Identify Potential Attacks from Simulated Log Analysis

Cinzia Bernardeschi*, Andrea Domenici*, Francesco Mercaldo† ‡, Antonella Santone§

*Department of Information Engineering, University of Pisa, Pisa, Italy
†Department of Medicine and Health Sciences “Vincenzo Tiberio”, University of Molise, Campobasso, Italy
‡Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy
§Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy
{cinzia.bernardeschi, andrea.domenici}@unipl.it
francesco.mercaldo@unimol.it
francesco.mercaldo@iit.cnr.it
antonella.santone@unimol.it

Abstract—Modern vehicles contain an ecosystem of several electronic units able to exchange data using the serial communication provided by the CAN bus. This protocol can be afflicted by a plethora of attacks that can expose the driver and the passengers to risks for their safety. In this paper we propose a method to detect potential attacks in automotive networks. We start from the analysis of a log obtained from a simulation and we consider a formal verification environment to verify whether the formal model we built from the log is safe. As a proof of concept, we evaluate the proposed method in a case study related to adaptive cruise control, to preliminarily demonstrate its effectiveness.

I. INTRODUCTION

Modern vehicles exhibit a large number of networked components that as a whole are responsible for monitoring and controlling the state of the car [1]. Each component is responsible for communicating with neighboring components: this traffic generates a large amount of data [2]. As a matter of fact, automobiles contain upwards of fifty electronic control units (ECUs) networked together [3]. For this reason, the overall safety of the car relies on near real-time communication between the ECUs [4]. While ECUs are communicating, they must detect skids, predict crashes, perform anti-lock braking and a lot of other safety-critical tasks. [5], [6].

Usually, ECUs are networked using one or more buses complying with the Controller Area Network (CAN) standard. CAN† is a high-integrity serial data communication technology developed in the early 1980s by Robert Bosch GmbH with the aim to provide efficient communication between several automotive applications [7]. The ECUs communicate with each other by sending CAN packets. These packets are broadcast to all components on the bus and each component checks if a packet is intended for it, although segmented CAN networks exist [8], [9]. In CAN packets there is no built-in source identifier or authentication. These are the reasons why it is easy for malicious components [10] to both sniff the CAN network as well as masquerade as other ECUs and send CAN packets. The lack of a source identifier also makes it difficult to monitor traffic because it is impossible, a priori, to know which ECU is sending or receiving a particular packet.

The safety of drivers and passengers relies critically on the code running in their automobiles and the threat to their physical well-being is real [11], [12], [3].

In this paper, a method is exposed to identify conditions suggesting the possibility of an impending attack, based on the analysis of logs from simulated attacks. Once these conditions have been identified and characterized, it would be straightforward to implement algorithms that recognize them at run-time, enhancing system security and safety. The method relies on a tool that, given a simulation log, creates an abstraction of that log in the form of a labeled transition system, and a model-checking prover for temporal logic is used to check if a given condition on the system state (e.g., distance and relative speed of vehicles) is compatible with an attack.

Any kind of simulation environment can be used with this method. In automotive applications, systems are usually modeled and simulated with a block-based graphical tool, such as Simulink [13]. Such tools model a system with a set of interconnected functional blocks realizing mathematical or logical operations on system quantities, such as sampling, amplification or attenuation, or integration. Discrete components are modeled as automata. In this work, systems are assumed to be modeled in Simulink.

As a proof of concept, the method has been applied to a case study related to adaptive cruise control (ACC), a widely used system that automatically adjusts vehicle speed to maintain a safe distance from vehicles ahead. Such systems may use a radar or laser sensor or a camera setup to measure the distance and relative speed of the vehicle ahead, allowing the controller to decelerate or accelerate accordingly. ACC is already available in many cars and will be a key component in the future driverless cars.

