A Framework for Global Trust and Reputation Management in 6G Networks

Position Paper

Bac Trinh-Nguyen[1,2], Sara Berri[1], Sin G. Teo[3],
Tram Truong-Huu[4], Arsenia Chorti[1]

[1] ETIS UMR 8051 / CY Paris University, ENSEA, CNRS, France
[2] CNRS, IPAL, Singapore
[3] Agency for Science, Technology and Research (A*STAR), Singapore
[4] Singapore Institute of Technology (SIT), Singapore
`{trinh-nguyen.bac, sara.berri, arsenia.chorti}@ensea.fr,`
`teosg@i2r.a-star.edu.sg, truonghuu.tram@singaporetech.edu.sg`

Abstract. The evolution of the sixth generation (6G) of wireless will enable the full-scale deployment of systems of cyber-physical systems (CPS) such as autonomous vehicles, robots, and drones. Currently, a paradigm shift in security and trust-building approaches for 6G is being discussed within the communications and networking communities, to tackle novel and more sophisticated behavioral attacks of malicious CPS agents. As an example, such autonomous agents will use vast amounts of data – generated from and exchanged amongst heterogeneous network devices – for artificial intelligence-driven decision-making. As a result, standard device authentication approaches and current agent reputation models will not suffice for trust building; the verification, traceability, and trustworthiness of the exchanged data themselves become critical issues. In this paper, we highlight related open issues using as an example a specific behavioral attack in a vehicular ad-hoc network (VANET), to open up the broader discussion regarding potential trust issues in 6G CPS networks. In light of this, we propose a comprehensive framework for trust and reputation management in 6G. We discuss alternative framework architectures and provide high-level details of the framework components with involved technologies.

Keywords: Cyber-physical systems · 6G · trust and trustworthiness · bogus information attacks · machine learning · trust and reputation management.

1 Introduction

The rapid development of communication technologies in recent years paves the way for the evolution of 6G networks, which are expected to contribute to the interconnected world of the future. These future networks will be deployed in various application domains such as smart cities, smart healthcare, industrial automation, and more, forming cyber-physical systems (CPS) where agents seamlessly interact with both the physical and digital worlds and are also known as *multi-agent cyber-physical systems* [16]. In such a dynamic ecosystem, autonomous agents, vehicles, robots, and other intelligent devices will be able to communicate and make decisions independently of humans. However, the increased connectivity and interaction between autonomous entities will open up novel potential security risks and challenges [30,34].

One key challenge stems from the fact that these agents will exchange data continuously, including regarding their *physical states* such as location, speed, acceleration, and environmental conditions, to assist in making accurate real-time decisions. Multiple challenges and opportunities arise as a result. For example, apart from the obvious location privacy issues, a key challenge concerns how to verify that the exchanged information is accurate, not modified or intentionally falsified as it propagates through multiple agents. To capture related issues, the ITU 2018 report [26] defines trustworthy networking as *a set of methods to provide reliable and secure communications among any pair of network elements that have trust relationships.* Subsequent works [22] identified six 6 attributes for trustworthy systems, namely *resilience, security, privacy, safety, reliability, and availability.* In addition, a vast amount of literature is dedicated to trustworthy AI [21], while high-level views of the requirements for 6G trust were also presented [32]. Without robust trust-building mechanisms, malicious agents can conduct adversarial attacks, inject false information that leads to privacy concerns, or cause potentially widespread corruption across the network.

In this paper, we take the position that trust and reputation models should *use both views of the physical world* – as provided through sensing and the lower communication layers – *and the cyber world*, as provided through semantics and upper communication layers, unlike existing trust and reputation models. To highlight the importance of incorporating physical views in trust models, in a preliminary investigation, we exemplify security attacks in VANETs, a typical example of CPS where network operations take place in a highly dynamic environment. The mobility of vehicles and constant changes in network topology highlight the need for trust and reputation management to maintain network security. One of the common threats in VANETs is the bogus information attack [14] or a false data injection attack [9], where rogue vehicles send false data with the intention of causing delays, traffic congestion, accidents, or gaining priority. As a proof of concept, we discuss the practicality and effectiveness of a multi-view trust framework incorporating location and speed information.

Motivated by the previous discussion, we highlight the potential open issues in 6G trust and trustworthiness and propose a novel framework for global trust and reputation management (TRM). We present two potential architectural approaches that take into account cross-layer information from the physical layer to the application layer, as well as sensing and semantic plane inputs.

The rest of the paper is organized as follows. In Section 2, we present our preliminary investigation with VANETs, to motivate the follow-up discussion. In Section 3, we discuss the new trends in 6G and describe the paradigm shift in trust building and management to meet the evolving requirements of 6G networks of CPSs. In Section 4, we present possible trust management framework architectures with detailed descriptions of the framework components. In Section 5 we conclude the paper and present future directions.

2 Experiments and Motivation

Along with the development of 6G networks, the need to expand VANETs is also constantly increasing, leading to security risks. Attacks in VANETs focus on different directions to undermine confidentiality, integrity, availability, identity, authentication, and non-repudiation properties [33]. Fig. 1 shows the network stack layers including attack vectors classified as single-layer or multi-layer attacks, along with the unique characteristics of each layer used for current detection solutions. At each layer, we identify the features that can be abused by adversaries to perform attacks but also be used by defenders to detect attacks or abnormal behavior.

