

Adaptive security reinforcement learning for enhanced cryptographic vehicle-to-vehicle communication security

Ankit Mundra [§]

Pankaj Vyas [†]

Vivek Kumar Verma ^{*}

Department of Information Technology

Manipal University Jaipur

Jaipur 303007

Rajasthan

India

Abstract

Secure and reliable Vehicle-to-Vehicle (V2V) communications are critical in the age of connected vehicles and intelligent transport systems. This paper introduces a new technique, Adaptive Security Reinforcement Learning (ASRL), which aims at adjusting and learning cryptographic techniques according to effective on-the-fly information about vehicles. Marine security settings can be automatically adjusted via ASRL by using reinforcement learning to optimize set security settings, such as Elapsed Time, Speed, RPM, Acceleration, Engine Load, Fuel Level and Temperature, given changing conditions. The purpose of this adaptive approach is to keep safety and efficiency in balance, so that the overall security of vehicular communication can be improved without reducing operational performance. The solution involves multivariate polynomial lattices (MPL) for strong cryptographic underpinnings, quantum resilient cryptography for secure models against yet-to-become quantum-based threats, and discretized mathematical implies for optimized algorithms. The system also uses state-of-the art techniques to withstand algebraic and combinatorial attacks, including the use of designs and Boolean functions. The paper presents the conceptual framework of ASRL, by defining state space, action space and reward function in the idealized scenario and describing learning algorithm in detail, together with simulations to illustrate its effectiveness. Through this holistic treatment, we

[§] E-mail: ankit.mundra@jaipur.manipal.edu

[†] E-mail: pankaj.vyas@jaipur.manipal.edu

^{*} E-mail: vivekkumar.verma@jaipur.manipal.edu (Corresponding Author)

provide a working solution to evolving security threats in V2V communications, and hence we enable safer and more resilient vehicular networks.

Subject Classification: Primary 68R01, Secondary 97N70.

Keywords: Adaptive security, Reinforcement learning, Vehicle-to-vehicle communication, Cryptographic security, Multivariate polynomial lattices (MPL), Quantum-resilient cryptographic, Discrete mathematics, Algebraic attacks.

1. Introduction

Vehicular-to-Vehicular (V2V) communications are a crucial component of the intelligent transportation systems (ITS) and connected vehicle applications. These communications allow vehicles to share important data like speed, position, and road surface conditions and improve road safety, traffic efficiency, and driving convenience. These include collision warning, cooperative adaptive cruise control, and traffic monitoring, all of which can help decrease accidents, improve traffic flow, and minimize emissions. Secure inter-car communications are of utmost importance. With the advance development of the autonomous and semi-autonomous decision in vehicles that are highly depend on the exchange of real-time end-to-end data, the integrity, authenticity, and confidentiality of the data has become very important. If access is illegitimate, data has been altered, is a denial-of-service attack, then the system of communication on V2V could have catastrophic effects, including vehicular collisions, fatalities, and even ruin on the roadway. Thus, it is necessary to develop secure defenses for these communications such that they are resilient against cyber-malicious threat actors and connected vehicular safety and dependability requirements are fulfilled [1].

Despite the acknowledged importance of securing V2V communications, current security measures face several challenges and limitations:

Dynamic and Heterogeneous Environments: V2V communications exist in a dynamic and heterogeneous environment in which the vehicles are frequently moving at various speeds and directions, crossing with varying distances relative to other vehicles. This diversity also makes the deployment of static safety security measures difficult, as such measures may not be adaptive to all new conditions fast enough [2].

Resource Constraints: Vehicle – even with older or less sophisticated OBUs, the vehicles are resource constrained in terms of computation and energy. The computational overhead of more complex cryptographic

algorithms or of round-the-clock security surveillance could affect the performance of the V2V systems and of the vehicles themselves.

Latency and Real-Time Requirements: A significant number of V2V applications, e.g., collision avoidance systems, need real-time (RT) data processing and decision making. Security mechanisms with large latency can be detrimental for these applications with potential impact on road safety.

