REDEM: Real-time Detection and Mitigation of Communication Attacks in Connected Autonomous Vehicle Applications*

Srivalli Boddupalli and Sandip Ray

Department of Electrical and Computer Engineering
University of Florida at Gainesville
bodsrivalli12@ufl.edu, sandip@ece.ufl.edu

Abstract. Emergent vehicles will support a variety of connected applications, where a vehicle communicates with other vehicles or with the infrastructure to make a variety of decisions. Cooperative connected applications provide a critical foundational pillar for autonomous driving, and hold the promise of improving road safety, efficiency and environmental sustainability. However, they also induce a large and easily exploitable attack surface: an adversary can manipulate vehicular communications to subvert functionality of participating individual vehicles, cause catastrophic accidents, or bring down the transportation infrastructure. In this paper we outline a potential direction to address this critical problem through a resiliency framework, REDEM, based on machine learning. REDEM has several interesting features, including (1) smooth integration with the architecture of the underlying application, (2) ability to handle diverse communication attacks within the same underlying foundation, and (3) real-time detection and mitigation capability. We present the vision of REDEM, identify some key challenges to be addressed in its realization, and discuss the kind of evaluation/analysis necessary for its viability. We also present initial results from one instantiation of REDEM introducing resiliency in Cooperative Adaptive Cruise Control (CACC).

Keywords: Vehicular communication · Automotive security · Machine learning · Anomaly detection

1 Introduction

Recent years have seen rapid transformation of automotive systems from being primarily human-operated, electro-mechanical systems to complex electronic systems with hundreds of connected Electronic Control Units (ECUs), a variety of sensors and actuators, several in-vehicle networks, several miles of cable, and several hundred megabytes of software code. Much of this transformation has

* This research has been partially supported by the National Science Foundation under Grant CNS-1908549.

been towards increasing autonomy, *i.e.*, augmenting and replacing human functionality with electronics and software. Autonomous features hold the promise of dramatically increasing road safety, by reducing and eventually eliminating human errors [27]. However, an unfortunate effect of this trend is a corresponding increase in the vulnerability of these systems to a variety of cyber-attacks. Recent research has shown that it is possible, — even relatively straightforward, — to compromise a vehicle and get control over its driving function [25, 26, 11, 21]. The trend towards increasing autonomy will only exacerbate this situation: the increasing dependence of critical vehicular operations on complex electronics and software will result in an increased attack surface as well as the increasing ability of an attacker to create catastrophic impact from a compromise. ***Consequently, the proliferation or even adoption of autonomous vehicles critically depends on our ability to ensure that they perform securely, in a potentially adversarial environment.***

A critical feature of emergent autonomous vehicles is *connectivity*, *i.e.*, the ability to communicate with other vehicles (V2V), with the infrastructure (V2I), and with other devices connected to the Internet (V2IoT). Vehicular communications, referred to as V2X, are performed through a variety of protocols, *e.g.*, DSRC, and form a fundamental enabler for autonomous driving by enabling cooperative information sharing for streamlining traffic movement, improving road safety, and efficiently utilizing traffic and transportation infrastructure. V2X forms the foundation for critical applications like platooning [8], cooperative route management [14, 12], intersection management [29], etc. Unfortunately, V2X is also a highly vulnerable feature that can be exploited by an adversary to disrupt traffic movement and cause catastrophic accidents. A key problem with V2X is that it obviates the need for an adversary to actually hack a vehicle: sending misleading or malformed V2X communications is often sufficient to disrupt the connected car ecosystem. For example, in platooning, an adversary may cause an accident simply by sending a misleading message with an acceleration directive while braking [1]. Unsurprisingly, in a recent survey by the world's second-largest reinsurer Munich Re, 55% of the surveyed corporate risk managers named security of vehicular communications as their top concern for autonomous vehicles [17]. Perhaps even more alarming, 64% of the companies surveyed mentioned that they were completely unprepared to address this threat.

In this paper, we present the vision of a potential approach to address this critical problem. Our proposed solution is REDEM (for "REal-time DEtection and Mitigation"), a novel resiliency architecture that can be integrated with a variety of cooperative autonomous applications to detect and mitigate communication attacks. A key component of REDEM is an anomaly detection system (ADS) based on machine learning to detect malicious V2X communications in real time. The central idea is to build models that can learn normal behavior corresponding to benign V2X communication and detect anomalous behavior in order to sense potentially malicious communication. On detecting an anomaly, REDEM performs real-time mitigation, also using machine learning to estimate

the appropriate driving decisions. A unique feature of REDEM is its flexibility: the same infrastructure can address an elaborate set of adversaries in the connected car ecosystem, including man-in-the-middle (MITM), wormhole, Sybil, Denial-of-Service (DoS), etc. Furthermore, it accounts for the natural differences in communication patterns among a variety of driving scenarios, road conditions, etc. *This is in contrast to most related work on V2X security [29, 2, 15] that require detailed, continuous models of vehicular and adversarial functionalities.*

REDEM is early work in progress. We are currently realizing the REDEM vision in introducing resiliency to a specific but foundational connected car application, Cooperative Adaptive Cruise Control (CACC). We provide initial results on resilient CACC to demonstrate the viability of REDEM.

The remainder of the paper is organized as follows. Section 2 provides the relevant background on connected car applications and related research. In Section 3 we discuss challenges and design constraints involved in the development of resilient connected car applications, and REDEM's approach to addressing them. Sections 4 and 5 discuss REDEM's envisioned architecture and Section 6 discusses evaluation challenges. In Section 7 we present initial results from our current efforts on realizing REDEM on CACC. We conclude in Section 8.

