Attacker model for Connected and Automated Vehicles

Jean-Philippe Monteuuis
Groupe PSA / Telecom ParisTech
Velizy, France
jeanphilippe.monteuuis@mpsa.com

Jonathan Petit
OnBoard Security
Wilmington, MA, 01887, USA
jpetit@onboardsecurity.com

Jun Zhang
Telecom ParisTech
Paris, France
jun.zhang@telecom-paristech.fr

Houda Labiod
Telecom ParisTech
Paris, France
houda.labiod@telecom-paristech.fr

Stefano Mafrica
Groupe PSA
Velizy, France
stefano.mafrica@mpsa.com

Alain Servel
Groupe PSA
Velizy, France
alain.servel@mpsa.com

ABSTRACT

Connected and Automated Vehicle is the next goal for car manufacturers towards traffic safety and efficiency. While researchers deceived range sensors and vehicular communication, few analyzed the inside and the outside of the vehicle surface. As a result, current attacker models are too network-oriented or sensor-oriented. Therefore, we propose an attacker model which details attacks occurring in *Ground Truth* environment and data fusion processes. Then, we define a new security model with the perspective of achieving a secured automotive perception.

CCS CONCEPTS

• **Security and privacy → Security requirements**; **Embedded systems security**; • **Computer systems organization → Embedded and cyber-physical systems**; Dependable and fault-tolerant systems and networks;

KEYWORDS

V2X, Sensors, Attacker Model, Security Goals Model, Automotive Perception

1 INTRODUCTION

Original Equipment Manufacturers (OEMs) plan to commercialize Connected and Automated Vehicles (CAVs) by 2020. To achieve safe automation, OEMs enhance automotive range sensors perception with Vehicle-to-X (V2X) communication. Indeed, the OEMs race towards full automation is nowadays about conceiving reliable perception systems to ensure passenger safety. Therefore, using data fusion mechanisms to cope with V2X and sensors weaknesses is mandatory [1]. However, such a system assumes that the surrounding environment perceived by sensors is trustworthy [2] or that communicating nodes are benevolent [3]. Thus, it is still unclear what are the possible attacks outside and inside the vehicle that can fool the perception system. To achieve reliable automation, attackers action(s) and target(s) need to be defined using an attacker model. Although several attacker models exist, they are either too V2X-centric [4–6] or vehicle-centric [7], and thus, do not consider the whole perception lifecycle. As a result, they fail to capture the entire attack space. For instance, an attacker can target road signs to deceive camera perception without interacting with the CAV [8]. Hence, the definition of a new attacker model is necessary. The remainder of this paper is organized as follows. First, Section 2 presents the perception lifecycle and allows the identification of assets. Then, Section 3 presents our attacker model derived from the identified assets. Accordingly, Section 4 proposes the corresponding security goals model. We discuss the feasibility of some attacks in Section 5. Finally, Section 6 concludes the paper.

CSCS 2018, September 13–14, 2018, Munich, Germany

ACM ISBN 978-1-4503-6616-8/18/09…$15.00
https://doi.org/10.1145/3273946.3273951

2 PERCEPTION LIFECYCLE

This section identifies the assets within the perception lifecycle. Figure 1 outlines the perception lifecycle which has two main components.

The first one is *Objects* which regroups:

- the perceiver of the perception system named *ego-vehicle*,
- the perceived entities named *Road Object*.

The second component is *Data Stages* which are the stages followed by the data through the perception lifecycle defined as follows:

- *Data Acquisition* is the transition of the physical signal (e.g., light intensity, radio wave, pulsed laser light, sound waves) between a detected road object and the ego-vehicle and its acquisition processes. It includes communication signals for V2X and measurement signals for ranging sensors. The acquisition processes include message encoding/decoding, security mechanisms (e.g., cryptographic verification) [9], object detection (e.g., Doppler Shift [10]), and object classification (e.g., dots and pixels clustering).
- *Data Processing* regroups the data fusion mechanisms applied to the acquired data such as association and/or tracking [11]. Their localization and their implementation within the Perception Lifecycle model vary among OEMs [2, 12, 13].
- *Data Storage* contains the data stored temporarily (e.g., tracks) or permanently (e.g., algorithms). Indeed, these data are a keystone in ensuring the monitoring (e.g., tracks) or the operation of the perception system (e.g., association algorithm).
- *Phenotype Data* is the observable traits of a *Road Object*, such as its morphology (e.g., dimensions), physiological properties

(e.g., color), behavior (object state over time), and behavior actions (e.g., human-made tags).

Figure 1 works independently of any communication protocols, sensors, or data fusion algorithms. Such abstraction exhibits the primary assets of a perception system to derive our attacker model.

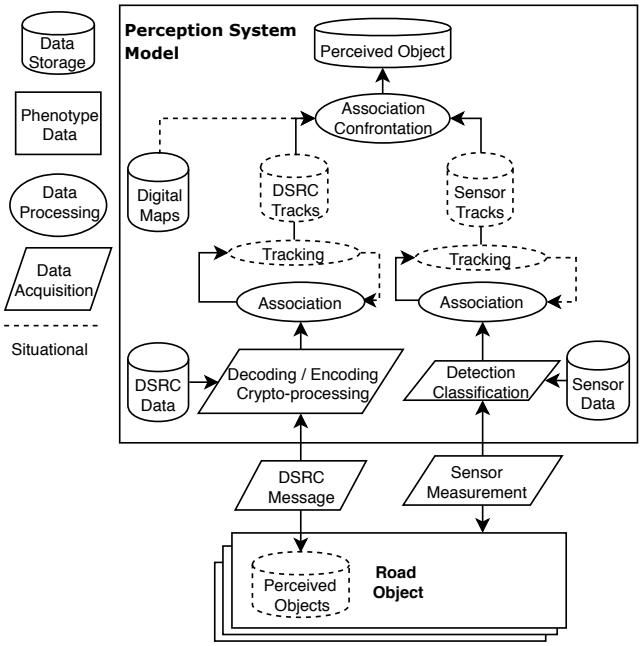

Figure 1: Perception Lifecycle Model

3 ATTACKER MODEL

This section defines the attacker model for the perception system of a CAV. First, we define a generic attacker model. Next, using the assets identified in Section 2, we describe specific attacker models for the perception system. To do so, we assume that cryptographic mechanisms yield against cryptanalytic attacks (e.g., message forgery or side channel attacks).

