University of Pisa
School of Engineering

Formal Methods for Secure Systems

An ADVISE model for attacks on autonomous vehicles back-end servers

Supervisors
Prof. Cinzia Bernardeschi
Ing. Maurizio Palmieri

Students
Yuri Mazzuoli
Francesco Iemma
Marco Pinna

July 18, 2021

Contents

1 Introduction **2**

2 Overview **3**
2.1 Actors 3
2.2 Attacker's Profile 4
2.3 Goals 5
2.4 Countermeasures 5

3 Attack tree **7**
3.1 Attacks 8

4 Simulation **10**
4.1 General Results 10
4.2 Hacker Attacks 11
4.3 Physical Intruder and Insider Attacks 13

5 Appendices **14**

Chapter 1

Introduction

In what follows the study of different adversaries trying to attack the back-end servers related to autonomous vehicles is carried out.

The work is organized as follows:

- Firstly, chapter 2 gives an initial overview and a presentation of the scenarios.
- Secondly, in chapter 3 the complete attack tree is presented.
- In chapter 4 the development of the Mobius simulation is described and its results are presented.

The entire codebase is available at `https://github.com/YuriMzz/vehiclesADVISE`.

Chapter 2

Overview

In this chapter a general overview of the scenarios and the involved actors is given.
The final goal for each one of the attackers it to gain access to the back-end servers related to self-driving vehicles. This goal can be reached by means of different knowledges, skills and types of attacks, which all differ from one adversary to the other.

The attack steps and the goals analyzed in this report are based on *"Agreement Concerning the Adoption of Harmonized Technical United Nations Regulations for Wheeled Vehicles, Equipment and Parts which can be Fitted and/or be Used on Wheeled Vehicles and the Conditions for Reciprocal Recognition of Approvals Granted on the Basis of these United Nations Regulations - 4 March 2021"* and in particular on Table A1 page 20 (see Appendix A).

2.1 Actors

Three different types of actor have been identified:

Insider

An insider is defined as someone who belongs to the company or works for it. They already have, therefore, access to some facilities (depending on their role in the company) and the equipment they contain (PCs, laptops, routers, etc.), the knowledge necessary to perform some tasks (internal network passwords, website and/or database credentials, VPN access, etc.) and possibly the trust of the rest of the staff.
They might also be aware of potential flaws in the company's systems (vulnerabilities in information system, defective equipment, weaknesses of some employees, etc.).
On the other hand, an insider does not necessarily have the skills needed to perform some of the intermediate attack steps needed to reach the final goal.

Physical intruder

A physical intruder is different from an insider since they have less knowledge about the internals of the company. However, they have advanced *physical* penetration testing skills and tools, as well as the knowledge about the company facilities location, security systems and possibly the timetables of employees and/or security personnel.
Therefore they might have access to a bigger number of facilities and spaces (e.g. they might have – or be able to gain – direct access to a server room).

Hacker

A hacker is an external actor with no prior knowledge of the company's systems and no access to its facilities. On the other hand, they have advanced security and penetration skills, in-depth knowledge of the most important and widespread technologies and attacks, as well as the tools needed to perform these attacks on the target systems.

Actor	**Skills**	**Knowledges**	**Accesses**
Insider	Basic/medium IT skills, social engineering	Login credentials (website, LAN, VPN, DB), employees info, vulnerabilities in company's security system	LAN, VPN, DB, facilities, PCs, workstations, routers, switches
Physical intruder	Lockpicking, ???	Facilities location, personnel timetables	Company's external premises
Hacker	Reverse engineering, social engineering, advanced attack and penetration testing	-	Vehicle (i.e. firmware)

Table 2.1: Attackers' skills, knowledges and accesses

2.2 Attacker's Profile

As we have seen each actor has its own set of knowledges, accesses and skills in table 2.1 it is possible to see the arrangement of them among the attackers.
For what concern the attention of the actors about the probability of detection, the cost of the attacks and the expected payoff, the weights for each of these factors are shown in the table 2.2.
Thus we can see that the choices that attackers done are guided by different things, for instance an insider will prefer to avoid an attack with an high probability of detection, as well as a physical intruder, instead an hacker has not this preoccupation and he will choice the attack with the higher payoff.

Attacker Name	Cost Weight	Detection Weight	Payoff Weight
Hacker	0.1	0.1	0.8
Physical Intruder	0.2	0.3	0.5
Insider	0.1	0.4	0.5

Table 2.2: Weights of Attacker Profile

2.3 Goals

In the following we will describe the goals that an attacker can achieve.