2 Background and Related Work

2.1 Connected Car Applications and Security Challenges

We present a brief overview of a few connected car applications to explain the scope and spectrum of security challenges in V2X communications. The following are representative examples.

- *Platooning.* Platooning involves a group of autonomous vehicles (referred to as a string or platoon) traveling with relatively small headway distance and very small relative velocity [9]. The goal is to improve the operational efficiency of the transportation infrastructure by improving highway capacity. The vehicles must brake or accelerate simultaneously to ensure safety of the platoon and optimal usage of the highway infrastructure. In emergent, distributed platooning systems, a vehicle uses V2V messages to communicate its intent (e.g., to brake or accelerate), as well as its relative distance with its neighbor; every vehicle in the platoon accounts for this information to compute its course of action.
- *Smart Intersection Management.* This application is developed for smart cities, with the goal to enable smart and efficient control of (autonomous) vehicles approaching an isolated intersection. In this case, vehicles communicate with an intersection manager through V2I communications to notify estimated arrival time to the intersection. The intersection manager uses this information to schedule vehicles for crossing the intersection.
- *Cooperative Collision Detection.* The goal of this application is for vehicles approaching an intersection from directions to coordinate through V2V messages and avoid collision. Vehicles broadcast their speed, direction of motion,

and position relative to the intersection. A vehicle $\mathcal{E}$ receiving this communication from other vehicles $\mathcal{T}$ computes its relative distance, angle, and speed and determines if a collision is possible. In recent CCD systems, vehicles communicate, in addition to their own information, data about other vehicles within their V2V communication range; each vehicle accounts for this additional information to increase precision of its calculation and facilitate fault tolerance in sensor measurements.

– *Dynamic Cooperative Route Management.* Augmenting dynamic routing strategies with Co-operative communication enables improved traffic management, faster recovery from an unforeseen disturbance in the traffic flow, better congestion control as well as improved safety of the vehicles [13]. The co-operative application is proven to be more efficient and accurate, than mapping services that rely purely on satellite imagery. In this application, vehicles constantly broadcast their mapping and localization and in turn utilize the information shared by other vehicles driving along the desired routes.

Clearly, viability of all the above applications critically depends on the trustworthiness of the V2V and V2I communications. A rogue vehicle participating in the application can send misleading, malicious, or confusing messages designed to cause accidents or disrupt the transportation infrastructure. Such messages can easily result in catastrophic accidents or disruption of the entire connected car infrastructure. Since vehicles are consumer items, an adversary can simply buy a car, hack its vehicular communication components, and use such a compromised vehicle to disrupt a connected car application. Correspondingly, adversarial activity in V2I application may entail a compromised or hacked infrastructural component, or one that is "confused" by a compromised vehicle participating in the application. Finally, it is not necessary for an adversary to actually compromise a vehicle: connected car applications are vulnerable to rogue intermediary agents performing man-in-the-middle (MITM) attack, Sybil attack, and many others. In traditional secure communications, such problems are addressed through strong message authentications; however, this requires computationally intensive algorithms which may not be practical with the limited computational resources of automotive ECUs under aggressive real-time requirements.

2.2 State of the Practice and Related Research

In today's industrial practice, detection of security vulnerabilities in connected car applications primarily entails manual *penetration testing*. Human validators with deep insight into the application, the implementation of the vehicular functionality, and potential vehicle responses to various V2X communication, conceive various adversarial scenarios with elaborate simulation models, physical prototype of vehicles, or field testing environments. Such methods obviously depend crucially on the insight of the experts. Furthermore, since elaborate prototypes of vehicular functionality are only available late in the design lifecycle, mitigation of security vulnerabilities identified precludes complex changes

in the overall system architecture of individual vehicles; instead, workarounds are employed, including functionality reduction, patches, and point-fixes, which themselves may lead to further vulnerabilities and in-field attacks.

Given the importance of security in automotive systems, there has been significant research interest for security mechanisms to ensure resiliency of vehicular functionality. In related work, vehicle intrusion detection systems (IDS) are largely divided by the targets for security assurance. IDS for in-vehicle network considers intrusion and anomaly detection on CAN [10, 20], while IDS for vehicular adhoc network (VANET) considers security of V2X communications [4, 19]. These works do not consider security of V2X communications across connected car applications in a comprehensive framework considered in this work. In research on connected cars in particular, there have been works on security of platooning and cooperative adaptive cruise control [7]. Proposed approaches include a variety of techniques based on control theory to address targeted adversary models [16], and application-specific techniques assuming certain adversary properties such as rationality [23]. However, control-theoretic approaches require detailed models precisely specifying the adversary operation, machine learning techniques suffer from the unavailability of sufficient data for training the models, and rationality-based techniques need assumptions that may be violated by in-field adversaries. Furthermore, these works do not explore the viability of realizing the approaches with on-board computational resources.

There are also related machine learning research relevant to our work. Levi [22] provides a data abstraction approach that converts raw vehicle data to events that helps in filtering noise and reducing data dimensionality. For automotive systems, machine learning has been used for computer vision modules to improve on-board perception [32, 30]. Tiwari [31] describes attack features that are undetectable at each time instance but can be detected from sequence data. There has also been related work on adversarial attacks on these systems [24, 33].

3 REDEM Vision

3.1 Design Constraints

REDEM is an anomaly detection system (ADS) based on machine learning, that can be installed in autonomous vehicles involved in connected car applications; it will enable the vehicle (referred to as *ego vehicle*) to detect adversarial communications in real time, and perform mitigation. For such a system to be viable, it must satisfy the following requirements.