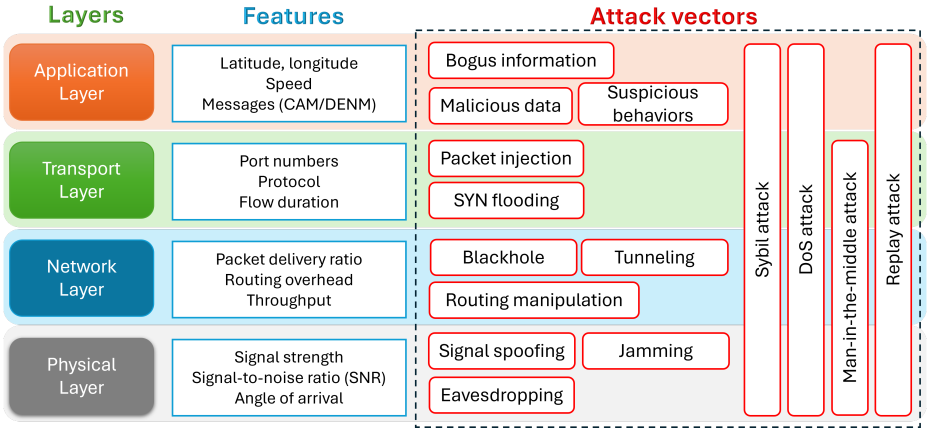


Fig. 1: Features and attack vectors in each layer of the network stack.

- **Physical layer.** In the study by Alzahrani *et al.* [4], received signal strength indicator (RSSI) and angle of arrival (AoA) are used as trusted features to determine the reliability of cooperative awareness messages (CAM) received from adjacent vehicles. However, we note that the RSSI and the AoA have been shown to be vulnerable to certain types of attacks, e.g., when using analog array processing chains [29] as opposed to digital array processing [24].
- **Network layer.** Shamim *et al.* [35] used features such as packet delivery ratio (PDR), routing overhead (ROH), throughput, packet loss rate (PLR), packets generated, packets dropped, energy consumption in their collaborative detection method of black hole and gray hole attacks.

- **Transport layer.** Hassan *et al.* [17] used the features at the transport layer to interpret the decision-making process of the random forest classifier, thereby understanding the characteristics of different types of attacks in VANETs. For example, "destination-port" represents the port number where the packet will arrive, "flow-packets-sec" is the packet rate in a flow, and "min-seg-size-forward" is the minimum segment size in the forward direction. However, it has been shown that destination ports could be vulnerable to a certain type of attacks, such as web-attack, infiltration, and port scan (PS) [23].
- **Application layer.** So *et al.* [28] calculated velocity, acceleration, and distance traveled values from GPS location information and then converted them into plausible checks to detect anomalies. Several attacks on GPS location have been identified, along with several mitigation methods [10,31]

To highlight the novel, complex security challenges in 6G networks of CPSs, in this section, we illustrate a specific use case of anomaly detection in a VANET. First, we use the Eclipse MOSAIC simulation platform [27] to set up different VANET scenarios. Then, we examine the limitations of current approaches and present a proposed solution using positioning and speed information to identify a bogus information attack. As an example, an attacker attempts to broadcast a fake message about obstacles on the current route using a type of vehicle-to-everything (V2X) communication called decentralized environmental notification message (`DENM`) to disperse vehicles away from the optimal route to gain an advantage [19].

2.1 Scenario setup

Simulation framework. We used Eclipse MOSAIC [27] to develop a small-scale attack scenario. This simulation framework allows for the realistic simulation of vehicle communications, providing a controlled environment and entities for testing different types of attacks and mitigation approaches. It also enables the programming of applications for vehicles and roadside units (RSUs).

Simulation setting. The experimental scenario takes place in an urban setting around the area of the ETIS campus in Cergy, France, with a simple road topology, including a main route and an alternative route. One of the vehicles in the middle of a convoy plays the role of an attacker, broadcasting fake `DENM` messages to notify about non-existing obstacles on the main route, causing vehicles within a certain radius to redirect to an alternative (sub-optimal) route. During the simulation, the system collects real-time data exchanged between vehicles and infrastructure. The collected data includes information on location, speed, and acceleration, and all communication is monitored for analysis.

Attack mitigation. We implemented a centralized system with a misbehavior detection mechanism, specifically focusing on detecting bogus information attacks within this scenario. The simulation system includes:

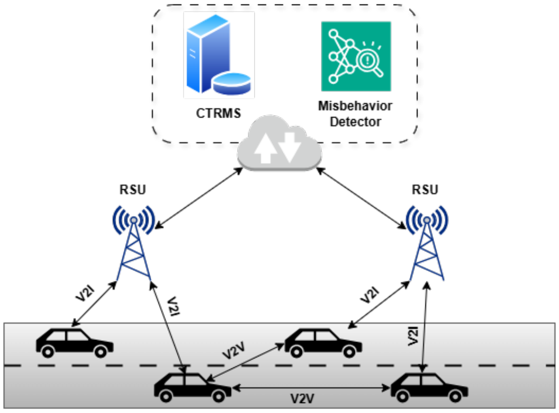


Fig. 2: Illustration of simulation scenario including V2X communications: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I).

- **A centralized trust and reputation management system (CTRMS)**: The system collects logs from the network and then transfers them to the central server for storage and processing. The system includes a rule-based anomaly detector and vehicle trust management, which is quantified as a trustworthiness score to describe the truthfulness of the vehicle in the network. The management model is shown in Fig. 2.