Vehicle Undesidered Behaviour

The back-end servers can be used as a means to attack a vehicle and/or extract data from it e.g. its position, the destination the driver is heading to etc.
Moreover having accesses to the back-end server allow the attacker to cause an undesidered behaviour of the vehicle.
This is the most rewarding attack (reward equal to 300 in the ADVISE model) because allow the attacker to completely control the vehicles.

Data breach

Servers can be also attacked to extract sensitive data related to customers.
This is the second most rewarding attack (reward equal to 150 in the ADVISE model) because the data can be very useful and valuable to the attacker, but he has no control over the vehicles.

Back-end server service disruption (DoS)

An attacker could also target the back-end server just to take them down and disrupt their service (Denial of Service), causing issues to all the vehicles whose proper functioning relies on it.
It is the the attack with the lower reward value (150) because it allows the attacker only to stop the service.

2.4 Countermeasures

We have considered the following countermeasure in order to cope with the attacks that the attackers wants to perform.

Intrusion Detection System

It is a software that inspects the network in order to detect unauthorized intrusion or unauthorized changes in the level of privileges, it affects the attack step *Privilege Esca-*

lation and so the Insider and the Physical Intruder. We consider two level of sensitivity: 0 (disabled) and 1 (enabled).

Code Obfuscation

It is the technique that consists in the deliberate act of creating source code that is difficult for humans to understand. It affects the *Firmware Reversing* attack and so the Hacker. We consider two level of sensitivity: 0 (disabled) and 1 (enabled).

Firewall

We consider different qualities of firewall, it affect the *Port Scan* and so the Hacker. We consider three level of sensitivity: 0 (no firewall), 1 (bad quality firewall), 2 (very good firewall).

Chapter 3

Attack tree

The figure 3.1 represents the attack tree with the paths that the attackers can follow.

