemination cycle where the secret keys are locally generated at the vehicles based on the chosen quantization intervals.

$$\max_{\delta_i} \sum_{i=1}^{N} \Pr\{\delta_1, \delta_2, \cdots, \delta_i\}, \tag{3}$$

where $\delta_i = [\delta_l^-, \delta_l^+]^*$, $l \in [1, L]$. With Pr representing probability, the uniformity among the secret bits occurs when they are quantized to the same optimal l-th CQI quantization interval in $[\delta_l^-, \delta_l^+]^*$. Secret Key Generation is done by fixing 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 vehicular system. 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. A gray codeword is generated among two neighbouring 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.

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

$$f_i^-(l) = \lfloor ((l+1) \mod L)/4 \rfloor. \tag{5}$$

Thus, $K^-i(l) \in 0, 1^G$ can be defined as the $f^-i(l)$-th Gray codeword. In our system model, adjacent vehicles V_i and $Vi+1$ adopt the same private key following the execution of our algorithm, i.e., $K^{\text{private}}i = K^{\text{private}}i+1$. Consequently, we have $K^{\text{public}}i = K^{\text{public}}i+1$. Vehicle V_i computes $K^{\text{private}}i * K^{\text{public}}i+1$, and vehicle $Vi+1$ computes $K_{i+1}^{\text{private}} * K^{\text{public}}i$, thereby both vehicles reaching a consensus on a shared secret key with identical bits, denoted as $K_S = K^{\text{private}}i * K^{\text{public}}i+1 = K^{\text{private}}i+1 * K^{\text{public}}i = K^{\text{private}}i * K^{\text{private}}i+1 * G = K^{\text{private}}i+1 * K_i^{\text{private}} * G$.

IV. A2C-LSTM BASED SECRET KEY GENERATION

In our architecture, the LSTM layer periodically feeds the partially observable states to enrich the environment. The A2C structure consists of an actor A, state S, cost C, and underlying policy. An actor chooses the action strategy corresponding to

the state output from the environment. Based on the choice of action, a reward 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. Now considering the channel quality $\Psi_{t,i,j}$ and the quantization intervals between the vehicles 0 and 1 which are observable O states, for a cycle time of dissemination t, the network state S_α is defined as:

$$S_{\alpha,t} = \{\Psi_{\alpha,i,j}, \delta_{\alpha,v_{i,j}}\}. \tag{6}$$

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

$$A_{\alpha,t} = \{\boldsymbol{\delta_{l(i,j)}}\}. \tag{7}$$

Here, S^o_α 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, $S^o_\alpha \in S_\alpha$ is given by:

$$S^o_\alpha = \{(\Psi^o_{\alpha,i,j}, \delta^o_{\alpha,V_{i,j}})\}. \tag{8}$$

The cost function ensures uniformity in key generation across the vehicles, aiming to minimize overall cost, denoted by the action-value function. $\varphi_{(i,i+1)}$ represents the difference in key bits between adjacent vehicles. Minimizing this difference ensures equal distribution of maximum secret bits among vehicles, promoting unanimity in key bits. With $(i-N)$ vehicles, minimizing differences in secret bits is essential. Therefore, the cost of the POMDP is defined as:

$$C\{\beta|\alpha, A_\alpha\} = \sum_{1}^{N} |\varphi_{(\mathbf{i},\mathbf{i+1})}|. \tag{9}$$

The critic network approximates the action value function, which can be defined as follows.

$$Q_{\pi_\theta}(S^o_\alpha, A_\alpha) = \min_{\pi\in\Pi}\{E^{\pi_\theta}_{S^o_\alpha}\Big\{\sum_{n=0}^{\infty}\gamma^n C\{\beta|\alpha, A_\alpha\}\}\Big\} \tag{10}$$

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 (10) 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

$$Q_{\pi_\theta}(S^o_\alpha,A_\alpha) = \min_{\pi_\theta\in\Pi} E^{\pi_\theta}_{S_\alpha}\Big\{C\{\beta|\alpha, A_\alpha\} + \gamma Q_{\pi_\theta}(S_{\alpha'},A_{\alpha'})\Big\}. \tag{11}$$