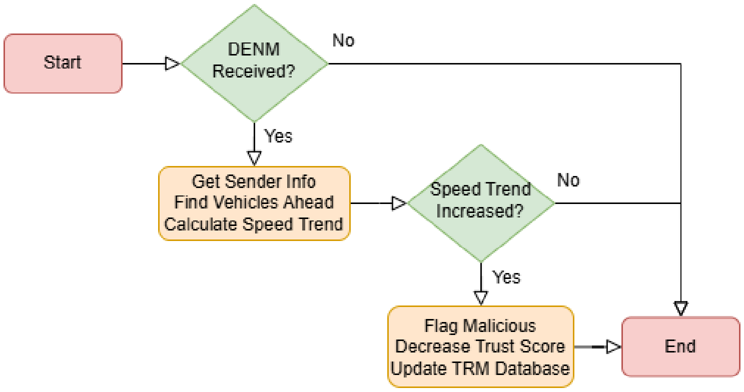


Fig. 3: Flowchart for rule-based detection of bogus information attack.

- **Misbehavior detector**: This module plays a role in analyzing the collected data of vehicles in the network and finding anomalies, identifying malicious vehicles. In the scope of this scenario, we implemented a simple rule-based detection mechanism. This rule-based processing flow is described in Fig. 3. When a vehicle receives a DENM message from the network, it requests CTRMS to check the sender's trustworthiness. The system then verifies whether there are any abnormalities based on the collected logs and identifies the location of the vehicles and the behavior (speed and movement trends) of nearby vehicles. If any abnormalities related to the sender are detected, the system will flag the sender as malicious, decrease their trust score, notify the requesting vehicle, and take necessary measures. In this scenario, to simplify, all vehicles start with a trust score of 1. Once a vehicle is identified as malicious, its trust score will decrease to 0. Based on this, vehicles receiving the DENM message will decide whether to change routes according to the trust scores provided by the CTRMS system.

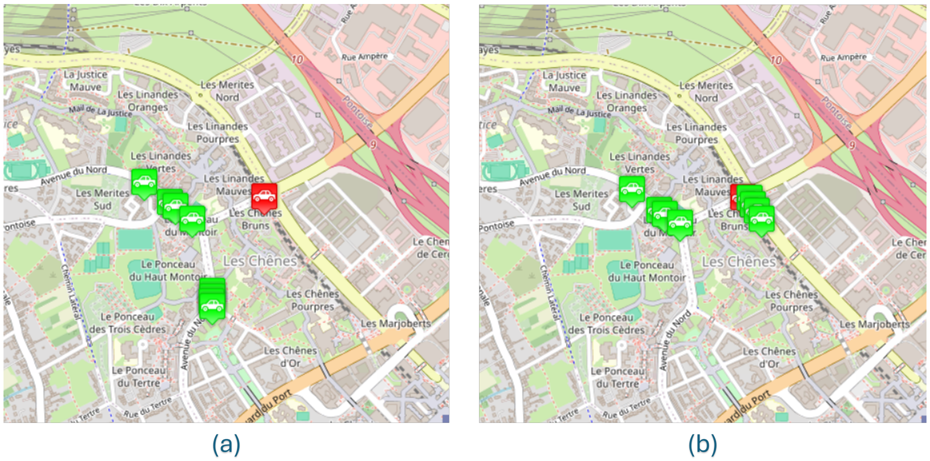
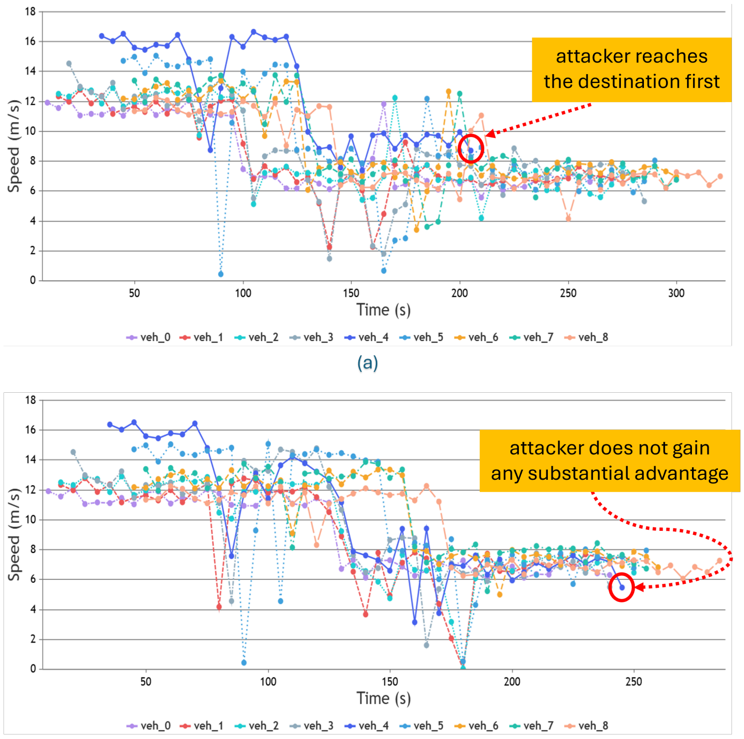


Fig. 4: Map visualization for two test cases **(a) successful bogus information attack successfully** and **(b) unsuccessful attack – detected and prevented**. The red marker indicates the malicious vehicle while the green markers represent benign vehicles.

2.2 Analysis

We conducted experiments through three test cases: the first simulates an attack without CTRMS, the second demonstrates the detection capability of the CTRMS combined with a misbehavior detector, and the third highlights the scalability issue of the system when there is a large number of vehicles in the network. We measure the processing delay for each request at CTRMS.

- **Case 1: Successful attack.** Due to the lack of a trust framework, the attacker can succeed in spreading bogus information about the optimal route



(a)

(b)

Fig. 5: Vehicle speed graph for two test cases **(a) bogus information attack successfully propagated** and **(b) attack detected and prevented**.

