Electromagnetic Signal Injection Attacks on Embedded Systems: Modeling and Detection

Youqian Zhang

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

This thesis is dedicated to
Ms. Huishu Ji and Ms. Tieou Song

Acknowledgements

When I was a junior student, my teachers, Ms. Huishu Ji and Ms. Tieou Song, always encouraged me to study hard and think big. They mentioned the world-leading universities very often, and it was where I learned about the University of Oxford. Since then, studying there has become my dream, motivating me to pursue the highest degree - Doctor of Philosophy (DPhil). Years passed, and here came the ultimate realization of the dream. Conducting research and converting the results into papers are challenging and tedious, but I am grateful that I have always received great support while studying at Oxford. I want to express my appreciation to those who helped and spurred me on.

First and foremost, I want to thank all my family members, particularly my parents and grandparents, who taught me tenacity and supported me with their endless love.

I would like to extend my sincere gratitude to Rasmussen's research group for their helpful feedback on my work and for sharing cultural diversities. In particular, Professor Kasper Rasmussen, not only criticized my work critically and pointed out my problems timely but also encouraged me when I felt down. During Covid-19, Professor Kasper Rasmussen provided me with both mental and financial help to ensure my research progressed smoothly. I would also like to acknowledge Dr. Ilias Giechaskiel's guidance in my first-year DPhil study, and co-authoring a paper with him was an amazing experience, from which I learned how to use the devices in the laboratory and how to convert results into papers. Mr. Lukáš Halgaš, a genius and my best friend, helped me get familiar with fabulous local restaurants and figure out where to find the best coffee beans all over the world.

Especially, I would like to thank my undergraduate supervisor Professor Chao Lu at the Hong Kong Polytechnic University (PolyU hereafter), for always giving me great support and constructive suggestions while exploring my academic interests. Moreover, I owe a big thank you to Dr. Grace Ngai, Dr. Stephen Chan, Mr. Kenneth Lo, Mr. Chi Kin Lau, and many other colleagues from the Service Learning and Leadership Office (SLLO) of PolyU, where I learned the importance of social responsibility. This strengthened my determination to choose my current research topic so as to develop more secure systems to serve all people in the world.

A huge thank you to all my friends worldwide, and it is impossible to finish this thesis without their caring, encouragement, and love. Triple big shout-outs to

the New College Boat Club and the Oxford University Lawn Tennis Club; they make my life in Oxford full of joy.

Last but not least, I would like to thank PAG (formerly Pacific Alliance Group), New College, Univerisity of Oxford, Department of Computer Science, and the China Scholarship Council for providing scholarships and funding to cover my university fees and living expenses, as well as conference fees and equipment purchases.

Abstract

Embedded systems are ubiquitous in our lives, from smart locks in home automation to robotic arms in industrial equipment, playing key roles in many safety- and security-critical applications. An embedded system can interact with the external world through three interfaces: it uses *sensors* to sense environmental changes, controls *actuators* to cause physical impacts, and exchanges information with others through *transmission lines*. In recent years, studies have demonstrated using electromagnetic interference (EMI) to wirelessly manipulate signals in these interfaces. Such manipulation can maliciously control the embedded systems, threatening users' privacy and safety, for example, unlocking a smart lock or raising the temperature of infant incubators.

Detecting such attacks is becoming increasingly essential, but proposed detection methods in the literature are designed for specific applications. Thus, this thesis proposes two novel detection methods that can protect various systems regardless of their types, filling the gap of generalized detection methods. The first detection method is for the sensors, and its core idea is to modulate the sensor power in a secret pattern unknown to the attacker. To bypass the detection, the attacker must guess the secret correctly; however, this detection method provides a strong security guarantee, where the probability of a correct guess is negligible. The second detection method is designed for the actuators, and its detection principle is to compare a signal to be protected with a reference, between which the difference can indicate whether an attack occurs. This method can guarantee that any attack effectively impacting a victim system will be detected. This thesis will demonstrate that these detection methods do not only provide strong security guarantees but are also lightweight and flexible to be integrated with different systems. In addition to these detection methods, this thesis presents a pioneering study about how to corrupt the signal integrity of differential signaling. Since many popular protocols such as USB, Ethernet, HDMI, and CAN derive their electromagnetic noise immunity from differential signaling, many people believe it can make communications immune to external interference, whereas the study challenges this assumption and shows a state-of-the-art attack that allows an attacker to use fine-tuned EMI to inject arbitrary messages into differential signaling.

Contents

List of Figures **ix**

List of Tables **xi**

List of Abbreviations **xii**

1 Introduction **1**
1.1 Motivation and Research Scope 2
1.2 Research Goals and Contributions 5
1.3 Published Results 7
1.4 Statement of Originality 8

2 Background **9**
2.1 An Overview of Embedded Systems 9
2.2 Fundamentals of Electromagnetic Signal Injections 11
2.3 Summary 13

3 Detection Method for Sensor Systems **15**
3.1 Approach, System-, and Adversary Models 17
3.1.1 Three Types of Sensors 17
3.1.2 Randomized Sensor Output 18
3.1.3 System Model 21
3.1.4 Adversary Model 22
3.2 Attack Detection 23
3.2.1 Detection Rule 23
3.2.2 Interfering with the Bias 24
3.2.3 Security Analysis 25
3.3 Non-constant Physical Quantity 25
3.3.1 Attack Detection for Non-constant Physical Quantities . . . 27
3.3.2 Security Analysis 27
3.3.3 The Sampling Rate of the ADC 28
3.4 Implementation 28
3.4.1 Setup 29

3.4.2 Microphone System 30
3.4.3 Temperature Sensor System 37
3.4.4 Summary of Implementation 41
3.5 Discussion 41
3.5.1 Guaranteeing the Security with Small n for Constant Physical Quantities 41
3.5.2 Trade-off between Security and Speed 42
3.5.3 The Approach for Non-Powered Passive Sensors 43
3.5.4 Difference between PyCRA and the Approach 44
3.6 Summary 46

4 Detection Method for Actuator Systems **47**
4.1 System Model and Adversary Model 49
4.1.1 System Model 49
4.1.2 Adversary Model 51
4.1.3 Two Injection Points 51
4.2 Attack Detection 53
4.2.1 Modeling Differential Amplifier Output 55
4.2.2 Detection Rule and Choice of Parameters 56
4.2.3 Security Analysis 58
4.2.4 Differences Between Injection Points 59
4.2.5 Attacks on Detection Circuit 60
4.3 Extended Maximum Detectable Frequency 61
4.4 Implementation 63
4.4.1 Setup 63
4.4.2 Speaker System 65
4.4.3 Motor Control System 71
4.4.4 Summary of Implementation 73
4.5 Discussion 74
4.5.1 Different Detection Strategies 74
4.5.2 Adaptive Threshold 76
4.5.3 Choice of Differential Amplifiers 77
4.5.4 Difficulty of Canceling Attacking Signals 77
4.5.5 Difficulty of Drive Signal Injection 78
4.6 Summary 79

5 Message Injection into Differential Signaling Systems 80
5.1 System Model and Adversary Model . . . 82
5.1.1 System Model . . . 82
5.1.2 Adversary Model . . . 86
5.2 Bit Injection Attack . . . 87
5.2.1 Bypassing Subtractor . . . 88
5.2.2 Bit Detected Incorrectly in Receiver . . . 89
5.3 Analysis of Success Rate . . . 90
5.3.1 Parameterization . . . 91
5.3.2 Success Rate of Bit Injection . . . 92
5.3.3 Success Rate of Message Injection . . . 95
5.4 Experiments . . . 96
5.4.1 Testbed . . . 96
5.4.2 Subtractor . . . 97
5.4.3 Receiver . . . 101
5.5 Message Injection into CAN . . . 106
5.5.1 CAN Basics . . . 106
5.5.2 Message Injection . . . 108
5.6 Discussion . . . 110
5.6.1 Gaining Knowledge . . . 110
5.6.2 Restricted Attack Power and Distance . . . 111
5.6.3 Future Countermeasures . . . 111
5.7 Summary . . . 111

6 Related Work 113
6.1 Detection . . . 114
6.1.1 Methods for Sensor Systems . . . 114
6.1.2 Methods for Actuator Systems . . . 115
6.1.3 Other Anomaly Detectors . . . 117
6.2 Attenuation . . . 117
6.2.1 Shielding . . . 117
6.2.2 Miniaturization . . . 118
6.2.3 Filtering and Mitigation . . . 118
6.2.4 Robust Hardware . . . 119

7 Conclusion 121

References 124

List of Figures

2.1 Structure of embedded systems. 10

3.1 Model of a sensor system. 18
3.2 An example of a sensor measurement. 19
3.3 A sensor system equipped with the detection method. 21
3.4 A sensor output of a constant physical quantity. 24
3.5 A malicious change in non-constant physical quantity. 26
3.6 A testbed for a microphone system. 29
3.7 A temperature sensor circuit. 30
3.8 One sound signal and one malicious signal with the same frequency. 31
3.9 Measure properties of a microphone output. 33
3.10 Microphone outputs in different cases. 34
3.11 Reconstruction of an audio signal. 36
3.12 Attack power and measured temperature. 37
3.13 Microphone outputs in different cases. 39
3.14 Configuration for non-powered passive sensors. 43

4.1 Model of an actuator system. 50
4.2 A schematic of the detection method for actuator systems. 53
4.3 The minimum detectable power and other parameters. 56
4.4 Extended detectable frequency. 61
4.5 A setup of actuator systems. 64
4.6 Outputs of a differential amplifier. 66
4.7 A concrete injection into a speaker system. 68
4.8 Power ratios between malicious and legitimate signals. 69
4.9 The DC offset of a differential amplifier output. 71
4.10 A diagram of a motor driver. 71
4.11 The DC offset of a differential amplifier output for a motor system. 72
4.12 A timeline of an attack. 74

5.1 System model of differential signaling. 83
5.2 Working principles of a subtractor and a receiver. 84
5.3 Input circuits of differential signaling receivers. 87

5.4 A testbed for a subtractor. 96
5.5 Outputs of a subtractor chip TJA1050. 98
5.6 Power of bypassed injected signals. 99
5.7 Frequency domain of a bypassed injected signal. 100
5.8 The signal to noise-and-distortion (*SINAD*) ratios. 101
5.9 A testbed for a receiver. 102
5.10 Success rates of bit injections in nRF52833. 102
5.11 Pairs of u and v characterizing the chip's responses. 104
5.12 Optimal pairs outstripping other pairs. 105
5.13 A schematic diagram of a CAN bus. 107
5.14 An injected malicious message and its corresponding attacking signal. 107
5.15 A practical setup of message injection attacks on a CAN bus. 108

List of Tables

3.1 Detection results in a microphone system. 35
3.2 Detection results in a temperature sensor system. 41

List of Abbreviations

ADC . . . . . Analog-to-Digital Converter
CPU . . . . . Central Processing Unit
CAN . . . . . Controller Area Network
CM . . . . . Common Mode
CRC . . . . . Cyclic Redundancy Check
CMOS . . . . . Complementary Metal-Oxide-Semiconductor
CIED . . . . . Cardiac Implantable Electrical Device
DAC . . . . . Digital-to-Analog Converter
DC . . . . . Direct Current
DUT . . . . . Device Under Test
DPI . . . . . Direct Power Injection
DLC . . . . . Data Length Code
EMI . . . . . Electromagnetic Interference
ESD . . . . . Electro-Static Discharge
HDMI . . . . . High-Definition Multimedia Interface
ID . . . . . Identity
LED . . . . . Light-Emitting Diode
LPF . . . . . Low-Pass Filter
NTC . . . . . Negative Temperature Coefficient
OPAMP . . . Operational Amplifier
PCB . . . . . Printed Circuit Board
PCC . . . . . Pearson's Correlation Coefficient
RF . . . . . Radio Frequency
SINAD . . . . Signal to Noise-and-Distortion
SNR . . . . . Signal-to-Noise Ratio
USB . . . . . Universal Serial Bus

1 Introduction

Contents

1.1 Motivation and Research Scope 2
1.2 Research Goals and Contributions 5
1.3 Published Results 7
1.4 Statement of Originality 8

The last decades have seen embedded systems promoting various devices' automation and computerization. Unlike a computer for general purposes, an embedded system is a microcontroller-based system that is merely for specific tasks [1, 2]. For instance, a fridge's embedded system is designed to maintain cabinet temperature at a low level; an automatic door's embedded system senses pedestrians' movement and opens/closes a door. Embedded systems automate tasks, helping people live and work smarter and easier.

To achieve such automation, embedded systems need to interact with the external world, and the interaction is associated with *sensors* that measure the physical environment, *actuators* that cause physical impacts, and *transmission lines* that connect different devices. Since these sensors, actuators, and transmission lines (and by extension, embedded systems) guide so many safety- and security-critical functions in different applications in our daily lives, attacking them could directly

threaten users' privacy and safety, possibly leading to severe consequences. Indeed, a large number of researchers have successfully shown that they can remotely manipulate embedded systems' operation by different attack modalities, such as sounds [3–8], lights [7, 9, 10], magnetism [3, 11], and particularly, electromagnetic waves, which this thesis will focus on hereafter. Note that in this thesis, attacks using electromagnetic waves to interfere with signals in embedded systems are called *electromagnetic signal injection attacks*.

1.1 Motivation and Research Scope

Starting from sensors, there are many different types, from thermometers in a fridge and motion detectors in home security systems, to altimeters in drones and pressure sensors in industrial applications. Many previous studies have done thorough research on exploiting "antenna-like" behaviors of metal conductors (e.g., electrical cables or PCB traces) [12, 13] to conduct electromagnetic signal injection attacks to manipulate sensor measurements: in 2013, Kune et al. [14] wirelessly injected an adversarial signal into the leads of cardiac implantable electrical devices (CIEDs) to inhibit pacing and induce defibrillation shocks, which threaten patients' health; in 2015, Kasmi and Esteves [15] modulated malicious voice commands on electromagnetic waves and wirelessly injected them into a smartphone's earphone cable so as to control the smartphone; later on, researchers used similar signal injection attacks to cause ghost touches onto smartphone's touchscreens [16–18], tamper temperature sensor readings in baby incubators [19], and manipulate frames captured by image sensors [20]. The attacks are becoming prevalent. Regarding defenses, traditional methods such as wrapping vulnerable places with Radio-Frequency (RF) shielding materials [14, 15, 21–23] or using Electromagnetic Interference (EMI) filters [24, 25] to attenuate adversarial electromagnetic waves can provide finite protection, but they are not always feasible. Hence, extra defenses are becoming more and more essential. Detection is vital because it allows the embedded system to be alerted when an attack occurs and take further actions to mitigate the attack. Many previous studies have proposed and demonstrated various

detection methods for sensors (please refer to Chapter 6 for details). However, there are still challenges that these methods could not overcome: first, since these methods are devised for specific applications, the solutions cannot be generalized for different applications, leaving many other embedded systems unprotected; second, these methods need complicated circuits or algorithms, meaning that extra hardware or software overheads need to be carefully considered in practice, especially resources/budgets are constrained for deployment. To get rid of these challenges, this thesis brings a novel detection method in Chapter 3, which fills the gap of generalized and lightweight solutions.

Similar attacks can also maliciously control actuators, which are embedded in our daily lives to such an extent that it is hard to find an example of an electronic system that does not have actuators in some form: they are from motors in smart locks and loudspeakers in webcams, to solenoids in pneumatic valves and light-emitting diodes (LEDs) on dashboards of industrial equipment. Yet, only a few studies conducted electromagnetic signal injection attacks on actuators: in 2018, Selvaraj et al. [26] demonstrated that they could inject an electromagnetic wave into the input wire of a servo, which can be used to control the rudder and the aileron of a drone, leading to a crash; in 2020, Dayanıklı et al. [27] demonstrated that they can inject a signal to manipulate the switching state of individual transistors, thus causing irreparable damage to power converters for electric vehicles. Note that it is more complex to manipulate signals that control actuators (or actuator control signals) than sensor measurements. This is because the actuators work at tens or hundreds of watts (e.g., motors) while the sensors operate at a much lower power level, so attacking the actuators consumes more power. Also, the actuators may need complicated control signals: for example, a three-phase motor works by three signals being 120 degrees out of phase, and accordingly, the attacking signals need to be crafted more carefully and delicately. Indeed, it is arduous to manipulate actuators, but it is essential to conduct research on defenses before such attacks are abused. A few studies proposed detection methods for actuators, but they are application-specific (see details in Chapter 6), making them difficult to fit into

different actuator systems. Since there is no generalized detection method, a novel solution is proposed in Chapter 4 to fill this gap. Please note that the detection methods for the sensors do not apply to the actuators because they work in different ways (i.e., different working principles and circuits, and see details in Chapter 4).

Note that the sensor measurements and the actuator control signals above are essentially analog signals. Although so-called digital sensors/actuators can communicate with others by digital signals, these devices have a built-in analog-to-digital/digital-to-analog converter (ADC/DAC). Unless stated otherwise, the sensor measurements and actuator control signals are analog signals hereafter.

Besides these two analog interfaces, digital signals are widely used within/between embedded systems for information exchange as it has a better noise immunity [28]. Still, it is feasible to use electromagnetic signal injection attacks to manipulate digital signals, but the manipulation is a substantially different process and has not received much attention. Please note that early work mainly focused on studying how the electromagnetic interference impacts communication qualities [29–31], and it is more recently that a few studies started to use electromagnetic signal injection attacks to achieve arbitrary manipulation. Previous work showed that an attacker could flip bits [26], and further inject arbitrary messages [32]. In the latest study, Köhler et al., [33] demonstrated an electromagnetic signal injection attack, namely *Brokenwire Attack*, which can interrupt necessary control communications between a vehicle and a charger, causing charging process to abort. It is essential to point out that integrity checks such as checksum and Cyclic Redundancy Check (CRC) can help to spot erroneous bits caused by the attacks. To avoid violating the checks, an attacker must manipulate the whole message. It particularly requires the attacker to know what the original bits are and when they are transmitted, which are arduous and challenging tasks in practice [32]. Although some researchers showed they could realize such an attack in single-end signaling [26, 32], it is easy to block those attacks by differential signaling [32], which transmits information as the difference between two complementary signals. Differential signaling generally works well against electromagnetic noise. Many protocols such as Universal Serial

Bus (USB), Ethernet, High-Definition Multimedia Interface (HDMI), and Controller Area Network (CAN) use differential signaling to achieve a robust communication channel in a noisy environment. This has led many to believe that it is infeasible to remotely inject attacking signals into such a differential pair. However, in Chapter 5, a state-of-the-art attack is shown to challenge this assumption: the feasibility of injecting arbitrary messages into differential signaling lines is studied and demonstrated systematically and experimentally, making it the pioneering work that fills the gap of such attacks.

1.2 Research Goals and Contributions

This thesis focuses on security issues lying in the signal integrity of sensors, actuators, and differential signaling. It aims to propose *two generalized detection methods to protect sensor- and actuator systems from electromagnetic signal injection attacks* and *one tricky attack method that can inject arbitrary messages into differential signaling systems*.