- *Basic Safety:* Any driving decision generated from an automated source must be *safe*, *i.e.*, should not increase the risk of accident. This applies particularly to any system that performs real-time mitigation in response to detected anomalies: road safety should not be compromised by the mitigating action irrespective of whether the response is to a message classified as anomalous as the result of a real attack or imprecision/inaccuracy in the detection algorithm.

- *Reusability/Extensibility:* Connected car applications are proliferating rapidly. Furthermore, new, previously unknown, attacks are being discovered every day in research as well as in practice. *It is critical for a viable ADS mechanism to be easily extensible for a variety of new adversarial operations.* Note that ADS approaches based on control theory depend on detailed mathematical models that precisely define the adversarial activity: solutions for Denial-of-Service (DoS) and data corruption attacks typically require different mathematical models and independent analysis. This makes it difficult to deploy such solutions to practical automotive applications.
- *Limited Computation:* Any solution integrated within an automotive system architecture must operate within the constraints imposed by that architecture and the real-time response requirements of connected car applications. Consequently, it must be realizable by smooth, disciplined extension of the system functionality without significant design overhaul. Furthermore, it must be possible to perform the computation with automotive ECUs in real time. This rules out any solution that requires installation of sophisticated, computation-intensive algorithms implemented within ECUs.
- *Small Data Problem and Machine Learning Attacks:* Any ADS system based on *machine learning* must additionally cope with two critical challenges. First, machine learning solutions targeted towards learning anomalies suffer from the so-called "small data" problem: assuming that the number of adversarial in-field examples is limited, there is only a small amount of field data exhibiting anomalous behavior. Furthermore, unlike traditional machine learning targets (*e.g.*, recommendation systems), it is generally impossible for security training sets to get progressively sophisticated through accumulation of years of anomaly data. Recent research has shown that it is also possible for an adversary to target the machine learning system itself [28, 3], resulting in degradation in prediction accuracy that renders the system useless.

3.2 REDEM Approach and Viability

REDEM addresses the above constraints by exploiting a number of critical observations as described below.

- REDEM on-board Mitigator includes an explicit *Plausibility Checker* to determine whether the mitigation response can potentially compromise safety of the application (see 4). Consequently, *basic safety* is preserved by construction.
- *We address real-time requirements by separating the training of prediction models from on-road prediction.* The key observation is that the computation-intensive component of the machine learning solutions is in training predictor models to be used in the ADS; once the model is created, detection can be performed within the limited resources of automotive ECUs. Our system includes a cloud-based methodology for training prediction models, while

the on-board architecture is responsible for collecting data and performing real-time prediction.

- *We do not require detailed adversarial model beyond the assumption that the adversary affects V2X communications outlined in our threat model described in Section 3.3.* This makes the same approach applicable for diverse connected car applications, *e.g.*, we are applying the same framework for platooning, cooperative route management, and cooperative collision detection. Furthermore, our on-board architecture is designed to account for compatibility with automotive electronic system architectures from the ground up.
- *To address the small data problem, we observe that while the data concerning anomalous behavior is limited, data on normal behavior is typically plentiful.* Consequently, we train prediction algorithms to learn *normal behavior model* (NBM), *i.e.*, the normal (benign) pattern of V2X communications relevant to a connected vehicle application rather than the anomalous behavior; the on-board anomaly detector then operates by calculating the degree of deviation from NBM as a measure of anomaly. Furthermore, the NBM training uses data collected from *all* vehicles with the ADS architecture integrated, in addition to the ego vehicle.
- *We enable resilience against adversarial machine learning attacks by noting that such attacks require sustained, consistent deviation of predicted behavior from actual for a continued period of time.* Consequently, the system can be resilient to the *effects* of such attacks by appropriate choice of prediction parameters such that attacks on prediction system have no perceptible effect on the safety of the application beyond tolerable degradation in performance. Furthermore, the prediction parameters can be tuned to minimize the effects of adversarial machine learning attacks on performance.

3.3 Threat Model and Design Assumptions

We consider connected cooperative applications that make use of V2X communications to augment information obtained from sensors to make various on-road decisions. We assume that the application can still function in the absence of V2X by relying on sensory information alone, albeit with significantly lower efficiency. For example, in Cooperative Adaptive Cruise Control (CACC) [6], the ego vehicle in the absence of V2V messages from the leading car can fall back on Adaptive Cruise Control (ACC), where the basic functionality (*i.e.*, following the leading car at a safe distance) is maintained, albeit at a much higher time headway. One key goal of REDEM is to ensure resiliency while enabling targeted applications to enjoy the higher efficiency induced by V2X as much as possible.

Given our focus on V2X security, our threat model assumes that the attacker can tamper arbitrary V2X messages. This includes (1) message mutation, *i.e.*, arbitrary modification of a V2X message packet while in flight, resulting in either malformed communications or misleading/erroneous messages; (2) denial of delivery of a message packet to its receiver; (3) masquerading as a legitimate vehicular or infrastructure entity; and (4) fabrication and transmission of arbitrary

new (legitimate or malformed) message packets. Note that the last component also covers flooding or jamming attacks. Our framework is also oblivious to the source of the attack: it can be a rogue car, a compromised transportation infrastructure component, a compromised V2X mechanism, or an intermediate networking component, *e.g.*, denial of message delivery is possible by compromising the software/hardware component of the ego vehicle or interfering with the communication protocol. We assume that our on-board ADS architecture *in the ego vehicle*, as well as the actuarial/control components it controls, are not compromised. We also assume the sensor data in the ego vehicle (*e.g.*, data captured through radar, LIDAR, camera, etc.) is not compromised. Note that there has been significant work on attacks to automotive sensors; nevertheless, in the context of our approach, assuming fidelity of sensor data is reasonable since it is unlikely that the same adversary can concurrently manipulate both sensor and V2X inputs. Finally, our infrastructure does not require real-time, on-road communication between the ego vehicle and cloud: transfer of trained models and on-road V2X data can be performed periodically offline when the vehicle is connected to secure communication channels.

