Enhancing privacy in VANETs through homomorphic encryption in machine learning applications

Yulliwas Ameur, Samia Bouzefrane

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 238 (2024) 151–158

Procedia Computer Science

www.elsevier.com/locate/procedia

The 15th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 23-25, 2024, Hasselt, Belgium

Enhancing privacy in VANETs through homomorphic encryption in machine learning applications

Yulliwas Ameur*[a], Samia Bouzefrane[a]

[a]*CEDRIC Lab, Conservatoire National des Arts et Métiers (Cnam), Paris, France*

Abstract

This paper presents a novel framework for enhancing privacy in Vehicular Ad Hoc Networks (VANETs) by integrating homomorphic encryption with machine learning applications. VANETs, essential for Intelligent Transport Systems (ITS), face significant challenges in privacy and security due to their highly dynamic and heterogeneous nature. Our framework addresses these challenges by employing a simplified but effective machine learning algorithm, the K-nearest neighbors (KNN), to ensure the security and privacy of the network. The flexibility of the framework allows for the incorporation of other machine learning algorithms, enhancing its adaptability and efficiency in various VANET scenarios.

Key to this framework is the use of homomorphic encryption (HE), a cryptographic technique that enables computations on encrypted data without the need for decryption. This feature preserves data confidentiality and allows for secure third-party computations. Our paper discusses the evolution and types of homomorphic encryption, emphasizing the importance of Fully Homomorphic Encryption (FHE) for its ability to evaluate complex polynomial functions.

The paper also highlights the different domains of cybersecurity concerns in VANETs, including in-vehicle systems, ad-hoc and infrastructure networks, and data analysis. The proposed framework aims to mitigate these vulnerabilities, particularly focusing on preventing common attacks like DoS and location tracking.

A significant advantage of our approach is its general nature, making it applicable to various privacy issues in VANETs. We propose the potential integration of homomorphic encryption with other privacy-preserving techniques, such as differential privacy or secure multi-party computation, to enhance computation times while ensuring robust privacy protection.

Keywords: Privacy Machine Learning, PPML, Vanet, Homomorphic encryption .

* Corresponding author. Tel.
E-mail address: yulliwas.ameur@lecnam.net

10.1016/j.procs.2024.06.010

1. Introduction

Vehicular Ad Hoc Networks (VANETs) are pivotal for the advancement of Intelligent Transport Systems (ITS), facilitating not only vehicle-to-vehicle (V2V) communication but also extending to vehicle-to-infrastructure (V2I), vehicle-to-cloud (V2C), and vehicle-to-pedestrian (V2P) interactions. These communications are essential for enhancing road safety, traffic efficiency, and the overall driving experience. However, the expansion of VANETs introduces significant privacy and security challenges, primarily due to their highly dynamic and heterogeneous nature. The dense urban settings and high mobility of vehicles, combined with the inherent openness of wireless communication, create a complex network environment where scalability issues and privacy concerns become increasingly difficult to manage[6].

Amidst these challenges, the privacy of communication and data within VANETs emerges as a critical issue. Ensuring the confidentiality and integrity of data is paramount for the trustworthiness and reliability of ITS applications. Traditional security measures often fall short in addressing the unique challenges posed by VANETs, necessitating innovative approaches that can adapt to their dynamic environment.

This paper introduces a novel framework aimed at enhancing the privacy and security of VANETs by leveraging the synergy between homomorphic encryption (HE) and machine learning (ML) algorithms as in [5]. Homomorphic encryption offers a groundbreaking approach to secure data processing, allowing for computations on encrypted data without the need for decryption. This capability not only preserves data confidentiality but also facilitates secure third-party computations, a crucial feature for the decentralized nature of VANETs.