In Chapter 3 (based on [34]), a detection method regarding sensor systems is presented. The basic idea is to encode sensor power in a secret way unknown to the attacker. The detection method can immediately spot the attack if a wrong guess is made to cause malicious changes to sensor measurements. The generality of this method makes it flexible to different types of sensors, and its simplicity only requires tiny changes in software and hardware for implementation. Although the detection method is lightweight, it provides a strong and provable security guarantee for embedded systems. This detection method is further validated on a microphone system and a temperature sensor system to show its effectiveness and robustness.

Next, actuator systems are considered in Chapter 4 (based on [35]). The detection principle is to monitor whether the actuator control signal is what it should be. Specifically, this is realized by comparing the signal with a reference, where both should be identical, and any unexpected difference will indicate attacks. The novelty of the detection method is using a comparator in an uncommon way: it deliberately captures the attacking signals rather than get rid of them. Since the

comparator handles its input signals regardless of their types, the detection method is applicable to different types of actuators. Moreover, this detection method can be quickly implemented with cheap off-the-shelf electronics. The detection method will be validated by a speaker system and a motor control system to show its generality, feasibility, and robustness. As mentioned in Section 1.1, only a few studies showed electromagnetic signal injection attacks on actuators, and this chapter will present more scenarios (i.e., the speaker system and the motor control system), which on the one hand, reflect how severe consequences the attacks can cause, and on the other hand, imply the importance and value of the detection method.

In Chapter 5 (based on [36]), it studies how to inject arbitrary messages into differential signaling by exploiting hardware imperfection of receiver circuits. It will systematically and experimentally show differential signaling is not sufficient enough to prevent electromagnetic signal injection attacks, where an attacker can inject malicious signals from a distance, purely using common-mode (CM) injection, i.e., injecting into both wires simultaneously. Also, the success rates of the attacks will be thoroughly analyzed, and critical factors that determine a high success rate will be carefully discussed. Further, this chapter will show, by finely tuning attacking signals, an attacker can manipulate any transmitting bit to whatever bit she wants with a success rate up to 90%, where it is unnecessary to know the value and the timing of the transmitting bit. This chapter will also present a case study on how to inject arbitrary commands into a CAN bus, highlighting electromagnetic signal injection attacks pose threats to many critical applications, such as automotive and aviation sectors.

Let's abstract an overall picture of the whole thesis. As mentioned previously, sensors, actuators, and transmission lines are the critical interfaces through which an embedded system interacts with its surroundings. In this thesis, Chapter 3, Chapter 4, and Chapter 5, which form the main topics, give in-depth insights into the security issues regarding these interfaces, which are under the threat of electromagnetic signal injection attacks. In short, Chapter 3 shows a generalized solution to handle the attacks on the sensors, and Chapter 4 presents a generalized

detection method for the actuators. These two methods help the embedded system resist attacks on its analog interfaces (i.e., sensors and actuators). Further, Chapter 5 investigates the feasibility of malicious message injections into digital interfaces, and it focuses on differential signaling, which has not been studied yet until this work. In addition to studying the injection attacks, it also implies and discusses the necessity of deploying proper defenses against such attacks. In the rest of this thesis, Chapter 2 presents background on embedded systems and fundamentals of electromagnetic signal injections; the purpose of this chapter is to help the readers to understand the security issues that the following chapters will handle. In Chapter 6, related work of defenses against the attacks is summarized; please note that attacks are not repeated in Chapter 6 because Chapter 1 and Chapter 2 already give sufficient introduction and explanation. At last, a conclusion of this thesis, as well as an outlook for future work, is drawn in Chapter 7.

1.3 Published Results

The research that I conducted during my DPhil study have resulted in the following publications (shown in chronological order):

1. Youqian Zhang and Kasper Rasmussen. "Electromagnetic Signal Injection Attacks on Differential Signaling". In: *arXiv preprint arXiv:2208.00343* (2022)

2. Youqian Zhang and Kasper Rasmussen. "Detection of Electromagnetic Signal Injection Attacks on Actuator Systems". In: *25th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2022)*. ACM. 2022 (Best Paper Award)

3. Youqian Zhang and Kasper Rasmussen. "Detection of Electromagnetic Interference Attacks on Sensor Systems". In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2020, pp. 203–216

4. Ilias Giechaskiel, Youqian Zhang, and Kasper Rasmussen. "A Framework for Evaluating Security in the Presence of Signal Injection Attacks". In: *European*

Symposium on Research in Computer Security. Springer. 2019, pp. 512–532 (Best Paper Award)

These papers have been peer-reviewed in computer security conferences, except for the latest work about "differential signaling", which is being submitted to a top-tier conference, and a pre-print has been published in an open-access archive.

1.4 Statement of Originality

As the first author of the publications produced throughout my DPhil study, I have been responsible for the conception, experiments, analysis, and write-up. The co-author has provided criticism and constructive feedback on my research, but all writing in this thesis is mine.

Please note that in the paper [37] that I co-authored with Ilias Giechaskiel and Kasper Rasmussen, I analyzed experimental results and converted them into meaningful figures, as well as got my hands on successful electromagnetic signal injection attacks. Since I am not responsible for the idea, this work is therefore not discussed further.

2
Background

Contents

2.1 An Overview of Embedded Systems 9
2.2 Fundamentals of Electromagnetic Signal Injections . . 11
2.3 Summary 13

This chapter introduces imperfections of circuits that can be exploited to break signal integrity. Starting with an overview of embedded systems, Section 2.1 depicts the structure and functions of essential components. Section 2.2 then details the fundamentals of electromagnetic signal injections.

2.1 An Overview of Embedded Systems

Recall that an embedded system is a microcontroller-based system built to control a function or a range of functions. It can interact with the physical world through sensors and actuators. Such interactions are illustrated in Figure 2.1. The sensor converts physical quantities into electrical signals. Then, it transmits the electrical signals, or sensor measurements, to the microcontroller. After receiving the sensor measurements, the microcontroller starts to process them (e.g., filtering and digitizing) for further tasks. Note that a wire/trace connects the sensor to the

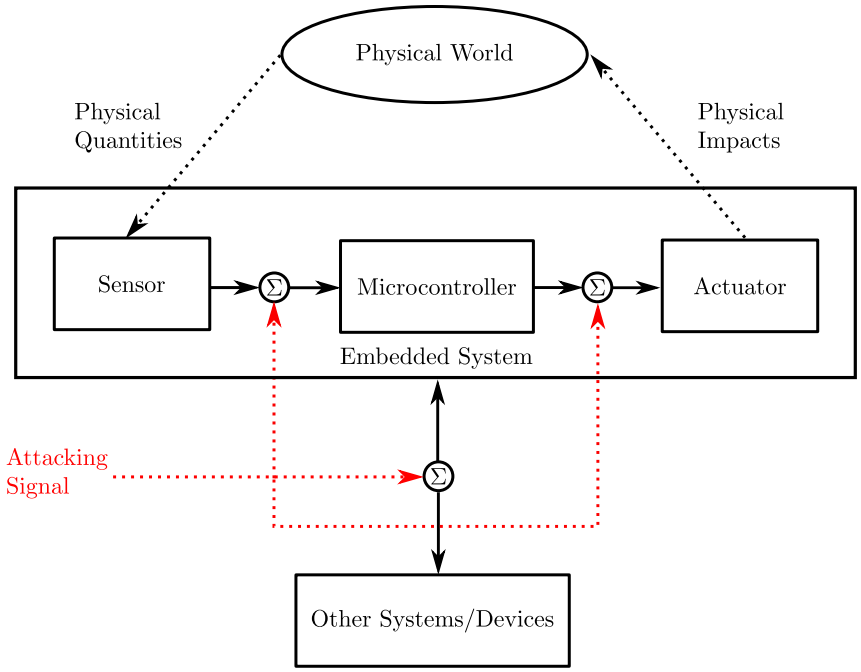


Figure 2.1: An embedded system consists of microcontrollers, sensors, and actuators. The embedded system interact with the world by the sensors and the actuators, and communicate with other systems/devices by transmission lines. These interfaces are susceptible to electromagnetic signal injection attacks.

microcontroller. When the embedded system needs to cause physical impacts on the physical world, the microcontroller sends out actuator control signals, which are also transmitted by a wire. After receiving the signals, the actuator transduces electricity into other forms of energy (e.g., kinetic energy or heat), impacting the external world physically. Moreover, an embedded system can communicate with other systems/devices through transmission lines. Such communications allow a bunch of systems/devices to form a more extensive system or network, completing more complicated tasks (e.g., automobiles, airplanes, robots). There are various transmission media for such inter-system communications, and only electrical cables are considered in this thesis; indeed, electrical cables are widely used in practice.

It is essential to emphasize that the operations of the embedded system rely on the signal integrity of these interfaces. The communications between systems use digital signals where integrity check mechanisms can spot erroneous bits. Still, as mentioned previously, it is possible to bypass the integrity check by manipulating

each bit of a message carefully [32]. Compared with the digital interfaces, it is harder to defend the analog interfaces (i.e., sensors and actuators), as they are energy transducers with no computational capability to guarantee their signal integrity. Specifically, the sensor intrinsically trusts the physical quantities they gauge, and the microcontroller trusts whatever the sensor measurements are. The actuator directly responds to whatever it receives without any authentication.

As mentioned in Section 1, the wires are vulnerable places where an attacker can inject adversary electromagnetic signals, and these three interfaces allow the attacker to interfere with sensor measurements, actuator control signals, and information exchanged between different systems, which are illustrated in Figure 2.1. In the next section, more details about the signal injection will be presented.

2.2 Fundamentals of Electromagnetic Signal Injections

Electromagnetic waves consist of a magnetic field and an electric field. Maxwell's equations explain the relationship between the electric field and the magnetic field as they vary with time: the changing electric field produces a magnetic field; simultaneously, the varying magnetic field also generates an electric field. Electromagnetic fields can affect a metal conductor by inducing voltage changes, and this has been thoroughly studied in the area of "Electromagnetism". Specifically, when an electromagnetic wave flows through the metal conductor, electrons start to oscillate with the same frequency as the electromagnetic wave. Thus, an electric current is generated in the metal conductor. Besides antennas for wireless communications, the metal conductor also exists in circuits in the form of wires (or traces) connecting electronic components. These wires can act like antennas to capture environmental electromagnetic waves [12, 13].

A disturbance that affects the circuits due to electromagnetic waves is called electromagnetic interference (EMI). The EMI can be either unintentional or intentional. Unintentional radiation can be produced from medical devices, vehicles, and power lines. Many studies [38–41] reported that the radiation can cause severe

negative influence to electronic devices. In addition, researchers also found that such radiation from cryptographic devices [42, 43] and screens [44] can lead to information leakage. Regarding intentional EMI, which has significantly higher power than unintentional EMI, researchers [45] summarized that it means generating intentional malicious electromagnetic energy introducing noise or signals into electric or electronic systems, thus disrupting, confusing, or damaging systems for terrorist or criminal purposes. On the one hand, the intentional EMI can disrupt or damage victim systems [45–48]; on the other hand, by fine-tuning the electromagnetic waves, attackers can maliciously manipulate the victim systems rather than simply destroy them [14, 15]. In this thesis, only the fine-tuned intentional EMI is considered.

The injection process is rather complicated, and many factors affect it, but the attack power and the attack frequency are the basic ones that an attacker tunes, as they determine the effectiveness and efficiency of the injection [49]. To cause effective impacts on the circuits, the injected voltage needs to be strong enough. Since the injected voltage is proportional to the attack power [50], the more powerful the attacking signal is, the higher the injected voltage will be, and it is more likely that the attack is effective. In addition, in order to maximize the injected voltage, the attack frequency must be the resonant frequency of the wire, and many studies [51–57] have shown the feasibility of exploiting the resonant frequency of the target wire to inject attacking signals. At other frequencies, it will cost more attack power to achieve the same amount of injected voltage [14]. Note that the resonant frequency of the wire can be approximated by its length [58]; nonetheless, a better way to determine the resonant frequency is to have a copy of the receiving circuits and sweep through a range of frequencies [14]. To put it simply, by properly tuning the attack power and the attack frequency, the attacker can inject arbitrary signals into the wires. Note that since the lengths of the wires in the victim systems usually range from as short as several millimeters to meters long, the frequencies of the attacking signals are typically in the MHz and GHz frequency bands.

When the attacking signal enters the wire, it is superimposed with the signal that is transmitted in the wire, and hence, the signal is maliciously changed. After

the injection, a successful attack depends on how the circuits respond to injected signals. On the one hand, the injected signal can be within the frequency band in which the circuits are designed to operate, namely the operational band (in-band). Since the malicious voltage changes are within the operational band, the circuits respond to them directly. On the other hand, the injected signal can also be out of the operational band (out-of-band). However, in order to affect the circuits in an effective and predictable way, it is essential to cause voltage changes within the operational band. A well-studied method of transferring the out-of-band changes to the operational band is exploiting the nonlinear properties of electronic components in the victim circuits: the attacker first exerts an in-band malicious signal onto an out-of-band radio-frequency (RF) carrier to form the attacking signal; next, after the signal injection, the malicious signal is extracted from the attacking signal due to nonlinearities of electronic components such as amplifiers [14, 15, 19], electro-static discharge (ESD) circuits [26, 27], and analog-to-digital converters (ADCs) [37, 59]. As a result, the in-band malicious signal appears in the operational band. Either way, the attacker can maliciously change signal waveforms, thus manipulating sensor measurements or actuator actions.

Note that Chapter 3 and Chapter 4 focus on detection rather than attacks, and the background above is sufficient to understand these chapters. Readers are referred to previous studies such as [49, 59], which well summarized, modeled, and explained the circuits' responses to the injected signals, if they are interested in more details. Besides, it is essential to emphasize that how the differential signaling circuits respond to the injected signals is a different process, and it will be detailed in Chapter 5.

2.3 Summary

This chapter depicts an overview of embedded systems, which consist of microcontrollers, sensors, actuators, and wires connecting these components. It explains the working mechanism and functions of these components at a high level, explicitly pointing out susceptible interfaces on which the following chapters

will focus. Further, this chapter covers how an adversary signal can be injected into a system wirelessly and how the imperfections of circuits allow the injected signal to cause negative impacts, providing sufficient information to understand the following chapters.

3

Detection Method for Sensor Systems

Contents

3.1 Approach, System-, and Adversary Models 17
3.1.1 Three Types of Sensors 17
3.1.2 Randomized Sensor Output 18
3.1.3 System Model 21
3.1.4 Adversary Model 22
3.2 Attack Detection 23
3.2.1 Detection Rule 23
3.2.2 Interfering with the Bias 24
3.2.3 Security Analysis 25
3.3 Non-constant Physical Quantity 25
3.3.1 Attack Detection for Non-constant Physical Quantities . 27
3.3.2 Security Analysis 27
3.3.3 The Sampling Rate of the ADC 28
3.4 Implementation 28
3.4.1 Setup 29
3.4.2 Microphone System 30
3.4.3 Temperature Sensor System 37
3.4.4 Summary of Implementation 41
3.5 Discussion 41
3.5.1 Guaranteeing the Security with Small n for Constant Physical Quantities 41
3.5.2 Trade-off between Security and Speed 42
3.5.3 The Approach for Non-Powered Passive Sensors 43
3.5.4 Difference between PyCRA and the Approach 44
3.6 Summary 46

Recall that a sensor transforms a physical quantity into an analog signal, and without an authentication scheme, the microcontroller has no choice but to trust the measurement. This allows an attacker to use EMI to remotely, using readily available radio equipment, inject an attacking signal into the sensor system and change the sensor output, regardless of the sensor type. As a result, the attacker can manipulate the microcontroller into believing that the measurement was obtained by the legitimate sensor. In this chapter, a novel method is proposed to detect electromagnetic signal injection attacks on various sensor systems. The method is based on the idea that when a sensor has its power switched off, the output of the sensor should be "quiet". If an attacking signal is maliciously induced into the sensor system during the "quiet" period, the microcontroller can detect it immediately. The contributions of this chapter are as follows:

- It proposes a novel approach to detect attacks by modulating the sensor power and monitoring the sensor output. It abstracts a universal system model from practical circuits and clarifies an attacker's capabilities and limitations. (Section 3.1)

- It details the method and analyzes the detection method's security, proving that it can be bypassed only with a negligible probability (Section 3.2).

- It shows how to maintain some security guarantee even if the measured quantity becomes non-constant in the measuring period (Section 3.3)

- It deploys the detection method on an off-the-shelf microphone system and a temperature sensor system, demonstrating the feasibility and robustness of discovering an attacking signal for both constant and non-constant signals. (Section 3.4)

In the rest of this chapter, a few additional points are discussed in Section 3.5. Finally, the whole chapter is summarised in Section 3.6.

3.1 Approach, System-, and Adversary Models

This section first introduces three classes of sensors on which the proposed method is effective. Then, it explains the core idea and next, presents the system- and adversary models. Note that the details of the defense scheme and a careful security analysis will be presented in Section 3.2.

3.1.1 Three Types of Sensors

Sensors are classified into three main types: active sensors, powered passive sensors, and non-powered passive sensors. An active sensor consists of an emitter and a receiver. The emitter sends out a signal to be reflected by a measured entity, and the receiver gathers information from the reflected signal. Examples of active sensors are ultrasonic sensors and infrared sensors. A powered passive sensor or a non-powered passive sensor has no emitter, and the sensor directly senses the physical phenomenon such as vibration or radiation of the measured entity. A powered passive sensor needs an external excitation signal or a power signal when it works. Examples of such sensors are microphones, light-dependent resistors, and thermistors. A non-powered passive sensor does not need any external power signal. When the non-powered passive sensor is exposed to an entity expected to be measured, the sensor generates an output, which can be a voltage signal or a current signal. Sensors such as piezoelectric sensors, photodiodes, and thermocouples are non-powered passive sensors.

This detection method modifies the way that the powered/non-powered passive sensor works; since the receiver of an active sensor is a powered/non-powered passive sensor, it also works for the active sensor. To simplify the exposition, the powered passive sensor is used as an example to explain the approach in the rest of this chapter. In Section 3.5.3, how to suit this approach to the non-powered passive sensor will be illustrated. Unless otherwise stated, sensors represent powered passive sensors hereafter.

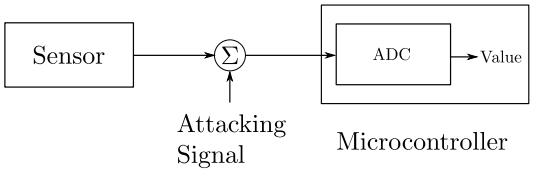


Figure 3.1: A sensor system consists of a sensor, a microcontroller, and a wire connecting them. The microcontroller uses an ADC to digitize the analog sensor outputs.

3.1.2 Randomized Sensor Output

Before introducing the detection method, how an attacker can change a sensor output of a sensor system is briefly recapped. A sensor system consists of two essential modules: a sensor and a microcontroller, as shown in Figure 3.1. The sensor measurements are transmitted to the microcontroller through a wire connecting the output of the sensor and the input of the microcontroller. Since the wire is sensitive to EMI, which can affect the sensor system by inducing voltages on the wire, an attacker can utilize the wire to inject an attacking signal into the sensor output. The malicious sensor output is digitized by an ADC in the microcontroller, and finally, an incorrect digitized sensor output is processed by the microcontroller.