II. RELATED WORK

In last years, several researches investigated about the safety and security in automotive networks, for instance Checkoway et al. [14] show that remote exploitation is feasible through
expressions and whose transitions correspond to transformations
an automaton whose states correspond to process algebra ex-
composed of atomic symbols representing system’s actions,
given transformation (or inference) rules. The expressions are
expressions that can be transformed algebraically according to
is representing the structure of concurrent processes as ex-
In this work we adopt Milner’s Calculus of Communicating
formal language, and the effects of attacks are studied with co-simulation
resources: radar, traffic and neighboring cars with the aim
the average beacon time for vehicular networks.
practical attacks in real-time, based on performance metrics
Authors in [16] design an anomaly detector considering a
New Century (CWB-NC) [27] formal verification environ-
Lyamin et al. [19] propose an algorithm to detect denial
of service attacks in real-time, based on performance metrics
nodes that are the source of the incorrect information
node able to detect incorrect information and identify the
Alheeti et al. [17] propose an intelligent IDS-based system
to safeguard cars from attacks. Their system computes the
Lyolle et al. [20] discuss a sensor-based approach that makes
Liu et al. [18] propose a system to secure the communi-
ranges over system variables:
I. Model Checking Background
In the following, preliminary notions on formal methods
are recalled. To apply formal methods we need a formal
model representing the system under study, and a formal
system to express and prove properties of the system. In
this work, we adopt Milner’s Calculus of Communicating
Systems (CCS) [25] to build a formal model, and modal mu-
calculus [26] as the formal system.

The Calculus of Communicating Systems is one of the best
known process algebras. The basic idea of process algebras
is representing the structure of concurrent processes as ex-
pressions that can be transformed algebraically according to
given transformation (or inference) rules. The expressions are
composed of atomic symbols representing system’s actions,
and operators to build and combine sequences of actions in
various ways. The semantics of an expression of a process
algebra are given by a labeled transition system (LTS), i.e.,
an automaton whose states correspond to process algebra ex-
pressions and whose transitions correspond to transformations
of expressions. In this way, a process in its initial state is
represented as an expression describing all potential evolu-
tions of that state, depending on the occurrence of different
actions, which in turn determine which transformation rules
are applicable. A transition from a state \( s \) to a state \( s' \), due to
an action \( \alpha \), is represented as \( s \xrightarrow{\alpha} s' \).

Temporal logics are logical formalisms designed to express
properties of the possible evolutions of LTSs. A model-
checking prover is used to verify if a temporal formula holds
for a given LTS.

The syntax of the extended mu-calculus is the following,
where \( K \) ranges over sets of actions (i.e., \( K \subseteq A \)) and \( Z \)
ranges over system variables:

\[
\phi ::= \top \mid \bot \mid Z \mid \phi \land \phi \mid \phi \lor \phi \mid [K]\phi \mid (K)\phi \mid \nu Z.\phi \mid \mu Z.\phi
\]

A fixpoint formula may be either \( \mu Z.\phi \) or \( \nu Z.\phi \), where \( \mu Z \) and \( \nu Z \) binds free occurrences of \( Z \) in \( \phi \). An occurrence of \( Z \) is free if it is not within the scope of a binder \( \mu Z \) (resp. \( \nu Z \)). A formula is closed if it contains no free variables. \( \mu Z.\phi \) is the least fixpoint of the recursive equation \( Z = \phi \), while \( \nu Z.\phi \) is the greatest one. From now on we consider only closed
formulas.

Scopes of fixpoint variables, free and bound variables, can
be defined in the mu-calculus in analogy with variables of first
order logic.

The satisfaction of a formula \( \phi \) by a state \( s \) of a transition
system is defined as follows: each state satisfies \( \top \) and no state satisfies \( \bot \); a state satisfies \( \phi_1 \lor \phi_2 \) (\( \phi_1 \land \phi_2 \)) if it satisfies \( \phi_1 \) or (and) \( \phi_2 \); \( [K] \phi \) is satisfied by a state which, for every
performance of an action in \( K \), evolves to a state obeying
\( \phi \). \( (K) \phi \) is satisfied by a state which can evolve to a state
obeying \( \phi \) by performing an action in \( K \).

For example, \( (a) \phi \) denotes that there is an \( a \)-successor in
which \( \phi \) holds, while \( [a] \phi \) denotes that for all \( a \)-successors \( \phi \)
holds.

The precise definition of a closed formula satisfaction \( \varphi \) by
a state \( p \) (written \( p \models \varphi \)) is given in Table I.

A transition system \( \mathcal{T} \) satisfies a formula \( \phi \), written \( \mathcal{T} \models \phi \),
if and only if \( q \models \phi \), where \( q \) is the initial state of \( \mathcal{T} \).