4 REDEM On-Board Architecture

We design the on-board architecture of REDEM with three goals: (1) reusability across different connected car applications, (2) compatibility with existing automotive system architecture, and (3) realizability within the limited computation resources of automotive ECUs. We assume the existence of a cloud-based infrastructure for NBM generation for the targeted applications (which will be considered in Section 5).

The key insight behind our on-board design is that the architecture of most connected car features follow a standard template with two major components, a *Decision Computation Module* and an *Actuarial Controller*. Given the sensory and V2X inputs pertaining to the application, the Decision Computation Module computes the desired actuarial actions of the vehicle, and the Actuarial Controller generates the control commands for the actuators. For CACC, [6], the V2X messages for any control cycle t are the intended acceleration/deceleration information $a^t_{\mathcal{L}}$ provided by the leading car $\mathcal{L}$, the sensory information is the distance $d^t_{\mathcal{E},\mathcal{L}}$ between $\mathcal{L}$ and the ego vehicle $\mathcal{E}$, the desired actuarial action is the corresponding response of the ego vehicle, *e.g.*, acceleration $a^t_{\mathcal{E}}$ computed as a function of $a^t_{\mathcal{L}}$ and $d^t_{\mathcal{E},\mathcal{L}}$, and the actuarial controller manipulates the motor output torque and braking pressure to achieve $a^t_{\mathcal{E}}$. REDEM augments this template with additional components to account for resilience of the ego vehicle to malicious V2X communications. Consequently, the same architecture would work on a variety of connected car features with little reconfiguration; and it will be compatible with the on-board system architecture for most emergent autonomous vehicles.

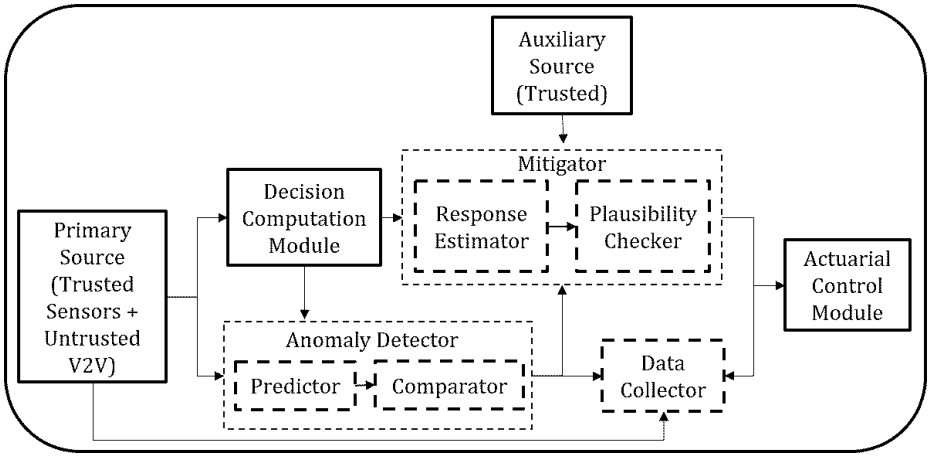

Fig. 1. REDEM On-board Architecture. The subsystems bordered with dashed lines are components introduced by REDEM.

Fig. 1 provides a high-level view of REDEM on-board architecture. Roughly, it introduces three additional system components (on top of the underlying connected application architecture).

1. *Anomaly Detector* is responsible for detecting suspicious V2X communications;
2. *Mitigator* is responsible for adjusting the actuarial action of the vehicle in response to a detected anomaly; and
3. *Data Collector* captures real-world on-road data for improving prediction accuracy of the anomaly detector and mitigator components.

The data from the *Data Collector* is periodically transferred to trusted cloud server to retrain the machine learning components in Anomaly Detector and Mitigator (*e.g.*, the Predictor and Response Estimator respectively). Recording real-world data in this manner facilitates curating a database of different communication anomalies, eventually improving anomaly detection. The roles of Anomaly Detector and Mitigator are described in more detail below.

Anomaly Detector

The Anomaly detection subsystem comprises of two components, Predictor and Comparator. The Predictor implements a machine learning model trained to learn normal behavior of the Decision Computation Module of a conventional application, as discussed in Section 5. The output of the Predictor is compared by the Comparator against the (real) output of the Decision Computation Module. A deviation beyond a pre-defined threshold is classified as an anomaly. If no anomaly is detected, the output of the Decision Computation Module is applied to the vehicle; otherwise, the Mitigator is triggered (see below).

Mitigator

When an anomaly is detected, the Mitigator overrides the output of the Decision Computation Module and computes a different decision that relies solely on the trusted source of information, *viz.*, sensors. It includes two components, Response Estimator and Plausibility Checker. Analogous to the Predictor, the Response Estimator also implements a machine learning model trained to predict the expected response of the Decision Computation Module, but it only uses sensory data in training and prediction. The Plausibility Checker determines whether the output of the Response Estimator, if applied in place of the output of the Decision Computation Module, can potentially compromise the safety of the application. If the check fails (*i.e.*, the safety of the application cannot be guaranteed), then the system falls back to a more conservative, non-cooperative mode of the application; otherwise the output of the Response Estimator is applied by the Mitigator in place of the output of the Decision Computation Module.