3.1 Generic Attacker Model Definition

Firstly introduced in VANETs [4] then extended for automotive sensors [1], a general attacker model defines attacker actions and potential targets. However, previous works assume that the attacker always reaches its goal directly which is false. Indeed, a malicious node can badmouth to neighboring nodes to provoke its victim exclusion from the network [14]. Also, the alteration of road sign impacts the vehicle perception indirectly [8, 15]. Thus, we propose a new generic attacker model with a five-dimensional set as follows:

- **Membership** stands for an *Insider* or an *Outsider* attacker. An *insider* attacker is an authenticated member of one or multiple CAV networks (e.g., CAN, LIN, VANET). Therefore, she can mount a diverse set of attacks using her given credentials. Whereas, an *outsider* is an unauthenticated member who can mount a limited set of attacks due to her restricted network access.
- **Motivation** stands for *Malicious* or *Rational*. A *malicious* attacker seeks no personal benefits from the attacks and aims to harm an asset. Whereas, a *rational* attacker seeks profit and thus is predictable regarding her attack means and target(s). Such attribute may help to define the financial severities of an attack in security risk analysis process [16]. For instance, a *rational* attacker will aim the perception algorithms contained in a CAV to sell them to hackers on the black market.
- **Scope** stands for *Local* or *Extended*. A *local* attacker controls few entities (e.g., car or traffic light [17]) within a limited scope (e.g., road intersection). However, an *extended* attacker controls several entities scattered across an extended scope (e.g., university campus [18]).
- **Method** stands for *Active* or *Passive*. While an *active* attacker must act to attack, a *passive* attacker simply listens or observes its target (e.g., network eavesdropping). For instance, in the context of standardized efforts towards the cooperation between safety and security risk analysis [16], a meteorological hazard could be a *passive* attacker.
- **Goal** stands for *Direct* or *Indirect*. A *direct* attacker reaches its primary target directly, whereas an *indirect* attacker reaches its primary target through secondary targets.

Table 1: Examples of similar attacks with different goal

Attacks \ Attacker Model	Membership	Motivation	Scope	Method	Goal
Alter road signs to fool sensors	Outsider	Malicious	Local	Active	Indirect
Alter road signs for "fun"					Direct
Camera blinding towards unperceived stop sign	Outsider	Malicious	Local	Active	Indirect
Camera Blinding for "fun"					Direct
Communication Badmouthing	Insider	Malicious	Both	Active	Indirect
Faulty Safety Message			Local		Direct

As depicted in Table 1, the attack goal helps to define attackers in the perception domain. However, *Goal* does not situate wherein the perception domain the attacker may perform an attack. Therefore, we need to specify *Goal* explicitly. To do so, we derive each sub-attacker model from the *Data Stages* identified in Section 2 (Table 2). We define these sub-attacker models and their attacker profiles in the following sections.

Table 2: Sub-Attacker Models in the Perception Lifecycle

Data Stage	Sensor Disrupter	Evil Mechanic	Malicious Communicator	Fusion Persuader
Phenotype	✓			✓
Acquisition	✓		✓	
Storage		✓		✓
Processing		✓	✓	✓

3.2 Sensor Disrupter

Sensor Disrupter is an attacker that aims at vehicle sensors. Indeed, CAV perception relies on the acquisitions of exteroceptive sensors (e.g., camera, lidar, or radar) to perceive the surrounding environment. Thus, sensors are assets which a *Sensor Disrupter* can disturb through various attack means.

3.2.1 ***Sensor Illusionist***. *Sensor Illusionists* target sensors *directly* during the *acquisition* stage. During this stage, ranging sensors (e.g., lidar) provide a closely real-time, trusted, and more or less accurate depiction of the surrounding by measuring the reflected physical signal [3]. However, at signal impact, an *Illusionist* can capture, delay, and replay it forcing the sensor to produce erroneous measurements. For instance, from the position of sensors target (the detected object), Petit et al. [19] captured, delayed, and replayed the lidar signal. Also, they relayed the signal and replayed it from a different position leading the way towards signal forgery attacks. Thus, *Illusionist* attack means include signal delay, relay, replay, and forgery.

3.2.2 ***Sensor Blinder***. Similarly, *Sensor Blinders* target exteroceptive sensors *directly* during the *acquisition* stage. During this stage, *Blinders* can alter the physical signal trajectory transiting between ego-CAV sensors and its surrogating environment. For instance, a camera cannot detect a facing traffic light state due to the parked vehicle blocking the view. Or, *Blinders* can maximize or minimize signal intensity to emit signals outside the sensing domain of the sensor [19, 20]. For instance, using fog light against an automotive camera is a realistic and accessible attack to perform.

3.2.3 ***Evil Sensor Calibrator***. *Evil Sensor Calibrators* target exteroceptive sensors *directly* during the *storage* stage.*Evil Calibrators* aims to modify sensor settings to provoke incorrect/missing measurements. Indeed, range sensors measure the distance between the *Road Object* and itself. Then, the measurement system of the sensor computes the absolute position by moving from the local referential base of the sensor to a global referential base. However, *Evil Calibrators* can modify the local referential base by changing the physical position, orientation, or internal settings of the sensor. Such actions lead to an incorrect perception of the *Road Object*. For instance, taking the case of *Lenticular Printing* attack which is an optical process used to create road signs that look different when viewed from different angles [15]. Sitawarin et al. demonstrated that if the localization of the camera used for road signs recognition is at a different height from the human controller, then the camera classifier performances are diminished while appearing to be correctly positioned to the human operator. Thus, an *Evil Sensor Calibrator* can drastically modify the sensor orientation to provoke an absence of measurements. Rarely mentioned, *Evil Sensor Calibrator* attacks remain easy to perform physically and may extend to other in-vehicle hardware (e.g., *Evil Mechanic* attacks).

3.2.4 ***Ground Truth Falsifier***. *Ground Truth Falsifiers* target exteroceptive sensors *indirectly* through *Road objects* at *phenotype* stage. *Falsifiers* physically alter *Road objects* (e.g., road signs) to provoke incorrect sensors measurement. For instance, *Falsifiers* can forge counterfeit road marks [21]. Therefore, an automotive camera can detect fake road marks as real ones which may influence vehicle trajectory. Also, the alteration of road signs known as ***D**eceiving **A**utonomous ca**R**s with **T**oxic **S**igns* (DARTS) leads to camera misclassification from the camera which may affect vehicle dynamic [8, 15]. Thus, mentioned attacks are *indirect Illusionists* attacks.