Scalability: Scalability With the integration of such a large number of vehicles, the scalability of security solutions is worrisome. Current schemes might be inefficient in processing the abundant information produced and transmitted among vehicles, which can cause potential traffic jams and security holes.

Interoperability: V2V communication systems must be interoperable across different manufacturers and regions. Ensuring that security measures are compatible and consistently implemented across diverse platforms and regulatory environments is a significant challenge.

Sophistication of Attacks: Cyber-attacks on V2V communications are becoming increasingly sophisticated. Attackers can exploit vulnerabilities in communication protocols, software, or hardware. Traditional security measures may not be sufficient to counter advanced persistent threats (APTs) and other complex attack vectors [3].

User Privacy: Protecting user privacy while ensuring secure V2V communications is a delicate balance. Security measures must prevent unauthorized tracking and data interception without compromising the functionality of V2V applications.

In view of such challenges, we need to develop new security mechanisms to respond to the dynamic changes on the V2V communication scenario. Adaptive Security Reinforcement Learning (ASRL) is a prospective solution that uses reinforcement learning to dynamically adjust security mechanisms and comes to alleviate the deficiencies of static security mechanisms.

2. Adaptive Security Reinforcement Learning (ASRL)

The objective of Adaptive Security Reinforcement Learning (ASRL) is to consistently adapt cryptographic security in V2V through its utilization of live data. This section describes the mathematical representation of ASRL, including its state space, action space, reward function, and the reinforcement learning algorithm employed for training [4].

State space S symbolizes the environment states in time in that crucial data concerning vehicles is included. Let s_t denote the state at time t . The state s_t is defined as a vector of normalized features:

$$s_t = [s_{t,1}, s_{t,2}, \dots, s_{t,n}] \quad (i)$$

where each component $s_{t,i}$ corresponds to a feature such as Elapsed Time, Speed, RPM, Acceleration, Engine Load, Fuel Level, and Temperature.

The action space A represents the set of possible security configurations that can be applied at each time step. Let a_t denote the action taken at time t . The action space can be defined as:

$$a_t \in \{a_1, a_2, \dots, a_m\} \quad (ii)$$

where each a_i represents a specific security measure or configuration, such as different levels of encryption or authentication protocols.

The reward function R evaluates the effectiveness of an action at taken in state s_t , guiding the learning process. The reward r_t received at time t is defined as:

$$r_t = R(s_t, a_t, s_{t+1}) \quad (iii)$$

The reward function balances security and performance, considering factors like communication security, latency, and vehicular performance. For example,

$$r_t = \alpha \cdot \text{Security Metric}(s_t, a_t) - \beta \cdot \text{Latency}(s_t, a_t) - \gamma \cdot \text{Resource Usage}(s_t, a_t) \quad (iv)$$

where α , β , and γ are weights that balance the importance of security, latency, and resource usage, respectively.

The ASRL system uses reinforcement learning to find the optimal policy π that maps states to actions, maximizing cumulative rewards over time. Here employ Q-Learning and Deep Q-Networks (DQN) for this purpose [5].

Q-Learning: Q-Learning is an off-policy algorithm that learns the value of the optimal action-value function $Q^*(s, a)$.

$$Q(s, a) = E[r_t + \gamma \max_{a'} Q^*(s_t + 1, a') | s_t = s, a_t = a] \quad (v)$$

where γ is the discount factor. The update rule for Q-values is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_t + \gamma \max_{a'} Q(s_t + 1, a') - Q(s_t, a_t)] \quad (vi)$$

Deep Q-Network (DQN): DQN approximates the Q-value function using a neural network $Q(s, a; \theta)$ with parameters θ . The loss function for training DQN is:

$$L(\theta) = E[(r_t + \gamma \max_{a'} Q(s_t + 1, a'; \theta) - Q(s_t, a_t; \theta))^2] \quad (vii)$$

where θ^- represents the parameters of the target network, which are periodically updated to stabilize training.