5 Prediction Models

The central component of REDEM is the construction of the machine learning models to be used in the predictor (and response estimator). These components are implemented and trained on a trusted cloud platform, and are refined with training data from subscribed vehicles through on-board data collector component. Obviously, the quality of the models, in addition to the training data, depends crucially on the model parameters. The communication pattern among vehicles or between vehicles and infrastructure depends on a variety of parameters, including terrain (*e.g.*, hilly, rural highway, city), time of day or night, ambient weather, etc. Furthermore, there is trade-off between the quality of prediction induced by the model and the complexity of computation and storage requirements induced by a high-precision model.

There is no reason to believe there is a unique, uniform deep learning model that will be suitable across all cooperative connected car applications. Nevertheless, some models can be easily ruled out, *e.g.*, simplistic prediction models that depend on linearity assumptions are clearly unsuitable, and so are highly complex computation-intensive or storage-intensive models which might be difficult to implement within the limited resources of automotive ECUs. For our initial CACC work, we have found a multilayer perceptron (MLP) model sufficient for both the Predictor and the Response Estimator. We suspect that such a model provides the sweet spot between accuracy needs and computation cost for most major applications.

Another key question to address is whether (for a specific application) one unified model NBM is sufficient or whether a different model is necessary for each driving scenario (*e.g.*, terrain, weather, time of day, etc.). If effective prediction requires a custom model for each specific driving scenario, then there must be facility to download/switch models as the vehicle drives from one scenario to another (*e.g.*, moving from highway to city or sunny to cloudy weather). It would

appear that such custom models might have higher accuracy than a single model that has to predict normal behavior under all potential driving scenarios. Consequently, the question of one unified model vis-a-vis custom models for different scenarios might appear to be a trade-off between prediction accuracy and cost of switching. On the other hand, our very recent experiments suggest that this trade-off might actually be spurious. In a recent experiment we found that in fact a single unified model may turn out to be *more accurate* than custom models, at least for some specific applications. The reason for this apparent paradox is that a single global model that predicts normal behavior for all scenarios usually has more in-field data for training, causing it to be a better source of prediction than custom models for different scenarios trained with less data.

6 REDEM Evaluation

The success of research targeted at connected car applications critically depends on effective evaluation framework that enables clear comprehension of the effects of different architectural trade-offs on the resiliency and efficiency of the applications. Roughly, there are two critical requirements in addressing the evaluation needs as described below, *e.g.*, effective adversary models and realistic datasets on vehicle behavior.

- *Adversary Modeling.* As discussed in Section 3.1, a key requirement for an automotive resiliency solution is that it must provide protection against the spectrum of (known and unknown) potential adversaries. On the other hand, evaluating this requirement requires developing a set of adversary models that can be justified as comprehensive, and demonstrating the robustness of a proposed solution against this set. Unfortunately, no such comprehensive set of adversaries exist for communication attacks. Indeed, determining adversaries is typically a reactive process: given a specific resiliency solution, one comes up with an adversary to subvert the specific solution. The result of this process is typically a collection of specific "point adversaries". For communication attacks, specific adversaries include masquerade, man-in-the-middle, Sybil, wormhole, etc. However, simply evaluating the solution against a collection of specific, known adversaries does not provide any confidence on its resiliency against unknown, zero-day attacks.
- *Evaluation Platform.* Since autonomous vehicles are complex, safety-critical systems, it is essential to evaluate the performance, safety, and effectiveness of any resiliency solution before deployment. Since REDEM is a machine learning solution, this additionally implies training and evaluation of the proposed prediction models. Unfortunately, this is challenging because of the lack of available datasets. Existing benchmark driving datasets do not comprehensively represent different driving environments, nor do they provide sufficient data corresponding to rare driving scenarios that are particularly important for evaluation of security attacks.

We address the challenges above in REDEM as follows. To address the problem of adversary models, we are developing an adversary taxonomy to enable comprehensive evaluation of the resiliency architecture. The key idea is to eschew specific adversary types (*e.g.*, Sybil, MITM, wormhole, etc.) but focus on adversary capabilities given the threat model of Section 3.3. In particular, we classify adversaries along three vectors, *e.g.*, (1) *stealth* or frequency of malicious communication, ranging from independent discrete attacks at very infrequent time instants to a continuous sequence of attacks at each instant over a time interval; (2) the effect of the attack on V2X, *e.g.*, message mutation, injection of fabricated message, delivery prevention; and (3) potential effect on the target vehicle, *e.g.*, accident, string instability, inefficient use of infrastructure, etc. The attack taxonomy incorporates a diverse spectrum of attacks considered in wireless and networking communities, including wormhole, masquerade, misdirection, Sybil, and man-in-the-middle attacks. On the other hand, since the focus of the taxonomy is on *effects* rather than on specific point adversaries, we can be confident that a system demonstrated to be resilient to adversaries across this taxonomy is also resilient against other zero-day adversaries.

In addition to the above classifications, the possibility of adversarial machine learning attacks can be addressed by considering a special class of attacks, which we refer to as *Predictor Subversion Attacks*. These attacks involve an adversary with complete knowledge of the REDEM architecture (including the anomaly detection threshold) and the trained predictor configurations. Predictor subversion attacks can bypass the REDEM's anomaly detection system and go undetected. The application will be considered robust against adversarial machine learning if the Predictor Subversions cannot create a perceptible impact (*e.g.*, compromising safety or loss of efficiency of the cooperative application) on the target vehicle.