The detection method turns the sensor on and off. Turning on means that the sensor is biased at a high voltage; turning off means that the sensor is biased at 0 V (or other known voltage levels). When the sensor is on, the sensor measures the physical quantity and the sensor output carries the information of the physical quantity. As the sensor is off, the sensor output becomes a constant signal at a specific voltage level. Suppose that the attacker injects an attacking signal to the sensor system when the sensor is off, a disturbance will appear in the flat sensor output. The microcontroller can easily detect such disturbances, and hence the attacking signal is discovered. If the sensor system can randomly turn off the sensor, the attacker has to guess when the sensor is off so that she can avoid sending an attacking signal to the sensor system; otherwise, a mistake of causing an uneven sensor output when the sensor is off will directly unveil the attacker herself to the sensor system.

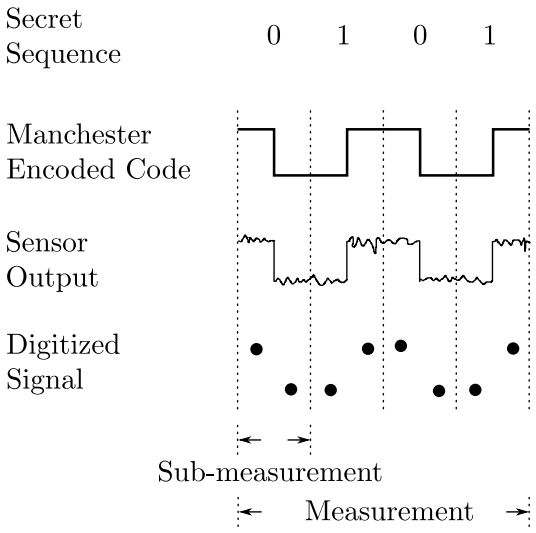


Figure 3.2: An n-bit ($n = 4$) secret sequence of zeros and ones is converted to a Manchester encoded code, which is toggled between a high voltage level and a low voltage level (0 V). The sensor output carries the information of the physical quantity and the noise. After digitization, a digitized signal is obtained.

Such an idea requires that the microcontroller can measure the physical quantity and monitor the attacking signal by turns. Hence, the sensor should be switched between the on and the off states. This approach uses a Manchester encoded code [60] as the bias voltage for the sensor because the Manchester encoded code toggles between a high voltage level and 0 V at the midpoint of each clock cycle (see Figure 3.2). In this approach, the Manchester encoded code is encoded from an n-bit randomized secret sequence of zeros and ones. Because the secret sequence is randomized, the sensor is switched on and off randomly, and hence the sensor output has a randomized on-and-off pattern. In this approach, it is assumed that the physical quantity is constant (see details in Section 3.1.3). Since the physical quantity is constant, as shown in Figure 3.2, the waveform of the sensor output is similar to the Manchester encoded code.

A built-in ADC digitizes the sensor output, and the microcontroller decides whether an attack occurs by checking the digitized sensor output. As shown in Figure 3.2, the secret sequence has n bits, and thus the Manchester encoded code has n clock cycles. Accordingly, the sensor output has n clock cycles. Each clock cycle of the sensor output is defined as a sub-measurement, and all n sub-

measurements form a measurement. Note that each sub-measurement is digitized into two samples by the ADC: one is sampled when the sensor is biased at the high voltage, and the value of the sample is non-zero volt; the other sample is digitized when the sensor is biased at 0 V, and the value of the sample is 0 V. The microcontroller can align the digitized signal with the secret sequence precisely, and hence, given any sample, the microcontroller knows whether it should be zero or non-zero. Hereafter, based on the microcontroller's knowledge of the secret sequence, a sample that should be non-zero is called a "non-zero sample", and a sample that should be zero is called a "zero sample".

Under an attack, either a zero or a non-zero sample in a sub-measurement can be influenced by the attacking signal. If the attacker alters a zero sample, the microcontroller can spot the attack immediately, as the voltage level of the zero sample is not 0 V. Conversely, if the attacker alters a non-zero sample, she will also be detected quickly. This is because the physical quantity should remain unchanged during a measurement, and all non-zero samples should be equal; however, the changed non-zero sample has a different voltage level from the other non-zero samples, and hence the attack is detected. This detection approach are detailed in Section 3.2.

If the sensor system does not detect any attacking signal, the quantification of the physical quantity is the value of a non-zero sample. In practice, noise must be considered. As shown in Figure 3.2, since the sensor output is noisy, the non-zero samples vary slightly in a small range. Thus, the quantification is an average of all non-zero samples. To simplify the exposition, noise is ignored in Section 3.2 and Section 3.3. How to handle noise will be detailed in Section 3.4.

Note that researchers [61] have proposed a defense strategy named PyCRA, which detects sensor spoofing attacks by turning off the emitter in an active sensor. Details of the working principle of PyCRA and a comparison between this approach and PyCRA are presented in Section 3.5.4.

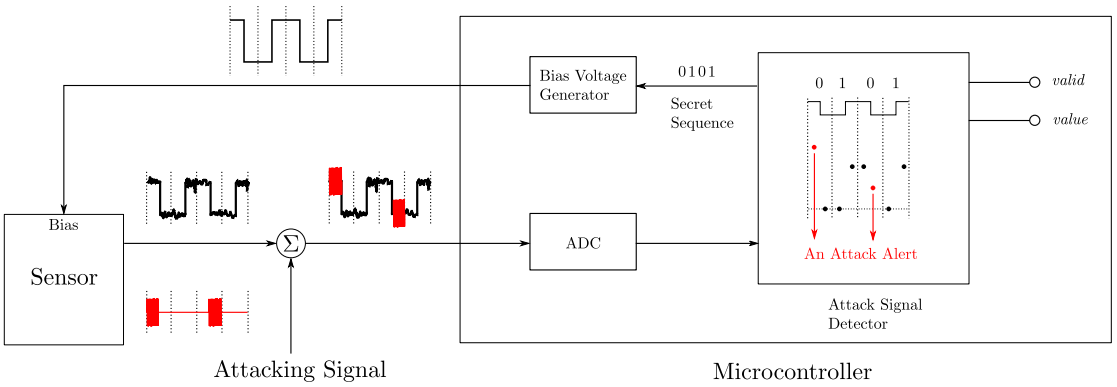


Figure 3.3: A sensor system that is equipped with the detection method consists of a sensor and a microcontroller. The bias voltage of the sensor is controlled by the microcontroller. In the attack signal detector, unequal non-zero samples imply an attack. Also, a changed zero sample indicates an attack.

3.1.3 System Model

Figure 3.3 presents a sensor system that is equipped with the detection method. The sensor is driven by a bias voltage that is controlled by the microcontroller. An output of the sensor is used to send a measurement to the microcontroller, which checks the existence of attacking signals and recovers the physical quantity from the measurement.

The microcontroller has three blocks, including a bias voltage generator, an ADC, and an attack signal detector. The bias voltage generator encodes an n-bit secret sequence into a Manchester encoded code, which is the bias voltage for the sensor. The ADC digitizes the sensor output and transmits the digitized data to the attack signal detector to check whether an attacking signal exists. The attack signal detector has two outputs: *value* represents a measurement of the physical quantity; *valid* indicates whether *value* is ready to be read. If no attacking signal is detected, the measurement is assigned to *value*, and then *valid* is set to true. Hence, the sensor system knows that *value* is valid to be further processed. However, if an attacking signal is detected in a measurement, *valid* is set to false throughout that measurement, which means that *value* is invalid to be read. Also, the microcontroller will be alerted that the sensor system is under attack.

In the system model, it is assumed that the physical quantity remains unchanged in a measurement. Even though the physical quantity varies, if the duration of a measurement is short enough, the physical quantity can also be regarded as constant. An example of a constant physical quantity is room temperature. The temperature changes slowly over a long period; however, in a short time, such as 0.01 s, the temperature is unchanged.

For each measurement, the microcontroller generates an n-bit secret sequence, and accordingly, the Manchester encoded code has n clock cycles. Two samples are digitized from each clock cycle or sub-measurement, and hence, the sampling rate of the ADC is two times larger than the clock rate of the Manchester encoded code. In practice, the sampling rate of the ADC has an upper limit, and thus the clock rate of the Manchester encoded code also has a maximal value, which is half of the fastest sampling rate. The shortest duration of n clock cycles is determined by the fastest sampling rate of the ADC. To apply this detection method, it is essential to ensure that the physical quantity is unchanged within the n clock cycles.

3.1.4 Adversary Model

The objective of the attacker is manipulating the waveform of the sensor output without being detected by the sensor system. It is supposed that the attacker cannot access the sensor system physically. Also, it is assumed that the attacker has no information about the n-bit secret sequence. Given any sub-measurement, it is assumed that the attacker knows voltage levels, but she does not know whether the voltage level transitions from the high voltage to 0 V or from 0 V to the high voltage in the midpoint of the sub-measurement (see Figure 3.2). Thus, the attacker has to guess the direction of the voltage level transition in each sub-measurement. Moreover, the attacker can deliberately inject a crafted signal into the sensor system, and thus the attacker can change the waveform of the sensor output as she wishes. Also, the attacker knows when the sensor module starts and stops transmitting the measurement, and she can align the crafted signal with the sensor output precisely.

Note that if such a strong adversary cannot avoid being detected by the approach, it is much more difficult for any other attackers who are weaker than this strong adversary to avoid the detection.

3.2 Attack Detection

After receiving the digitized sensor output, the attack signal detector aligns it with the corresponding secret sequence. As shown in Figure 3.2, each digit in the secret sequence corresponds to two samples in the digitized sensor output. A digit 1 means that the corresponding two samples are zero and non-zero in a consecutive order; a digit 0 indicates a non-zero sample and a zero sample in a consecutive order. Thus, the microcontroller knows the order of all samples.

3.2.1 Detection Rule

When no attacking signal exists, the digitized sensor output satisfies two requirements:

1. All non-zero samples are equal.

2. All zero samples are zero.

Once an attack occurs, either sample in a sub-measurement can be altered. The attack signal detector first checks non-zero samples. As shown in Figure 3.4, if the attacker only changes several non-zero samples in the measurement, the signal formed by all non-zero samples becomes non-constant. Unequal non-zero samples imply that an attack occurs. To bypass the detection, the attacker is forced to increase or decrease all non-zero samples to the same voltage level. It is possible for the attacker to make a mistake and change a zero sample. Once a zero sample is altered by the attacker accidentally, the attack will be detected.

After checking the digitized sensor output, if an attack is discovered, the measurement is discarded. In contrast, if no attacking signal is detected, a quantification of the physical quantity can be obtained. As it is discussed in Section 3.1.2, the quantification is the value of non-zero samples; however, in

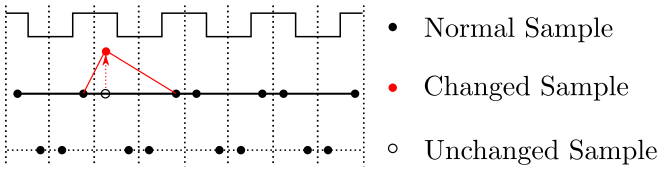


Figure 3.4: A sensor output of a constant physical quantity. An attacker shifts one non-zero sample, and the signal formed by all non-zero samples becomes non-constant.

practice, considering the existence of noise, it can be calculated by averaging all non-zero samples.

A smart attacker must guess whether a sample is zero or non-zero. To avoid being detected, the attacker must not affect any zero sample, and she must alter all non-zero samples to keep them the same. In Figure 3.3, it presents an example of detecting an attacking signal in the sensor system. The attacker aims to alter the first and the third sub-measurements of the sensor output. In the first sub-measurement, the attacker makes a correct guess, and a high-frequency signal is added to the non-zero half cycle. However, in the third sub-measurement, the attacker makes a wrong guess and adds the high-frequency signal to the zero half cycle. After digitization, two samples are shifted up: the non-zero sample in the first sub-measurement and the zero sample in the third sub-measurement. Compared with other non-zero samples, the non-zero sample in the first sub-measurement has a different value, and the attack signal detector can discover the attack immediately. In the third sub-measurement, the second sample should have been zero; however, it is shifted to a non-zero value, and the microcontroller can notice the change. As a result, the attacking signal can be detected.

3.2.2 Interfering with the Bias

As described above, the detection method is used to spot attacking signals that are injected into the sensor system through the wire connecting the sensor output and the ADC. However, in practice, the wire controlling the bias of the sensor may also be an unintentional antenna. An attacking signal that is injected into this wire may alter the voltage levels of several specific periods of the Manchester encoded

code. Further, the corresponding periods of the sensor output are impacted. For example, some periods that should have been at a certain voltage level are at other voltage levels; some periods that should have been 0 V are not zero. After digitizing the sensor output, the microcontroller may spot that non-zero samples are unequal and some zero samples are lifted. Therefore, this method can also detect attacks affecting the bias. For simplicity, the wire connecting the sensor and the ADC is regarded as the injection point of an attacking signal hereafter.

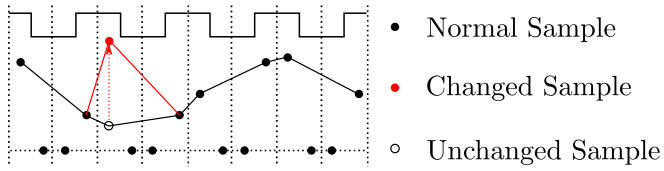
3.2.3 Security Analysis

Only when the attacker changes all non-zero samples without influencing any zero sample, can she avoid being detected by the sensor system. This section proves that the attacker can bypass the detection method with a negligible probability.

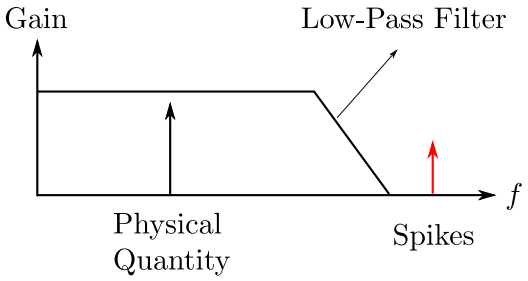
For a constant physical quantity, all non-zero samples in a measurement have the same voltage level. To avoid being detected by the sensor system, the attacker must change all non-zero samples to the same voltage level. Thus, the attacker must correctly guess the order of the zero and the non-zero samples in every sub-measurement. There are two combinations of the order of samples in a sub-measurement, and the probability of correctly guessing the order is $\frac{1}{2}$. Considering a measurement with n sub-measurements, the probability of correctly guessing the orders is $\frac{1}{2^n}$. In other words, the probability of bypassing the detection method in one measurement is $\frac{1}{2^n}$, which is negligible. The larger the n is, the more difficult it is for the attacker to achieve the attack.

3.3 Non-constant Physical Quantity

In the previous section, it describes the approach regarding constant physical quantities. However, there are physical quantities such as sounds that oscillate rapidly; even though the sampling rate of an ADC reaches the maximum, the digitized non-zero samples may have different values in a measurement. Such a physical quantity is called a non-constant physical quantity, and an example is shown in Figure 3.5a.



(a) A sensor output of a non-constant physical quantity.



(b) A digital low-pass filter removes the spikes.

Figure 3.5: The attacker alters an non-zero sample in the digitized sensor output.

If the attacker affects either a non-zero sample or a zero sample in a constant physical quantity, this approach can detect the attack (see details in Section 3.2). For a non-constant physical quantity, unequal non-zero samples do not indicate an attack anymore. This means that if the attacker plans to alter one sample only, she can bypass the detection with a probability of $\frac{1}{2}$. For example, as shown in Figure 3.5a, the attacker wants to affect the third clock cycle: if she changes the non-zero sample, she succeeds; otherwise, changing the zero sample still leads to an alert of the attack. Compared with the detection method for a constant physical quantity, the one for the non-constant source gives a weak security guarantee. In order to achieve a strong security guarantee, the sampling rate of the ADC must be large enough so that the physical quantity can be regarded as constant, and thus the approach for a constant source applies.

However, in practice, a sensor system may have to handle non-constant scenarios subject to multiple limitations (e.g., sampling rates of ADCs). Then, it is necessary to revise the approach for non-constant physical quantities to detect attacks affecting either non-zero or zero samples. In this section, it describes the revised method. Also, it shows that the negative impacts that are caused by attacking signals can be mitigated. It analyzes the security of this detection method. Finally, an additional

requirement for the ADC is discussed.

3.3.1 Attack Detection for Non-constant Physical Quantities

An attacker can change any numbers of non-zero samples. Without loss of generality, it assumes that the attacker plans to change k ($1 \leq k \leq n$) out of n samples. She can achieve the modification without being detected with a probability of $\frac{1}{2^k}$ (see details in Section 3.3.2). When a few samples are changed, as shown in Figure 3.5a, the modified sample leads to a spike in the measured signal. Without knowing any information about the measured signal, nothing can be done to detect the change. However, if concrete characteristics that can describe the behavior of the non-constant signal are known, modified samples can be recognized as outliers. As depicted in Figure 3.5b, if the bandwidth of the measured signal is known, it can recognize the sample that causes a spike beyond the band as an outlier. Moreover, if a model of the measured signal is attainable, it can recognize the sample that fails to fit the model as an outlier. Despite that a few modified samples form spikes in the measured signal, the major information of the physical quantity may still be retained. For example, regarding an audio signal, a spike in the measured signal sounds like a chirp; however, a listener can still understand the information that is conveyed in the audio signal. A digital low-pass filter can be used to filter out the spike so that the negative impacts can be mitigated.

If the attacker changes many samples, the modified samples dominate, and she may bypass the detection of outliers. However, the probability of avoiding affecting zero samples is $\frac{1}{2^k}$, which exponentially decreases with the number of samples that the attacker wishes to change. Therefore, changing more samples increases the difficulty of bypassing the detection.

3.3.2 Security Analysis

It has been assumed that the attacker plans to change k ($1 \leq k \leq n$) out of n non-zero samples. When $k = n$, the probability of bypassing the detection method is

the same as the one for a constant physical quantity. When $1 \leq k < n$, the attacker needs to guess the orders of samples in k sub-measurements. The probability of bypassing the detection method is $\frac{1}{2^k}$, which is negligible. If k is small, the attacker can easily achieve an attack, but the impacts of the modified samples are small; while k is large, it is difficult for the attacker to bypass the detection method.

3.3.3 The Sampling Rate of the ADC

To ensure that the measurement contains complete information of the physical quantity, according to the Nyquist-Shannon sampling theorem, the clock rate of the Manchester encoded code should be at least twice larger than the bandwidth of the non-constant physical quantity. Since the sampling rate of the ADC is twice larger than the clock rate of Manchester encoded code, the sampling rate is at least four times larger than the bandwidth of the physical quantity.

3.4 Implementation

In this section, the detection method is implemented on two sensor systems: a microphone system (in Section 3.4.2) and a temperature sensor system (in Section 3.4.3). A microphone can convert sound into an electrical signal. At present, microphones can be found in many different devices, such as smartphones, headphones, and laptops. In a microphone system, a wire is used to connect a microphone module and a microcontroller. The wire allows an attacker to radiate electromagnetic waves to inject an attacking signal into the microphone system. For example, an attacker can inject voice commands into a smartphone, and the voice assistant system can be asked to execute malicious tasks. Note that since human beings cannot hear any electromagnetic waves, the user will not notice the attack at all. Similarly, in a temperature sensor system, the wire transmitting electrical signals from a temperature sensor to a microcontroller can also become an injection point, allowing an attacker to manipulate temperature readings.