In the sequel we will use the following abbreviations:

\[
\langle \alpha_1, \ldots, \alpha_n \rangle \phi = \langle \{\alpha_1, \ldots, \alpha_n\} \rangle \phi \\
\neg \phi = \langle \bot \rangle \phi \\
\neg (K) \phi = \langle A - K \rangle \phi \\
[\alpha_1, \ldots, \alpha_n] \phi = \langle [\alpha_1, \ldots, \alpha_n] \rangle \phi \\
[\neg] \phi = \langle \neg A \rangle \phi \\
[\neg (K)] \phi = \langle A - K \rangle \phi
\]

In this work we resort to the Concurrency Workbench of
New Century (CWB-NC) [27] formal verification environ-
ment, which supports several different specification languages,
among which CCS. In the CWB-NC the verification of tem-
poral logic formulæ is based on model checking [28].
with ϕ (depicted in Figure 1) and verification and testing in the last years. Techniques, that demonstrated to be effective in software and/or sensors. To this aim, we resort to formal verification the vehicle under analysis i.e., without additional hardware expected behavior by simply analyzing the log retrieved from target the adaptive cruise control) from the legitimate criminate behind a driving session with an attack (for instance, some different values if compared with the legitimate behavior. generated. The effect of the attack will result in a log with effect of a possible attack and the logs of the simulation are car is violated, brakes must be activated. Expected behavior are identified on the system, as expected to be true. For example, if a safe distance between requirements for the expected behavior have been identified, conditions and logs of sensed data and data sent to actuators are generated. Safety requirements are identified on the system, as well as the operating conditions and logs of sensed data and data sent to actuators. To this aim, we resort to formal verification environment is used to verify whether an attack is potentially happening. The LTS is then built from the discretized log. Each log entry is the (abstract) state of the system at each simulation step, represented by the tuple of feature values. The set of values for each feature is taken as the set of actions of the process at step i, and for each step, represented by the tuple of feature values, all terminated by the process at step i + 1. More details about the model construction can be found in [29]. An example of formal model generation from a log with two features (i.e., F1 and F2) is shown in Figure 2.

IV. IDENTIFICATION OF ATTACKS

Our analysis starts from a Simulink model of the system under analysis. The system is simulated under operating conditions and logs of sensed data and data sent to actuators are generated. As hinted in Figure 1, the values of relevant variables (features) are stored in comma-separated-values (CSV) format. The logs are then converted to a so-called discretized form, where each actual value is mapped to a symbolic value in the set {low, medium, high}, with respect to given thresholds. For example, the values of some feature could be mapped to low if less than or equal to −0.1, medium if between −0.1 and 0.1, and high if greater than or equal to 0.1.

The LTS is then built from the discretized log. Each log entry is the (abstract) state of the system at each simulation step, represented by the tuple of feature values. The set of values for each feature is taken as the set of actions of the process at step i, and for each step, represented by the tuple of feature values, all terminated by the process at step i + 1. More details about the model construction can be found in [29]. An example of formal model generation from a log with two features (i.e., F1 and F2) is shown in Figure 2.

TABLE I
SATISFACTION OF A CLOSED FORMULA BY A STATE

<table>
<thead>
<tr>
<th>Condition</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>p ≠ t f t</td>
<td>p = t f</td>
</tr>
<tr>
<td>p = ϕ ∧ ψ</td>
<td>p = ϕ and p = ψ</td>
</tr>
<tr>
<td>p = ϕ ∨ ψ</td>
<td>p = ϕ or p = ψ</td>
</tr>
<tr>
<td>p = (K)r ϕ</td>
<td>∃p',rα ∈ K.p →p'/R.p' and p' = ϕ</td>
</tr>
<tr>
<td>p = νZ.ϕ</td>
<td>p = νZ.ϕ for all n</td>
</tr>
<tr>
<td>p = µZ.ϕ</td>
<td>p = µZ.ϕ for some n</td>
</tr>
</tbody>
</table>

| 
| 
| 
| where for each n, νZ^n.ϕ and µZ^n.ϕ are defined as: |
| νZ^n.ϕ = tt |
| μZ^n.ϕ = ff |
| νZ^{n+1}.ϕ = ϕ[νZ^n.ϕ/Z] |
| μZ^{n+1}.ϕ = ϕ[μZ^n.ϕ/Z] |
| with ϕ(ψ/Z) denoting the substitution of ψ for every free occurrence of the variable Z in ϕ. |

IV. IDENTIFICATION OF ATTACKS

Our analysis starts from a Simulink model of the system under analysis. The system is simulated under operating conditions and logs of sensed data and data sent to actuators are generated. Safety requirements are identified on the system, as expected to be true. For example, if a safe distance between car is violated, brakes must be activated.