Finally, the massive alteration of a *Road Object* can provoke an acquisition absence. Indeed, *Falsifiers* can destroy, remove, or severely deface road infrastructures. Therefore, mentioned attacks are *indirect Blinder* attacks.

3.3 Evil Mechanic

As depicted in Figure 1, the perception lifecycle takes place mostly within the *ego-CAV*. Each ECU performs an automotive function (e.g., powertrain, infotainment, body, chassis, safety) by collecting and *processing data* from various sources such as sensors and ECUs. Therefore, attacking *processing data* is valuable for an attacker willing to force the CAV into a wrong assessment or to extract valuable data (e.g., data fusion algorithms). Related attack sets are *In-vehicle Manipulator* and *In-vehicle Miner*.

3.3.1 ***In-vehicle Manipulator***. aims to add, modify, or remove automotive components or data contained in it. Indeed, an attacker with elevated physical access (e.g., mechanic) could easily replace a smart camera by one with a dysfunctional detection algorithm. Although the camera is recording, its detection capabilities are abnormal which may catch off-guard the driver. Besides safety, the removal or injection (e.g., odometer manipulation [7]) of vehicle history permits data repudiation. Therefore, a vehicle owner can repudiate facts in case of fraud insurance, resale, or crime investigation because the falsified vehicle history confirms her statement. Moreover, the intentional manipulation of tamper-resistant automotive equipment [7, 22] may activate defense mechanisms that erase all the data contained in such hardware which, thus, benefits to the attacker. Finally, a mechanic can flash equipment with a modified firmware to increase her attack range [23]. Therefore, a malware installation in this equipment allows the injection of CAN message with incorrect content without requiring the mechanic to remain plugged into the vehicle.

3.3.2 ***In-vehicle Miner***. eavesdrops in-vehicle data for personal deeds. For instance, a *Miner* can sell the vehicle history to third parties (*rational attacker*). Indeed, robbers can use the sole localization history to identify the driver routine and rob her house. Moreover, eavesdropping *Storage* and *Processing* steps help to analyze the behavior of perception algorithms. Once reviewed, this information is valuable to *Sensor Disrupter*, *Malicious Communicator*, or *Fusion Persuader*.

3.4 Malicious Communicator

As introduced, V2X communications aim to improve vehicular automation reliability, safety, and traffic efficiency. Like in all social group, some participants behave against the interest of the community. Such behaviors threaten communication. We define such attacker as *Malicious Communicator* which regroups *Fully Adversarial Networking*, *Voyeur* and *Communication Deceiver*.

3.4.1 ***Fully Adversarial Networking***. is an attacker who inserts arbitrary messages and performs selective Denial Of Service attack [7].

3.4.2 ***Voyeur***. is an attacker that surveys anonymous public data exchanged in cooperative ITS to obtain confidential data (e.g., car owner identity). For instance, in VANET, localization and trajectory of the vehicle are willingly broadcast. Indeed, cooperative

awareness through communication requires a frequent update of surrounding vehicles localization. Therefore, it is mandatory to be able to track vehicles locally. However, *Voyeurs* can use tracking to track broadcasting vehicles in a neighborhood or a campus [18]. By tracking vehicle localization contained in the messages, the attacker extracts private data such as preferred driving path, house localization, children localization, or health status (e.g., hospital, gym, fast-food). After being processed, the anonymous data allow extracting confidential information such as vehicle owner identity using its house localization [24].

3.4.3 ***Communication Deceiver.*** refers to authentic messages with erroneous content. For instance, a malicious traffic light can send incorrect Signal Phase and Timing (SPaT) messages with a color state which differs from the *phenotypic* state. At best, it creates two different outputs which confuse the automated driving system. At worst, if the real state color is unavailable (e.g., NLoS), the system relies on a single incorrect output from the SPaT. Another example of erroneous message content is the definition of a node dimension for the standardized *Cooperative Awareness Message* [25]. Indeed, the absence of correlation between the class of a V2X node (e.g., pedestrian) and the node dimensions could allow *Communication Deceivers* to emit a message defining an object with an implausible size. Therefore, a pedestrian node may have a length that is between 10 centimeters and 102 meters. Despite some standards recommendations, the choice of plausibility mechanisms regarding V2X Data are left open. Thus, if these erroneous content remain unchecked that may lead to some mis-associations between a V2X message and a sensor measurements.

3.4.4 ***OTA Poisoner.*** refers to an entity that sends any malicious updates Over-The-Air (OTA). Indeed, CAVs will update OTA their software, firmware, *Data Storage* to fix vulnerabilities, inaccurate information, bugs [26]. A malicious update can alter the integrity of the *Data Storage* by modifying the processing algorithms (e.g., cryptographic algorithms) or the perception data (e.g., cartography data).

3.5 Fusion Persuader

Persuaders disrupt the *processing* and *storage* stages to disable or to deceive the perception system. *Persuaders* can perform the followings attacks:

3.5.1 ***Misbehaving Ground Truth.*** is a road object behaving against the CAV mission (e.g., pedestrian crossing the road at red or traffic light blocked on a red state). These attacks have safety, functional, financial, and privacy impacts on the system. Indeed, a pedestrian faking a collision can block the CAV, extort money from the car company/driver, or provoke an emergency braking threatening passenger safety [27]. Such a behavior questions the need to register and report such actions using the camera recording as juridical proof. Indeed, the recording and storage of identifiable traits of an individual may imply some privacy issues.

3.5.2 ***Sybil Gating.*** The *Gating* process is a filtering/screening mechanism to determine which objects observations (e.g., V2X messages or sensor measurements) are valid candidates to update existing objects tracks. *Gating* aims primarily to reduce unnecessary computation during data association and tracks maintenance processes [11]. Therefore, an attacker could create valid virtual candidates to increase the computation load of *Data Processing*. Although lidar spoofing is possible [1], its feasibility in dense or/and highly dynamic scenario may be unrealistic. Indeed while the targeted vehicle is moving fast or is highly surrounded by *Road Objects*, aiming its lidar to achieve a detection is challenging. However, the creation through V2X communication of ghost vehicles [28] fitting the gate conditions is achievable. Therefore, the attacker could create *Sybil Attacks* to disable the filtering benefits of the *Gating*. We define such attack as *Sybil Gating* (Figure 2).