We address the problem of evaluation platform by using a physical research simulator. The specific simulator we use in REDEM is RDS1000® (`https://www.faac.com/realtime-technologies`), but other physical simulators that provide similar functionalities will also be sufficient. The simulator gives us the flexibility to define and simulate driving environments at a fine detail, and capture realistic driving data pertaining to normal behavior models. Note that we generally need more sophisticated platforms than desktop simulators used in previous research, *e.g.*, VENTOS [5] or Carla (`http://carla.org`). In particular, RDS1000 enables flexible programming and simulation of virtually any environmental, terrain, or traffic conditions, and any (autonomous) maneuver of the vehicle. Data pertaining to each of these environments can be used for training and testing machine learning components of REDEM. It also enables gathering data that reflects the real-time behavior of a vehicle. Data pertaining to each of these environments can be used for training and testing machine learning components of REDEM. Driving environments are classified based on various major parameters that impact the driving patterns: (i) Road terrain (Highway, Suburban and Urban); (ii) Weather (Clear, Windy, Snowy, Rainy);

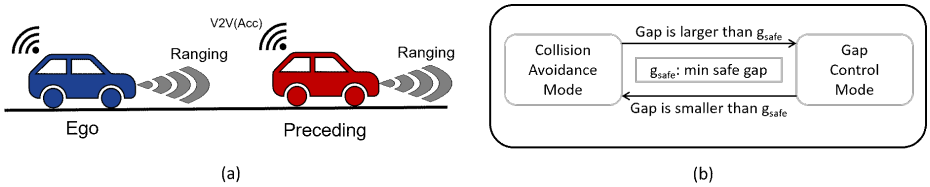

Fig. 2. (a) represents two vehicles engaged in CACC; (b) represents the modes of operation of a conventional CACC decision computation module.

and (iii) Time of day (Day, Night). Different combinations are considered with these factors and the environments are simulated accordingly.

7 Case Study: Secure CACC

To determine viability of the REDEM vision, we are realizing it on a specific connected car application, *viz.*, Cooperative Adaptive Cruise Control. CACC forms the basis for several connected car applications such as platooning, cooperative on-ramp merging etc. In CACC, the ego vehicle autonomously adapts its velocity in accordance to the acceleration of the vehicle in front (received through V2V communication), as well as the relative velocity and gap between the two vehicles (obtained from the ranging sensor readings). CACC enables improved road safety and efficiency (*e.g.*, a much smaller headway) compared to its non-cooperative counterpart, Adaptive Cruise Control (ACC) which does not utilize V2V communication.

7.1 CACC Functional Overview

Fig 2(a) depicts vehicles engaged in CACC. Fig 2(b) is used to demonstrate the high-level functionality of a CACC decision computation module implementing a constant time headway policy. The specific CACC implementation considered here [18] targets a constant time headway of 0.55secs from the preceding vehicle. The safety goal of CACC is to maintain a space gap that is greater than a safety threshold g_{safe} computed as a function of relative velocity between the vehicles. CACC operates in two modes: collision avoidance and gap control, based on the instantaneous space gap between the vehicles. The vehicle normally operates in gap control mode where it follows the leading car as closely as possible while maintaining a space gap greater than g_{safe}; if space gap is less than g_{safe}, it switches to collision avoidance mode and the vehicle is decelerated at its maximum value.

Obviously, CACC is susceptible to attacks targeting the V2V communication. Consider the following attack scenarios.

1. The preceding vehicle reports falsified acceleration values that are *higher than actual* for a continued period of time. The ego vehicle operates in gap

control mode, and is misled to accelerate, until the gap g falls below g_{safe} switching to collision avoidance mode. If the speed v_f of the victim is sufficiently high, a sudden deceleration may result in a collision, a precarious skid, or at the least, a highly uncomfortable jolt.
2. The preceding vehicle reports falsified acceleration values that are *lower than actual* for a continued period of time. The ego vehicle would decelerate and fall behind, resulting in degraded fuel efficiency and travel time. In extreme cases, the vehicle might switch to collision avoidance mode, resulting in sudden deceleration, jolt, or even a collision with the vehicles behind.
3. The leading vehicle stops reporting acceleration values completely, or communicates a random sequence of values, with the intent to mislead or confuse the CACC Decision Computation Module of the follower vehicle. One effect could be for the V2V messages to become uncorrelated with the sensor data,e.g., positive acceleration of the leading car accompanied with reduced distance. Depending on the CACC controller implementation, this can result in vehicle stall, sudden deceleration, downgrading of CACC to ACC, etc.

7.2 REDEM for CACC

We developed a realization of REDEM for CACC. Here we discuss some of the initial experimental results from that effort, primarily as a demonstration of viability of REDEM as a means to introduce resiliency in cooperative connected car applications. As shown below, our results are promising. Nevertheless, they should be taken with the caveat that the work is still early at the time of this writing and much more experimentation is necessary to thoroughly vet the REDEM architecture even for this specific application.

Simulation Setup and Training Data Generation

As discussed in Section 6, we used a physical automotive simulator for creating the various driving environments and traffic conditions, and recording the necessary data parameters required for training the machine learning based global predictor component. For this analysis, we considered three road parameters (*e.g.*, highway, suburban, and city), four weather parameters (*e.g.*, rain, snow, clear, and windy), and two diurnal parameters (*e.g.*, day and night). A unique model is created and trained for each combination of parameters, resulting in 24 unique models. Each dataset corresponding to about 15 mins of driving time and constitutes approximately $90,000$ samples collected at a frequency of 100Hz. 80% of the data is used to train the machine learning models to learn normal behavior (in each driving environment) while the rest is used for testing and evaluation purposes.