In this section, first, the experiment setup is introduced. Next, in each sensor system, it first shows how an attacker can remotely modify sensor readings by EMI,

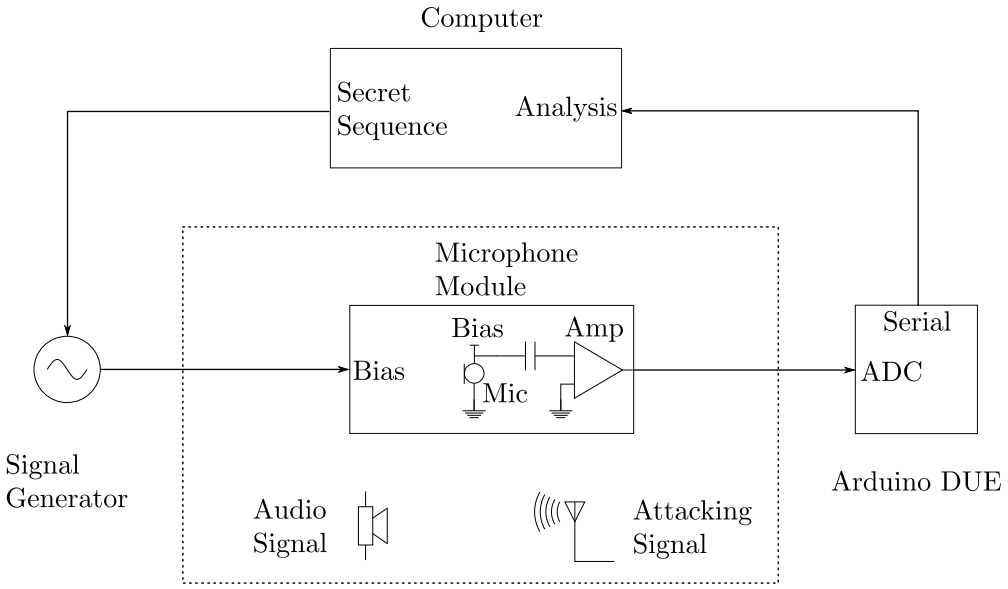


Figure 3.6: A testbed is built to test a microphone system. A signal generator, which is controlled by a computer, provides the microphone module with a bias voltage. An Arduino DUE is used to collect the signal from the microphone module. The computer is used to analyze the signal.

and then it presents the effectiveness and robustness of the detection method. At last, a brief summary of the implementation is drawn.

3.4.1 Setup

A testbed that can be quickly configured into a microphone system or a temperature control system is built. In Figure 3.6, a setup of the microphone system is presented. The microphone system consists of a computer, a signal generator, an off-the-shelf microphone module, and an Arduino DUE. The computer controls a RIGOL DG4062 signal generator to generate a bias voltage for the microphone. The microphone converts the sound into a voltage signal, which is further amplified by the amplifier. The output of the amplifier is biased at 1.65 V. Then, the output of the microphone module is digitized by a built-in ADC in the Arduino DUE at a sampling rate of 666.8 kHz. Next, the Arduino DUE sends the digitized data to the computer through a serial port. Finally, the computer is used to analyze the digitized signal.

Note that the sampling rate that is chosen is higher than the minimum theoretical sampling rate required. According to Section 3.3.3, the sampling frequency should

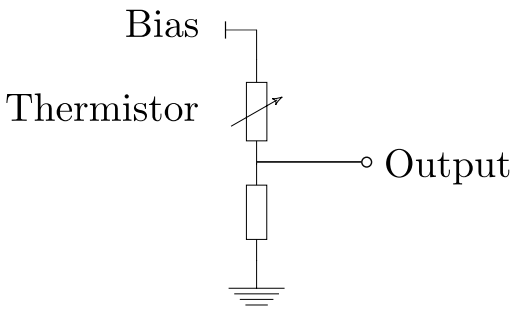


Figure 3.7: A thermistor circuit is a voltage divider. When the temperature increases, the output voltage of the circuit increases accordingly.

be at least four times larger than the bandwidth of the physical quantity. Since the microphone in the experiment can measure up to 20 kHz, the sampling frequency is 80 kHz in theory. However, in practice, it needs to consider samples that are digitized from signal edges, and hence the sampling rate is higher than the theoretical one. Details are discussed in Section 3.4.2.

In the temperature sensor system, a thermistor is selected to sense room temperature. In Figure 3.7, a diagram of a thermistor circuit is presented. The thermistor circuit is a voltage divider, which is formed by connecting an NTC thermistor and a resistor in series. The output voltage of the thermistor circuit increases with increasing the temperature. The thermistor circuit is tested using the setup shown in Figure 3.6, i.e., the microphone module is replaced with the thermistor circuit. This setup is placed in a laboratory with a constant temperature at around 25.0 °C. Since the room temperature can be regarded as a constant physical quantity, digitized samples that should be non-zero are supposed to be approximately equal. The sampling rate is set to 284 Hz, which is much lower than the one in the microphone system.

3.4.2 Microphone System

In the experiment, there are two signal sources: one is a legitimate sound from a speaker of a Motorola XT1541 Moto G3 smartphone, and the other is an attacking signal from the attacker. The attacker uses an R&S SMC 100A signal generator

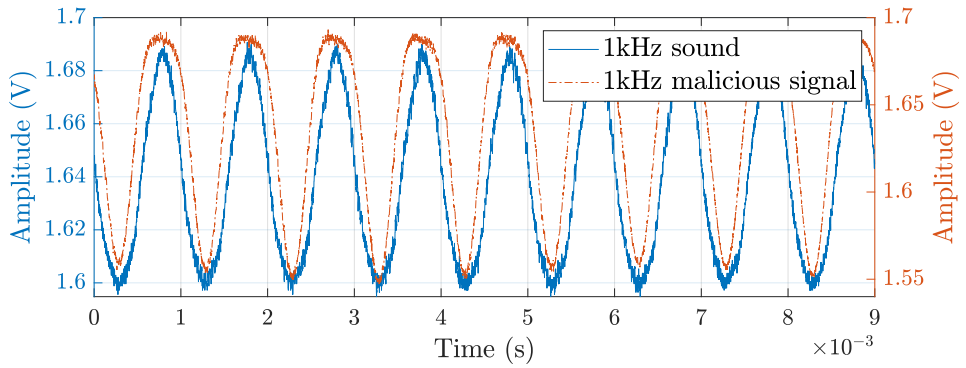


Figure 3.8: One 1 kHz signal is the sound, and the other 1 kHz signal is from the attacker, who injects it by EMI. The similarity of these two signals is above 0.93.

to amplitude-modulates a malicious signal on a 144 MHz carrier signal to form the attacking signal. Then, the attacking signal is radiated through a 144 MHz omnidirectional vertical antenna. The reason why 144 MHz is chosen as the carrier frequency of the attacking signal is that, by experiment, the 144 MHz signal can be received by the unintentional antenna in the microphone module effectively. Both the antenna and the speaker are placed 10 cm away from the microphone module.

Without the Detection Method

Without the detection method, the microphone system cannot determine whether the signal is legitimate or malicious. In the following parts, they will show that the attacker can remotely inject a malicious signal that is similar to the audio signal into the microphone system.

The signal generator is configured to output a constant 300 mV signal, and thus the microphone is biased at 300 mV. A 1 kHz audio signal is played through the speaker of the mobile phone at the maximal volume. Next, the speaker turns off, and an attacking signal, which is generated by modulating the 1 kHz malicious signal on a 144 MHz carrier signal, is emitted through the antenna at −5 dBm. The attacking signal is demodulated by the nonlinear electronic components (e.g., amplifiers and ADCs) in the microphone system, and a 1 kHz digitized malicious signal is obtained.

In Figure 3.8, two 1 kHz signals that are reconstructed by the computer are presented: one is the signal from the speaker; the other one is induced by the

attacker. It can be observed that, without this detection method, it is difficult to tell whether a received signal is from the speaker or the attacker: both the sound and the malicious signal are 1 kHz, and they have similar amplitudes. It is known that Pearson's correlation coefficient (PCC) can be used to measure the linear correlation of two signal[62, 63], and PCC is a suitable metric to show the similarity of two signals in the experiments. The PCC of the 1 kHz audio signal and the 1 kHz malicious signal is above 0.93, which means that these two signals have high similarity. Above all, the attacker can control the output of the microphone system and deceive the microcontroller.

Applying the Detection Method

From the experimental results above, the microphone system may regard the malicious signal as the legitimate audio signal. In this part, it illustrates how to deploy the detection method to the microphone system to detect the attacking signal.

When the detection method is applied to the microphone system, the computer repeatedly transmits a secret sequence of [1100] to the signal generator, and the signal generator encodes the secret sequence into a Manchester encoded code with a clock rate of 40 kHz. The Manchester encoded code toggles between 0 mV and 300 mV. Note that the bias voltage is for the microphone, which is denoted as "Mic" in Figure 3.6, instead of the amplifier[1]. In Figure 3.9, without any audio signal or attacking signal, it presents the output of the microphone module that is captured by a RIGOL DS2302A Digital Oscilloscope, which has a sampling frequency of 2 GHz.

When the computer receives the digitized signal from the Arduino DUE, three practical challenges in the microphone system need to be considered before checking the existence of an attack. The first challenge is synchronizing the digitized signal with the secret sequence. Each digit in the secret sequence corresponds to one sub-measurement, and the value of the digit decides the direction of the voltage level transition at the midpoint of the sub-measurement. Only if the digitized signal

[1]If the Manchester encoded code is used to bias the amplifier, when the amplifier is off, an attacking signal that is injected before the amplifier does not affect the output of the amplifier. This means that attacks that affect zero samples cannot be detected.

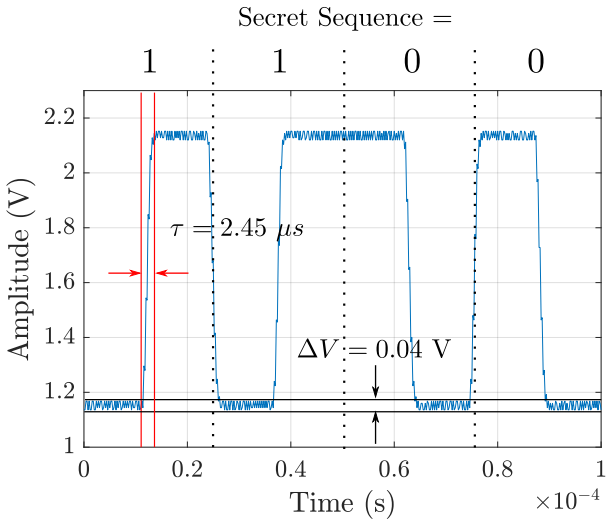
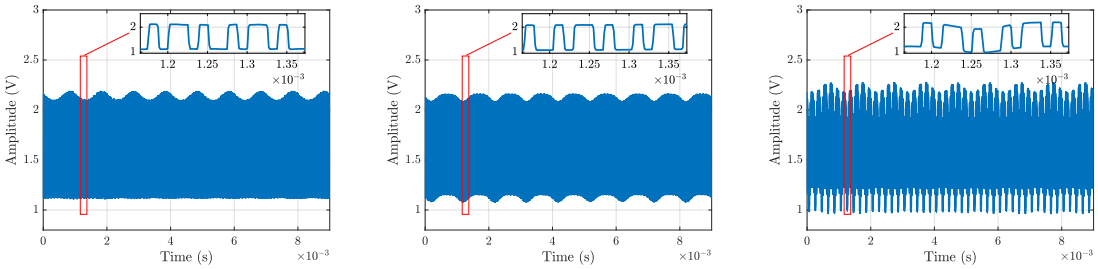


Figure 3.9: Measure the bound of zero samples and the time of the signal edges by an oscilloscope with a sampling frequency of 2 GHz.

is aligned with the secret sequence precisely will the computer knows whether a specific sample is zero or non-zero. In practice, the signal generator is configured in a way that there is always a voltage level transition from high to low at the beginning of the first sub-measurement so that it is easy to identify the start point of the digitized signal. Further, it is straightforward to align the digitized signal with the secret sequence.

Another practical challenge is how to handle samples from the rising or the falling edges of the microphone module's output. The samples from the edge can lead to a false positive alert of attack or an inaccurate measurement of the physical quantity. As shown in Figure 3.9, the time of the signal edge is $\tau = 2.45\,\mu s$. The sampling period of the ADC is $\frac{1}{f_s} = \frac{1}{666.8\,\text{kHz}} \approx 1.50\,\mu s$, and hence at most two samples emerge from the signal edge. Also, given the sampling rate and the clock rate, it can be found that there are 16 samples in each sub-measurement. Thus, to eliminate the negative impacts of the edge samples, the first and the last samples are removed in each half cycle.

The third practical challenge is to determine the voltage level of zero samples. Because the output of the microphone module is centered at 1.65 V, the zero samples are shifted to a non-zero level. As shown in Figure 3.9, the mean value of the zero samples is 1.15 V. However, it can be observed that the zero samples fluctuate



(a) Case 1 **(b)** Case 2 **(c)** Case 3

Figure 3.10: When detection method is applied, (a) the speaker plays a 1 kHz tone; (b) the attacker transmits an attacking signal, which is generated by modulating 1 kHz signal on a 144 MHz carrier signal at the power of −5 dBm; (c) the attacker transmits an attacking that is generated by modulating a 5 kHz signal on a 144 MHz carrier signal at a transmission power of 0 dBm, and the speaker plays 1 kHz tone at the same time.

around 1.15 V, and the range of the fluctuation is $\Delta V = 0.04\,\text{V}$. Note that ΔV is also the noise tolerance of zero samples. When there is no attacking signal, the zero samples are within a range of $[1.15\,\text{V} - \frac{1}{2}\Delta V, 1.15\,\text{V} + \frac{1}{2}\Delta V] = [1.13, 1.17]$ V. If a zero sample is outside $[1.13, 1.17]$ V, the microphone system will be alerted with an attack.

After obtaining a measurement from the microphone module, the computer synchronizes the corresponding secret sequence with the measurement, and removes samples from the edges. According to the bounds of zero samples, which is $[1.13, 1.17]$ V, the computer can determine whether an attack occurs in the measurement. To evaluate the performance of the detection method, the following three cases are considered:

Case 1

A 1 kHz audio signal is played from the speaker at its maximal volume, and there is no attacking signal. In Figure 3.10a, the amplitude envelope that is formed by non-zero samples of the digitized sequence represents the 1 kHz component. Since no attacking signal exists, this case is a reference for the following two cases.

Table 3.1: Detection results of Case 2 and 3 in a microphone system.

Case No.	Sound	Attacking Signal (modulating signal, carrier)	True-positive Rate
2	-	(1 kHz, 144 MHz)	100%
3	1 kHz	(5 kHz, 144 MHz)	100%

Case 2

Turn off the speaker, and the attacker radiates an attacking signal at -5 dBm. To inject a 1 kHz signal into the microphone system, the attacking signal is generated by modulating the 1 kHz signal on a 144 MHz carrier. As Figure 3.10b shows, it can be noticed that both zero and non-zero samples carry the information of the 1 kHz signal.

Case 3

Turn on the speaker, and the attacker radiates an attacking signal at the same time. The frequency of the audio signal is still 1 kHz, and the volume is unchanged. To insert a 5 kHz signal into the system, the attacker modulates the 5 kHz signal on a 144 MHz carrier, and the transmission power of the attacking signal is 0 dBm. As it is shown in Figure 3.10c, the 5 kHz signal dominates in both zero and non-zero samples.

In each case, 100 measurements are recorded. Because the physical quantity is non-constant in a measurement, the detection criteria of a non-constant physical quantity are used to check whether an attacking signal exists in each measurement. Accordingly, in Case 2 and Case 3, the true-positive rate of detecting the attacking signal is calculated. The detection results are presented in Table 3.1. In Case 2 and Case 3, the computer finds that some zero samples are outside the bounds, and thus the attacking signal can be detected. The true-positive rates of detecting the attack are 100% in both Case 2 and Case 3. The results mean that the attacking signals exist in every measurement in these two cases.

The experiments also show that, when there is no attacking signal (Case 1), all zero samples are within the bounds, and the detection method does not give any false positive alarm of an attack. Once the attacker accidentally increases or

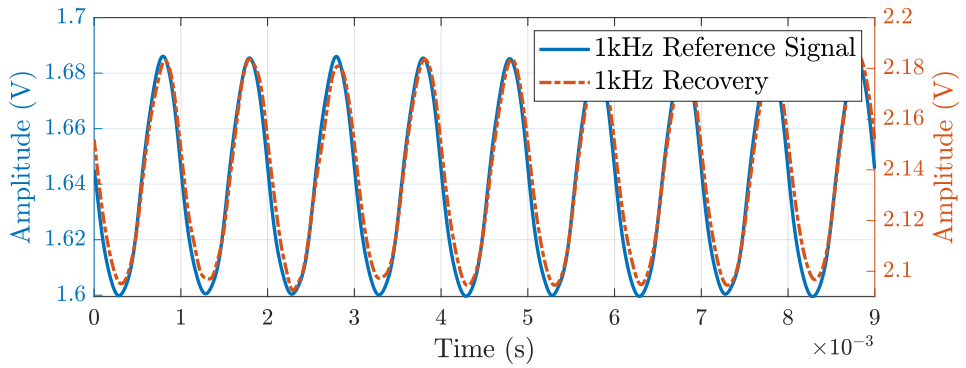


Figure 3.11: Remove zero samples and edge samples to reconstruct the 1 kHz audio signal. As a comparison, the 1 kHz reference signal is presented.

decreases the value of the zero samples to a value that is outside the bounds (e.g., Case 2 and 3), the attack is detected immediately.

Note that, in Case 2 and 3, the attacker initiates "dumb" attacks, which mean that the attacker does not guess when the sensor is on or off. In other words, the dumb attacking signal affects every sample in the measurement. This is the reason why the true-positive rate is 100% for these two cases. In practice, it is difficult to conduct "smart" attacks that allow the attacker to do the guessing and align the attacking signal with the sensor output. In the experiment of a temperature sensor system in Section 3.4.3, smart attacks are simulated from real sensor data.

Signal Reconstruction

When no attack is detected, the final step is to recover the physical quantity. Because measurements in Case 2 and 3 are detected with attacking signals, it cannot recover the physical quantity from these two cases. In Case 1, no attacking signal is detected, and it can recover the 1 kHz signal by excluding zero samples and edge samples in the measurement. Then, a digital second-order Butterworth low-pass filter with a cut-off frequency of 5 kHz is used to get rid of high-frequency components in the digitized signal. The recovered 1 kHz signal is shown in Figure 3.11. As a comparison, a 1 kHz audio signal with the same ADC is digitized as a reference signal, and it is filtered by the same low-pass filter. The reference signal is depicted in Figure 3.11.