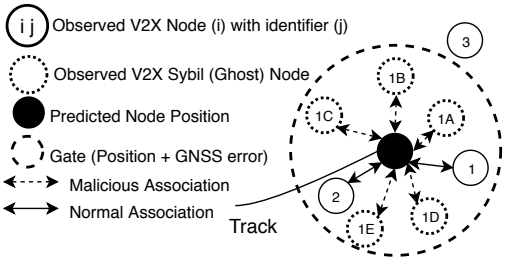

Figure 2: Sybil Gating

3.5.3 ***Tracking Poisoner.*** Tracking algorithms aim to predict the state of an object at the next step according to the measurement of object state at the current step. Thus, the system must update each track of its tracking database to ensure the next prediction [11]. However, it remains unclear how to perform the track management in pseudonymous V2X communication [29]. According to the European Certification Policy [30], a vehicle can contain simultaneously valid pseudonymous certificates. As mentioned, a vehicle can create a ghost vehicle per pseudonymous certificates. Without proper trustworthiness mechanisms, the ego-CAV will have its tracking database poisoned by tracks of ghost vehicles (Figure 3).

Thus, we called such attack *Tracking Poisoning*. Also, such attacks require to adapt existing tracks update mechanisms. Indeed without proper tracks update, these attacks could impact the association process which aims to find the most plausible acquisition-to-track association.

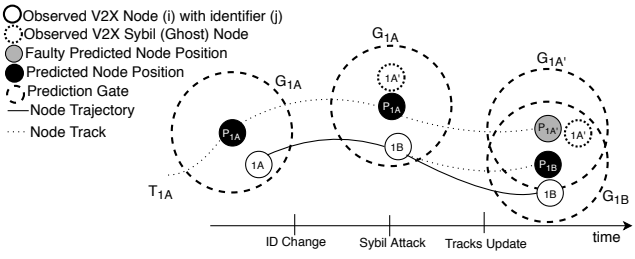

Figure 3: Tracking Poisoner

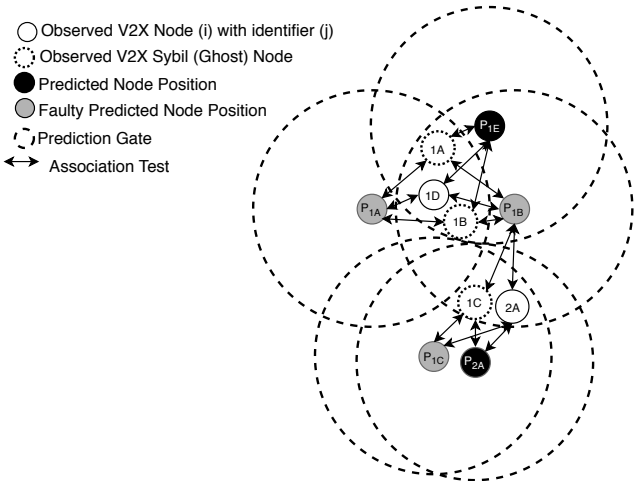

Figure 4: Association Manipulator Attack

3.5.4 ***Fusion Manipulator.*** Association algorithms aim to search the most likely acquisition from each source observation set (e.g., sensor measurements or V2X messages) that share the same detected object as subject. Therefore, a *Fusion Manipulator* has multiple ways to manipulate the association process.

First, an attacker can increase the computation time by increasing the number of potential measurements (*Sybil Gating* or/and *Tracking Poisoner*). For instance, Merdrignac et al., [3] proposed a perception system associating lidar measurements and V2X messages. Despite working with few connected pedestrians, their system does not scale in dense scenarios. Indeed, the perception system associates each lidar observations within the *Gating* area, which is the Global Navigation Satellite System (GNSS) error defined as a circle of 5 meters radius around the GNSS position of the pedestrian emitting V2X messages. Thus, we can assume that the increasing number of potential associations between lidar measurements-V2X Tracks in an urban scenario leads to an increase in the association time.

Second, an attacker can provoke conflicting *acquisition* between two acquisitions. For instance, let us consider a Green Light Optimal Speed Advice (GLOSA) system that uses camera acquisition to verify the content correctness of a SPaT message [2]. A *Misbehaving Ground Truth* attacker alters the physical signal state of a traffic light [17]. Hence, the perception system would disapprove all SPaT message thinking that the camera acquisition represents the correct state.

4 SECURITY GOALS MODEL

To secure a CAV perception, both identifying and defining proper security goals against identified attackers are mandatory. This section defines such security goals. First, we identify security countermeasures based on the attackers defined in Section 3. Then, we derive the security goals model from the identified countermeasures. Finally, we evaluate the model against standardized models.

4.1 Attackers Countermeasures

This section identifies security countermeasures for each attacker defined in Section 3.

4.1.1 ***Security Goals for Sensor Disrupter.*** To do so, we define the security goals for a *Sensor Disrupter* which regroups *Sensor Illusion*, *Sensor Blindness*, and *Evil Sensor Calibrator*.

Sensor Illusion requires mechanisms which assess the trustworthiness of sensor measurements. Approaches checking the measurement *consistency* assume that a ghost (e.g., Radar) or spoofed measurements are not or hardly repeatable. Therefore, the use of metrics such as *Object Existence* [13] computed on the object past detections allows to down-weight newly appearing and inconsistent objects during the fusion process.

Sensor Blindness attacks target the *availability* of sensor measurements. Therefore, it is crucial to ensure the *redundancy* of *Data Acquisition*. Current solutions include hardware redundancy (redundant sensor) or data redundancy (different sensor type). For instance, SPaT messages can provide an accurate state of a traffic light while the camera is under *Sensor Blindness*.

Evil Sensor Calibrator targets the sensor *integrity*. Therefore, the security goal to ensure is physical *integrity*. Indeed, a sensor should not be easily manipulated or moved. For instance, tamperproof hardware can store valuable data (e.g., detection algorithms). Also, the access to the sensor settings must be restricted. Thus, *Access Control* is mandatory to identify and to authenticate authorized personnel. Hence, binding authorized actions to a person profile limits its actions on the sensor according to its function (e.g., developer, mechanic).

Ground Truth Falsifier requires to harden the physical structure of *Road Objects* to ensure their *Phenotype integrity* and *availability*. For instance, the use of anti-graffiti coatings is a solution to avoid the alteration of road signs.