Our proposed framework utilizes the K-nearest neighbors (KNN) algorithm as a foundational ML technique for identifying and mitigating security threats within VANETs. While KNN is highlighted for its simplicity and effectiveness, our framework's design is inherently flexible, allowing for the integration of more sophisticated ML algorithms, such as decision trees, random forests, and neural networks. This adaptability enhances the framework's capacity to address a wide range of security and privacy challenges in VANETs.

Furthermore, we delve into the evolution and types of homomorphic encryption, with a particular focus on Fully Homomorphic Encryption (FHE). FHE's ability to evaluate complex polynomial functions on encrypted data makes it an invaluable tool for preserving privacy in VANETs. By combining FHE with ML, our framework addresses key cybersecurity concerns, including the protection against common attacks like Denial of Service (DoS)[1] and unauthorized location tracking[8].

In summary, our paper presents a comprehensive framework that not only addresses the immediate privacy and security challenges in VANETs but also lays the groundwork for future research in integrating advanced cryptographic techniques with machine learning for enhanced ITS applications.

2. Background and overview

2.1. VANET

Vehicular Ad Hoc Networks (VANETs) are a specialized subset of Mobile Ad Hoc Networks (MANETs) tailored to enable communication among moving vehicles and between vehicles and roadside units (RSUs). Their primary aim is to enhance road safety and optimize traffic management by facilitating the exchange of vital information such as traffic conditions, safety alerts, and accident notifications in real-time. Unlike traditional MANETs, VANETs feature high mobility dynamics with rapidly changing network topologies, requiring robust and efficient communication protocols to ensure reliable connectivity. The unique characteristics of VANETs, such as high mobility, dynamic network topology, and real-time constraints, pose distinct challenges in terms of security, privacy, and communication efficiency. To address these challenges, VANETs leverage advanced cryptography and network security technologies, as well as sophisticated communication architectures that incorporate vehicles as mobile nodes and RSUs to provide extensive and reliable network coverage. In summary, VANETs play a crucial role in the development of Intelligent Transportation Systems (ITS) by improving traffic efficiency and enhancing road safety for all users.

Fig. 1: Vehicular Ad Hoc Networks (VANETs): Simplified Communication Diagram

2.2. Homomorphic Encryption

Homomorphic encryption [4] is an encryption technique that supports a particular time-consuming evaluation algorithm. This algorithm allows certain types of operations to be carried out on the ciphertext without requiring access to a secret key. Moreover, this algorithm generates an encrypted result in which the decryption matches the result of the computation on the plaintext. For instance, consider two plaintexts x and y; we want to compute $3xy + x$ without leaking x and y. Thus, we first use the homomorphic encryption algorithm Enc to encrypt x and y. $\mathrm{Enc}(x)$ and $\mathrm{Enc}(y)$ are ciphertexts of x and y, respectively. Then, we compute $\mathrm{Enc}(3xy + x) = 3 \times \mathrm{Enc}(x) \times \mathrm{Enc}(y) + \mathrm{Enc}(x)$, where $3 \times \mathrm{Enc}(x) \times \mathrm{Enc}(y) + \mathrm{Enc}(x)$ denotes homomorphic operations. The final ciphertext is $\mathrm{Enc}(3xy + x)$, and the plaintext is $3xy + x$. To gain a better understanding of homomorphic encryption, we present related definitions and the current state of homomorphic encryption in this section.

Definition 1 (Homomorphic Encryption. A homomorphic encryption scheme $HE = (\mathrm{KeyGen}, \mathrm{Enc}, \mathrm{Dec}, \mathrm{Eval})$ consists of four probabilistic polynomial algorithms. The detailed definition of homomorphic encryption is described as follows:

- $HE.\mathrm{KeyGen}(1^\lambda)$: The security parameter λ is taken as input. Output parameters include a public key pk, a secret key sk and an evaluation key evk, namely $(pk, sk, evk) \leftarrow HE.\mathrm{KeyGen}(1^\lambda)$.
- $HE.\mathrm{Enc}(pk, m)$: The public key pk and a plaintext m are taken as inputs. Then, the ciphertext c is output, namely $c \leftarrow HE.\mathrm{Enc}(pk, m)$.
- $HE.\mathrm{Dec}(sk, c)$: The secret key sk and the ciphertext c are taken as inputs. The decryption result m^* is output, namely $m^* \leftarrow HE.\mathrm{Dec}(sk, c)$.
- $HE.\mathrm{Eval}(evk, f, c_0, \ldots, c_{l-1})$: Input parameters include the evaluation key evk, a function f and ciphertexts $c_0, \ldots, c_{l-1}$, where the plaintext of c_i is m_i, $i = 0, \ldots, l-1$, l is the number of ciphertexts. Then, the final ciphertext c_f is output, namely $c_f \leftarrow HE.\mathrm{Eval}(evk, f, c_0, \ldots, c_{l-1})$, where $HE.\mathrm{Dec}(sk, c_f) = f(m_0, \ldots, m_{l-1})$, f is an operational circuit over the plaintext space.

3. Framework for Enhancing Privacy in VANETs through Homomorphic Encryption in Machine Learning Applications

Vehicular Ad Hoc Networks (VANETs), crucial for advancing Intelligent Transport Systems (ITS), face significant security and privacy challenges. These networks' dynamic and heterogeneous nature makes them susceptible to various cyber threats[9], from Denial of Service (DoS) and Jamming [10] to more sophisticated Eavesdropping and Traffic Analysis attacks[7], each posing risks to the network's availability, confidentiality, and data integrity. Overcoming these challenges is essential for ensuring the seamless operation and reliability of VANETs in promoting road safety and efficiency.

Our proposed framework introduces a comprehensive security strategy that leverages the strengths of homomorphic encryption and the K-nearest neighbors (KNN) algorithm [3] within machine learning applications to safeguard VANET communications. Homomorphic encryption is particularly noteworthy for its ability to perform calculations on encrypted data, thus maintaining data privacy while enabling valuable data analytics. This capability is complemented by the KNN algorithm's efficiency and low computational overhead, making it well-suited for the real-time processing needs of VANETs, where swift data analysis can facilitate immediate and potentially life-saving decisions.

By integrating these technologies, our framework not only addresses the immediate threats to privacy and security in VANETs but also sets a foundation for robust, adaptive security mechanisms. These mechanisms can efficiently handle the complex security demands of VANETs, ensuring the integrity and confidentiality of data transmissions and significantly enhancing the overall resilience of Intelligent Transport Systems.

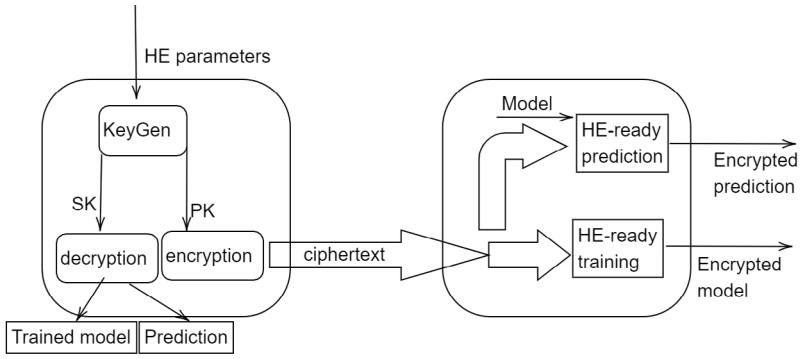

Fig. 2: Framework for Enhancing Privacy in VANETs through Homomorphic Encryption in Machine Learning Applications [2]

While the current implementation focuses on KNN, our framework is designed with adaptability in mind. It can incorporate other machine learning algorithms such as decision trees, random forests, or neural networks, depending on the specific requirements and challenges of the VANET environment. This adaptability not only enhances the framework's robustness against various security threats but also ensures scalability and flexibility in response to evolving technological landscapes.