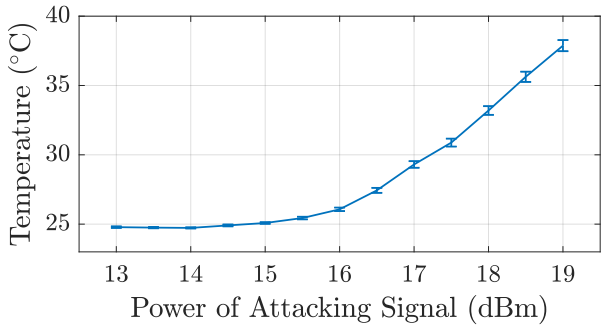


Figure 3.12: The power of attacking signal is increased from 13 dBm to 19 dBm with a step of 0.5 dBm. Under the attack, the temperature is changed from 24.9 °C to 37.9 °C.

The quality of the recovered signal is analyzed in two aspects: similarity and Signal-to-Noise Ratio (SNR). As discussed in Section 3.4.2, PCC can be used to measure the similarity between two signals. The PCCs between 100 recovered signals and the reference audio signal are calculated. The averaged PCC in Case 1 is above 0.99, which implies that the recovered signal is similar to the audio signal in the time domain. The averaged SNR of all 100 recovered signals in Case 1 is 30.6 dB ± 0.1 dB at a 99% confidence level; the SNR of the reference signal is 29.9 dB. It can be concluded that the recovered signal has an equivalent quality as the reference signal.

3.4.3 Temperature Sensor System

In the experiment on the temperature sensor system, the attacking signal has a frequency of 144 MHz, and it is radiated from a 144 MHz omnidirectional antenna. The antenna is placed 1 cm away from the thermistor circuit. Note that the distance between the antenna and the thermistor circuit is small because the remote injection can be realized with low power of the R&S SMC 100A signal generator. In the following sections, it first demonstrates how an attacking signal affects a sensor reading. Then, it shows that the detection method can detect the attacking signal.

Without Detection Method

The thermistor circuit is biased at 1 V. When no attacking signal is radiated, the temperature sensor system outputs 24.9 °C ± 0.1 °C at a 99% confidence level.

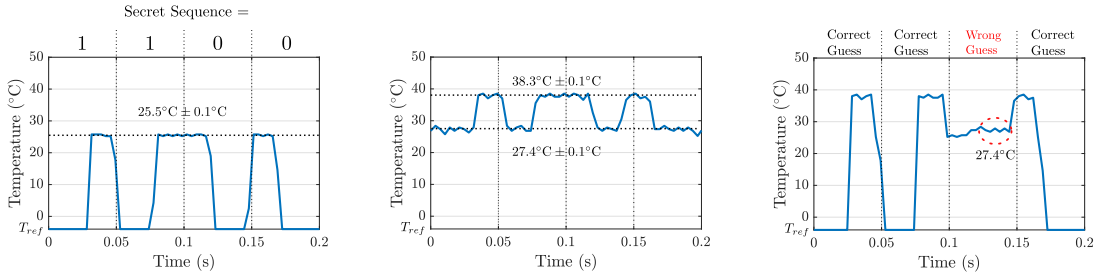
Next, the attacker radiates an attacking signal, and the power of the attacking signal is increased from 13 dBm to 19 dBm with a step of 0.5 dBm. For each power level, 100 temperature measurements are recorded. The 99% confidence interval around the mean of the 100 measurements is calculated, and the results are presented in Figure 3.12. Below 14 dBm, the attacking signal has no significant effect on the temperature measurement. When the power of the attacking signal is increased above 14 dBm, the temperature measurements increase. The 19 dBm attacking signal results in a temperature measurement of $37.9\,°\text{C} \pm 0.4\,°\text{C}$, which is approximately 13 °C higher than the true room temperature of $24.9\,°\text{C} \pm 0.1\,°\text{C}$. The curve in Figure 3.12 shows that the attacker can change the temperature reading of the sensor to any values as she wishes. Without any detection method, the temperature sensor system cannot detect the existence of the attacking signal.

Applying Detection Method

The secret sequence that is used is also [1100], and the clock rate of the Manchester code is set to 20 Hz. An oscilloscope is used to measure the time of signal edge, and the width of signal edge is around 2 ms. Regarding that the sampling period is $\frac{1}{284\,\text{Hz}} = 3.5\,\text{ms}$, at most one sample is digitized from signal edges. In order to eliminate the negative influence caused by samples from signal edges, the first and the last sample in each half clock cycle are abandoned.

The oscilloscope is used to measure the bound of non-zero samples, which is 0.03 V; the bound of zero samples has the same value. When no attacking signal is radiated, fluctuations of non-zero samples are within 0.03 V; note that zero samples swing between 0 V and $\frac{1}{2} \times 0.03\,\text{V} = 0.015\,\text{V}$, as the ADC in the microcontroller can only read positive voltages. Because the room temperature is a constant physical quantity, the requirements are as follows:

- The standard deviation of all non-zero samples is smaller than or equal to $\frac{1}{2} \times 0.03\,\text{V} = 0.015\,\text{V}$.
- All zero samples are within $[0, 0.015]$ V.



(a) Case 1 **(b)** Case 2 **(c)** Case 3

Figure 3.13: The detection method is deployed to the temperature sensor system, and the outputs of the thermistor circuit are presented. In (a), no attacking signal exists, and the non-zero samples are approximately equal, which indicates a temperature of 25.5 °C. In (b), a dumb attacking signal is radiated, and the non-zero samples indicate a room temperature of 38.3 °C, and the zero samples corresponds to a temperature of 27.4 °C. In (c), a smart attack is simulated, and a wrong guess is made in the third clock cycle.

In the following parts, a reference case (Case 1) is presented, in which no attacking signal exists. A dumb attack (Case 2) is conducted on the temperature sensor system, and then a smart attack (Case 3) is simulated from data that are collected from Case 1 and 2. In the following parts, the thermistor circuit's voltage outputs are converted into temperature. Note that when the bias voltage is 0 V, the output is also 0 V. Since 0 V corresponds to a temperature that is beyond the measurement range of the thermistor circuit, this temperature is denoted as T_{ref} (see Figure 3.13).

Case 1

No attacking signal is radiated from the antenna, and the microcontroller records the output of the thermistor circuit. In Figure 3.13a, a measurement is presented. The measured temperature is 25.5 °C ± 0.1 °C.

Case 2

In order to change the sensor reading to a significant high temperature, the antenna radiates an attacking signal with a power of 19 dBm. The microcontroller records the output of the thermistor circuit. A measurement is shown in Figure 3.13b. Note that such an attack is a dumb attack, as the attacker radiates the attacking signal continuously. The mean of the non-zero samples corresponds to a temperature of

38.3 °C ± 0.1 °C, which is around 13 °C higher than the true room temperature. The zero samples are lifted to 27.4 °C ± 0.1 °C, which indicates an attack.

Case 3

(A simulation of a smart attack) The attacker has a fair coin that has a probability of 50% showing a head and 50% showing a tail every time it is tossed. The attacker selects a measurement from Case 1, and each measurement contains 4 clock cycles or 8 half clock cycles (see Figure 3.13a). For each clock cycle, the attacker tosses the coin to decide whether to send an attacking signal. A head means that the attacker radiates an attacking signal in the first half cycle and remains silent in the second half cycle. Accordingly, the first half cycle is replaced by a half cycle that corresponds to 38.3 °C ± 0.1 °C from Case 2. Conversely, a tail means that the attacker remains silent in the first half cycle and radiates an attacking signal in the second half cycle. Accordingly, the second half cycle is replaced by a half cycle that is 27.4 °C ± 0.1 °C from Case 2. After tossing the coin for all four clock cycles, there is a new measurement that is affected by a smart attack (see Figure 3.13c).

As shown in Figure 3.13c, except for the third clock cycle, the attacker's guesses in the other three clock cycles are correct. The attacker accidentally radiates the attacking signal during the second half cycle of the third cycle: the temperature of that half cycle is enhanced from T_{ref} to 27.4 °C. After digitization, non-zero samples form a non-constant signal, and thus an attack can be detected. Also, since samples that should be T_{ref} in the third clock cycle are lifted, the attack is alarmed.

In each case, 100 measurements are recorded, and a summary is presented in Table 3.2. In Case 2, the true-positive rate is 100%, which implies that an attacking signal is detected in each measurement. Also, the simulation of smart attacks is repeated 100 times, and the true-positive rate is 93%. In theory, since the number of digits of the secret is four, the attacker has a probability of $\frac{1}{2^4}$ guessing the secret of each measurement correctly. Among 100 measurements, the expectation of correct guesses is $\frac{100}{2^4}$. Therefore, the theoretical true-positive rate is $1 - \frac{100}{2^4}/100 = 93.75\%$. The real true-positive rate is approximately equal to the theoretical one.

Table 3.2: Detection results of Case 2 and 3 in a temperature sensor system.

Case No.	Guess	Attacking Signal	True-positive Rate
2	No	144 MHz	100%
3	Yes	144 MHz	93%

3.4.4 Summary of Implementation

When the microphone system and the temperature sensor system are not protected by the detection method, it is hard to tell whether the sensor measurements are manipulated or not. The implementation of the detection method on these two systems shows a true positive rate up to 100% and a false positive rate of 0%, explicitly showing the detection method's generality, feasibility, and robustness. The implementation also demonstrates the flexibility of the detection method regarding non-constant and constant physical quantities; moreover, deploying the detection method does not degrade the signal quality of the sensor measurements.

3.5 Discussion

In this section, how to guarantee the security with a short secret for constant physical quantities will be discussed. Also, a trade-off between security and speed is weighed up. This section will also show how to suit the approach for non-powered passive sensors, and a thorough comparison with PyCRA.

3.5.1 Guaranteeing the Security with Small n for Constant Physical Quantities

In Section 3.2.3, it has discussed that increasing the length of the secret sequence n leads to increasing the difficulty of bypassing the detection method. A larger n results in a more secure system. Given a fixed duration of a measurement, a larger n requires a faster sampling rate of the ADC. Because of the hardware limitations, the sampling rate has an upper limit, and thus n also has a maximal value. Although the sampling rate reaches the highest, it is possible that n is a small number (e.g., $n = 8$). However, in this detection method, a small n can also guarantee the security of the sensor system.

For each measurement, the number of combinations of the n-bit secret sequence is 2^n, and the attacker can find the correct secret sequence to bypass the detection method by trying all combinations. However, in practice, it is impossible for the attacker to try 2^n times, and the attacker has only one chance to change the measurement. The probability of successfully attacking the measurement without being detected is $\frac{1}{2^n}$, and this means that the expected number of successful attacks in attacking 2^n measurements is only one. In the other $2^n - 1$ measurements, the attacking signal is discovered by the sensor system. Imagine that the microcontroller receives $2^n - 1$ invalid measurements before one valid measurement. Because the $2^n - 1$ invalid measurements imply that the sensor system is currently under attack, the valid measurement is still untrustworthy, and hence the microcontroller rejects further processing of the valid measurement.

In general, we suggest using a large n (e.g., $n = 128$) to guarantee the security of the sensor system. However, limited by the sampling rate, although a substantial n may be impractical, a relatively small n is still effective to prevent an attacker from bypassing the detection method, and further, the security of the sensor system is guaranteed.

3.5.2 Trade-off between Security and Speed

In some applications, the sampling rate of an ADC is fixed. To increase the security, the duration of one measurement can be lengthened, and thus more sub-measurements are included. If the physical quantity keeps constant after lengthening the measurement, the number of sub-measurements that the attacker must change increases. As a result, it is more difficult for the attacker to change all sub-measurements without being detected. For non-constant physical quantities, to change the waveform of the sensor output effectively, the attacker has to alter more sub-measurements after lengthening the measurement. Consequently, the difficulty of bypassing the detection method also increases. Above all, without changing the sampling rate of the ADC, the security of the sensor system can be further improved at the cost of lengthening the measurement.

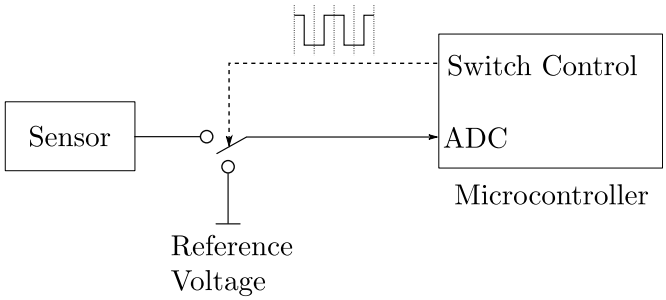


Figure 3.14: A switch that is controlled by the microcontroller is added between the output of the sensor and the ADC.

In summary, to achieve a more secure sensor system, it has to sacrifice the speed, which is either the speed of sampling or the speed of obtaining a measurement. In real applications, designers need to consider the constraints of their sensor systems to choose the proper option to enhance the security.

3.5.3 The Approach for Non-Powered Passive Sensors

In order to deploy this approach to a sensor system with a non-powered passive sensor, a switch can be added between the sensor output and the ADC. Figure 3.14 depicts a configuration for the non-powered passive sensor. The microcontroller can "turn on" and "turn off" the sensor by controlling the switch. When the microcontroller "turns on" the sensor, the switch connects the sensor output and the ADC; thus, the microcontroller can read the sensor output. When the microcontroller "turns off" the sensor, the switch disconnects the ADC from the sensor output. When the ADC is disconnected from the sensor output, in order to ensure that inputs to the ADC settle at a specific level, the ADC should be connected to a reference voltage.

Note that the location of the switch needs to be selected carefully. In short, the switch must be installed very close to the sensor output, as the wire between the switch and the ADC must act as an unintentional antenna to capture attacking signals so that the security is the same as powered passive sensors. Otherwise, if the wire between the sensor output and the switch is long enough, this wire works as an unintentional antenna. As a result, when the ADC is disconnected from the sensor output, the readings of the ADC will not be affected by the attacking signal; since the

zero samples will not be affected by the attacking signal, no attack will be detected, significantly weakening the security of the detection method. It is suggested to make the wire between the sensor output and the switch as short as several millimeters, as such a length of an antenna corresponds to a resonant frequency of hundreds of gigahertz, which is far beyond the operating frequency band of the system; in other words, such a high-frequency attacking signal rarely impacts the system effectively.

Besides, it is essential to point out that the switch frequency is the same as the clock rate of the secret sequence. In practice, when the switch toggles between the "turn-on" state and the "turn-off" state, it will introduce extra noise to sensor measurements. This can be easily handled by abandoning samples at the edges of the sensor measurements as the implementation demonstrates in Section 3.4. In high-frequency applications, the switch frequency also needs to increase accordingly, and it is essential to select a switch that can still work properly at that high frequency. Otherwise, any instatbility (e.g., significant fluctuations during the "turn-on/off" state) will cause disturbance to the sensor measurements, increasing the false positive rate of detection and negatively affecting the usability of the detection method. Also, a higher switch frequency implies a higher sampling rate of the ADC and a faster CPU, challenging the hardware of the system. Therefore, more future studies need to be conducted to ensure the usability of the detection method in high-frequency applications.

3.5.4 Difference between PyCRA and the Approach

Shoukry et al.[61] proposed a generalizable sensor spoofing detection method named PyCRA for sensors such as ultrasonic sensors and infrared sensors, which consist of emitters and receivers. As described in Section 3.1, the emitter sends a challenge signal to the measured entity, and the receiver gathers information from the reflected signal. In a spoofing attack, an attacker manipulates the reflected signal. To detect such attacks, PyCRA turns off the emitter randomly, and hence the receiver should receive nothing during the shutdown of the emitter; if a reflected signal is received when the emitter is off, an attack is detected.

This approach differs from PyCRA in the following aspects. In this chapter, it shows that this approach works for powered/non-powered passive sensors. Because the powered/non-powered passive sensors are the receivers of active sensors, this approach also applies to the active sensors. Hence, it is applicable to all three types of sensors that are defined in Section 3.1. In PyCRA, since an emitter is necessary, this method is designed for active sensors only. Thus, the approach outperforms PyCRA as the approach covers two more types of sensors.

PyCRA counts on the secrecy of the timing of voltage level changes in the challenge signal. In PyCRA, for an attacker in real life, there is a non-zero physical delay between capturing the challenge signal and radiating an attacking signal. This means that the attacker cannot align the attacking signal with the reflected signal. Researchers [64] showed that PyCRA could be entirely bypassed: suppose that the attacker has a faster sampling rate than the sensor system, when the challenge signal starts falling, the attacker can quickly spot the change and stop generating attacking signals. Because the attacker does not influence the periods that are used to detect attacks, she will not be noticed by PyCRA. However, such an attacker cannot bypass this detection method. Note that the attacker has full information of the timing as it is assumed in Section 3.1.4. In other words, this detection method allows the attacker to precisely align the attacking signal with the legitimate sensor output. Even so, the attacker still must guess whether the sensor turns on or off, and a wrong guess will expose the attacker herself to the sensor system.

Regarding the threat model in this chapter, the attacker can stay far away from the sensor system, as the attacker uses EMI to remotely interfere with the sensor readings. In PyCRA, the attacker must stay in a specific area near the sensor system and the measured entity so that she can capture the challenge signal and produce a malicious reflected signal. Therefore, this detection method has a stronger threat model.

For the working principle, this method detects attacks by examining both non-zero and zero samples; however, PyCRA monitors attacking signals by checking

zero samples only. In other words, PyCRA cannot recognize attacks affecting non-zero samples.

3.6 Summary

In this chapter, a novel method is proposed to detect electromagnetic signal injection attacks for sensor systems that match the system model. In the detection method, a sensor system turns off the sensor to monitor the attacking signal in the sensor output. This detection method can prevent the sensor system from processing an attacking signal: once the microcontroller detects an existence of an attacking signal, the microcontroller rejects to handling the sensor output further. Compared with other detection methods, this approach is not only low-cost and space-saving but also can be quickly deployed.

Regarding the security of the sensor system, it proved that the detection method can be bypassed with a negligible probability. The security of the sensor system is based on that the n-bit secret sequence is unknown to the attacker. The longer the secret sequence is, the more secure the sensor system is. Also, this detection method can guarantee the security with a small n.

In practice, the detection method is deployed to a microphone system and a temperature sensor system. The high true-positive rates show that this detection method is effective and robust in detecting the attacking signal. It is essential to highlight that succeeding studies [20, 65] further adapted the detection method to more practical applications, showing its generality and effectiveness.

4

Detection Method for Actuator Systems

Contents

4.1 System Model and Adversary Model 49
4.1.1 System Model 49
4.1.2 Adversary Model 51
4.1.3 Two Injection Points 51
4.2 Attack Detection 53
4.2.1 Modeling Differential Amplifier Output 55
4.2.2 Detection Rule and Choice of Parameters 56
4.2.3 Security Analysis 58
4.2.4 Differences Between Injection Points 59
4.2.5 Attacks on Detection Circuit 60
4.3 Extended Maximum Detectable Frequency 61
4.4 Implementation 63
4.4.1 Setup 63
4.4.2 Speaker System 65
4.4.3 Motor Control System 71
4.4.4 Summary of Implementation 73
4.5 Discussion 74
4.5.1 Different Detection Strategies 74
4.5.2 Adaptive Threshold 76
4.5.3 Choice of Differential Amplifiers 77
4.5.4 Difficulty of Canceling Attacking Signals 77
4.5.5 Difficulty of Drive Signal Injection 78
4.6 Summary 79

Since an actuator is simply an energy transducer, it cannot authenticate its

input signals and will respond to whatever it receives, in the worst case resulting in the adversary being able to fully control the state of the actuator. It is easy to see how electromagnetic signal injection attacks can be used, e.g., to rotate the motor in the smart lock to unlock a door; or force to close a fuel injection valve in a car to stop the car's engine. When the target system is complex and important, these attacks can be incredibly powerful and dangerous. For example, imagine the potential harm if an adversary could control critical industrial applications (e.g., robotic arms) or medical devices (e.g., pacemakers), or say, move the control surfaces of an airplane without pilot input.