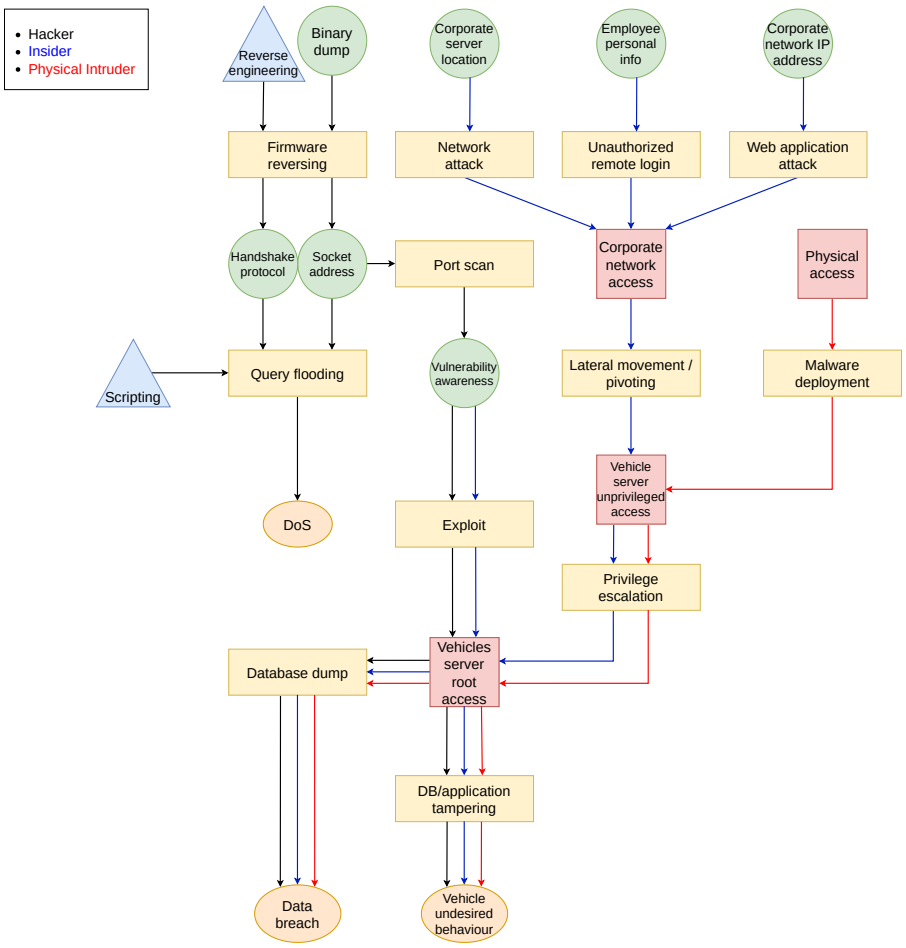

Figure 3.1: Attack tree

3.1 Attacks

All the attacks has its own probability of achieving results and each of them needs a different time to be executed: in the attack tree we have considered *hours* as unit of time and a deterministic distribution for all the attacks except for the *Query Flooding* (rate equal to 0.1) and *Exploit* (rate equal to 0.2).
Below there is a brief description of every attack in the tree with the reference to the table used as starting-point (see Appendix A):

- **Firmware reversing**: a hacker in possession of a vehicle and with reverse engineering skills can dump the vehicle firmware, reverse engineer it and acquire information about the IP/port of the server, as well as the protocol used for the communication and/or the structure of the messages sent.
- **Query flooding**: a hacker can perform a (D)Dos attack by sending a large number of queries on the server (possibly by means of a botnet) and overwhelming it so that it stops being responsive. (2.1)
- **Port scan** with the knowledge of the server's IP, a hacker can scan the server for open ports and possibly find out the OS, which services are running and their versions.
- **Exploit** If an attacker knows a vulnerability inside the system he can exploit it obtaining a root access to the vehicle server. (1.2)
- **Network attack** An insider supposedly already has access to the corporate network but, in case they don't, they can either discover the password in a variety of ways (dictionary attack, password re-use, unattended workstations or laptops) or exploit some vulnerabilities in the Wi-Fi protocol. (1.2)
- **Unauthorized remote login**: an attacker could access the corporate VPN by either discovering some employee's login credentials (e.g. phishing, typosquatting, OSINT, social engineering), exploiting some vulnerabilities in the login procedure itself or by directly having access to credentials (i.e. password dumps found or bought online). (1.2)
- **Web application attack**: an attacker could gain unauthorized access to a reserved area of the company's website by means of some web-application attacks (SQLi, XSS, XXE, etc.) or backdoors. (1.2) (3.3)
- **Lateral movement/pivoting**: an attacker who has already gained access to a part of the network which is not directly the vehicle server, can exploit further vulnerabilities and "move laterally" i.e. move deeper inside the network to find more vulnerabilities and entry points that could lead them to the final target server. (1.1) (1.2) (3.1) (3.3)

- **Malware deployment**: an attacker can deploy a malware in different ways e.g. dropping an infected USB drive and wait for some employee to plug it in a corporate machine. The malware can guarantee them the access to a server with a varying level of privilege. (1.3) (3.4)
- **Privilege escalation**: once an attacker has gained the so-called "foothold" into the target server, they will try to increase their privilege level on the system as to access to a greater number of resources. (1.2) (3.4)
- **Database dump**: once the attacker has gained access to the database, the latter can be dumped and the *Data Breach* goal is reached. For the sake of simplicity the `database dump` attack in the tree is considered to have constant time duration and probability of detection. In reality this is not the case as there are different ways to perform data exfiltration, which can be more or less stealthy (simple `scp` or `wget` commands – or their equivalent on non-Unix systems – will be faster but they will also have a high probability of detection while looking at the network traffic and the server logs, while other techniques such as ICMP exfiltration will be stealthier at the expense of a longer time required). (3.2) (3.5)
- **DB/application tampering**: if the goal of the attacker is not to access sensitive data but to cause undesired behaviour of the vehicles, once they have access to the DB or the application they can tamper the data in the DB, or attack the application/framework that is running on the server and which interacts with the vehicles. (2.1) (3.2)

Chapter 4

Simulation

4.1 General Results

We decide to simulate 200 units of time (200h) with Mobius in order to evaluate each attacker results for different configuration of mitigations. We obtain those general results:

- the **Hacker** will always try to achieve the most remunerative goals, despite the possibility to be discovered;
- the **Physical intruder** will try to achieve only the most remunerative goal, and will ignore others because of the high probability of being discovered (risk-reward ratio is too low)
- the **Insider** has the same behaviour of the Physical intruder, but for him the probability to reach the goal increases faster because his attack path has higher success probability.