Furthermore, considering the computational intensity of homomorphic encryption, a hybrid approach that combines it with other privacy-preserving techniques could be explored. Such a hybrid solution aims to optimize computation times while maintaining the highest standards of privacy. The integration of these technologies in our framework represents a significant step forward in addressing the complex privacy and security challenges inherent in VANETs.

In summary, our framework for enhancing privacy in VANETs through homomorphic encryption and machine learning applications stands as a testament to the innovative integration of cryptographic and computational techniques. It offers a scalable, adaptable, and efficient solution to the pressing privacy and security challenges in the dynamic environment of vehicular networks.

4. PERFORMANCE EVALUATION

In this section, we discuss the experiments of our solution. First, we describe the technical and the setup of the environment. Then, we will evaluate the performances of our solution according to different criteria: execution time, accuracy, bandwidth consumption.

4.1. Test Environment

4.1.1. Setup

Our solution is implemented using the TFHE scheme in C/C++ and Python for training k-NN in clear text and for tests. To test the effect of parallelism, we used OpenMP to do some parallelization. The source code is available in the

Table 1: TFHE Parameters:
λ for the overall security, N for the size of the polynomials,
σ for the Gaussian noise parameter.

λ	**N**	σ
110	1024	10^{-9}

Table 2: HE-*k*NN Parameters:
the number of operations m without needing a bootstrapping,
the bootstrapping base b, and the rescaling factors v and p.

m	v	p	b
64	4	1000	4*m-4

following github "https://github.com/Yulliwas/HE-kNN-V". Our solution is tested on Linux Ubuntu 64-bit machine with i7-8700 CPU 3.20GHz.

Table 1 shows the parameters used to setup TFHE scheme.

4.1.2. Datasets

To test our solution, we choose to use 6 datasets: Iris, Breast Cancer, Wine, Heart, Glass and MNIST as in Table 3. The goal is to test the performances of our algorithm in different distributions of data, so that to confirm that our solution works with any dataset and that has performances that are equivalent to those of clear-text domains.

Table 3: Datasets:
number of individuals(n), the size of the model (d), and number of classes

Dataset	**n**	**d**	**classes**
Iris	150	4	3
Wine	178	13	3
Heart	303	75	5
Breast Cancer	699	10	2
Glass	214	10	6
MNIST	1797	10	10

4.1.3. Simulation procedure

First, we preprocess the data by rescaling each attributes to a value between 0 and 1. Our dataset and the query should be rescaled by a factor of v as seen above. We must also multiply the dataset vectors by the precision factor τ and then rounded. In the other hand, the query vector is divided by this same factor.

4.2. Performance results

To position our approach according to existing works, and especially regarding the voting step that is performed without information leakage, we compare in Table 4 our solution with Zuber's solution[11] and with a clear-text version based on the Iris dataset and a fixed k=3. The comparison is done in terms of complexity (C), Information Leakage (L), accuracy (A), interactivity (I) and execution time (T). The accuracy and the prediction time are indicated only when it is possible.

Table 4: Comparison between solutions for Iris Dataset:
complexity (C), Information Leakage (L), accuracy (A), interactivity (I) and execution time (T).

Work	**C**	**L**	**I**	**A**	T
HE-*k*NN-V	$O(n^2)$	N	N	0.97	1.72s
HE-*k*NN-VP	$O(n^2)$	N	N	0.97	0.46s
Zuber[11]	$O(n^2)$	Y	Y	0.98	1.74s
Clear k-NN	$O(n)$	Y	N	0.95	1.8ms