Even though such attacks are complicated to perform in practice, and as a result are still rare, it is essential to find effective detection and mitigation strategies to deal with them before they become common. This chapter focuses on detecting attacks on actuators, which is quite a bit harder. The reason is that when a sensor is attacked, the receiving device is a microcontroller that has the ability to run filters and algorithms, or use redundant measurements for added security. For actuators, that is not as easy. When actuators are attacked the receiving device is the actuator itself, and since actuators are "dumb devices" (it might just be a coil of wire, like in a motor), they do not have the ability to ignore malicious signals, even if such signals deviate from some usual pattern.

In this chapter, it provides a novel detection method that uses common and inexpensive electrical components, making it possible to apply this method at scale. The basic idea is to compare the signal to be protected with a reference, in order to identify when any external interference is present. However, this is not as easy as it sounds. First of all, an adversarial signal will affect any reference signals as well, and there are challenges with the sampling rate, bandwidth limits, and signal processing efforts that can make a trivial scheme unusable in practice. The detection method solves all those problems, and it can provide strong detection guarantees for most actuator systems. The contributions of this chapter are as follows:

- It creates a universal system model from practical circuits, and presents an adversary model, clarifying an attacker's capabilities and limitations. (Section 4.1).
- It proposes a general and lightweight detection method that uses differential amplifiers to detect electromagnetic signal injection attacks, and shows that it can provide the actuator system with a strong security guarantee (Section 4.2 and Section 4.3).
- It implements the proposed detection method on a speaker system and a motor control system, demonstrating the generality, feasibility, and robustness of this detection method (Section 4.4).

In the remaining parts of this chapter, additional important issues are discussed in Section 4.5, and a summary is drawn in Section 4.6.

4.1 System Model and Adversary Model

In this section, it introduces a general and flexible system model that fits most actuator systems. This allows to capture the needs of any specific system by tuning the model parameters. In addition, a comprehensive attacker model is presented, which, together with the system model, forms a flexible tool to describe signal injection attacks and defence mechanisms on actuator systems.

4.1.1 System Model

A system that controls an actuator is called an "actuator system". In an actuator system, a microcontroller is the device used to regulate, command, and manage the behaviors of the actuator. Between the microcontroller and the actuator, there are circuits transforming the microcontroller output signal into a suitable signal to drive the actuator. For instance, such circuits may be for signal amplification or waveform conversions. To capture all the characteristics of the circuits, a new device is defined, called the signal conditioner. How this device works will differ from circuit to circuit, but it is treated as a black box. Therefore, the system model

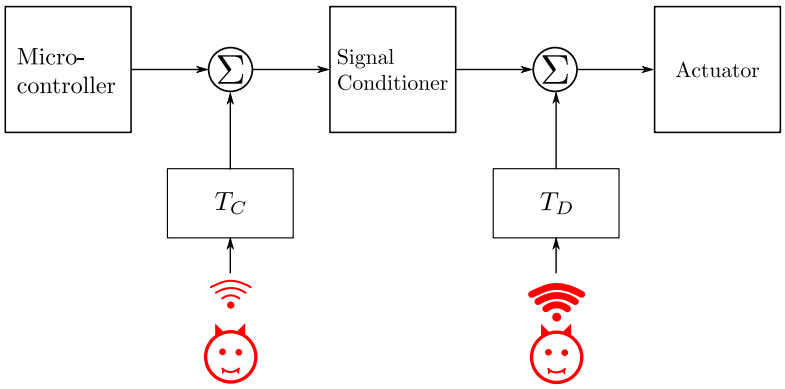


Figure 4.1: The actuator system consists of a microcontroller, a signal conditioner, and an actuator. The transfer functions T_C and T_D explain the control signal wire and the drive signal wire capturing the attacking signals, respectively.

consists of three devices: a microcontroller, a signal conditioner, and an actuator; a block diagram of the system model is presented in Figure 4.1.

In the system model, wires are used to connect these devices: as shown in Figure 4.1, one wire is used to transmit the microcontroller output signal, which is called the *control signal*, to the signal conditioner; the other transmits the signal conditioner output signal, which is called the *drive signal*, to the actuator. Note that, in practice, the control signal wire often carries comparatively little power compared to the drive signal wire, as the voltage and the current of the microcontroller outputs are constrained to a few volts and milliamperes, respectively. Whereas the drive signal wire can carry high-power signals because some actuators consume significant power while working.

The operational frequency of the control signal and drive signal can vary significantly from system to system. Still, it is generally possible to define a normal operating range to which signals are confined in normal operation. This is important because while low/high pass filters can filter out adversarial injections at extreme frequencies, it is more difficult to filter out attacks in the operational range without affecting the valid control/drive signal. This solution assumes that such an operational range can be defined and it is called the upper limit of this range f_{max}. Note that no assumptions about the value of this limit is made, only

that it exists. In Section 4.3, ways of extending this range way beyond the design limits of the electrical components are discussed.

4.1.2 Adversary Model

The attacker aims to affect the actuator by EMI, i.e., injecting an attacking signal into the system. The attacker can inject attacking signals into the actuator system remotely but cannot physically access or modify the actuator system. The attacker is granted full knowledge of the actuator system; specifically, the attacker can predict the waveform and timing (phase) of signals in the actuator system. The attacker can also craft any (physically possible) signal she wants.

In practice, a signal injection can be rather complicated, especially from far away. Still, the adversary is deliberately granted extremely strong power to make sure the detection method works in every case. This complexity is managed using a transfer function that encapsulates any changes to the attacking signal caused by the injection process, e.g., frequency selectivity, attack distance, attenuation, spreading and convolution, etc., as shown in Figure 4.1. The power available to the adversary is not limited. However, it is assumed that a lower limit exists, below which any injected signal no longer has a meaningful effect on the target system. This lower limit is denoted as P_{min}. This power limit is set by the system designer to make sure that any injected signal above this limit is detected. It can be set arbitrarily low, but in order to successfully attack the system, the attacker must inject a signal with power higher than P_{min}.

The reason why the attacker has ideally strong abilities is that if such an attacker cannot avoid being detected by the proposed detection method, it is impossible for any other attackers who are no better than this ideally strong attacker to bypass the detection method.

4.1.3 Two Injection Points

In a particular physical system, there could exist multiple injection places through which attacking signals enter the system. Therefore, many electronic components

will also be affected by the injections. However, only when these injections lead to effects on signals that directly determine system responses will the system be successfully manipulated by the attacker. This has also been considered and shown in previous studies [19, 66]. Therefore, regardless of where the signal injection happens in the system, even if currents are induced in many places at once, it is possible to find an input signal that, when applied to one of the two wires in the system model, produces the same effect. This means that without loss of generality, it can model any signal injection as if the attacking signal was injected into the control or drive signal wire through an appropriate transfer function. In practice, the signal injection does in fact almost exclusively happen via these wires because these are the most efficient "antennas" in the system, and thus where most of the energy is transferred.

There is an important difference between these two injection points. As mentioned previously, the power of the control signal is comparatively weak, so the adversary can more easily overshadow any valid signal on the control signal wire, and it will generally take less power to make changes that affect the actuator through this injection point. Such an injection is defined as a *control signal injection*.

The second injection point is the drive signal wire. This wire will generally carry signals with higher power and more specialized waveforms. For some actuators, e.g., brushless electric motors, the drive signals are not only high-powered but also somewhat complex, and the timing between the different phases of the signal is very important to the operation of the motor. This means that an injection into this wire is more difficult and requires much more power from the adversary. Such an injection is defined as a *drive signal injection*.

Regarding the control signal injection, as indicated in Figure 4.1, only if the signal conditioner reacts to malicious changes in the control signal will the actuator be impacted. Thus, there is no possibility of performing a faster control signal injection where the circuits would not react. However, the drive signal injection impacts the actuator directly, meaning that such an injection can manipulate the actuator without any reaction from the circuits.

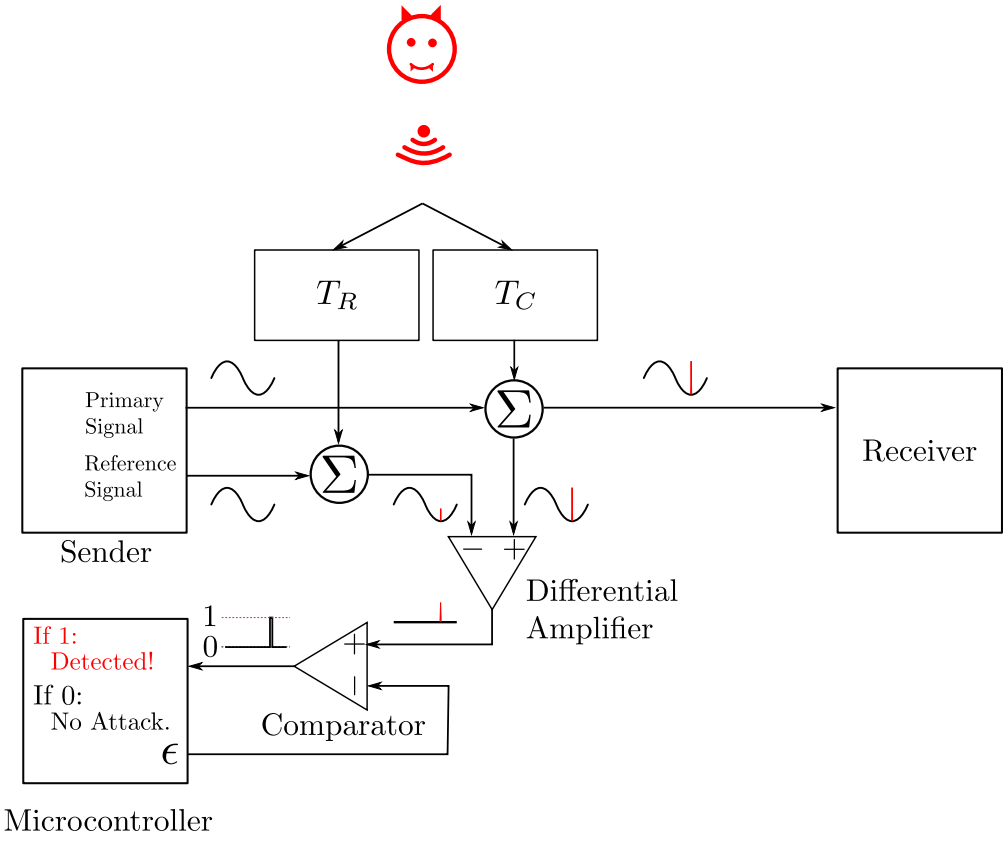


Figure 4.2: A differential amplifier compares the primary signal that is transmitted from the sender to the receiver with the reference signal. A comparator circuit further compares the differential amplifier output with a threshold ϵ, and the microcontroller determines whether attacks happen according to the binary results of the comparator.

Despite the differences in attacker capabilities between the two injection points, the detection system, described in the following sections, works for both the control signal wire and the drive signal wire.

4.2 Attack Detection

As mentioned previously, the detection approach works for both the control signal wire and the drive signal wire. Rather than choose one of the two injection points for description, each wire is instead treated as having a "sender" and a "receiver". Thus, the sender device is either the microcontroller or the signal conditioner, with the corresponding receiver being the signal conditioner or the actuator. Any minor differences between the injection points are discussed in Section 4.2.4.

In order to detect a signal injected into the wire, the sender generates two identical signals, one primary signal (sent to the receiver) and one reference signal (used for detection). This idea is illustrated in Figure 4.2. A differential amplifier is then used to amplify any differences between the primary and reference signals. In the absence of an attack, these two signals should be identical and thus produce no output from the amplifier. However, if the primary wire is affected by an external signal, the difference will be amplified and can be detected using a comparator and a microcontroller. A very important requirement is that the reference wire is sufficiently different from the primary wire to make sure that the adversary cannot modify both in the same way. This can be easily accomplished by simply making the wires different lengths [67] in order to make them sensitive to different frequencies, but a more significant difference can be achieved, e.g., with additional Radio Frequency (RF) shielding materials on the reference wire.

When no attack signal is present, the two input signals of the differential amplifier (the primary and the reference) are the same, and the differential amplifier output is zero. In reality, there will be a non-zero amount of noise, but the output is essentially zero. When an attack happens, the primary wire and the reference wire both pick up the attacking signal. But, because the two wires cannot be modified in the same way, captured by the two different transfer functions T_C and T_R shown in Figure 4.2, the two inputs to the differential amplifier will be different. This results in a non-zero signal on the output of the differential amplifier, and allows the microcontroller to detect the attack. It is essential to emphasize that simultaneously radiating two attacking signals, each carefully matched with the characteristics of a wire, will not cause the same injected signals into the two wires or avoid the detection. This is because the transfer functions essentially guarantee different frequencies of the injected signals, meaning the voltages in the wires change at different rates and eventually result in a voltage difference.

Please note that the differential amplifier is used in a novel way that is substantially different from how it is commonly used in analog electronics, in which the differential amplifier is used to reduce equal interference (common-mode

interference) onto its two inputs [68]. However, in this detection method, the two input channels are deliberately crafted such that the differential amplifier captures the attack interference rather than mitigates it.

4.2.1 Modeling Differential Amplifier Output

The differential amplifier amplifies the difference between its input signals. This is modeled as the difference between the primary and the reference $\delta(t)$, plus additive white Gaussian noise $n(t)$, amplified by a constant gain G. To simplify the notation, t is omitted hereafter. The output of the differential amplifier becomes:

$$o = G(\delta + n)$$

It is essential to point out that the noise n explains the random fluctuations in o; the noise could be very weak, but it always exists regardless of the difference between the input signals. Furthermore, please note that the amplitude of the noise can be easily obtained: in Section 4.4, examples will be shown, where an oscilloscope is used for the observation and the measurement. Since the amplitude is sufficient enough to reflect the strength of the noise, we can regard n as the amplitude hereafter.

Given an attacking signal s, the signal that is injected in the primary wire is $T_C(s)$, and the signal that is injected in the reference wire is $T_R(s)$. In order to obtain a simple relationship between these two injected signals, a simplifying assumption is made: T_R can be expressed as being K times weaker than T_C, and therefore:

$$\delta = T_C(s) - T_R(s) = T_C(s) - \frac{1}{K}T_C(s) = \frac{K-1}{K}T_C(s)$$

Thus, the output of the differential amplifier becomes

$$o = G\left(\frac{K-1}{K}T_C(s) + n\right) \tag{4.1}$$

Finally, taking advantage of the fact that the power that is absorbed by the receiving antenna (the primary wire) is proportional to the attack power P [50], the final model for the detection system is obtained:

$$o = G\left(\frac{K-1}{K}P + n\right) \tag{4.2}$$

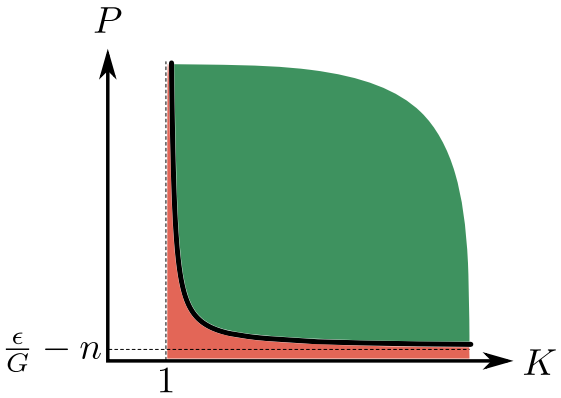


Figure 4.3: The minimum detectable attack power is expressed as a function of K. The detection method can detect attacks on and above the curve (green), but it cannot detect attacks below the curve (red). By decreasing (increasing) ϵ or increasing (decreasing) G, the horizontal asymptote can be moved down (up).

This equation gives the output of the differential amplifier as a function of the main parameters of the detection system, namely the noise n, the gain of the differential amplifier G, the "difference" between the primary and reference wires K, and the power of the adversarial signal P.

4.2.2 Detection Rule and Choice of Parameters

According to Equation (4.2), when no attack signal is present, i.e., the attacker's power is 0, thus $o = Gn$. To make sure that small amounts of noise do not cause false positives, a threshold ϵ is defined, which the output of the amplifier must exceed in order to be detected as an attack. The actual detection is done by a comparator whose output is high when $o \geq \epsilon$ and low otherwise. This allows the output of the comparator to be fed into an interrupt pin of the microcontroller, as shown in Figure 4.2, and ensures that even attack signals with a very short duration can be detected efficiently without requiring the microcontroller to sample at a high rate. Moreover, regarding the detection latency, such a design can detect an attack immediately when the detection circuit captures the waveform difference.

Since the attack is detected if:

$$G\left(\frac{K-1}{K}P + n\right) \geq \epsilon$$

it can be rearranged to see that the detection method can detect any attack with power that fulfills the following requirement:

$$P \geq \left(\frac{\epsilon}{G} - n\right) \cdot \frac{K}{K-1} \tag{4.3}$$

From Inequality (4.3), it can be seen that the minimum detectable power can be made arbitrarily small with appropriate choices of K, G, and ϵ. In the following, the procedures for choosing these values are described.

The choice of K is relatively simple: bigger is better. A large K means that the difference between the two transfer functions T_C and T_R, which govern how an attacking signal affects the primary and reference wires, is as big as possible. To get a sense of how the choice of K affects the detection performance, the attacker's power P as a function of K is plotted in Figure 4.3. The detection method detects attacks on or above the curve (green region), while the attacks below the curve (red region) are not detected. It can be observed that K does not have to be very high for the detection method to be effective, but it does have to be above 1, i.e., the primary and reference wires do have to differ.

The amplification of the differential amplifier G is dictated by choice of amplifier used. Different amplifiers have different maximum gains, and typical values range from 100 to 300. Generally, G should be chosen as high as possible, although in noisy environments, it may be beneficial to reduce the amplification to reduce the sensitivity to noise.

As for choosing the detection threshold ϵ, it needs to be chosen such that environmental noise does not cause a detection event. Therefore, ϵ is chosen just high enough to make sure that false positives from noise are kept to a minimum; an example of this is shown in the implementation in Section 4.4. Moreover, because noise environments are often complicated and change significantly over time, it is emphasized that ϵ does not have to be constant. For example, it can be adaptively adjusted to accommodate lower levels of ambient electromagnetic noise during the night. This is further discussed in Section 4.5.2.

4.2.3 Security Analysis

Recall from the adversary model that the goal of the attacker is to affect the actuator. In order to achieve this goal, the attacker must inject a signal with power of at least P_{min}. Here it proves that such attacks are always detected by the detection method as follows.