4.2 Hacker Attacks

DoS

When no mitigation are applied, the Hacker will rarely try to obtain a DoS, but he will commit himself on other attacks in order to obtain high reward goals. Introducing **CodeObfuscation**, binary reversing become harder to complete with success, leading to a probability of reach DoS near to 0.
Increasing the firewall sensitivity will indeed increase the probability for the hacker to reach the DoS, because he will give up on trying to reach other goals, focusing only on this one.

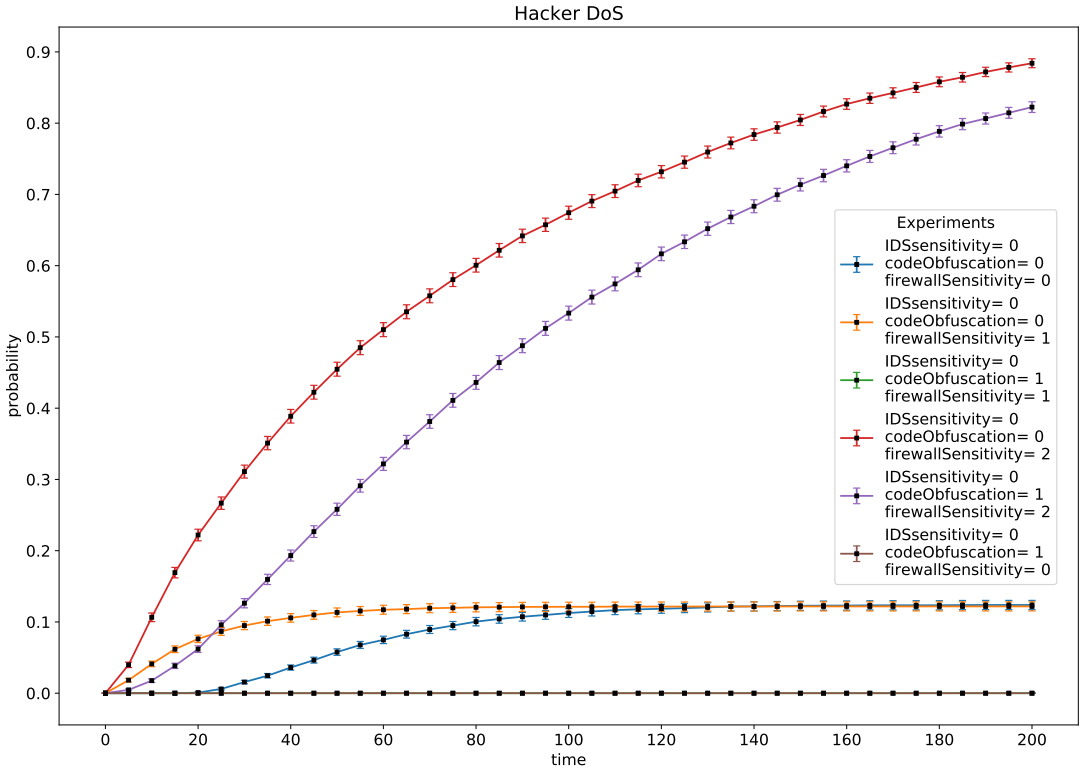

Figure 4.1: Hacker DoS

Data Breach and Vehicle Undesidered Behaviour

When no mitigation are applied, the Hacker will reach those goals pretty fast. Introducing **CodeObfuscation** will slow him down, but increasing **Firewall Sensitivity** we are able to reduce his will to reach them until he will give up.