Then the design of the system is instrumented with the effect of a possible attack and the logs of the simulation are generated. The effect of the attack will result in a log with some different values if compared with the legitimate behavior.

The idea behind our proposal is that it is possible to discriminate behind a driving session with an attack (for instance, targeting the adaptive cruise control) from the legitimate expected behavior by simply analyzing the log retrieved from the vehicle under analysis i.e., without additional hardware and/or sensors. To this aim, we resort to formal verification techniques, that demonstrated to be effective in software verification and testing in the last years.

Then the method we propose is composed of two steps: Formal model generation (depicted in Figure 1) and Potential attack detection (depicted in Figure 3).

Fig. 1. Formal model generation.

As hinted in Figure 1, the values of relevant variables (features) are stored in comma-separated-values (CSV) format. The logs are then converted to a so-called discretized form, where each actual value is mapped to a symbolic value in the set {low, medium, high}, with respect to given thresholds. For example, the values of some feature could be mapped to low if less than or equal to −0.1, medium if between −0.1 and 0.1, and high if greater than or equal to 0.1.

The LTS is then built from the discretized log. Each log entry is the (abstract) state of the system at each simulation step, represented by the tuple of feature values. The set of values for each feature is taken as the set of actions of the process at step i, and for each step, represented by the tuple of feature values, all terminated by the process at step i + 1. More details about the model construction can be found in [29]. An example of formal model generation from a log with two features (i.e., F1 and F2) is shown in Figure 2.

Fig. 2. An example of formal model generation.

Once the formal model has been generated and safety requirements for the expected behavior have been identified, a formal verification environment is used to verify whether an attack is potentially happening. Let S(P) be the automaton corresponding to the log for a driving session and ϕ be the formalization of a safety requirement. If S(P) does not satisfy the safety requirement,
we suspect that a potential attack is happening. Figure 3 describes the potential attack detection step.

\[ \varphi = \nu X.[b][c] \varphi_1 \land [-] X \]
\[ \varphi_1 = [a] \text{ff} \]

To verify a formal model in the CWB-NC environment, we use the \texttt{chk} command, with the formal model \( S(P) \) to verify and the \( \varphi \) property as inputs.

V. A CASE STUDY: ADAPTIVE CRUISE CONTROL SYSTEM

As a case study, an adaptive cruise control system is considered, taken from the Matlab/Simulink documentation, which includes several examples in the automotive field\(^2\). The Simulink scheme is shown in Figure 4.

The \textit{Lead car} (the box on the left of the figure) is parametric with respect to the acceleration function. The \textit{Ego car} reads the actual position and the actual velocity of the Lead car. Then these parameters are provided as input to the Controller, which together with the actual position and velocity of the Ego car, computes the new acceleration of the Ego car.

Actually, the schema in Figure 4 is a variant of the original model. In particular, an attack has been added that changes the value of the position of the Lead car, after this value has been read by the Ego car sensors and before this value has been provided as an input to the Controller. As a proof of concept, we assume that the attack increases the position of the Lead car by 60 units. Graphically, the attack is modeled by a block named \texttt{Attack} in the figure.

Figure 5 shows the system behavior, under some operating conditions. In particular, \( D_{safe} \) is the safe distance between the Lead and the Ego car. The figure shows that the relative position is always greater than the safe distance and that in the steady state the Ego car has the same behavior of the Lead car.

\(^2\)https://it.mathworks.com/help/mpc/ref/adaptivecruisecontrolsystem.html

Figure 6 shows the system behavior in case of attack. At time 14.5 sec, the relative position becomes negative, which means that there is a crash; at time 4.2 the safety requirement described below is violated. Assume the log stores the following information: \texttt{RelativePos}, the difference between the actual positions of the two cars (read by sensors); \texttt{RelativeSpeed}, the relative velocity of the two cars; \texttt{EgoAcc}, the acceleration of the Ego car and \texttt{RelativePos - Dsafe}, the difference between the actual distance and the safety distance of the two cars. An excerpt of the Log file is shown in Table II, where \texttt{State} is the timestamp. The duration of a step is 0, 1. State 145 corresponds to 14.5 sec in the simulation.