Substituting P_{min} into Inequality (4.3), if the following inequality holds the attack is detected:

$$P_{min} \geq \left(\frac{\epsilon}{G} - n\right) \cdot \frac{K}{K-1}$$

To show that it is always possible to find values of K, ϵ, and G to make this inequality true, for any value of n, the first observation is that K can be made arbitrarily high independent of noise. Since $K/(K-1)$ approaches 1 for high enough values of K, a high value can be picked and reduce the above inequality to

$$G(P_{min} + n) \geq \epsilon$$

In addition, as mentioned in the previous subsection, it is a functional requirement that the detection threshold must not be triggered by the noise alone, i.e., the following must hold:

$$Gn < \epsilon$$

Both of the two inequalities above must be true in order to have a functional detection system. That gives the following constraint:

$$G(P_{min} + n) \geq \epsilon > Gn$$

$$P_{min} + n > n$$

$$P_{min} > 0$$

Thus for any value of $P_{min} > 0$, it is possible to find values of K, G, and ϵ that allow the detection system to detect any adversarial signal with power above P_{min} and at the same time do not trigger false positives from noise.

4.2.4 Differences Between Injection Points

Recall that one injection point is the control signal wire, and the other is the drive signal wire. The first difference is between the differential amplifiers at these two injection points. Since the control signal has a low voltage level, it is sufficient for the differential amplifier to have an input voltage range of several volts. However, the drive signal's voltage can go up to hundreds of volts, e.g., 380 V industrial motors. Thus, a differential amplifier with a large enough voltage input range is needed such that the tapped signal will not cause any damage to the differential amplifier. It is not hard to find such a differential amplifier in the market. Note that since the differential amplifier has a much higher impedance than the actuator, the tapping only draws a tiny portion of the control/drive signal, causing negligible impacts on the signal conditioner/actuator.

Another difference between these two injection points is that the drive signal can be much more complex than the control signal, and thus, it may be more complicated while deploying this approach for the drive signal. In the previously mentioned example of a brushless electric motor, the microcontroller produces one signal for controlling, while the signal conditioner needs to convert this solitary control signal into three different signals to drive the motor. In general, it is essential to deploy this approach to each signal to guarantee the security, which means one for the control signal and three for the drive signals. However, in many cases where the physical properties of the multiple wires are the same or very similar and they are put close to each other, it is tricky that protecting one wire is sufficient enough, and doing so can significantly reduce the complexity of deploying this approach. This is because the attacker cannot selectively choose a wire to affect in these cases, and in other words, all of these identical or similar wires will be impacted by the attack. In the example of the brushless DC motor, since its three drive signal wires are almost identical and are put very close to each other, protecting any one of the wires with this approach is equivalent to protecting all three wires.

4.2.5 Attacks on Detection Circuit

This defense mechanism has added circuitry to the system that could itself be the target of an injection attack. In this section, it demonstrates that this circuitry cannot be exploited by the attacker to achieve the injection.

First, note that there is no path from the detection circuit to the actuator, so the only malicious action that needs to be considered is whether an adversary could inject a signal that would be hidden from detection because of interference in the detection circuit itself.

To analyze this, a new transfer function T is defined for the main wire in the detection circuit. The resulting signal that is injected into the detection wire is then $T(s)$ when the adversary sends s. Please note that s also explains multiple attacking signals that are radiated by the adversary simultaneously because the superimposition of these attacking signals makes them into one attacking signal. Note that there may also be multiple injection points, as discussed in Section 4.1.3, but they can be modeled to the main wire, as o directly determines whether an attack happens or not. Therefore, the injected signal is superimposed onto o, described in Equation 4.1, making the modified differential amplifier output o':

$$\begin{aligned} o' &= T(s) + G\left(\frac{K-1}{K}T_C(s) + n\right) \\ &= \frac{G(K-1)}{K}\left(\frac{K}{G(K-1)}T(s) + T_C(s)\right) + Gn \end{aligned}$$

If the attacker wants to avoid detection o' must be zero (technically just less than ϵ, but basically zero). That means that the value in the parentheses must be zero, which in turn requires the following equation holds:

$$\frac{K}{G(K-1)}T(s) = -T_C(s) \tag{4.4}$$

The negative sign in Equation 4.4 implies that $T(s)$ and $T_C(s)$ must be 180 degrees out of phase, and this requires that the physical distance between the two corresponding wires is exactly half of the wavelength of the attacking signal s [67]. This is already a good argument for why an attacker cannot inject a signal that

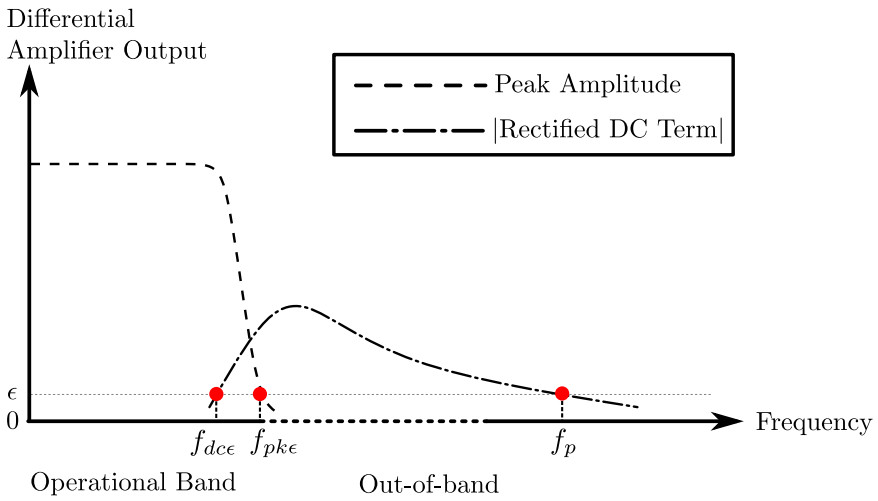


Figure 4.4: With constant attack power, the peak amplitude (dashed line) and the rectified DC term (dash-dotted line) of the differential amplifier output signal change along with the frequency. The maximum detectable frequency is extended to f_p.

affects the actuator, and simultaneously cancel it out in the detection system, since the frequency would have to be in the 10-100s of GHz to get a half wavelength short enough. Such a high-frequency signal is way above what affects most actuators.

However, just to make the point extra clear, let's assume that the attacker could in fact send a signal with a high enough frequency to make this work. Even then, the constant $\frac{K}{G(K-1)}$ in Equation 4.4 means that the signal injected into the detection system, $T(s)$, must be 100s of times stronger than $T_C(s)$, the signal injected into the actuator control signal itself. However, given that the two wires are so close to each other, and that the smaller of the two needs more power injected into it, it is impossible to achieve such a $T(s)$ in practice. As a result Equation 4.4 can never hold in practice.

For those two reasons (phase difference and relative power), no adversarial signal s can ever prevent its own detection due to interference with the detection circuitry itself.

4.3 Extended Maximum Detectable Frequency

This detection method relies on a differential amplifier to help detect injected signals. Like all electronic components, a differential amplifier is designed to

work within a particular frequency range. When choosing parameters for the detection system, a suitable differential amplifier should be used, which covers the entire range where adversarial signals are likely to be able to affect the actuator. However, on rare occasions, e.g., for very high-frequency applications or if cost is a significant concern, it might be difficult to get a differential amplifier that fully covers the desired frequency range. For such cases, a method is brought about to extend the maximum detectable frequency f_{max} beyond the normal upper bound of the differential amplifier.

Many previous studies [69–72] have shown that a differential amplifier will still respond beyond its normal operational band, although the response is entirely different from the normal amplification within its design parameters. As the frequency increases beyond f_{max}, the peak amplitude of the differential amplifier output starts to decline, as the gain plummets to almost zero [73, 74]. This is shown in Figure 4.4 where the dashed curve depicts the change of the peak amplitude.

Although the peak amplitude decreases to nothing, the output will gain a DC offset with respect to the normal ground state, shown in Figure 4.4 as the Rectified DC Term. This happens as the differential amplifier rectifies the high-frequency signals [69, 75, 76]. The phenomenon is also known as radio-frequency (RF) rectification, and it is attributed to the nonlinear voltage-current characteristic of transistors that make up the differential amplifier [75]. Further increasing the frequency will eventually decrease the rectified DC term, which will ultimately become negligible when the frequency is high enough [70–72]. While this effect does eventually disappear, it allows us to extend the detection by hundreds or thousands of times higher than the upper bound of the operational band.

It is important to note that this phenomenon is not limited to a specific differential amplifier, but is true for many different designs, which have been experimentally verified in the literature [69, 77].

For the detection system to provide firm guarantees, it is essential to ensure no gap in the protected frequency band. Therefore, it has to be ensured that the DC offset rises enough to be detected before the normal peak amplitude of

the differential amplifier goes to zero. In Figure 4.4, the frequency at which the magnitude of the rectified DC term exceeds ϵ is denoted as $f_{dc\epsilon}$, and the frequency at which the peak amplitude falls below ϵ is $f_{pk\epsilon}$. In Section 4.4, it shows that $f_{dc\epsilon} < f_{pk\epsilon}$ can be easily achieved in practice.

4.4 Implementation

This detection method is implemented on two practical and distinct actuator systems: a speaker system (in Section 4.4.2) and a motor control system (in Section 4.4.3). The objective of the implementation is to validate the feasibility of the detection method in practice. One of the reasons why they are chosen is that they are widely deployed in many critical applications: the speaker system can be found in applications in which sound information needs to be broadcast, such as mobile phones and car satellite navigation; the motor control system can be found in those that need to drive some mechanical structures, such as smart locks and insulin pumps. Implementing the detection method on these two systems also verifies its capabilities of handling different actuator systems regardless of types of signals: sinusoidal signals (analog) are used in the speaker system, while pulses (digital) in the motor control system.

First, this section introduces how to build an actuator system on which the detection method can be quickly implemented. Then, it shows how to detect various attacking signals in each actuator system. Only the control signal injection is demonstrated, as the drive signal injection is power-consuming and difficult to achieve with our equipment (please see detailed discussion in Section 4.5.5). Finally, a summary of the implementation of these two actuator systems is given in Section 4.4.4.

4.4.1 Setup

Based on the system model, a setup that can be easily configured into a speaker system or a motor control system is built, as shown in Figure 4.5. A signal generator is used to produce the control signal and the reference signal. The signal

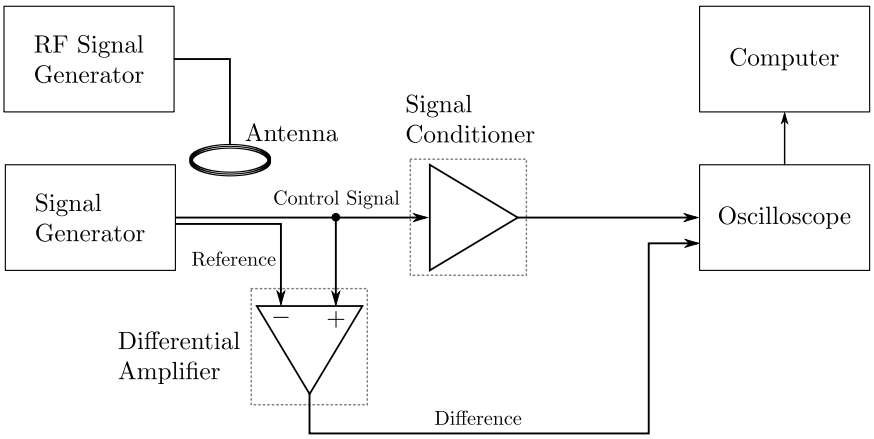


Figure 4.5: A setup of the actuator system. Devices in the dotted squares differ from system to system, and others are the same.

generator is functionally equivalent to the microcontroller. The benefit of using the signal generator is having easier control of signals regarding their frequencies, amplitudes, synchronization, etc.

The control signal is fed into a signal conditioner. The signal conditioner is different in these two systems: an audio power amplifier LM386 is used in the speaker system, and a brushed DC motor driver chip DRV8833 is used in the motor control system.

Regarding the actuator (either a loudspeaker or a motor), since its responses are deterministic and its input signal (i.e., the drive signal) sufficiently reflects the responses, the actuator in the setup is simply omitted but an oscilloscope is used to monitor and record the drive signal. An advantage of doing so is that different actuator systems can be quickly tested without extra work of using different methods to sense and process the actuator responses (e.g., a microphone to measure sound played by the speaker, or a hall-effect sensor to measure the speed of the motor). Moreover, a computer is used to process the data that is recorded by the oscilloscope. Note that the oscilloscope and the computer are used for the purpose of demonstrating the feasibility of the detection method, and are not used in the proposed applications.

Based on such an actuator system, the detection method is deployed. The control signal and the reference signal are fed into a differential amplifier, as shown in

Figure 4.5. In the speaker system, an AD623 with a gain of around 150 is chosen as the differential amplifier because it is specifically designed to amplify small differences between its two inputs. As for the motor control system, a unity-gain differential amplifier AD629 is selected as the differential amplifier, as it can handle high-voltage inputs. The output of the differential amplifier is monitored and recorded by the oscilloscope, and the recorded data are sent to the computer for attack detection.

To achieve a large K, i.e., difference between the transfer functions of the control signal wire and the reference wire, a loop is formed on the control signal wire to make it easier to pick up the attacking signal, and choose a short cable as the reference wire. Thus, the control signal wire is much more sensitive to the attacking signal than the reference wire. Note that it does not matter which wire is more sensitive because the detection method only requires the transfer functions to be different. Moreover, to guarantee that the control signal and the reference signal arrive at the differential amplifier at the same time, the tapping point is carefully chosen to ensure that the paths that feed these two signals into the differential amplifier have the same length.

The setup is extremely flexible and allows us to easily experiment with different actuator types without having to build dedicated systems for each one. Despite being a lab setup, the experimental results accurately reflect the response of real commercial products.

4.4.2 Speaker System

In a speaker system, an audio signal is amplified and then broadcast. The objective of the attack is maliciously manipulating the waveform of the audio signal, and in the extreme case, can lead to the speaker system broadcasting any messages the attacker wishes.

Determining Threshold

The differential amplifier output is measured when no attack happens. The measurements show that the differential amplifier output signal amplitude is always

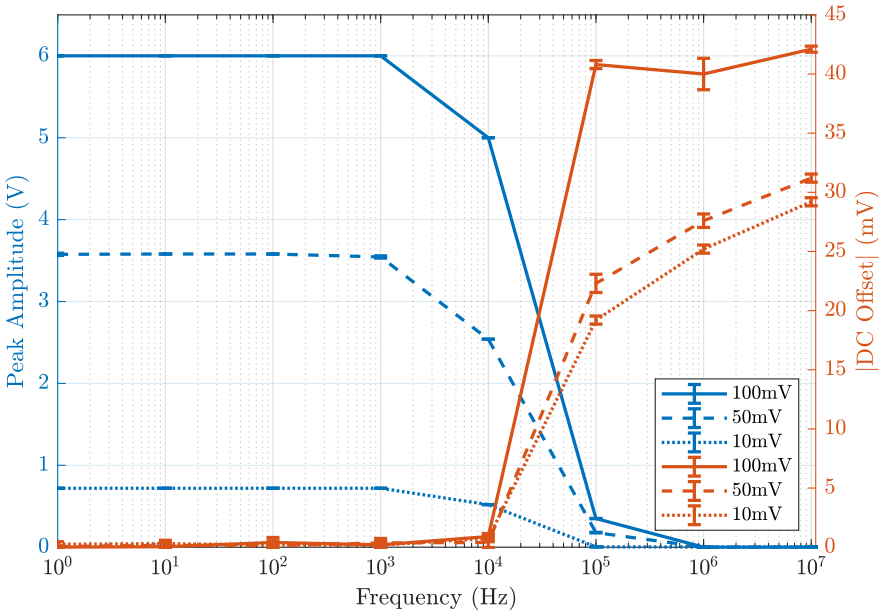


Figure 4.6: The peak amplitude (left y-axis) of the differential amplifier output drops to zero when the frequency of the attacking signal is far beyond the operational band of the audio amplifier; the DC offset (right y-axis) rises while increasing the frequency of the attacking signal. The peak-to-peak voltage of the attacking signal is from 10 mV to 100 mV

below 2.4 mV. Since this value already includes all noise sources in the experimental environment, it is chosen as the threshold. The benefit of choosing this value as the threshold is that, on the one hand, it significantly reduces the possibility of which the noise accidentally triggers the detection; the false positive rate remains at 0% as calculated from the measurements. On the other hand, this threshold is small enough to guarantee that the weakest attack that effectively impacts the actuator system is successfully detected, and this will be shown and explained in the experimental results as follows.

Direct Power Injection Attacks

The normal operational band of an audio amplifier is below the megahertz level, and low-frequency attacking signals are needed for in-band attacks. Due to the practical difficulty of injecting low-frequency attacking signals into the circuit wirelessly, direct power injection (DPI) [59] is used to demonstrate that the detection method can handle the in-band attacks. Note that in the following sections (Section 4.4.2

and Section 4.4.3), the attacking signals are injected wirelessly.

In order to show that any malicious frequency can be injected into the audio signal, the attack frequency is swept from 1 Hz to 10 MHz, and the peak-to-peak voltage of the attacking signal is from 10 mV to 100 mV. The reason why the highest attack frequency is set to 10 MHz, which is beyond the operational band of the audio amplifier, is to verify that no gap (as described in Section 4.3) exists in the frequency band. The reason why 10 mV is chosen as the weakest peak-to-peak amplitude of the attacking signal, is that the malicious change caused by an attack at this voltage is already around 49 dB weaker than the audio signal. Weaker attacking signals have little to no impact on the speaker system.

To demonstrate the impact of the attack on the differential amplifier in detail, both the peak amplitude and the DC offset are shown in Figure 4.6. Each point in the figure represents the averaged peak amplitude or the averaged DC offset with a standard deviation. The first observation of the experimental results is related to the attack power: the peak amplitude and the DC offset increase (decrease) while the attack power increases (decreases). Concerning the attack frequency, when it is lower than 1 kHz, the peak amplitude is significantly larger than the threshold, which reveals the existence of the attacking signal. When the frequency is between 1 kHz and 1 MHz, the peak amplitude plummets, but it is still above the threshold; meanwhile, the DC offset rises above the threshold. When the frequency of the attacking signal reaches 1 MHz and beyond, the DC offset is well above the threshold, indicating the existence of the attack.

The experimental results validate the capabilities of the differential amplifier to detect attacks in the entire frequency range from DC to 10 MHz. This experiment is repeated 240 times and all (240 out of 240) attacking signals are detected, making the true positive rate is 100%. This shows that even for practical systems, the detection method provides strong protection against both in-band and out-of-band attacks.

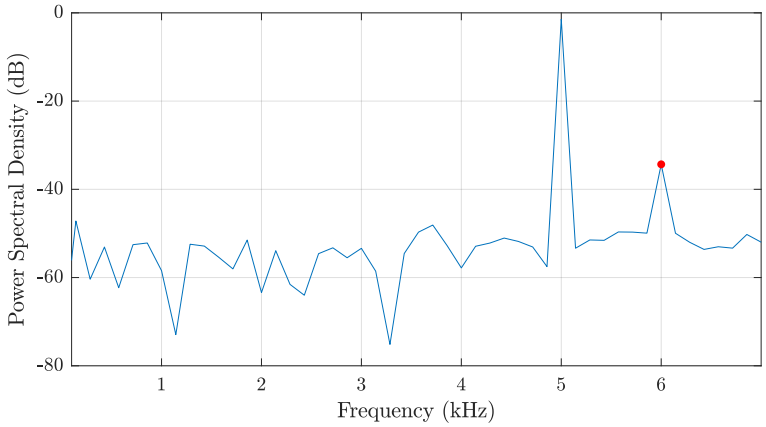


Figure 4.7: A 6 kHz malicious signal is successfully injected into the 5 kHz audio signal. The 6 kHz spike is highlighted by a red point in the frequency domain of the audio amplifier output. The power ratio between the 6 kHz frequency component and the 5 kHz is around −30 dB.