Overall, *Accountability* is a significant security goal against a *Sensor Disrupter*. Indeed, the sensor inability to perform its task must be recorded to understand the causes of misperception. For instance, if the radar detects an object forward but the camera does not, an analysis of the images recorded by the camera can explain that the mis-detection was due to the dense fog which blinded the camera. However, the global acceptance of this mechanism is unsure due to privacy concerns. For instance, the recording of a person face to identify and punish the author of road marks forgery is a possible option [21]. But, it requires a strict Privacy Policy regarding the data recorded by the camera of a road infrastructure or a CAV.

4.1.2 ***Security Goals for a Malicious Communicator.*** As mentioned and depicted in Figure 5, previous work defined current VANET security goals following *STRIDE* [31] or *CIA* models. Such goals (e.g., authenticity) are sufficient to define standardized security countermeasures in fully adversarial network conditions [9]. For instance, these mechanisms include cryptographic authentication of all messages to exclude unauthorized participation [32] or message semantic analyzer (e.g., ASN.1 encoding) to exclude unauthentic message formats [33].

However, against recent attackers such as a *Voyeur*, standardized countermeasures are ineffective. Thus, the need for new security

goals is necessary. Although mentioned [7, 19], the need for *Linkability* and *Anonymity* as distinct security goals remain unsettled. Despite *Privacy* recommendations from the European C-ITS Platform group [30], efficient countermeasures such as pseudonym change in V2X communication are still unsolved [34]. Despite common belief, it is the association algorithm and not tracking that decides whether two observations (e.g., track, sensor measurement or V2X message) belong to the same observed object [11]. In the sensor domain, range sensor measurements may be incorrect and anonymous. Therefore, it is essential to define the temporal window between two messages which disallows an association algorithm to match two observations based on just their dynamic state.

Communication Deceiver requires the use of *trustworthiness* countermeasures which regroups:

- *Consistency* mechanisms that check how often the emitting node state deviates from the predicted normal behavior (e.g., Kalman Filter [29]). Unlike *Voyeur Linkability*, V2X messages linkability is necessary for automotive perception. Indeed to track the V2X node state, the association algorithm must associate each V2X messages to its corresponding tracks.
- *Plausibility* mechanisms which rely on plausibility rules (e.g., maximal emitting distance) [29], multi-source checking [2], or single source various means checking [10]. The latter compares the object state contained in the V2X message to the measured object state from the communication radio wave (e.g., Doppler Shift).
- *Reputation* mechanisms which rely on the computation of trust score relative to a V2X node behavior [14]. The *Scope* of node trust can be global or local. *Local trust* implies that the trust value of a node is computed in the vehicle using trust mechanisms (e.g., *Consistency* and *Plausibility* mechanisms). *Local trust* defines a subjective opinion of the perception system towards a V2X node and therefore should not be extended in a cooperative system to avoid badmouthing attacks [14]. Whereas, global trust values are computed by a global authority (Public Key Infrastructure) and acknowledged by all authenticated VANET members. A specific authority of the Public Key Infrastructure (e.g., Misbehavior Authority) collects misbehavior reports and decides on revocation of node [35].

Overall, *Malicious Communicators* also require the following security goals:

- *Accountability* is mandatory to report and revoke malicious nodes.
- *Adaptability* is a major security goal for communication. Indeed, most of the related work assume that cryptographic algorithms will ensure security goals such as *Confidentiality* and *Integrity*. However, few questioned the algorithms obsolescence due to advances in quantum computing. Therefore, without a backup plan, communication system relying on Public Key Infrastructures based on these algorithms are vulnerable [36]. The need to define a system able to adapt by supporting other algorithms in case of such attacks becomes mandatory. For instance, the SCMS PKI uses such system thanks to the integration of specific authorities named Elector CAs [35].

4.1.3 ***Security Goals for an Evil Mechanic.*** *Evil Mechanics* are difficult to counter because non-expert can hardly detect the malicious actions of an expert. However, some security requirements can be implemented to prevent such attacks.

Security goals against *In-vehicle Manipulator* attacks include *Integrity*, *Availability*, *Access Control*, and *Non-Repudiation*. To perform *In-vehicle Manipulator* attacks, an *Evil Mechanic* will first try to access the hardware or the data. Therefore, *Access Control* mechanisms are important to ensure that only authorized personnel can access the data. For instance, such mechanisms include multiple *authentication* factors to ensure that the personnel or installed programs are authorized to access such data. *Authorization* mechanisms restrict actions from an *Evil Mechanic* or malware. Instructions to modify *Data Storage* should be signed using asymmetric cryptography to avoid communication alteration, hardware replacement or the spoofing of administrator session. Also, *Integrity* mechanisms mandatory to avoid the removal of any hardware components and ensure the overall *availability* of the perception system. Finally, *Accountability* mechanisms (e.g., events logs) is mandatory to monitor actions performed a hardware and its data. For instance, during a hardware replacement, the hardware logs indicate if it is new or already used.

Security goal against *In-vehicle Miner* attacks focus on *Confidentiality*. As mentioned, *Data Storage* contains valuable information such as private information or fusion algorithm. Therefore, they should be encrypted.

4.1.4 ***Security Goals for a Fusion Persuader.*** We define security goals for a *Fusion Persuader* that require the following *Trustworthy* mechanisms:

- *Consistency mechanisms* that detect a potential deviation between the estimated state and the observed state of a data source.
- *Plausibility mechanisms* that confront multiple data sources and detect disagreements among sources. For instance, the disagreement regarding a traffic light state between a camera recording and a SPaT message will raise an anomaly report [2].
- *Reputation mechanisms* that compute the opinion value of the perception system regarding a perceived *Road Object* by using sensor confidence and V2X node trust metrics. The former assigns a weight to the sensor observations based on sensor past performances such as the number of successful detection of a *Road Object*. The latter is the trust computed based on the detection number of malicious messages emitted by a V2X node.

Also, *Accountability* mechanisms require to record every conflict between data sources that occurs during the fusion process. The aftermath goal is to provide meaningful reports to the Misbehavior Authority [9]. For instance, law enforcement authorities or insurance companies can request these reports to verify the events occurred in an accident. But also, it could help OEMs to detect, understand, and improve vehicles automation. Experts can extract events logs and misbehavior reports to reconstruct the road scene and correct potential weaknesses in the cooperative perception.