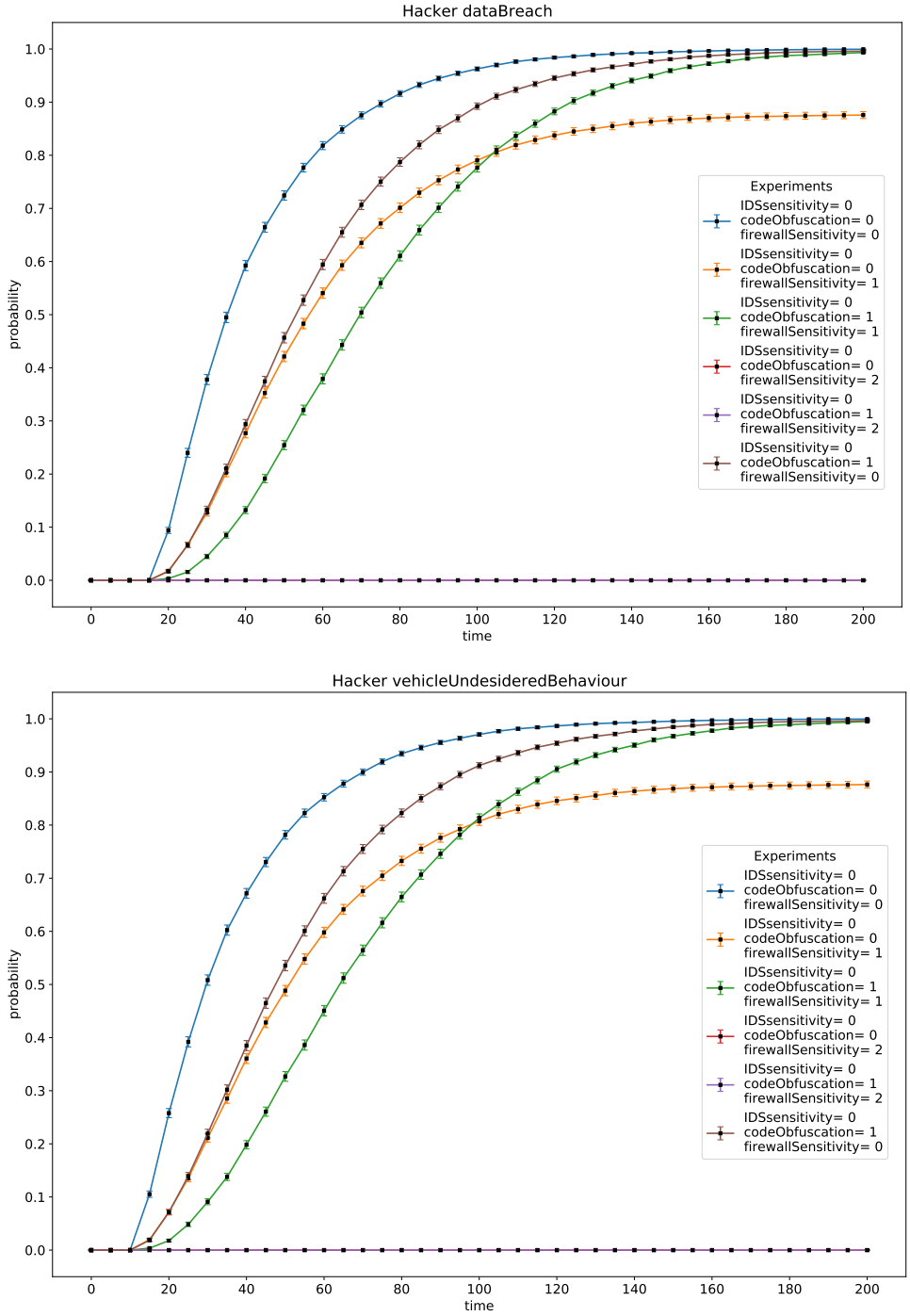

Figure 4.2: Hacker Data Breach and Vehicle Undesidered Behaviour

4.3 Physical Intruder and Insider Attacks

Vehicle Undesidered Behaviour

When no mitigation are applied, both attackers will reach this goal pretty fast. Increasing **Intrusion Detection System Sensitivity**, we are able to reduce their will to reach the goal until they will give up, because the probaility to be discovered will be to high.