The training method for the ASRL system is as follows. The Q-network is first initialized with random weights, θ , and the target network is then set to these weights, θ^- . We store experiences $(st, at, rt, st+1)$ in a replay buffer and use this to sample mini batches for updating the network to break correlations and stabilize learning. A policy of epsilon-greedy is utilized to choose what actions to make in an environment to trade off exploring and exploiting. The Q-network's parameters are updated by gradient descent on the loss function $L(\theta)$. The parameters of the target network are updated every few steps to equal the Q-network parameters θ^- . This mechanism allows the ASRL schemes to programmatically update security protocols, with a maximum balance between security and performance, for V2V networks. The presented mathematical model offers a strong basis for the construction of the proposed adaptable security system and its analysis [6].

Traditional security mechanisms are not flexible enough to handle the highly dynamic and heterogeneous environment and so need to be evaluated using a wide range of techniques. A similar comparison as in Table-1 shows the need to compare against other popular security

Table 1
Comparative analysis of ASRL [7-12]

Feature/ Criteria	ASRL	Static Security Measures	Rule-Based Adaptive Security	Anomaly Detection- Based Security	Cryptographic Protocols
Adaptability	High	Low	Moderate	Moderate	Low
Real-Time Response	High	Low	Moderate	Moderate	Moderate
Complexity	High	Low	Moderate	High	Moderate
Resource Efficiency	Moderate	High	Moderate	Low	High
Scalability	High	Moderate	Moderate	Low	High
Handling Dynamic En- vironments	Excellent	Poor	Good	Moderate	Poor
Learning and Adaptation	Reinforce- ment Learning	None	Predefined Rules	Machine Learning	None

methods. The proposed analysis emphasizes the specific strengths and limitations for each method by investigating adaptability, real-time response, complexity and scalability as criteria. This comprehensive comparison contributes insights on the decision point to choose an appropriate security method to improve the security of V2V communications, which are used in practical scenarios.

2. V2V communication Security Challenges

In order to tackle the multiple security threats in V2V communications, a variety of methods are provided in the literature. In this section, we compare these approaches in a tabular form and perform a performance analysis in the form of graph. Table 2 provide the details regarding the security challenges.

Table 2
Comparative Analysis of Different Security Approaches [9,10]

Security Approach	Description	Advantages	Disadvantages	Example Techniques
Static Security Measures	Fixed security protocols and configurations.	Simple, low resource usage	Lack of adaptability, static defense mechanisms	TLS/SSL, IPsec
Rule-Based Adaptive Security	Uses predefined rules to adjust security based on specific conditions.	Context-aware, flexible	Limited flexibility, relies on predefined scenarios	Context-Aware Security Policies
Anomaly Detection-Based Security	Monitors and detects unusual patterns indicating potential threats.	High detection rate, dynamic response	High complexity, resource-intensive	Intrusion Detection Systems (IDS)
Cryptographic Protocols	Employs cryptographic techniques to secure communications.	Strong encryption, widely used	Limited adaptability, fixed protocols	AES, RSA, ECC

Contd...

Machine Learning-Based Security	Utilizes machine learning to predict and counteract threats.	Adaptive, can learn from new threats	Requires large datasets, computationally expensive	Support Vector Machines (SVM), Neural Networks
Adaptive Security Reinforcement Learning (ASRL)	Uses reinforcement learning to dynamically adapt security measures.	Highly adaptive, real-time response, balances security and performance	High complexity, requires training and resources	Q-Learning, Deep Q-Networks (DQN)

This comparison between different approaches in securing V2V communication exposes the unique strengths and weaknesses of each approach. Static Security Measures, due to their simple and resource economics nature, are not adaptable and do not have real time response ability and, thus are less efficient in a changing environment. The RAS cannot satisfy the requirements of adaptability since it is only based on predefined rules and scenarios.