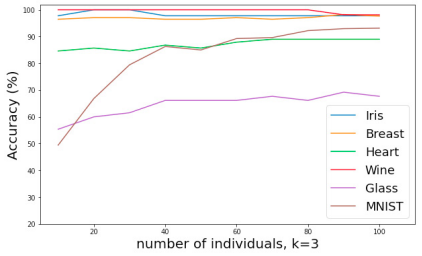

Fig. 3: Encrypted Accuracy vs number of individuals

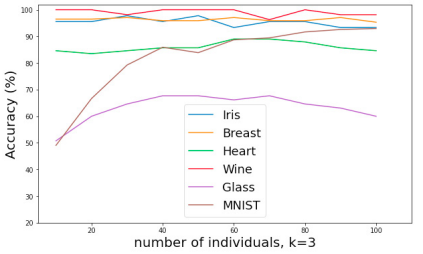

Fig. 4: Clear-text Accuracy vs number of attributes

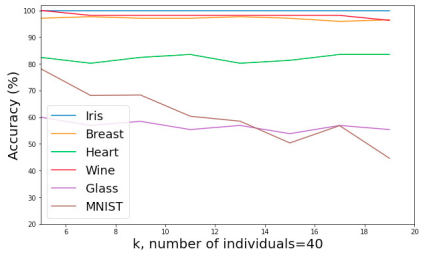

Fig. 5: Encrypted Accuracy vs k-parameter

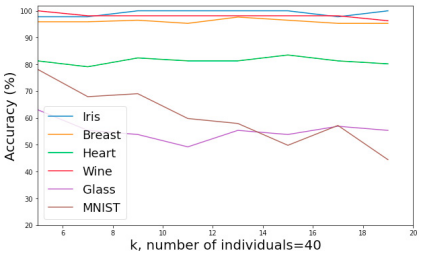

Fig. 6: Clear-text Accuracy vs k-parameter

4.2.1. Empirical study

Classification rate. To evaluate the classification rate, we have chosen the accuracy instead of other metrics like: recall or F1-score. We studied the accuracy according to two parameters: the number of data sampled from the dataset and the number k of neighbors. The goal is to choose the best points that represent the datasets and the best k parameters for each dataset.

We chose real-world datasets in order to see the evolution of the accuracy and compared it to clear-text accuracy.

In one hand, we know that the accuracy depends on the k parameter and we can confirm it easily in the graphs. On the other hand, the assumption that the accuracy depends on the number of data used is not complete. For the dataset where the data is well separated (like Iris), having a lot of data is not necessary, the best accuracy can be achieved using only few data. But, in the case where data is not well separated (like in Heart dataset), the accuracy seems to depend on the number of data.

According to our different simulations illustrated in Figure 3 and Figure 4, we do not lose accuracy when we apply our HE-kNN-V method on the encrypted data compared to the application of the kNN on the plain data. This is possible by varying the number of individuals and by fixing k to 3.

We also notice that by setting the number of individuals to 40 and varying k, (see Figure 5 and Figure 6) the accuracy behaves in the same way between the application of the kNN on the plain data and the application of our method HE-kNN-V on the encrypted data.

Execution time. In our solution, the execution time is independent of the content of the dataset, it does not depend on the values, but does depend on the content, since it depends on the number of tuples . We can use either simulated dataset or real world dataset. To visualize the evolution of the execution time according to k, n and d, we choose to use the Breast Cancer dataset instead of simulating a new dataset. We change n, k, d and we see the evolution of the execution time.

Our simulations, as depicted in Figure 7, illustrate that HE-kNN-V is parallelizable, and also that the number of individuals strongly impacts the execution time unlike the two simulations of Figure 8 and Figure 9 where the variation of respectively d the number of attributes and k does not impact the execution time.

Bandwidth. In our solution, the only thing that is communicated is the query in the ciphertext and the response in the ciphertext. The size of the query is proportional to the number of attributes d. Each attribute is a TLWE Sample with the size of 4 KB and the size of the response (number of classes)*4 KB. The bandwidth according to each dataset is illustrated in Table 5.