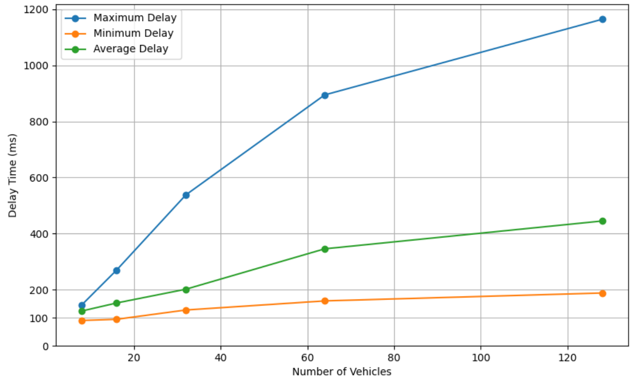


Fig. 6: Delay time measurement (maximum, minimum, average) for different vehicle counts (8, 16, 32, 64, 128).

to the neighboring vehicles, causing them to change to the alternative route, shown in Fig. 4(a). In this toy example, the attacker's gains correspond to reaching the destination first, at around 200 seconds after departure, while the other vehicles arrive within the time frame of 300 seconds. The impact on vehicle speed is illustrated in Fig. 5(a).

- **Case 2: The attack is detected and mitigated.** With the CTRM system deployed along with a misbehavior detector, the attack was successfully detected. Through the trust score verification from the server, the decision-making units on other vehicles will not make the rerouting decision, and the attack was successfully mitigated, as shown in Fig. 4(b). As a result, the attacker did not gain any substantial advantage, the speed graph is displayed in Fig. 5(b).
- **Case 3: Delay time measurement.** We continue with the above scenario but increase the number of vehicles in the network from 8 to 16, 32, 64, and 128, respectively. The purpose of this task is to simulate varying traffic densities in the network and, consequently, collect and analyze the processing time and response time (delay time) from the centralized server to the requested vehicles based on these density variations. The results are shown in Fig. 6, including the maximum, minimum, and average delay times. We notice that when the number of vehicles in the network increases, the server processing and response time slows down significantly. This exceeds the maximum requirement for low latency services, typically requiring from 1ms to 10ms [13,20]. The low latency or ultra-low latency is an important element in 6G applications such as self-driving cars, remote surgery, and industry automation. In the context of VANETs, failing to meet the strictly end-to-end latency requirements between vehicles, roadside units (RSU), and the traffic control center (TCC) can lead to serious consequences. VANETs are designed to support real-time communication for safety-critical applications such as collision avoidance, lane change assistance, and accident prevention.

Through this above example, we can observe that implementing a centralized system to manage the trustworthiness scores of nodes in the network shows both strengths and limitations in detecting anomalous behavior using a rule-based approach. *Firstly*, the system's dependence on application-layer data such as GPS, speed, and acceleration information can reduce its effectiveness in identifying complex attack patterns. If GPS spoofing is successful, the attack cannot be identified simply by this type of misbehavior detection mechanism. *Secondly*, using static rules to identify attacks is limited due to the difficulty of handling dynamic and evolving attack strategies, e.g., involving the coordinated action of multiple malicious entities simultaneously (colluding adversaries). *Thirdly*, as the size of the network increases, the rule-based system becomes less efficient in processing real-time data, leading to detection delays and an increased likelihood of false positives and false negatives. Thus, although similar approaches have been proposed in the literature [30], the limited trust and reputation model used proves to be inefficient for 6G networks of CPSs, where a vast number of autonomous nodes will connect, communicate, and actuate, using large volumes of complex data.

In such environments, the system must be able to adapt in real-time and handle multi-layer and multi-modality data to identify threats accurately. ML/DL models offer promising solutions for processing large amounts of data, identifying complex attack patterns and anomalies, and enabling the detection of both known and unknown (zero-day) attacks. Moreover, by combining the data from multiple layers, sensing and semantic layers, ML/DL models can comprehensively analyze the network and the node behaviors, providing the system with a vital view to detect the increasingly diverse and evolving threats in 6G networks.

3 Security Challenges and Opportunities in 6G

Below, we enumerate several key challenges that will impact the design of security solutions in 6G, notably with respect to latency, quantum resistance, and the role of AI. Subsequently, we outline novel opportunities that will be brought up in 6G, in terms of multiple security levels (quality of security), the introduction of physical layer solutions, and integrated communications and sensing.

3.1 New Security Challenges in 6G Networks

Low latency, low footprint, scalable security: Ensuring robust security for networks of autonomous agents requires operating within strict latency limits, handling massive connectivity, and maintaining low energy consumption, low footprint and low computational load, is an ambitious challenge.

Quantum resistance: Additionally, to remain secure against future threats, systems must incorporate quantum-resistant cryptographic methods. This approach aligns with the newly standardized post-quantum cryptographic algorithms established by NIST [5,6], the adoption of quantum-resistant algorithms into IoT devices is challenging due to their limited resources.

AI and ML: The widespread deployment of AI and ML in 6G systems will expand potential vulnerabilities, highlighting the need for defenses against adversarial AI (such as data-poisoning attacks). Additionally, the energy footprint needs to be contained to sustainable levels (green AI) and the AI outputs need to be interpretable, and unbiased (XAI) [8].

3.2 Opportunities for strengthened security in 6G

Alongside these challenges, new security opportunities emerge, including decentralized and democratized approaches to learning and computation, such as federated learning, the integration of sensing capabilities, and the ability to perform context and semantics distillation. These advancements open doors to innovative technologies, like physical layer security, as discussed below.