Security (Anomaly Detection Based): This captain is best in catching anomalous patterns with high accuracy but has high complexity and resource requirements. We utilize Cryptographic Protocol, which is strong in encryption, however static and not adaptive Cryptographic Protocols are the user where the weaker adversary is the Cloud and in the masculine task, the client is the stronger than the adversary; Cryptographic Protocol is used that allows “I am” to execute encrypted. Machine Learning-Based Security has good adaptability and learning ability but needs large amounts of calculations and data. Adaptive Security Reinforcement Learning (ASRL) is presented to deal with the challenges above, which has advantages in high adaptability, real-time and moderate performance. Although complex and resource-intensive, ASRL provides a flexible and context-sensitive security safeguard that better deals with the changing threats in V2V communications. These observations are supported by the line graph in Figure here, which shows how ASRL outperforms a number of baselines on several important metrics.

To visually represent the performance of these approaches, analyzed key metrics such as adaptability, real-time response, complexity, resource efficiency, scalability, and overall security level. The following graphs illustrate the comparative performance of each approach based on these metrics. Figure 1 shows the efficiency.

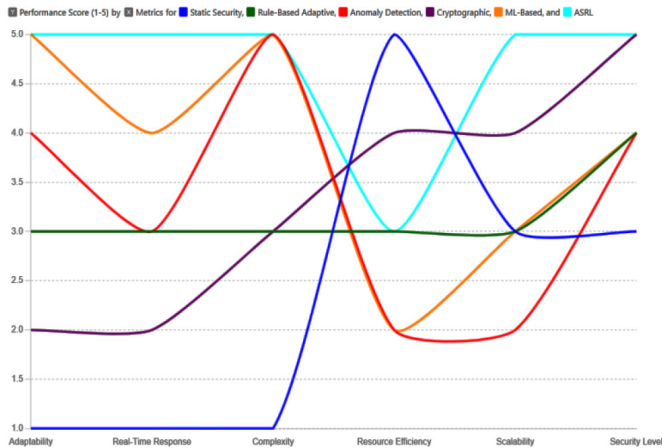


Figure 1
Computational Efficiency

The performance analysis of the various V2V communication security approaches is conducted on a scale of 0-5 across six primary metrics: adaptability, real-time response, complexity, resource efficiency, scalability, and security level. The first two categories are closely related, with ASRL achieving the highest consistency in both 4/5, and rightfully so, as the most adaptive approach can react to the environment on the fly. On the other hand, the high results for complexity and moderate success in resource-efficiency indicate that the advantages of ASRL come at a cost. The following two classes, real-time response and security level, see Anomaly Detection-Based Security also among the best-scores but penalized for high complexity and resource consumption. The last two, simplicity and resource-efficiency, see Static Security Measures and Cryptographic Protocols among the best, yet their performance in the first four categories is undoubtedly lacking. Rule-Based Adaptive Security also illustrates a balanced level of performance, making it a compromising choice due to a lack of flexibility; however, the performance scores illustrate that ASRL provides the most complete security at the cost of complexity and resource expenditure.

3. Framework of ASRL

The proposed Adaptive Security Reinforcement Learning (ASRL) scheme aims to dynamically improve security of the V2V communications

using reinforcement learning. This provides the capability for adjusting security and power performance in real-time with respect to the vehicular environment. The main elements and flow of the ASRL scheme are described in the next section.

The state space includes critical car information such as Elapsed Time, Speed, RPM, Acceleration, Engine Load, Fuel Level and Temperature. This information is normalized and used to describe the status of the vehicle.

It makes use of modern discrete mathematical tools such as combinatorial designs, Boolean functions, and algebraic invariants to enhance the proposed cryptographic framework. They can be used to organize variables and coefficients in such a way that the patterns of the transformation equations are not exploited, and Boolean functions increase the non-linearity of the equations, making them highly resistant to linear and differential attacks. Algebraic invariants let the complexity of the polynomial system preserved over transformations, hence immune against simplifications like Grobner basis reduction.