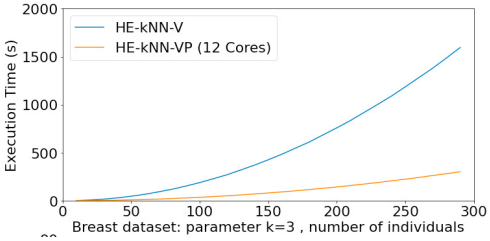

Fig. 7: Execution time vs number of individuals

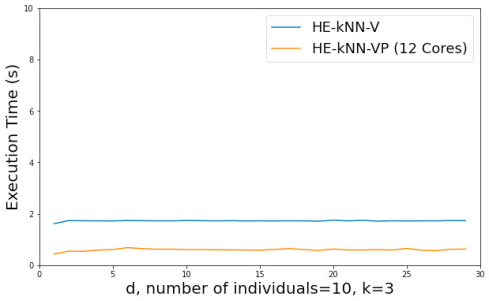

Fig. 8: Execution time vs number of attributes

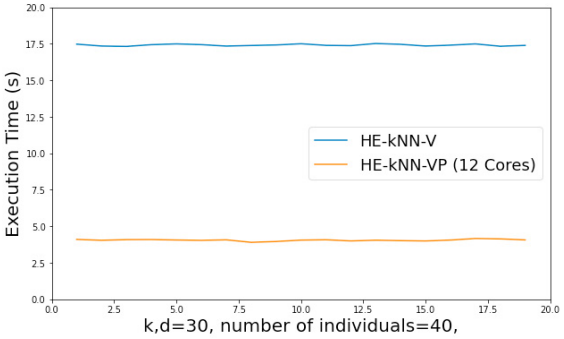

Fig. 9: Execution time vs k-parameter

Table 5: Bandwidth

Dataset	Bandwidth (KB)
Iris	28
Wine	64
Heart	64
Breast Cancer	128
Glass	60
MNIST	296

Discussion. According to our experiments, we can say that the accuracy in our case depends on three factors: the number of individuals, the representativity of these individuals and the k parameter. To have a better model that fits our dataset, we must select the individuals that are more representative of our dataset and the best k parameter. We also should take care of the number of individuals because most of the execution time depends on that number.

5. Conclusions and Perspective

This paper has introduced a comprehensive framework for enhancing privacy in Vehicular Ad Hoc Networks (VANETs) by leveraging the synergy between homomorphic encryption and machine learning algorithms. Our approach addresses the crucial need for robust privacy and security measures in VANETs, which are integral to the evolution of Intelligent Transport Systems (ITS).

The integration of homomorphic encryption (HE), a cryptographic technique that enables direct computations on encrypted data, is a cornerstone of our framework. This maintains data confidentiality and security while using the K-nearest neighbors (KNN) algorithm for its simplicity and effectiveness. The framework's flexibility allows for the incorporation of more complex machine learning algorithms, suitable for specific ITS applications.

One of the key advantages of this framework is its adaptability and general applicability in the dynamic and varied field of VANETs, where privacy concerns can rapidly evolve. Our framework is designed as a robust and adaptable solution, capable of addressing a wide array of privacy issues in VANETs.

Looking to the future, several exciting directions for research and development present themselves. Exploring the integration of advanced machine learning models, such as deep learning and ensemble methods, could enhance the accuracy and efficiency of privacy preservation. The potential of combining homomorphic encryption with other privacy-preserving techniques leads to the possibility of hybrid models that integrate HE with differential privacy or secure multi-party computation. These models could potentially improve performance and provide stronger privacy guarantees.

Real-world implementation and testing of our framework in VANET environments are crucial. This will validate the framework's effectiveness and help identify and address unforeseen challenges. Collaboration with industry and regulatory bodies is also essential to develop standards and policies for implementing privacy-preserving technologies in VANETs, ensuring technical, legal, and ethical compliance.

Interdisciplinary research at the intersection of cryptography, machine learning, and vehicular networks offers a fertile ground for innovation. Future studies could explore novel cryptographic techniques, advanced computational models, and innovative approaches to data security and privacy in ITS.