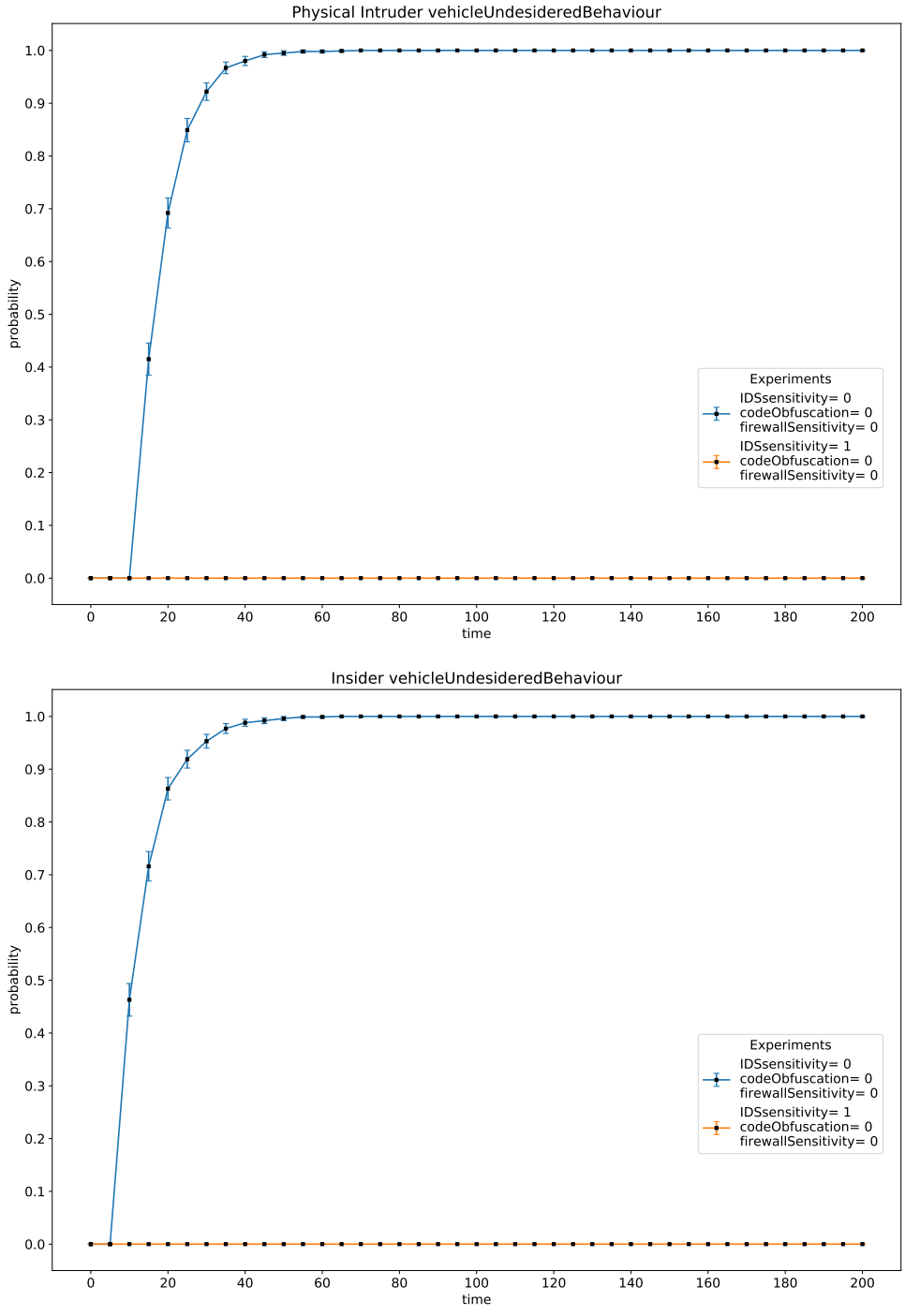

Figure 4.3: Physical intruder and Insider Vehicle Undesidered Behaviou

Chapter 5

Appendices

Appendix A

High level and sub-level descriptions of vulnerability/ threat			*Example of vulnerability or attack method*	
4.3.1 Threats regarding back-end servers related to vehicles in the field	1	Back-end servers used as a means to attack a vehicle or extract data	1.1	Abuse of privileges by staff (**insider attack**)
			1.2	**Unauthorized internet access** to the server (enabled for example by backdoors, unpatched system software vulnerabilities, SQL attacks or other means)
			1.3	**Unauthorized physical access** to the server (conducted by for example USB sticks or other media connecting to the server)
	2	Services from back-end server being disrupted, affecting the operation of a vehicle	2.1	**Attack on back-end server stops it functioning**, for example it prevents it from interacting with vehicles and providing services they rely on
	3	Vehicle related data held on back-end servers being lost or compromised ("data breach")	3.1	Abuse of privileges by staff (**insider attack**)
			3.2	**Loss of information in the cloud**. Sensitive data may be lost due to attacks or accidents when data is stored by third-party cloud service providers
			3.3	**Unauthorized internet access to the server** (enabled for example by backdoors, unpatched system software vulnerabilities, SQL attacks or other means)
			3.4	**Unauthorized physical access to the server** (conducted for example by USB sticks or other media connecting to the server)
			3.5	**Information breach** by unintended sharing of data (e.g. admin errors)