The introduction of sensing in 6G: Positioning and radar sensing will be default features in 6G networks [2]. This integration creates possibilities for using sensing and positioning for integrity checks and anomaly detection that takes into consideration the physical behavior devices. While Sybil cyberattacks in robotic systems have been found using AoA [15], a number of techniques that use location information as a second soft authentication factor have already been developed. Trusted positioning will be a vital aspect of assessing the trustworthiness of autonomous agents in 6G. However, this also introduces privacy challenges.

Quality of security (QoSec): In the long run, a safe and sustainable future requires adaptability to maximize limited resources. In this context, adaptive security protocols and algorithms could be developed to adjust configurations and parameters dynamically, based on multi-layered inputs from sensing, semantic, and contextual data. QoSec offers a scalable security framework for CPS networks, providing varying levels of security and trust to meet the needs of future networks.

Physical layer security: Within an adaptive, AI-driven security framework, PLS solutions [12], which leverage physical phenomena for security, are envisioned as complements to post-quantum cryptography, enhancing 6G's trustworthiness and resilience. PLS could enable secure keyless communication or the generation and distribution of symmetric keys through wireless channel properties. We will be able to systematically make use of such opportunities in 6G thanks to channel engineering and controllability (e.g., with the use of meta-surfaces, drones for multi-hop networks, and very narrow beamforming). Additionally, PHY and hardware layer authentication, e.g., utilizing physical unclonable functions and localization as a second-factor authentication, can allow for rapid, continuous verification of legitimate users, minimizing the need for upper-layer processing.

3.3 Paradigm Shift in Security for 6G Networks

The development of the new generation of networks has promoted the transition from a network of connected things to an intelligent network, where agents have more autonomous capabilities and more interactions with each other. However, without TRM systems in multi-agent networks (MAS) or IoT, such interactions will not ensure trustworthiness and will pose many security risks [36]. Therefore, 6G networks need to be designed with inherent trustworthiness to improve security effectiveness and enhance privacy protection capabilities [18].

Trust and trustworthiness: The behavior of devices, agents, and systems must be considered holistically when establishing trust and assessing trustworthiness. For a trustworthy 6G, multiple layers of trust must be guaranteed. While ongoing discussions continue, it is evident that a secure 6G network must ensure trust in both the AI intelligence and the infrastructure, including sensing, communication channels, and data processing. The initial anchors of trust can be reduced to trusting at a highly abstract level: (i) The sensing (radar, RF, camera, lidar, etc.) that gathers raw or processed data (high precision localization information is especially crucial in 6G) and drives actuation; (ii) The platforms for computation and processing (including learning and optimization) at various network nodes (on devices, edges, and core networks); (iii) The communication links that transport data and authenticate agents and devices, offering assurances of confidentiality, integrity, authentication, and availability; (iv) The AI algorithms that define the actions of autonomous agents, devices, and systems based on data inputs and sensor readings.

In summary, 6G will be the first generation inherently built on AI, connecting intelligent and autonomous cyber-physical systems, digital twins of physical entities, and the metaverse. This integration of the physical, digital, and human realms signifies a new era where the physical properties of interconnected systems are essential for security.

4 Trust and Reputation Management Framework

In 6G networks, AI is expected to be native, meaning it plays a vital role in the network infrastructure as a fundamental component of the operations, administration, and management process across all layers of the system [1]. A high-level perspective on the components and factors maintaining trustworthiness is presented in Fig. 7. The data plane, control plane, and sensing plane are accountable for data processing, network management, and environmental monitoring. The data plane handles forwarding and routing packets in the network, using AI to actively monitor data flows, identify inconsistencies, and automatically flag them, thus enhancing data security and integrity. The control plane manages routing decisions and overall network operations. AI enhances the control plane by offering real-time analytics and adaptive decision-making algorithms that continuously evaluate the trustworthiness of network entities such as routers, switches, or RSUs in VANETs. The sensing plane gathers data from the physical environment through sensors (e.g., GPS, speed, acceleration) and other contextual inputs. AI enhances trust at the sensing plane by validating sensor data in real-time, identifying anomalies, and quickly detecting faulty or compromised sensors. This reduces the risk of false information entering the network, improves sensor accuracy, and minimizes potential vulnerabilities. Besides, semantics is a key topic emerging in 6G with the semantic plane performing signal processing, and information filtering across all layers of the protocol stack by leveraging ML algorithms deployed on edge and core cloud platforms [25].

One of the critical challenges is how to quantify trustworthiness within a network. In this section, we present two possible approaches for measuring trustworthiness through the idea of the TRM framework. These frameworks play a crucial role in analyzing historical data, evaluating the behavior of network entities, and real-time assessment through the integration of AI-based techniques. The first

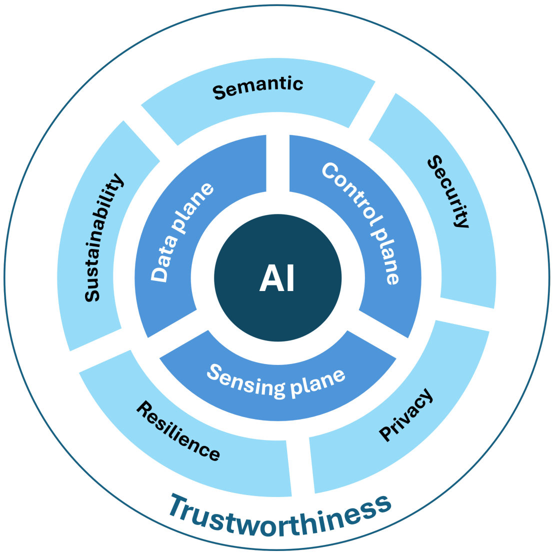


Fig. 7: High-level overview of trustworthiness in 6G.