As VANETs continue to grow, the sustainability and scalability of privacy-preserving solutions become paramount. Research into efficient, low-power, and scalable cryptographic solutions will be critical in ensuring the long-term viability of privacy-preserving techniques in VANETs.

In conclusion, this work marks a significant step towards realizing the full potential of VANETs in enhancing road safety and efficiency while rigorously protecting user privacy. We are optimistic that this framework will serve as a foundation for future innovations in the field, driving the development of secure, efficient, and privacy-preserving ITS applications.

References

[1] Alia Mohammed Alrehan and Fahd Abdulsalam Alhaidari. "Machine Learning Techniques to Detect DDoS Attacks on VANET System: A Survey". In: *2019 2nd International Conference on Computer Applications Information Security (ICCAIS)*. May 2019, pp. 1–6. DOI: 10.1109/CAIS.2019.8769454.

[2] Yulliwas Ameur, Samia Bouzefrane, and Vincent Audigier. "Application of Homomorphic Encryption in Machine Learning". In: *Emerging Trends in Cybersecurity Applications*. Springer, 2022, pp. 391–410.

[3] Yulliwas Ameur et al. "Secure and Non-interactive k-NN Classifier Using Symmetric Fully Homomorphic Encryption". In: *Privacy in Statistical Databases 2022*. Vol. 13463. Privacy in Statistical Databases. Paris, France: Springer International Publishing, Sept. 2022, pp. 142–154. DOI: 10.1007/978-3-031-13945-1_11. URL: https://hal-cnam.archives-ouvertes.fr/hal-03843608.

[4] Craig Gentry. "Fully homomorphic encryption using ideal lattices". In: *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*. Ed. by Michael Mitzenmacher. ACM, pp. 169–178. DOI: 10.1145/1536414.1536440. URL: https://doi.org/10.1145/1536414.1536440.

[5] Xian Guo et al. "Homomorphic encryption based privacy-aware intelligent forwarding mechanism for NDN-VANET". In: *Computer Science and Information Systems* 20.1 (2023), pp. 1–24.

[6] Sagheer Ahmed Jan et al. "A Survey on Privacy-Preserving Authentication Schemes in VANETs: Attacks, Challenges and Open Issues". In: *IEEE Access* 9 (2021), pp. 153701–153726. DOI: 10.1109/ACCESS.2021.3125521.

[7] Muhammet Ali Karabulut et al. "Inspecting VANET with various critical aspects–a systematic review". In: *Ad Hoc Networks* (2023), p. 103281.

[8] Shawal Khan et al. "Security Challenges of Location Privacy in VANETs and State-of-the-Art Solutions: A Survey". In: *Future Internet* 13.4 (2021). ISSN: 1999-5903. DOI: 10.3390/fi13040096. URL: https://www.mdpi.com/1999-5903/13/4/96.

[9] Bharat B Madan, Manoj Banik, and Doina Bein. "Securing unmanned autonomous systems from cyber threats". In: *The Journal of Defense Modeling and Simulation* 16.2 (2019), pp. 119–136. DOI: 10.1177/1548512916628335. eprint: https://doi.org/10.1177/1548512916628335. URL: https://doi.org/10.1177/1548512916628335.

[10] Sinan Ameen Noman and Travis Atkison. "Techniques to Overcome Network Attacks (Sybil Attack, Jamming Attack, Timing Attack) in VANET". In: *Journal of The Colloquium for Information Systems Security Education*. Vol. 10. 1. 2023, pp. 7–7.

[11] Martin Zuber and Renaud Sirdey. "Efficient homomorphic evaluation of k-NN classifiers". In: *Proceedings on Privacy Enhancing Technologies* 2021 (2021), pp. 111–129. URL: https://api.semanticscholar.org/CorpusID:231775630.