Also, *Freshness* mechanisms are mandatory to update the tracks database. For instance, the temporal freshness of tracks is a criterion

to remove ghost tracks caused by *Sybil Gating* attacks or outdated *Road Object* tracks that are out of the perception range.

Finally, *Adaptability* mechanisms require to patch the fusion algorithms against potential undiscovered weaknesses during the vehicle lifetime. Moreover, in the case of a detected faulty/malicious source of acquisition, the system can only rely on its communication mode or on its local sensors mode to achieve perception [37].

4.2 Proposed Model

Figure 5 presents our security goals model derived from the attackers security goals identified previously. This model includes standardized security goals used and already defined in VANET [4, 16] and threat models for computer networks[31]. But Also, newly identified security goals that we define as follow:

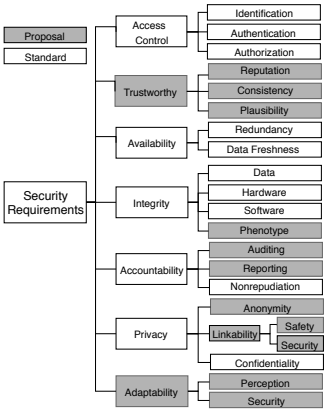

Figure 5: Security Goals Model for CAVs

Privacy is the degree to prevent unauthorized parties to obtain sensitive information. Note that *Privacy* includes *Confidentiality* because sensitive information does not only imply private data but also confidential data (e.g., source code) [38].

- *Anonymity* is the degree of identity disclosure of data users. Thus, *Pseudonymity* is one degree of anonymity that uses pseudonyms (e.g., pseudonym certificate) to identify users.
- *Linkability* is the degree of linking anonymous or pseudonymous data to their owner risking a potential disclosure of its private identity (e.g., home localization).

Trustworthy is the degree of trust assessed by the system regarding perceived *Road Objects* and perception data (Section 2). Trustworthy mechanisms rely on reputation, consistency, plausibility security goals.

- *Reputation* is the perception system opinion of a V2X system entity. This opinion is subjective. Its validity domain ranges from local to global.
- *Consistency* is the degree of temporal plausibility of a *Road Objects* behavior or products of behavior assessed by the perception system along the perception lifecycle (Section 2).
- *Plausibility* is the degree to which the system verifies that the perceived data are consistent with the ground truth (Section 2). As mentioned, other acquisition sources, *Road Objects* model, maximum-minimum thresholds, or Highway Code can be system ground truth assuming they are trustworthy.

Phenotype Integrity is the degree of protection of the *Phenotype* of a *Road Object* from malicious alterations.

Accountability is the degree of mapping security-related events to system entities.

- *Non-repudiation* is the degree of actions recognition of the entity that performs it.
- *Reporting* is the degree of recording Non-repudiated actions.
- *Security Auditing* is the degree of prevention, analysis, and evaluation of occurring, occurred, and potential security-events within a system.

Adaptability is the degree of attack recovery and defense of a system against future similar attacks.

4.3 Model Evaluation

This section evaluates our security goal model through the comparison Table 3.

Table 3: Security Goals Model Evaluation

Object	Data Stage	Attacker Model	Security Goals	STRIDE	CIA	Proposal
Ego-CAV	*Acquisition*	*Voyeur, Fully Adversarial, Sensor Blindness, Sensor Illusion*	Access Control	≈	✘	✓
			Trustworthy	✘	✘	✓
			Availability	✓	✓	✓
			Integrity	✓	✓	✓
			Accountability	≈	✘	✓
			Privacy	≈	≈	✓
			Adaptability	✘	✘	✓
	Processing	*Fusion Manipulator, Communication Deceiver, Sybil Gating*	Availability	✓	✓	✓
			Trustworthy	✘	✘	✓
			Accountability	≈	✘	✓
			Adaptability	✘	✘	✓
	Storage	*In-vehicle Miner, In-vehicle Manipulator, Tracking Poisoner, Evil Sensor Calibrator, OTA Poisoner*	Access Control	≈	✘	✓
			Availability	✓	✓	✓
			Integrity	✓	✓	✓
			Accountability	≈	✘	✓
			Privacy	≈	≈	✓
Road Object	*Phenotype*	*Ground Truth Falsifier, Misbehaving Ground Truth*	Availability	✓	✓	✓
			Integrity	✓	✓	✓

≈: the full security goal is not covered as depicted in Figure 5

First, we build this table. We define the *Target of Evaluation* which is the perception domain (Section 2). Then, we set the involved entities which are *Objects* which regroup *Ego-CAV* and *Road Object*. Then, we link each *Object* to its *Data Stages* (Figure 1). This approach avoids speculating on the chosen architecture for data fusion. Indeed, acquisitions tracking is either decentralized (acquisitions source) or centralized (fusion ECU) [11]. Accordingly, we relate each *Data Stages* to a sub-attacker (Table 2). Finally, we match to each sub-attackers its security goals (Section 4.1).

Second, we compare our proposal to standardized security goal models such as *STRIDE* and *CIA*. Where *STRIDE* stands for *Spoofing*, *Tampering*, *Repudiation*, *Information Disclosure*, *Denial of Service*, and *Elevation of Privilege*. *CIA* stands for *Confidentiality*, *Integrity*, and *Availability*. Therefore, both do not consider *Trustworthiness* and *Adaptability* as security goals. An explanation is that both models were designed for traditional IT environment and not for the CAV domain. Also, both do not distinguish *Authentication* and *Identification* which is not acceptable in the CAV domain. Indeed, in the case of Sybil attack on V2X nodes, the system allows a single identity to authenticate itself using multiple authenticators (e.g., Pseudonyms). Finally, *STRIDE* refers to *accountability* only through

non-repudiation. However, CAV domain will rely on trust between entities and therefore will need security reports from CAVs to report malicious *Data Objects*.

That is, we showed that current models despite being standardized do not answer accurately to the current threats in the CAV domain, unlike our proposal.

5 DISCUSSION

This paper presented an attacker model and attack types. Our attacker model can also help designers and testers to verify the security requirements of their perception system. As mentioned, while current work demonstrated their feasibility, Section 3.5 highlighted potential attacks in the domains of tracking and data fusion that have to be demonstrated.