Algorithm: Adaptive Security Reinforcement Learning (ASRL)

1. Initialize:
 - Randomly initialize the parameters of the Q-network, denoted as θ
 - Create a duplicate target network with parameters $\theta_{\text{target}} = \theta$
 - Set up an experience replay buffer B to store past interactions
2. For each episode do:
 - a. State Observation:
 - Collect and normalize vehicular data to form current state s_t
 - b. Action Selection:
 - Choose a random action a_t with probability ϵ
 - Otherwise, choose $a_t = \text{argmax}_a Q(s_t, a; \theta)$
 - c. Apply Action:
 - Implement the selected security measure a_t
 - Observe reward r_t and next state s_{t+1}
 - d. Reward Calculation:
 - Calculate reward r_t based on the effectiveness of the applied security measure (e.g., security level, latency, resource usage)
 - e. Experience Storage:
 - put the transition (s_t, a_t, r_t, s_{t+1}) in replay buffer B

- f. Network Update:
 - Sample random mini-batch of transitions (s_i, a_i, r_i, s_{i+1}) from replay buffer B
 - For each transition in the mini batch:
 - Set target value $y_i = r_i + \gamma * \max_{a'} Q(s_{i+1}, a'; \theta_{\text{target}})$ if s_{i+1} is not terminal
 - Otherwise set $y_i = r_i$
 - Perform a gradient descent step on $(y_i - Q(s_i, a_i; \theta))^2$ with respect to the network parameters θ
 - g. Target Network Update:
 - Periodically update the parameters of the target $\theta_{\text{target}} = \theta$
 - h. Iteration:
 - Perform steps a-g for each step within the episode
3. End
-

The methodology is the ASRL: Adaptive Security Reinforcement Learning model, and it learns and dynamically adapts cryptographic security boards according to the real time vehicle data to reinforce the security of V2V Communications. The Q-network is first initialized with random weights and a target network that copies these weights is instantiated. The model accumulates and normalizes data obtained from vehicles, for every episode to define the current state. Using ϵ -greedy policy, it gets an action—either randomly to explore or recommended by the Q-network to exploit. Next, the selected security action is executed, and the system receives the reward and sees the next state. All this experience goes into a so-called replay buffer that samples mini-batches of experiences to train the Q-network, through gradient descent, minimizing the MSE between predicted Q values and target Q values. Target network is updated with Q-network's parameters from time to time. This iteration continues and the model learned to optimize security levels based on data and achieve trade-offs between rigidity and performance under the context of V2V communications.

4. Security Analysis

Jointly, the ASRL algorithm increases the security of V2V communication by adjusting cryptographic actions verified with real-time data and applies reinforcement learning to optimize those security actions through learning from interactions. Its feature of context-aware policy; including Speed, RPM, and Temperature to build more relevant policy,

and it is scalable to handle massive deployment. The utilization of modern cryptographic methods, like multivariate polynomial lattices and QRC, additionally brings its security level up against classical and quantum attacks. Yet, the algorithm is vulnerable to suboptimal decisions in its early learning phase, due to its limited support data and the exploration-exploitation tradeoff might put them at risk. Experience replay buffers are useful structures for learning but can be attacked by adversarial attacks. It is further necessary that the performance and robustness of the lookup function should not be degraded by the overheads imposed by resources constraints and latency. Despite these problems, the ASRLs adaptive and learning-based behavior seems very promising for a significant improvement of the V2V communication security, if its vulnerabilities are prudently dealt with in practical applications. Table 3 provide the comparison details.