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 $\Big(\alpha, \beta, A_{\alpha'}, C\beta|\alpha, A_{\alpha'}\Big)$
  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 (8) and updated to the environment
  Update the timestep $t \to t+n$
  Until The maximum number of iterations $T_{Learning}$ is completed
end for
Output: δ^*_L, K_t

The optimal action, A^*_α, which is in line with (11), can be given by

$$A^*_\alpha = \arg\min_{\pi_\theta\in\Pi} E^{\pi_\theta}_{S_\alpha}\Big\{C\{\beta|\alpha, A_\alpha\} + \gamma Q_{\pi_\theta}(S_{\alpha'}, A_{\alpha'})\Big\}. \tag{12}$$

On one hand, the LSTM network captures any variation regularity from the observation vectors in line with the quantization intervals. On the other hand, the A2C layer employs an advantage function to delineate the predicted outcome for all future actions. 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.

As presented in Algorithm 1, The algorithm begins with initialization steps, where the LSTM states, critic parameters, and actor parameters are set up. Additionally, an empty buffer length of a $T_learning$ is created to store observations during the learning process. The initial time step is established to

TABLE I
P VALUES OF THE KEYS GENERATED BETWEEN THE VEHICLES USING NIST STATISTICAL SUITE

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

track the progress of the algorithm. In the event loop for the first iteration, the LSTM network enriches observable network states by incorporating hidden states into the environment. Subsequently, the actor selects an action to generate the secret key based on the selected quantization level. The agent then receives observations of the environment at the end of the time step, appending them to the buffer for future reference. In subsequent iterations of the event loop, the agent samples the next action of generating the key based on the reduction of cost, updating the environment accordingly. The buffer is maintained by removing the oldest observation and appending the latest one. The critic approximates the optimal action-value function, and based on this approximation, the optimal action is determined and updated in the environment. This process continues iteratively until the maximum number of iterations is reached. Finally, the output of the algorithm consists of the optimal action and the generated secret key. The combination of LSTM networks and A2C enables the iterative generation of secret keys based on observations of the dynamic environment, with the ultimate goal of unanimous key agreement.

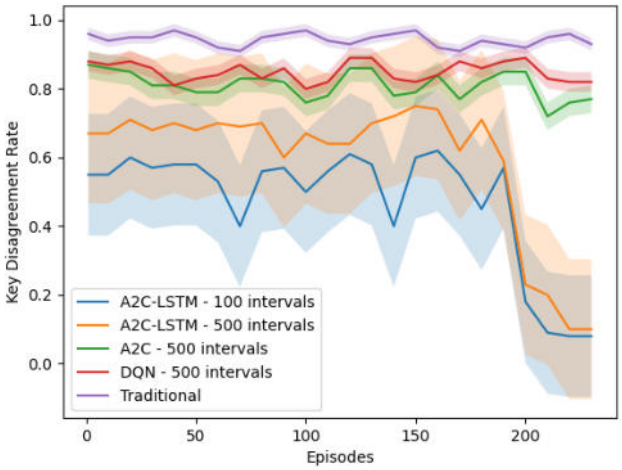

Fig. 3. Given 250 training episodes, we examine the KDR of the proposed combinatorial method in contrast to A2C and DQN.

V. PERFORMANCE ANALYSIS

The performance analysis of this work is done on two different fronts. Initially, we solidify the randomness amongst the secret bits generated between every vehicle. We employed the NIST standard randomness test suite [25] to validate the randomness of the keys generated through the DRL algorithm. This suite encompasses 16 tests designed by NIST specifically for assessing randomness. The P-value shows how likely it is that a well-adjusted random number generator could make a sequence less random than the input sequence being evaluated. To pass the test, all P-values must be higher than 0.01 [26]. Some tests in this suite needed an input bit stream longer than 106 bits. So, we chose tests that fit our dataset. In Table I, it's clear that the keys made by our method for each vehicle pair have P-values higher than 0.01, meaning the testing was successful. This means the secret bit streams we generated are entirely random with 99% confidence. Such strong randomness makes it tough for any eavesdropper to crack the keys.