An example of safety requirement is the following:

\textit{if the distance between the cars is less than the safety distance and the difference between the speeds of the two cars is negative, then the Ego car must decelerate.}

Below we show the representation of the formula in mu-
TABLE II

LOG FILE (AN EXCERPT).

<table>
<thead>
<tr>
<th>State</th>
<th>RelativePos</th>
<th>RelativeSpeed</th>
<th>EgoAcc</th>
<th>(RelativePos - Dsafe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>98</td>
<td>45.2629324696</td>
<td>-0.1734871587</td>
<td>-0.0543565760</td>
<td>0.0200504475</td>
</tr>
<tr>
<td>99</td>
<td>45.2459171683</td>
<td>-0.1668540851</td>
<td>-0.0544967007</td>
<td>0.0123214492</td>
</tr>
<tr>
<td>100</td>
<td>45.2295540452</td>
<td>-0.1604368351</td>
<td>-0.0538242986</td>
<td>0.0049424761</td>
</tr>
<tr>
<td>101</td>
<td>45.2138233922</td>
<td>-0.1542013153</td>
<td>-0.0526345966</td>
<td>-0.0020584492</td>
</tr>
<tr>
<td>102</td>
<td>45.1987079325</td>
<td>-0.1481315931</td>
<td>-0.0511222645</td>
<td>-0.0086762978</td>
</tr>
<tr>
<td>103</td>
<td>45.1841914158</td>
<td>-0.1422221882</td>
<td>-0.0468068141</td>
<td>-0.0149196477</td>
</tr>
<tr>
<td>104</td>
<td>45.1702569607</td>
<td>-0.1364976159</td>
<td>-0.0441163354</td>
<td>-0.0208397016</td>
</tr>
</tbody>
</table>

Fig. 5. Simulink output.

Fig. 6. Simulink output in case of attack.

Fig. 7. The output of the model checker.

Discrete-time calculus temporal logic:

\[ \varphi = \nu X. [\text{low\_RelativeSpeed}] \]

\[ [\text{low\_RelativePos \_ Dsafe}] \varphi_1 \land [-] X \]

\[ \varphi_1 = [\text{high\_EgoAcc}] \mathbf{ff} \]

The \( \varphi \) formula describes the following property: it must never happen that the relative speed and the (relative position - Dsafe) assume a low value and the acceleration in the next state exhibits a high value.

Figure 7 shows the output of the model checker when the legitimate and the attack models are verified. When the legitimate model is evaluated, the \( \varphi \) property results true (i.e., the system is safe) while, when the attack model is evaluated, the CWB-NC outputs false, symptomatic of an unsafe behavior (i.e., a possible attack).

We show also a variant of safety requirement than can be used to anticipate the potential attack:

\[ \chi = \nu X. [\text{low\_RelativeSpeed}] \]

\[ [\text{low\_RelativePos \_ Dsafe}] \chi_1 \land [-] X \]

\[ \chi_1 = [\text{medium\_EgoAcc}] \mathbf{ff} \]

The \( \chi \) formula describes the following property: it must never happen that the relative speed and the (relative position - Dsafe) assume a low value and the acceleration in the next state exhibits a medium value. This second formula can be helpful to alert that a potential attack can happen in a short
time (if the acceleration continues to increase). When the legitimate model is evaluated, the \( \chi \) property results true while, when the attack model is evaluated, the CWB-NC outputs false, symptomatic that an unsafe behavior can happen in the near future. One of the interesting points of the proposed approach is that it adopts variables currently measured in vehicles.

VI. CONCLUSIONS AND FUTURE WORK

This paper has introduced a method whereby simulated cyber-attacks to networks of ECUs in automotive systems can be identified from patterns in simulation logs. An algorithm has been designed and implemented that abstracts a system’s behavior by building an LTS defined in CCS. A temporal logic formula to check if the attack can drive the system into an unsafe condition, written in the mu-calculus language, can then be proved or disproved with a model-checking prover. This procedure is arguably useful at an early design stage, in the phase of risk analysis, when simulation can be used to assess the consequences of various types of attacks. Further, it makes it possible to characterize the conditions revealing a possible attack and have them detected at run-time with a simple monitoring algorithm.

ACKNOWLEDGMENT

Work partially supported by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Department of Excellence) and by the PRA 2018_81 project entitled Wearable sensor systems: personalized analysis and data security in healthcare funded by the University of Pisa.

REFERENCES