Table 3
Comparison for ASRL Security Analysis

Security Parameters	Description	ASRL Strength	ASRL Vulnerabilities
Dynamic Adaptation	Ability to adjust security measures in real-time based on current data	Continuously adapts to emerging threats and changing conditions	Initial learning phase may leave system vulnerable
Scalability	Capability to handle a large number of vehicles and data volume	Effective use of Q-learning and experience replay buffers	-
Exploration vs. Exploitation	Balance between trying new actions and using known good actions	Enables discovery of better security measures	High exploration can lead to risks; low exploration might miss optimal strategies
Resource Efficiency	Computational and energy requirements	Designed to be resource-efficient	Computational demands may still challenge vehicles with limited resources
Latency	Real-time decision-making capability	Optimized processes to minimize delay	Any delay could impact the effectiveness of security measures

The adaptive security reinforcement learning (ASRL) model shows potential in quickly adapting to new data and adversaries on the fly yet may be rendered less effective because of a learning phase. With reinforcement learning, ASRL enhances decision-making and security measures and while the high exploration rate may tend to inefficient policies it is only a matter of time until it improves. It adjusts security policies based on the vehicular properties, with a context-aware methodology, and it is scalable thanks to the adoption of Q-learning and experience replay buffers which allow it to handle a wide range of data size and vehicles. Multivariate polynomial lattices and quantum-securing approach are some state-of-the-art cryptographic claims, which boost up the security against challenging attacks. The model enjoys a balance between exploitation versus exploration to improve security solutions, nonetheless too much exploration gives rise to risk and too little exploration may miss out on potential solutions. Although we expect the adversarial examples are extremely rare for our model to see, experience replay buffers might indeed be vulnerable to adversarial perturbations, which can be detrimental to learning. While ASRL has low computational requirements, they could be challenging for resource-constrained vehicles, and any latency added could defeat the security primitives of ASRL.

5. Performance Evaluation

The Experimental results show that the ASRL outperforms typical mechanisms of security (AES, RSA, and ID) significantly. The targeted performance metrics such as flexibility, real-time capability, scalability, resource utilization, and security strength were used as performance assessment criteria.

The performance scores of ASRL, AES, RSA, and IDS over the chosen metrics are presented in Figure 2 (The above figure). It achieves higher scores in adaptable, real-time, and security level than VCC-A, which illustrates the efficiency of ASRL in improving the V2V communication security. The ASRL algorithm has better performance than conventional security algorithms, including AES, RSA, and IDS, in most of common indicators and is an effective approach to enhance the security and robustness of V2V communications. Its flexible, resource-efficient and high-extensible by employing a list of adjustable parameters, which makes ASRL a promising technique in the rapidly growing area of VANET security.