Current implementations of automotive perception that rely on V2X and sensor acquisitions never assumed V2X data as a threat [2, 3, 12]. Despite security work related to V2X tracking to detect inconsistencies [29, 39, 40], none studied the impact of malicious tracks in the tracking database and their impacts on the tracks managements and the perception fusion. Indeed unlike lidar [19], a location spoofing of V2X message can produce consistent tracks due to their easy repeatability [29, 39]. Therefore, unlike track management for sensors that removes echoes track after a defined duration, *Communication Deceiver* can maintain a false track by frequently emitting plausible V2X messages. Thus, it can be interesting to study the feasibility and scalability levels of a *Tracking poisoner* (e.g., number of false tracks, living duration of false tracks).

Also, the attacks towards fusion process (gating and association) are plausible assuming the absence of security counter-measurements in cooperative perception system. An explanation is that fusion algorithms are sensor-designed algorithms and do not consider V2X attacks. Thus, some fusion algorithms may not scale [41, 42] and an attacker could exploit such vulnerability by increasing the number of objects to process.

Additionally, these attacker models raise the need to define a secured framework for automotive perception relying on both sensors and V2X. However, current fusion processes are different for each acquisition source. Indeed, while the association of sensors observations relies on the observed object state (e.g., localization, velocity, class), V2X association relies on object meta-data (e.g., identifier [43, 44], trust [14]) which do not exist for sensors. Thus, the V2X-sensor association remains a challenge due to potential incorrect data in the V2X message that can lead to malicious misassociations.

Thus, our model validation requires the definition and implementation of a secured perception framework using multi-source observation. A starting point could be the V2X plausibility framework of Sun et al. [10] combined with the multi-source fusion framework of Van der Heijden et al. [45].

Finally, the generic attacker model (Section 3.1) can adapt to the attacker model of Ponikwar et al. [6] for VANET attacks. Both are extension of the same model [4]. However, for automotive perception attacks, our model is an extension of Petit et al. [7].

6 CONCLUSION

Secured automotive perception system is the next goal for reliable perception. However current attackers models did not capture the entire attack and security goal domains. Indeed, they did not describe the data lifecycle within a perception system. Therefore, it was unclear what were the most likely attack scenarios to fool automotive perception systems.

First, we described a data lifecycle within generic perception system model from which we identified its primary assets. Therefore, we derived an attacker model based on such assets and state of the art attacks. Following, we determined related countermeasures then accordingly we defined a security goals model. Finally, we evaluated such model against standardized models and highlighted missing security goals.

As a result, this paper showed the need for costless and straightforward countermeasures against attacks performed on the surrounding environment. Also, despite the use of pseudonym certificate, we explained the need to investigate privacy mechanisms furthermore against Voyeur attacker. Overall, we demonstrated that sensor and V2X data are untrustable and may lead to new attacks within data fusion processes which were not designed for an uncooperative environment. Therefore, we explained the need to revisit such processes which led to the identification of three trustworthy sub-goals. Also, we showed the current lack of adaptability countermeasures of a perception system which remains an unsettled issue. Indeed, few works analyzed the obsolescence of perception algorithms such as the break of cryptographic algorithms. Finally, by focusing on the automotive perception, we demonstrated that current tools for threat analysis are insufficient.

To conclude, we believe standardizing automotive perception will help security experts to deepen existing automotive security analysis. As a future work, we will define and implement a general trustworthy scheme for the automotive perception.

ACKNOWLEDGMENTS

This work was supported by a PhD Grant from French ANRT (Association Nationale de la Recherche et de la Technologie). The statements made herein are solely the responsibility of the authors. They have not been formally adopted and should not be considered as an official statement of *Groupe PSA*.

REFERENCES

[1] Jonathan Petit and Steven E Shladover. Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):546–556, 2015.

[2] Zaydounr Y Rawashdeh, Trong-Duy Nguyen, Anoop Pottammal, and Rajesh Malhan. Comfortable automated emergency brake for urban traffic light based on dsrc and on-board sensors. Technical report, SAE Technical Paper, 2017.

[3] Pierre Merdrignac, Oyunchimeg Shagdar, and Fawzi Nashashibi. Fusion of perception and v2p communication systems for the safety of vulnerable road users. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1740–1751, 2017.

[4] Maxim Raya and Jean-Pierre Hubaux. Securing vehicular ad hoc networks. *Journal of computer security*, 15(1):39–68, 2007.

[5] Tim Leinmuller, Robert K Schmidt, Elmar Schoch, Albert Held, and Gunter Schafer. Modeling roadside attacker behavior in vanets. In *GLOBECOM Workshops, 2008 IEEE*, pages 1–10. IEEE, 2008.

[6] Christoph Ponikwar, Hans-Joachim Hof, Smriti Gopinath, and Lars Wischhof. Beyond the dolev-yao model: Realistic application-specific attacker models for applications using vehicular communication. *arXiv preprint arXiv:1607.08277*, 2016.

[7] Jonathan Petit, Michael Feiri, and Frank Kargl. Revisiting attacker model for smart vehicles. In *Wireless Vehicular Communications (WiVeC), 2014 IEEE 6th International Symposium on*, pages 1–5. IEEE, 2014.

[8] Ivan Evtimov, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati, and Dawn Song. Robust physical-world attacks on machine learning models. *CoRR*, abs/1707.08945, 2017.

[9] IEEE standard for wireless access in vehicular environments–security services for applications and management messages - amendment 1. *IEEE Std 1609.2a-2017 (Amendment to IEEE Std 1609.2-2016)*, pages 1–123, Oct 2017.

[10] Mingshun Sun, Ming Li, and Ryan Gerdes. A data trust framework for vanets enabling false data detection and secure vehicle tracking. In *Communications and Network Security (CNS), 2017 IEEE Conference on*, pages 1–9. IEEE, 2017.

[11] Samuel Blackman and Robert Popoli. Design and analysis of modern tracking systems(book). *Norwood, MA: Artech House, 1999.*, 1999.

[12] Ting Yuan, Krishanth Krishnan, Qi Chen, Jakob Breu, Tobias B Roth, Bharanidhar Duraisamy, Christian Weiss, Michael Maile, and Axel Gern. Object Matching for Inter-Vehicle Communication Systems – An IMM-Based Track Association Approach With Sequential Multiple Hypothesis Test. *IEEE Transactions on Intelligent Transportation Systems*, 18(12):3501–3512, 2017.

[13] Michael Aeberhard. *Object-level Fusion for Surround Environment Perception in Automated Driving Applications*. VDI Verlag GmbH, 2017.