one is a multi-layer trust management system, as described in Section 4.1, where a global TRM serves as the highest authority, aggregating trust from all layers through their respective local TRM. The local trust evaluation is supported by ML/DL algorithms, which analyze features at each corresponding layer. On the other hand, the second one is a multimodal attention fusion approach, as outlined in Section 4.2, which leverages ML/DL to process and analyze data, extract the most important features at each layer, and synthesize them through a fusion module to support the process of assessing trustworthiness. This approach provides a comprehensive view of context and multiple behavior aspects, making trust assessments more reliable and accurate.

4.1 Multi-layer trust and reputation management framework

This framework is designed to maintain network security by combining multi-layer data from physical layer (PHY) signals and the information from the upper layers, applying ML and DL techniques to improve the capability of detecting attacks or misbehavior while enhancing the system's robustness. In Fig. 8, we present our first possible framework architecture, which is based on a multi-layer approach. The general framework architecture includes multiple TRM modules, each being responsible for collecting and processing data captured at each respective layer from PHY to higher network layers. Each layer plays an important role in the process of detecting potential attacks by providing unique features and observations at that layer to the inference module integrated with ML/DL models for analyzing complex attacks, handling vast amounts of data in real

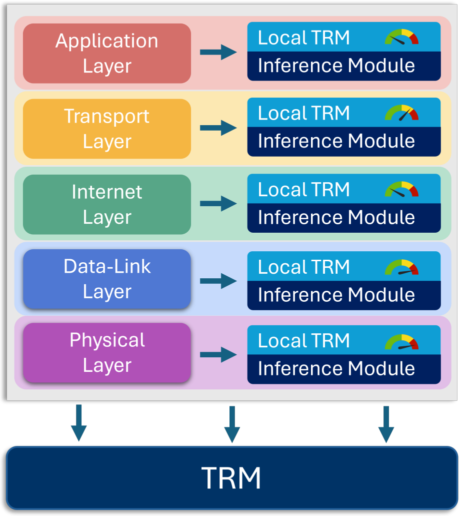


Fig. 8: Multi-layer architecture for TRM framework.

time, identifying subtle deviations in behavior that traditional rule-based systems often miss. At each layer, we propose to implement a specific local TRM based on the decision results from the inference module. Next, a global TRM is deployed to aggregate the decision from all local TRMs and compute an overall trust score for each node or vehicle. Depending on the nature of the attacks, a weightage vector can be applied to prioritize the decision of a particular local TRM. This multi-layer TRM architecture provides a holistic view and comprehensive trust score across the entire network while ensuring trust evaluation according to the data type and attack vector characteristics of each layer. The flexibility of this architecture is expected to be compatible with many types of cyber-physical systems while ensuring robustness in various contexts.

By combining the general model and the specific use case in the context of VANETs, Fig. 8 emphasizes the need for multi-layered security and trustworthiness in future autonomous and interconnected systems.

4.2 Multimodal attention fusion approach

Because of the variety of data types of wireless communication characteristics, it is not appropriate to apply a specific ML/DL algorithm to calculate trust. We propose to develop and deploy additional planes including communication, sensing, and semantic planes to collect further information from the environment and correlate them with the data collected from the network. For example, the data collected in VANET can include GPS coordinates, velocity, acceleration, weather sensor data, camera images from the sensing plane; signal strength, packet transmission rates from the communication plane, and the meaning and context behind the data being transmitted from the semantic plane. Therefore,

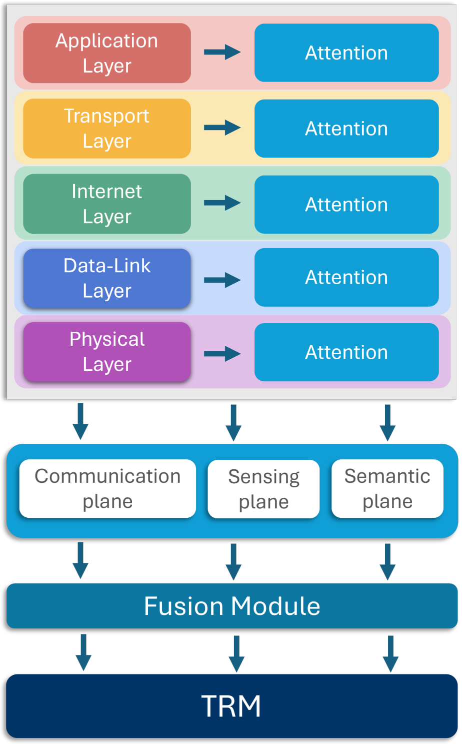


Fig. 9: Multimodal attention fusion architecture for TRM framework.

we propose to use a multimodal approach [7] which can use multiple types of input data for processing and analysis, providing a comprehensive view and contextual understanding, which has been applied and researched in security solutions [11,3]. Fig. 9 depicts our proposed framework based on the idea of combining the attention mechanism and multimodal models. In this multimodal security framework, TRM can benefit from the attention models by focusing on trustworthiness data at each layer. Unlike the architecture presented in Section 4.1, the main idea of this approach is to utilize the attention mechanism to highlight key features at each layer, and then combine (fuse) the features into a unified representation based on the correlation among features. This allows the model to consider multiple perspectives simultaneously, increasing the accuracy of decision-making. The proposed framework includes the following components:

- **Data collection and attention to various layers:** The model applies attention to the data at each layer, identifying and prioritizing the most relevant information that is most important for a particular task. For example, with the task of detecting attacks such as jamming or spoofing at the PHY layer, the attention model could focus on specific features from a diverse set of data collected at the physical layer such as signal strength (RSSI), signal-

to-noise ratio (SNR), bit error rates (BER), channel state information (CSI). By applying the attention mechanism, the model can focus on specific signal patterns that indicate fake signal transmission or jamming attempts.
- **Fusion module (multimodal models):** Once the important features at each layer are weighted using the attention mechanism, all features will be aggregated into a unified representation based on their correlation. This allows the model to consider multiple perspectives and contexts, thus detecting complex, multi-vector attacks.
- **Final decision-making:** After synthesizing the attention-weighted features from multiple layers, the model makes predictions or other decisions. The trust scores of nodes in the network are calculated based on insights from information across multiple layers. This allows for comprehensive and accurate decision-making.

The proposed framework is expected to form an effective trust and reputation management solution and enhance the ability to detect sophisticated, multi-layered attacks by fusing attention-weighted features from different layers of the network stack. In addition, combining it with AI-based methods will help improve the system's resilience against both known and unknown attack vectors in 6G networks.

5 Conclusion

In this paper, we introduced a global trust and reputation management framework for 6G networks through two possible architectural approaches, a multi-layer approach and a multimodal attention fusion approach. The proposed multimodal attention fusion framework is expected to enhance security in multi-layer networks by integrating attention mechanisms with a global trust and reputation management system. It is especially suitable in the context of 6G networks where massive connectivity and ultra-low latency, high security, and reliability are required. The scalability of the multimodal fusion approach allows it to handle vast amounts of collected from numerous entities across multiple network layers. Integrating a global TRM system will further enhance decision-making by continuously adjusting security responses based on the behavior, trustworthiness, and past actions of network entities. Based on the ideas proposed in this study, some potential directions of TRM framework development include:

- Development of a practical TRM framework: Since implementing trustworthiness in technical systems is still in its early stages, the next phase will focus on building a real-world TRM framework, where ML/DL algorithms play a crucial role in handling and analyzing data across multiple network layers. Besides, it also needs to ensure fairness and interpretability in the decision-making processes of these algorithms, as transparency is essential for trust in autonomous systems. Additionally, the developed framework will prioritize improving detection accuracy for anomalies while minimizing false

positives. This will involve refining the ML/DL models to better differentiate between normal variances in behavior and actual malicious activities, ultimately leading to a more robust and dependable TRM framework.
- Define and develop trust scoring mechanisms for the entities and their connections within the network: These mechanisms should be flexible and able to dynamically and continuously update based on changes in the entities' behaviors, enabling a more adaptive and sustainable trust management system.
- Considering the practical applicability and scalability of the system, deploying the TRM framework in large-scale cyber-physical systems presents a significant challenge. It will need to ensure scalability, meet latency requirements, and operate within resource constraints while maintaining accurate anomaly detection capabilities. Therefore, it is essential to implement realistic scenarios that can provide insights into the framework's performance across diverse conditions and environments. These scenarios will enable fine-tuning of the model to suit the specific requirements of each system.

Acknowledgements

A. Chorti and S. Berri have been partially supported by the EC through the Horizon Europe/JU SNS project ROBUST-6G (Grant Agreement no. 101139068), the EU HORIZON MSCA-SE TRACE-V2X project (Grant No. 101131204), the ANR-PEPR 5G Future Networks projects, and the CYU INEX-PHEBE and NESLI projects.

References

1. Transforming the 6g vision to action. White paper, Nokia Bell Labs (2024)
2. 3GPP: Study on integrated sensing and communication. Tech. Rep. TR 22.837, 3GPP (2022)
3. Agrafiotis, G., Kalafatidis, S., Giapantzis, K., Lalas, A., Votis, K.: Advancing Cybersecurity with AI: A Multimodal Fusion Approach for Intrusion Detection Systems. In: 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom) (2024)
4. Alzahrani, M., Idris, M.Y., Ghaleb, F.A., Budiarto, R.: Robust Misbehavior Detection Scheme for Vehicular Network. In: 2021 International Conference on Data Science and Its Applications (ICoDSA) (2021)
5. Avanzi, R., Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schanck, J.M., Schwabe, P., Seiler, G., Stehlé, D.: CRYSTALS-Kyber Algorithm Specifications And Supporting Documentation. NIST PQC Round (2017)
6. Bai, S., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schwabe, P., Seiler, G., Stehlé, D.: CRYSTALS–dilithium: Algorithm specification and supporting documentation (2021)
7. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(2), 423–443 (2019)