Attack Orchestration

As initial demonstration, we orchestrated a class of simple, independent discrete attacks on the REDEM augmented CACC system. Under these attacks, the

Table 1. Global Predictor Model Accuracy Evaluation: Testset Mean Absolute Error Values

Road Infrastructure	Day				Night			
	Rain	Snow	Clear	Windy	Rain	Snow	Clear	Windy
Highway	0.24	0.26	0.27	0.25	0.28	0.19	0.18	0.27
Suburban	0.12	0.29	0.16	0.21	0.18	0.08	0.11	0.23
City	0.08	0.33	0.06	0.02	0.11	0.05	0.16	0.04

Table 2. Anomaly Detector Accuracy Evaluation : % False positives and False Negatives

Road Infrastructure	Day								Night							
	Rain		Snow		Clear		Windy		Rain		Snow		Clear		Windy	
	%FP	%FN	%FP	%FN	%FP	%FN	%FP	%FN	%FP	%FN	%FP	%FN	%FP	%FN	%FP	%FN
Highway	6.2	0.09	2.11	4.42	4.23	0.92	1.07	5.2	6.4	2.8	0.4	5.9	7.2	5.5	7.4	8.8
Suburban	0.56	0.07	1.1	7.72	2.6	0.65	0.88	2.29	0.62	3.97	2.24	0.95	0.52	3.22	2.23	0.95
City	0.23	0.10	9.59	0.12	0.21	0.22	0	0.15	8.41	0.12	0.11	5.87	0.28	0.17	0.04	1.30

ego vehicle receives mutated V2X messages reporting false or anomalous vehicle acceleration values of the preceding vehicle. Discrete samples constituting 30% of the evaluation data are selected at random and a bias is added to the acceleration values such that the resultant headway between the two vehicles becomes smaller than the safe limit or large enough to cause inefficiency and string instability in the traffic.

Results

We evaluate the predictor models trained on data collected from each driving environment. The resiliency of REDEM depends both on the accuracy of the Predictor and the choice detection threshold of the Comparator. Table 1 shows the predictor accuracy indicated by the deviation from the expected acceleration prediction under normal operating conditions in the absence of malicious activity. The low mean absolute error indicates that the predictor models closely estimate the acceleration output of a conventional CACC Decision Computation Module. The Predictor accuracy under anomalous conditions are shown in Tables 2 and 3. False positive and false negative percentages indicate the percentage of normal samples falsely captured as anomalies and vice versa. Note that REDEM even with this initial realization still achieves a prediction accuracy of about 95%.

8 Conclusion and Future Work

With the trend towards increasing autonomy of automotive systems, cooperative connected applications will become increasingly crucial, together with the need to introduce resiliency in such applications against potential subversions targeting V2X communications. Clearly, a reactive approach to security, *i.e.*, point solutions/patches incrementally fixing the system as newer and newer attacks

Table 3. Anomaly Detector Accuracy Evaluation : % True positves and True Negatives

Road Infrastructure	Day								Night							
	Rain		Snow		Clear		Windy		Rain		Snow		Clear		Windy	
	%TP	%TN	%TP	%TN	%TP	%TN	%TP	%TN	%TP	%TN	%TP	%TN	%TP	%TN	%TP	%TN
Highway	93.78	99	97.88	95.57	95.77	99.07	98	94.8	93.6	97.2	99.6	94.1	92.8	94.5	92.6	91.2
Suburban	99.43	99.92	98.89	92.27	97.39	99.35	99.11	97.7	99.38	96.02	97.76	99.05	99.48	96.77	97.76	99.05
City	99.77	99.90	90.41	99.87	99.79	99.77	100	99.85	91.59	99.87	99.88	94.12	99.71	99.82	99.96	98.7

are discovered, is not viable in this space. In this paper we have introduced a novel vision for introducing resiliency in connected car applications, by providing an architecture to augment the application design with generic components. The architecture is reusable over different connected car applications, can be implemented within the computational/storage constraints induced by automotive systems, and can support real-time detection and mitigation. We have also introduced evaluation mechanisms to evaluate the viability of such resiliency architecture over a wide class of adversaries and driving scenarios. We provided initial evidence of viability of the approach in introducing resiliency in CACC.

Nevertheless, we have only scratched the surface of this vast research area. Even in the realization of REDEM in CACC, much evaluation is left to be done, *e.g.*, viability over the spectrum of attacks in our adversary taxonomy, efficiency of the models defined, effectiveness of the approach against a variety of adversarial machine learning attacks, quality of the dataset generated through our driving simulator, etc. Furthermore, we will work on realizing REDEM for other cooperative applications and consider extending it for scenarios where the sensor system (in addition to V2X) is compromised.

References

1. Cybersecurity for Autonomous Vehicle Platooning. `https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1559&context=researchweek`.
2. Z. Abdollahi Biron, S. Dey, and P. Pisu. Real-time detection and estimation of denial of service attack in connected vehicle systems. *IEEE Trans. Intelligent Transportation Systems*, 19(12):3983–3902, 2018.
3. N. Akhtar and A. Mian. Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. *IEEE Access*, 6:1441–14430, 2018.
4. K. M. A. Alheeti, M. S. Al-Ani, and K. McDonald-Maier. A hierarchical detection method in external communication for self-driving vehicles based on tdma. *PloS one*, 13(1):e0188760, 2018.
5. M. Amoozadeh, H. Deng, C.-N. Chuah, H. M. Zhang, and D. Ghosal. Platoon management with cooperative adaptive cruise control enabled by vanet. *Vehicular Communications*, 2(2):110 – 123, 2015.
6. B. Aygun, C.-W. Lin, S. Shiraishi, and A. Wyglinski. Selective message relaying for multi-hopping vehicular networks. In *IEEE Vehicular Networking Conference*, pages 1–8, 2016.
7. B. Aygun, C.-W. Lin, S. Shiraishi, and A. M. Wyglinski. Selective message relaying for multi-hopping vehicular networks. In *2016 IEEE Vehicular Networking Conference (VNC)*, pages 1–8. IEEE, 2016.