[14] Heng Chuan Tan, Maode Ma, Houda Labiod, Peter Han Joo Chong, and Jun Zhang. A non-biased trust model for wireless mesh networks. *International Journal of Communication Systems*, 30(9), 2017.

[15] Chawin Sitawarin, Arjun Nitin Bhagoji, Arsalan Mosenia, Mung Chiang, and Prateek Mittal. Darts: Deceiving autonomous cars with toxic signs. *arXiv preprint arXiv:1802.06430*, 2018.

[16] SAE Vehicle Electrical System Security Committee et al. SAE J3061-cybersecurity guidebook for cyber-physical automotive systems. *SAE-Society of Automotive Engineers*, 2016.

[17] Branden Ghena, William Beyer, Allen Hillaker, Jonathan Pevarnek, and J Alex Halderman. Green lights forever: Analyzing the security of traffic infrastructure. *WOOT*, 14:7–7, 2014.

[18] Jonathan Petit, Djurrre Broekhuis, Michael Feiri, and Frank Kargl. Connected vehicles: Surveillance threat and mitigation. *Black Hat Europe*, 11:2015, 2015.

[19] Jonathan Petit, Bas Stottelaar, Michael Feiri, and Frank Kargl. Remote attacks on automated vehicles sensors: Experiments on camera and lidar. *Black Hat Europe*, 11, 2015.

[20] Chen Yan, Wenyuan Xu, and Jianhao Liu. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle. *DEF CON*, 24, 2016.

[21] News from Elsewhere... Man fined for painting road signs to aid his commute, Nov 2017.

[22] Marko Wolf and Timo Gendrullis. Design, implementation, and evaluation of a vehicular hardware security module. In *International Conference on Information Security and Cryptology*, pages 302–318. Springer, 2011.

[23] Charlie Miller and Chris Valasek. Can message injection. *OG Dynamite Edition June*, 28, 2016.

[24] Björn Wiedersheim, Zhendong Ma, Frank Kargl, and Panos Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. In *Wireless On-demand Network Systems and Services (WONS), 2010 Seventh International Conference on*, pages 176–183. IEEE, 2010.

[25] EN ETSI. 302 637-2 v1. 3.2,âĂIJ. *Intelligent transport systems (ITS)*.

[26] Alex Brisbourne. Teslas over-the-air fix: Best example yet of the internet of things? *Wired Magazine February*, 2014.

[27] Didi Kirsten Tatlow. Scammers in china fake road injuries, but cameras capture the truth, Feb 2017.

[28] Philippe Golle, Dan Greene, and Jessica Staddon. Detecting and correcting malicious data in vanets. In *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 29–37. ACM, 2004.

[29] Attila Jaeger, Norbert Bißmeyer, Hagen Stübing, and Sorin A Huss. A novel framework for efficient mobility data verification in vehicular ad-hoc networks. *International Journal of Intelligent Transportation Systems Research*, 10(1):11–21, 2012.

[30] C-ITS Platform group. Certificate policy for deployment and operation of european cooperative intelligent transport systems (c-its). Technical report, C-ITS Platform, 2017.

[31] Michael Howard and David LeBlanc. The stride threat model. from the book writing secure code, 2002.

[32] JP Monteuuis, Badis Hammi, Eduardo Salles, Houda Labiod, Remi Blancher, Erwan Abalea, and Brigitte Lonc. Securing pki requests forC-ITS systems. In *Computer Communication and Networks (ICCCN), 2017 26th International Conference on*, pages 1–8. IEEE, 2017.

[33] Badis Hammi, Jean Philippe Monteuuis, Eduardo Salles Daniel, and Houda Labiod. Asn. 1 specification for ETSI certificates and encoding performance study. In *Mobile Data Management (MDM), 2017 18th IEEE International Conference on*, pages 291–298. IEEE, 2017.

[34] Jonathan Petit, Florian Schaub, Michael Feiri, and Frank Kargl. Pseudonym schemes in vehicular networks: A survey. *IEEE communications surveys & tutorials*, 17(1):228–255, 2015.

[35] Benedikt Brecht, Dean Therriault, André Weimerskirch, William Whyte, Virendra Kumar, Thorsten Hehn, and Roy Goudy. A security credential management system for V2X communications. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[36] William Whyte and Zhenfei Zhang. Quantum cryptanalysis, quantum-safe algorithms based on hard problems over lattices, and how we get there from here. 2016.

[37] Mohammad Y Abualhoul, Pierre Merdrignac, Oyunchimeg Shagdar, and Fawzi Nashashibi. Study and evaluation of laser-based perception and light communication for a platoon of autonomous vehicles. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 1798–1804. IEEE, 2016.

[38] Donald Firesmith. Specifying reusable security requirements. *Journal of Object Technology*, 3(1):61–75, 2004.

[39] Norbert Bismeyer, Sebastian Mauthofer, Kpatcha M Bayarou, and Frank Kargl. Assessment of node trustworthiness in vanets using data plausibility checks with particle filters. In *Vehicular Networking Conference (VNC), 2012 IEEE*, pages 78–85. IEEE, 2012.

[40] Hagen Stübing, Jonas Firl, and Sorin A Huss. A two-stage verification process for car-to-x mobility data based on path prediction and probabilistic maneuver recognition. In *Vehicular Networking Conference (VNC), 2011 IEEE*, pages 17–24. IEEE, 2011.

[41] Pierre Merdrignac. *Système coopératif de perception et de communication pour la protection des usagers vulnérables*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 2015.

[42] Karim Emara, Wolfgang Woerndl, and Johann Schlichter. On evaluation of location privacy preserving schemes for vanet safety applications. *Computer Communications*, 63:11–23, 2015.

[43] Stefan Dietzel, Rens van der Heijden, Hendrik Decke, and Frank Kargl. A flexible, subjective logic-based framework for misbehavior detection in v2v networks. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2014 IEEE 15th International Symposium on a*, pages 1–6. IEEE, 2014.

[44] Stefan Dietzel, Rens van der Heijden, Jonathan Petit, and Frank Kargl. Context-adaptive detection of insider attacks in vanet information dissemination schemes. In *Vehicular Networking Conference (VNC), 2015 IEEE*, pages 287–294. IEEE, 2015.

[45] Rens Wouter van der Heijden, Henning Kopp, and Frank Kargl. Multi-source fusion operations in subjective logic. *arXiv preprint arXiv:1805.01388*, 2018.