Having defined the randomness of the keys, we now compare the key disagreement rate against DQN uses neural networks with an experience replay to train the actions with an extended state and action spaces and the traditional method that does not employ any learning to reduce the KDR. The DQN algorithm combines neural networks with the Q-learning algorithm to learn a policy for maximizing a reward signal. The neural network approximates the Q-function, which measures the expected future reward for taking a particular action in a given state. The Q-learning algorithm updates the neural network by using a replay buffer to store past experiences and then sampling from this buffer to update the network.

A2C networks are typically more data and sample-efficient than the DQN operating in smaller action spaces. In our case, the action was to choose the respective quantization level, so A2C performed slightly better than the DQN. This is because A2C is an on-policy algorithm that learns from the current policy, whereas DQN is an off-policy algorithm that learns from a target policy. On-policy algorithms tend to require fewer samples to converge to an optimal policy. A2C uses a value function to estimate the expected return of a state, which helps to reduce the variance in the policy gradient updates. Our proposed method captures the temporal variations and adjusts the respective actions more efficiently. Thus, minimal KDR achieves convergence (around 220 episodes), we can achieve almost 40% better results than both DQN and A2C.

The length of the secret key plays a critical role in determining the level of security offered by a cryptographic system. Generally, a longer key length increases the difficulty for an attacker to decrypt the message and compromise the

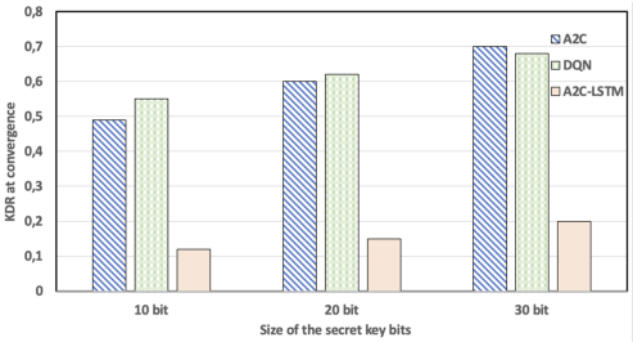

Fig. 4. Impact on the Key lengths on the KDR captured at the convergence of A2C-LSTM

encryption. However, there may be trade-offs between security and delay associated with longer keys. In this context, we compare the performance of DQN, A2C, and our proposed method, all trained to generate keys of identical lengths. As depicted in Figure 4, it becomes apparent that the KDR tends to increase with longer key lengths. With larger key sizes, achieving unanimity among numerous key bits becomes challenging. The results illustrated in the graph show the average KDR which is recorded at the event of convergence. It can be noted that our method consistently yields a KDR approximately 40% lower than that of both A2C and DQN.

VI. Conclusion

In this study, we introduced a method that combines A2C and LSTM to reduce disagreements in secret keys among vehicles, enhancing the security of vehicular communication. Our approach learns from changing network conditions, such as the quality of vehicles' communication channels, to create secure keys for transmission. Comparing our method with traditional techniques, A2C and DQN, we found it to outperform them in Key Disagreement Rate (KDR) across different key lengths. Additionally, the keys generated by our method successfully passed NIST's standard randomness tests, indicating increased difficulty in breaking them. As a future scope of this work, the impact of this method against adversarial and quantum attacks will be studied.

Acknowledgment

This work was supported by the CISTER Research Unit (UIDP/UIDB/04234/2020) and project ADANET (PTDC/EEICOM/3362/2021), financed by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology).

References

[1] Li, Kai, et al. "Confidentiality and timeliness of data dissemination in platoon-based vehicular cyber-physical systems." IEEE Network 35.4 (2021): 248-254.