8. C. Bergenhem, H. Pettersson, E. Coelingh, C. Englund, S. Shladover, and S. Tsugawa. Overview of Platooning Systems. In 19*th ITS World Congress*, 2012.
9. C. Bergenhem, S. Shladover, E. Coelingh, C. Englund, and S. Tsugawa. Overview of platooning systems. In *Proceedings of the 19th ITS World Congress, Oct 22-26, Vienna, Austria (2012)*, 2012.
10. I. Berger, R. Rieke, M. Kolomeets, A. Chechulin, and I. Kotenko. Comparative study of machine learning methods for in-vehicle intrusion detection. In *Computer Security*, pages 85–101. Springer, 2018.
11. S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno, et al. Comprehensive experimental analyses of automotive attack surfaces. In *USENIX Security Symposium*, volume 4. San Francisco, 2011.
12. L. Du, S. Chen, and L. Han. Coordinated Online In-Vehicle Navigation Guidance Based on Routing Game Theory. *Transportation Research Record: Journal of the Transportation Research Board*, 2497:106–116, 2015.
13. L. Du, S. Chen, and L. Han. Coordinated online in-vehicle navigation guidance based on routing game theory. *Transportation Research Record*, 2497(1):106–116, 2015.
14. L. Du, L. Han, and X. Li. Distributed Coordinated In-Vehicle Online Routing under Mixed Strategy Congestion Game. *Transportation Research Part B: Methodological*, 67:235–252, 2014.
15. R. G. Dutta, F. Yu, T. Zhang, Y. Hu, and Y. Jin. Security for safety: A path toward building trusted autonomous vehicles. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–6, Nov 2018.
16. R. G. Dutta, F. Yu, T. Zhang, Y. Hu, and Y. Jin. Security for safety: a path toward building trusted autonomous vehicles. In *Proceedings of the International Conference on Computer-Aided Design*, page 92. ACM, 2018.
17. C. Hempfield. Why a Cybersecurity Solution for Driverless Cars May be Found Under the Hood, 2017. `https://techcrunch.com/2017/02/18/why-a-cybersecurity-solution-for-driverless-cars-may-be-found-under-the-hood`.
18. M. Jagielski, N. Jones, C. Lin, C. Nita-Rotaru, and S. Shiraishi. Threat Detection in Collaborative Adaptive Cruise Control in Connected Cars. In *WISEC*, pages 184–189, 2018.
19. M. Jagielski, N. Jones, C.-W. Lin, C. Nita-Rotaru, and S. Shiraishi. Threat detection for collaborative adaptive cruise control in connected cars. In *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 184–189. ACM, 2018.
20. M.-J. Kang and J.-W. Kang. Intrusion detection system using deep neural network for in-vehicle network security. *PloS one*, 11(6):e0155781, 2016.
21. K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, et al. Experimental security analysis of a modern automobile. In *2010 IEEE Symposium on Security and Privacy*, pages 447–462. IEEE, 2010.
22. M. Levi, Y. Allouche, and A. Kontorovich. Advanced analytics for connected car cybersecurity. *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–7, 2018.
23. Y.-T. Lin, H. Hsu, S.-C. Lin, C.-W. Lin, I. H.-R. Jiang, and C. Liu. Graph-based modeling, scheduling, and verification for intersection management of intelligent vehicles. *ACM Transactions on Embedded Computing Systems (TECS)*, 18(5s):95, 2019.

24. A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. *ArXiv*, abs/1706.06083, 2018.
25. C. Miller and C. Valasek. A survey of remote automotive attack surfaces. *Black Hat USA*, 2014:94, 2014.
26. C. Miller and C. Valasek. Remote exploitation of an unaltered passenger vehicle. *Black Hat USA*, 2015:91, 2015.
27. National Highway Traffic Safety Association. Road Accidents In USA. See URL: `https://www.recalls.gov/nhtsa.html`.
28. N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami. The Limitations of Deep Learning in Adversarial Settings. In *IEEE European Symposium on Security and Privacy*, 2016.
29. M. O. Sayin, C.-W. Lin, S. Shiraishi, J. Shen, and T. Basar. Information-Driven Autonomous Intersection Control via Incentive Compatible Mechanisms. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):912–924, Mar. 2019.
30. Y. Tian, K. Pei, S. Jana, and B. Ray. Deeptest: Automated testing of deep-neural-network-driven autonomous cars. In *Proceedings of the 40th International Conference on Software Engineering*, ICSE '18, pages 303–314, 2018.
31. A. Tiwari, B. Dutertre, D. Jovanović, T. de Candia, P. D. Lincoln, J. Rushby, D. Sadigh, and S. Seshia. Safety envelope for security. In *Proceedings of the 3rd International Conference on High Confidence Networked Systems*, HiCoNS '14, pages 85–94, 2014.
32. M. Uricár, P. Krízek, D. Hurych, I. Sobh, S. Yogamani, and P. Denny. Yes, we GAN: applying adversarial techniques for autonomous driving. *CoRR*, abs/1902.03442, 2019.
33. H. Zhang, H. Chen, Z. Song, D. Boning, inderjit dhillon, and C.-J. Hsieh. The limitations of adversarial training and the blind-spot attack. In *International Conference on Learning Representations*, 2019.