8. Belle, V., Papantonis, I.: Principles and Practice of Explainable Machine Learning. Frontiers in Big Data **4** (2021)
9. Bennet Praba, M.S., Rathna, R.: False data injection attack detection in vanet using upgraded grey wolf optimization algorithm using lstm classifier. In: Sharma, H., Shrivastava, V., Bharti, K.K., Wang, L. (eds.) Communication and Intelligent Systems. pp. 703–713. Springer Nature Singapore, Singapore (2023)
10. Burns, J., Amiridis, D., Kar, D.C., Li, L.: On effectiveness of machine and deep learning algorithms for detection of gps spoofing attacks on unmanned aerial vehicles. In: 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE). pp. 2405–2410. IEEE (2023)
11. Chebbi, S., Jebara, S.B.: Deception detection using multimodal fusion approaches. Multimedia Tools and Applications **82**(9) (2023)
12. Chorti, A., Barreto, A.N., Köpsell, S., Zoli, M., Chafii, M., Sehier, P., Fettweis, G., Poor, H.V.: Context-Aware Security for 6G Wireless: The Role of Physical Layer Security. IEEE Communications Standards Magazine **6**(1), 102–108 (2022)
13. European Telecommunications Standards Institute: ETSI TS 122 261 V15.9.0 (2021-10): Service requirements for next-generation networks; Stage 1. Tech. Rep. V15.9.0, ETSI (October 2021)
14. Ghaleb, F.A., Zainal, A., Maroof, M.A., Rassam, M.A., Saeed, F.: Detecting Bogus Information Attack in Vehicular Ad Hoc Network: A Context-Aware Approach. Procedia Computer Science **163**, 180–189 (2019)
15. Gil, S., Kumar, S., Mazumder, M., Katabi, D., Rus, D.: Guaranteeing spoof-resilient multi-robot networks. Autonomous Robots **41** (2017)
16. Gil, S., Yemini, M., Chorti, A., Nedić, A., Poor, H.V., Goldsmith, A.J.: How Physicality Enables Trust: A New Era of Trust-Centered Cyberphysical Systems (2023)
17. Hassan, F., Yu, J., Syed, Z.S., Ahmed, N., Al Reshan, M.S., Shaikh, A.: Achieving model explainability for intrusion detection in vanets with lime. PeerJ Computer Science **9**, e1440 (2023)
18. Huang, X., et al.: 6G Trustworthiness Considerations. Tech. rep., NGMN Alliance (Oct 2023)
19. Iqbal, S., Ball, P., Kamarudin, M.H., Bradley, A.: Simulating malicious attacks on vanets for connected and autonomous vehicle cybersecurity: A machine learning dataset. In: 2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP). pp. 332–337. IEEE (2022)
20. Jun, S., Kang, Y., Kim, J., Kim, C.: Ultra-low-latency services in 5g systems: A perspective from 3gpp standards. Etri Journal **42**(5), 721–733 (2020)
21. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., Zhou, B.: Trustworthy AI: From Principles to Practices. ACM Comput. Surv. **55**(9) (Jan 2023)
22. Li, J., Mao, B., Liang, Z., Zhang, Z., Lin, Q., Yao, X.: Trust and Trustworthiness: What They Are and How to Achieve Them. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). Kassel, Germany (Mar 2021)
23. Loumponias, K., Raptis, S., Darra, E., Tsikrika, T., Vrochidis, S., Kompatsiaris, I.: Forecasting cyber-attacks to destination ports using machine learning. In: ICISSP. pp. 757–764 (2023)
24. Pham, T.M., Senigagliesi, L., Baldi, M., Fettweis, G.P., Chorti, A.: Machine learning-based robust physical layer authentication using angle of arrival estimation. In: GLOBECOM 2023-2023 IEEE Global Communications Conference. pp. 13–18. IEEE (2023)
25. Popovski, P., Simeone, O.: Start making sense: Semantic plane filtering and control for post-5g connectivity. arXiv preprint arXiv:1901.06337 (2019)

26. Recommendation ITU-T Y.3053: Framework of trustworthy networking with trust-centric network domains. Tech. Rep. ITU-T Y.3053, International Telecommunication Union (2018)
27. Schrab, K., Neubauer, M., Protzmann, R., Radusch, I., Manganiaris, S., Lytrivis, P., Amditis, A.J.: Modeling an its management solution for mixed highway traffic with eclipse mosaic. IEEE Transactions on Intelligent Transportation Systems **24**(6), 6575–6585 (2023)
28. So, S., Sharma, P., Petit, J.: 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 (2018)
29. Srinivasan, M., Senigagliesi, L., Chen, H., Chorti, A., Baldi, M., Wymeersch, H.: Aoa-based physical layer authentication in analog arrays under impersonation attacks. arXiv preprint arXiv:2407.08282 (2024)
30. Sullivan, S., Brighente, A., Kumar, S.A.P., Conti, M.: 5G Security Challenges and Solutions: A Review by OSI Layers. IEEE Access **9** (2021)
31. Wagner, M., Fröhlich, A.A.: Securing cyber-physical systems against gps spoofing attacks using confidence attribution. In: 2023 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). pp. 1–6. IEEE (2023)
32. Wang, Y., Kang, X., Li, T., Wang, H., Chu, C.K., Lei, Z.: SIX-Trust for 6G: Toward a Secure and Trustworthy Future Network. IEEE Access **11** (Oct 2023)
33. Xu, X., Wang, Y., Wang, P.: Comprehensive review on misbehavior detection for vehicular ad hoc networks. Journal of Advanced Transportation **2022**(1) (2022)
34. Yan, W.K., Chan, R., Truong-Huu, T.: 5G Core Security: An Insider Threat Vulnerability Assessment. In: Proc. 13th Conference on Information Technology and Its Applications (CITA). Danang, Vietnam (July 2024)
35. Younas, S., Rehman, F., Maqsood, T., Mustafa, S., Akhunzada, A., Gani, A.: Collaborative detection of black hole and gray hole attacks for secure data communication in vanets. Applied Sciences **12**(23), 12448 (2022)
36. Zeynalvand, L., Luo, T., Zhang, J.: COBRA: Context-Aware Bernoulli Neural Networks for Reputation Assessment. Proceedings of the AAAI Conference on Artificial Intelligence **34**, 7317–7324 (04 2020)