[2] Kurunathan, Harrison, et al. "Towards Safe Cooperative Autonomous Platoon systems using COTS Equipment." 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2023.

[3] Rafael, Rui, et al. "An RT-cloud solution towards security in Vehicular platooning systems." 2023 RT-Cloud workshop, 36th Euro Micro Conference on Real-Time Systems, 2023.

[4] Li, Kai, et al. "Optimal rate-adaptive data dissemination in vehicular platoons." IEEE Transactions on Intelligent Transportation Systems 21.10 (2019): 4241-4251.

[5] Li, Kai, et al. "Design and implementation of secret key agreement for platoon-based vehicular cyber-physical systems." ACM Transactions on Cyber-Physical Systems 4.2 (2019): 1-20.

[6] Xu, Peng, et al. "Group secret key generation in wireless networks: Algorithms and rate optimization." IEEE Transactions on Information Forensics and Security 11.8 (2016): 1831-1846.

[7] Jain, Prateek, et al. "Non-convex optimization for machine learning." Foundations and Trends® in Machine Learning 10.3-4 (2017): 142-363.

[8] Wang, Dong, et al. "A survey of optimization approaches for wireless physical layer security." IEEE Communications Surveys and Tutorials 21.2 (2018): 1878-1911.

[9] Vavasis, Stephen A. "Complexity issues in global optimization: a survey." Handbook of global optimization (1995): 27-41.

[10] Yang, Jie, et al. "RIS aided Physical Layer Key Generation of IoT network in Static Environments (March 2023)." IEEE Sensors Journal (2023).

[11] Shawky, Mahmoud A., et al. "Adaptive chaotic map-based key extraction for efficient cross-layer authentication in VANETs." Vehicular Communications 39 (2023): 100547.

[12] Shawky, Mahmoud A., et al. "An efficient cross-layer authentication scheme for secure communication in vehicular ad-hoc networks." IEEE Transactions on Vehicular Technology (2023).

[13] Han, Biao, et al. "LoRa-based physical layer key generation for secure V2V/V2I communications." Sensors 20.3 (2020): 682.

[14] Liu, Yiliang, et al. "Physical layer security assisted computation offloading in intelligently connected vehicle networks." IEEE Transactions on Wireless Communications 20.6 (2021): 3555-3570.

[15] Yang, Edwin, and Song Fang. "GPSKey: GPS-based secret key establishment for intra-vehicle environment." Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2022. 2022.

[16] Xu, Peng, et al. "Group secret key generation in wireless networks: Algorithms and rate optimization." IEEE Transactions on Information Forensics and Security 11.8 (2016): 1831-1846.

[17] Xu, Peng, et al. "Group secret key generation in wireless networks: Algorithms and rate optimization." IEEE Transactions on Information Forensics and Security 11.8 (2016): 1831-1846.

[18] Kurunathan, Harrison, et al. "Machine learning-aided operations and communications of unmanned aerial vehicles: A contemporary survey." IEEE Communications Surveys and Tutorials (2023).

[19] Olowononi, Felix O., Danda B. Rawat, and Chunmei Liu. "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 23.1 (2020): 524-552.

[20] Li, Kai, Wei Ni, and Falko Dressler. "LSTM-characterized deep reinforcement learning for continuous flight control and resource allocation in UAV-assisted sensor network." IEEE Internet of Things Journal 9.6 (2021): 4179-4189.

[21] 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.

[22] Li, Kai, et al. "Secret key agreement for data dissemination in vehicular platoons." IEEE Transactions on Vehicular Technology 68.9 (2019): 9060-9073.

[23] Li, Kai, et al. "Cooperative secret key generation for platoon-based vehicular communications." ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019.

[24] Li, Kai, et al. "Cooperative key generation for data dissemination in cyber-physical systems." 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS). IEEE, 2018.

[25] 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.

[26] Georgescu, Carmina, et al. "A view on NIST randomness tests (in) dependence." 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, 2017.