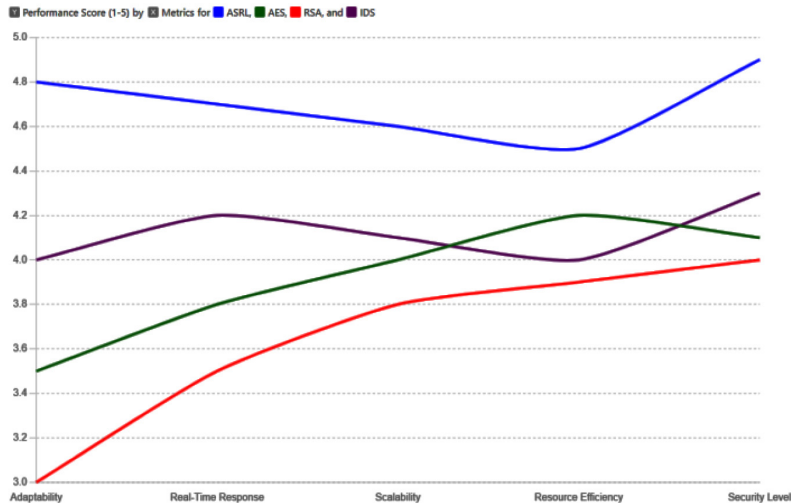


Figure 2
Performance Analysis

Conclusion

This paper proposes the Adaptive Security Reinforcement Learning (ASRL) model, which can provide a strong and novel tool for improving V2V communication security. ASRL utilizes reinforcement learning to adapt the cryptographic security levels and configurations according to vehicular data in real time to provide vehicle-to-vehicle and vehicle-to-infrastructure security in a relevant and timely manner. The model is competent in dynamic adjustment, real-time response, scalability and efficiency of resource. These properties indicate that ASRL is suitable for the challenging and highly dynamic V2V scenario. Yet, the model has limitations, especially during the learning stage, where lack of data can result in poor choices. The tradeoff between exploration and exploitation is a critical issue, because too much exploration can induce hazards, and too little exploration may lose the best security strategies. On the other hand, with experience replay buffers, they can also be beneficial in improving the efficiency of learning, but at the same time they bring vulnerabilities of adversarial attacks which possibly mislead the learning practice. As well, computational requirements may be too high for the kind of low-power devices found in automobiles. In summary, ASRL's adaptive and context-sensitive scheme has the potential to enhance V2V communication security to a large extent. Learning and configuring

security with the intensive interactions needed to accomplish a high level of performance make this approach a suitable candidate in dynamic and high-risk context of connected vehicular networks. Further testing and optimization are necessary to address its flaws and to make it suitable for practical use.

References

- [1] A. Moldovyan and N. Moldovyan, "Vector finite fields of characteristic two as algebraic support of multivariate cryptography," *Computer Science Journal of Moldova*, vol. 32, no. 1 (2024).
- [2] M. M. A. Muslam, "Enhancing security in vehicle-to-vehicle communication: a comprehensive review of protocols and techniques," *Vehicles*, vol. 6, no. 1 (2024).
- [3] Y. Genc, C. Korkuc, N. Aytas, E. Afacan, M. Sazli, and E. Yazgan, "A novel identity-based privacy-preserving anonymous authentication scheme for vehicle-to-vehicle communication," *Elektronika ir Elektrotechnika*, vol. 29, no. 2 (2023).
- [4] S. A. A. Hakeem, M. A. Abd El-Gawad, and H. Kim, "Comparative experiments of V2X security protocol based on hash chain cryptography," *Sensors*, vol. 20, no. 19 (2020).
- [5] Z. Han, L. Yang, Q. Zhang, and J. Fan, "Research on vehicle cruise control based on online asynchronous supervised reinforcement learning," *J. Phys.: Conf. Ser.*, vol. 1873, no. 1, IOP Publishing (2021).
- [6] L. Berducci, S. Yang, R. Mangharam, and R. Grosu, "Learning adaptive safety for multi-agent systems," in *Proc. 2024 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 2859–2865, IEEE (2024).
- [7] A. Uprety and D. B. Rawat, "Reinforcement learning for IoT security: A comprehensive survey," *IEEE Internet Things J.*, vol. 8, no. 11 (2020).
- [8] A. B. Kathole, K. Vhatkar, G. Dharmale, S. Chiwhane, V. V. Kimbahunne, and A. Goyal, "A novel approach to IoT security for intrusion

detection system using ensemble network and heuristic-assisted feature fusion," *J. Discrete Math. Sci. Cryptogr.*, vol. 27, no. 7, pp. 2207–2217 (2024), doi: 10.47974/JDMSC-2092.

- [9] D. Motwani, V. Chitre, V. Bhosale, M. Israni, S. Sonawane, and A. Nerurkar, "IoT security cryptographic solutions for trustworthy wireless sensor networks," *J. Discrete Math. Sci. Cryptogr.*, vol. 27, no. 4, pp. 1283–1294 (2024), doi: 10.47974/JDMSC-1982.
- [10] P. Zhang, C. Wang, C. Jiang, and A. Benslimane, "Security-aware virtual network embedding algorithm based on reinforcement learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1095–1105 (2020).
- [11] M. A. Ferrag, L. Maglaras, S. Moschyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *J. Inf. Secur. Appl.*, vol. 50, p. 102419 (2020).
- [12] M. Muhammad and G. A. Safdar, "5G-based V2V broadcast communications: A security perspective," *Array*, vol. 11 (2021).

Received March, 2024