Wireless Attacks

To test high-frequency attacks in a more realistic setting, a high-frequency carrier is modulated with an audio signal and inject it wirelessly into the control and reference wires. An RF signal generator is used to produce the attacking signals, and they are radiated by a coil antenna, as shown in Figure 4.5. The antenna is placed around 2 cm above the control signal wire for the best possible energy transfer. That way less power can be used to achieve the wireless attack in the experiments. If an attacker is further away from the victim system, she needs more powerful attacking signals to achieve the attack.

To present a concrete attack, a 6 kHz malicious frequency is picked to be injected into a 5 kHz audio signal. In Figure 4.7, an attack result is shown: in the frequency domain of the audio amplifier output, besides the legitimate 5 kHz frequency component, a malicious spike can be observed at 6 kHz. In order to quantify the impact of the attack, the power ratio between the malicious frequency component and the legitimate frequency component is measured, which can be expressed as the following equation:

$$impact = 10 \times \log_{10}\left(\frac{P_{malicious}}{P_{legitimate}}\right)$$

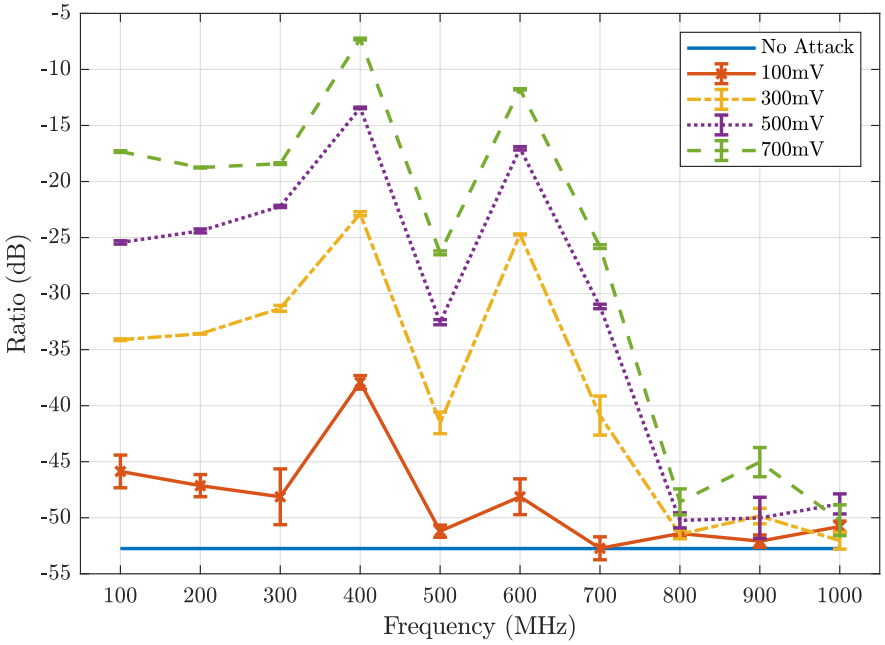


Figure 4.8: The power ratio between the malicious signal and the legitimate signal gradually decreases while increasing the frequency of the attacking signal. The peak-to-peak voltage of the attacking signal is from 100 mV to 700 mV.

where P represents the power. The bigger the ratio is, the stronger the injected signal is, and the larger the impact of the attack is. When no attacking signal is presented, the measurements show that the *impact* remains at around -52.7 dB.

Different attacking signals are generated to test the performance of the detection method: the peak-to-peak voltage of the attacking signal is changed from 100 mV to 700 mV, and the carrier frequency of the RF signal is changed from 100 MHz to 1000 MHz. The impact of the attacks are numerically represented in Figure 4.8. When the attack frequency reaches 800 MHz, the *impact* is close to -52.7 dB, which means that attacks beyond this frequency will have little practical significance. Experiments beyond 1000 MHz are also conducted, but the impact of the attacking signal beyond 1000 MHz are smaller, and hence this part only focuses on the frequency range within 1000 MHz.

Regarding the attack detection, the peak amplitudes of all measurements of the differential amplifier output are below the threshold. This is because the frequency of the attacking signal is already far beyond the operational band of the differential amplifier, as explained in Section 4.3. However, as shown in Figure 4.9, the DC

offset of the differential amplifier output is well above the threshold throughout the range for all attacker signals other than 100 mV, indicating the existence of the attack. It can be seen that the DC offset increases when the attack power is increased, so for attacking signals with peak-to-peak voltages of 300 mV, 500 mV, and 700 mV, the DC offsets are always above the threshold (solid blue line) regardless of the frequency. When the attacking signal is 100 mV, a few attacks fall below the threshold when the carrier frequencies reach 800 MHz and 900 MHz. Referring back to the impact of these two attacking signals in Figure 4.8, the ratios indicate that the impacts are so tiny that they are unlikely to have any significance for a practical system. Since the detection method successfully detects 389 out of 400 attacking signals, the true positive rate is 97.25%.

In Figure 4.9, the curves of DC offsets vary up and down along the attack frequency. This is because the attacking signal is injected wirelessly instead of through DPI. The transfer function of the wire accounts for the ups and downs of the curves: the attacking signal is efficiently injected into the wire at specific frequencies where local maximum values of the DC offset reaches, but less efficient at other frequencies.

The experiment results show that the frequency range covered by the differential amplifier is easily large enough to protect the frequency band that the speaker system is vulnerable to. This detection method shows the feasibility of detecting the attacking signals with frequencies from DC to far beyond the speaker system's operational band. Moreover, given the wireless injections, the detection method demonstrates its capabilities of handling real attack scenarios. Concrete attacking signals have been demonstrated that can precisely manipulate the audio frequencies, but it does not mean that the detection method can only handle these specific attacking signals. Any attacks that cause voltage changes of the differential amplifier output signal beyond the pre-determined threshold can be spotted immediately.

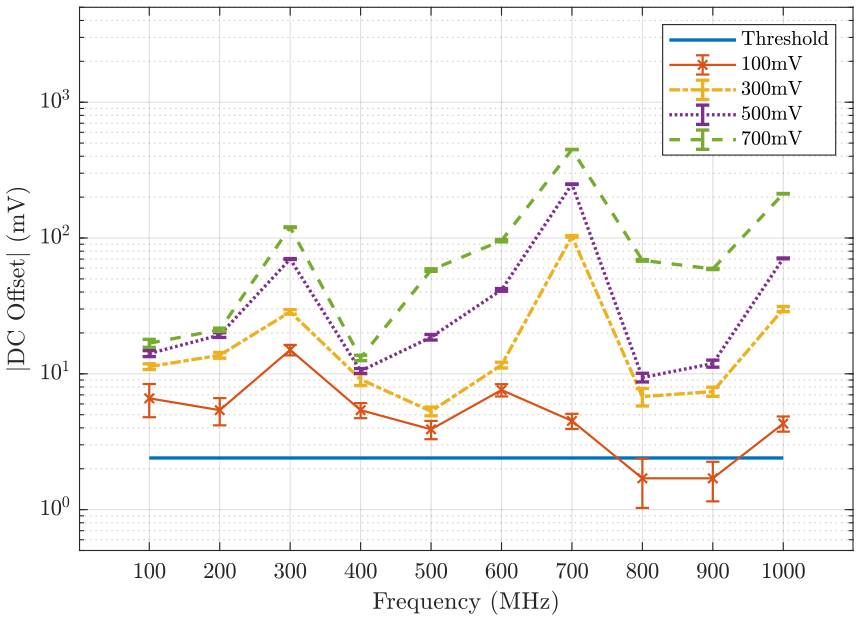


Figure 4.9: The DC offset of the differential amplifier output varies while changing the voltage level and the frequency of the attacking signal. The peak-to-peak voltage of the attacking signal is from 100 mV to 700 mV.

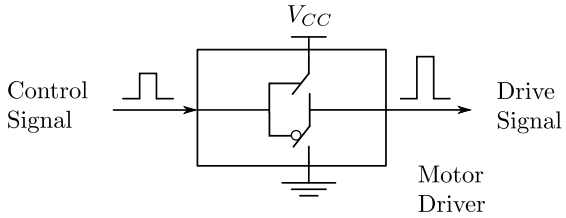


Figure 4.10: A motor driver is used to amplify a control signal to drive a motor.

4.4.3 Motor Control System

In the motor control system, a pulse signal is used to control the rotating speed of the motor. The duty cycle of the pulse signal describes the amount of time that the signal is at the high-voltage level as a percentage of the total time of a cycle. The larger the duty cycle is, the faster the motor's rotation speed is. As mentioned in the setup, a motor driver is used as a signal conditioner to amplify the control signal into a powerful drive signal to energize the motor. The motor driver is made of transistors, and for simplicity, as shown in Figure 4.10, they can be regarded as two switches that are connected in series and are controlled by the pulse signal. Since these two switches work in opposite ways, the output signal

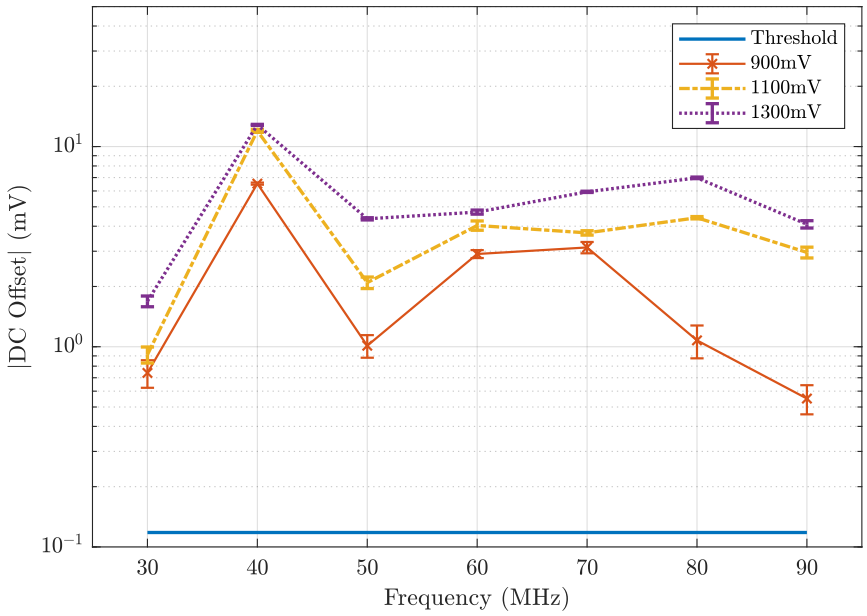


Figure 4.11: When an attack happens, the DC offset of the differential amplifier output is always above the threshold, implying detecting the attack.

toggles between V_{CC} and the ground in the same pattern as the input signal. The attacker's objective is to manipulate the duty cycle and impact the functionality of the motor. Previous studies [76, 78, 79] have shown that fine-tuned attacking signals can toggle a transistor. In order to change the duty cycle of the motor driver output signal, the attacker needs to control the duration of radiating an attacking signal so as to manipulate the rotating speed of the motor.

Determining Threshold

When no attack presents, the differential amplifier output signal is recorded, and the threshold is 0.17 mV. This threshold value is chosen as it makes the false positive rate to its minimum (0%) in the experimental environment; also, this threshold is sufficiently large to spot the weakest attacks, as shown as follows.

Detection of Attacks

Since the differential amplifier is specifically designed to handle the input difference in its operational band, it is not difficult to detect the in-band attacks. The in-band

attacks are no longer repeated but focus on the out-of-band attacks. Note that the out-of-band attacks are realized wirelessly.

In the experiments, the frequency of the attacking signal ranges from 30 MHz to 90 MHz, and the peak-to-peak voltage ranges from 900 mV to 1300 mV. The reason why the frequency of the attacking signal is below 90 MHz is that beyond this frequency, the motor driver never responds to the attacking signal, even though the peak-to-peak voltage of the attacking signal reaches its upper limit in the signal generator. The reason why the peak-to-peak voltage of the attacking signal is above 900 mV is that, below this voltage level, the attacking signal is too weak to affect the motor driver. In the experiments, the RF signals can cause the motor driver to output a low voltage level when a high voltage level should be outputted, thus reducing the duty cycle of the pulses. By precisely controlling when to start and stop radiating the attacking signals, the duty cycle of the control signal can be precisely manipulated, further controlling the motor speed. Note that using other types of attacking signals can also increase the duty cycle [26]; however, the purpose of the experiment focuses on attack detection, and thus not further show and discuss how to control the motor speed.

Regarding the attack detection, both the peak amplitude and the DC offset of the differential amplifier output signal are checked. Under these out-of-band attacks, the peak amplitude is always below the threshold. However, as shown in Figure 4.11, the DC offset is always above the threshold, indicating an attack. All (210 out of 210) DC offsets are above the threshold, indicating that all attacking signals are detected. Therefore, the true positive rate is 100%.

4.4.4 Summary of Implementation

The implementation of the detection method on the speaker system and the motor control system shows the generality of the method regardless of the type of signal. The deployments also demonstrate the simplicity of implementing the detection method in practice. The high true positive rates and low false positive rates in the

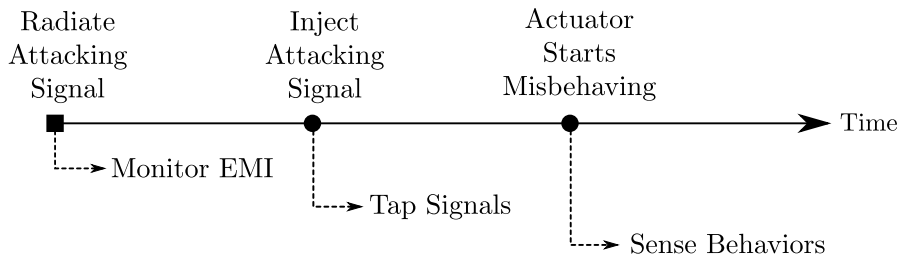


Figure 4.12: The timeline shows that an attack starts from radiating an attacking signal to actuator misbehaving. At different moments, the attack can be detected by different ways.

speaker system and the motor control system show the robustness of the detection method on different actuator systems.

Regarding the implementation overheads, a simple detection circuit that is composed of differential amplifiers is needed; in addition, two more channels of the microcontroller are required: one provides the reference, and the other (i.e., interrupt) receives detection results from the detection circuit and thus no significant computational power is needed. Compared with the detection methods such as EMI detectors requiring additional RF interfaces and detection methods counting on complicated algorithms to sense and process actuator behaviors (see details in Chapter 6), this detection method is much more straightforward and lightweight.

4.5 Discussion

This section discusses different detection strategies, how an adaptive threshold can handle varying environmental noise, and how to choose a differential amplifier. Moreover, the difficulties of canceling out an injected signal in circuits is discussed, as well as the difficulties of the drive signal injection.

4.5.1 Different Detection Strategies

An electromagnetic signal injection attack has several distinct phases, each of which gives rise to different detection strategies. In the attack timeline shown in Figure 4.12, three moments are highlighted: the first moment is when the attacking signal is radiated; the second is when the attacking signal is injected into a wire

in the target system; the third is when the actuator starts deviating from its intended activity. Each of the three moments marks the start of a new phase of the attack. The three phases should not be thought of as sequential since an attack of any meaningful duration will be in all three phases at once, but should rather be thought of as three opportunities to detect the attack.

In the first phase, when attacking signals are being radiated, the electromagnetic radiation can be detected in the environment using an antenna. Thus, the detection strategy is monitoring the environmental electromagnetic power level: if the power level is above a pre-defined threshold, or maybe outside a known noise profile, the attack is detected. This strategy has the potential to detect an attack early; however, it requires a monitoring device that can reliably detect adversarial interference at the frequency that would harm the target system, which is not always easy to achieve. Examples of this detection strategy are some of the anomaly detectors that will be introduced in Chapter 6.

In the second phase, when the attacking signals are successfully injected into the actuator system, the signals in the actuator system wires are changed. Since the attacking signals are not supposed to exist in the actuator system, these changes are a reliable indicator of an attack, if they can be measured. This detection method uses the second form of detection, i.e., it detects signals that are successfully injected into the actuator system.

In the final phase, the actions of the actuator will deviate from what the system expects, assuming the attack is powerful enough to result in a measurable change. If the system can detect this behavior change, this can be used to detect the attack. This might be an attractive detection strategy since no effort is wasted on attacks that do not have a measurable effect on the target actuator. However, detecting such attacks typically requires extra sensors. By the nature of the detection method, it will only detect attacks after they have already affected the system. An example of this detection method is Muniraj and Farhood's work [80] that will be discussed in Chapter 6.

4.5.2 Adaptive Threshold

In the implementation of the detection method, a proper threshold is determined by experiment and keep it constant while testing the performance of the detection method. The advantage of using a constant threshold is that once a proper threshold is found and determined, it is efficacious forever, and the designer never has to adjust it again. However, in some cases, the environmental noise varies significantly and complicatedly over time; for example, such noise may originate from the radiation of other complex circuits. To provide the actuator system with more flexibility, the designer can program the actuator system to adjust its threshold adaptively when necessary. Imagine a simple case: during the daytime, the noise is intense because of human activities (e.g., wireless communications, transportations), but at midnight when people sleep, the noise becomes relatively weak. During the daytime, the threshold can be slightly increased to allow more noise, and as such, it can avoid the noise frequently triggering the detection. At midnight, to restore the detection method to be more sensitive to attacks, the designer can program the actuator system to lower down the threshold.

Other environmental changes may also impact the detection circuits, and they can also be handled by the adaptive threshold. For example, in some harsh environments where the environmental temperature varies significantly, a temperature change will affect the resistance of metal conductors, further the voltages. Since the resistance of the two wires may not change consistently, the voltage difference between them increases when the temperature varies. To reduce the impact caused by the temperature, basically, materials with a small temperature coefficient should be chosen; for instance, regarding copper with a temperature coefficient of around 0.004, a change of 100 °C only lead to a change of 0.4 ohms in resistance. Furthermore, to cover the extra voltage difference that is caused by the temperature variation, the threshold can be adaptively set to a higher level in such an environment.

No matter how the designer adjusts the threshold adaptively, it is still essential to guarantee that the detection method meets the requirements as mentioned in

previous sections: first, no noise triggers the detection accidentally; second, no attack that effectively impacts the actuator is missed.

4.5.3 Choice of Differential Amplifiers

According to Section 4.2 and Section 4.3, while choosing a differential amplifier, its *gain*, *noise*, and *maximum detectable frequency* are essential factors to be considered. In short, the gain and the noise should make the minimum attack power that the system can detect as the system designer wants; the maximum detectable frequency should be sufficiently large enough to cover the frequency band within which the system is vulnerable. Only in this way can the detection method provide a strong security guarantee. Section 4.4 explains the reason why those two differential amplifiers meet the requirements of the speaker system and the motor control system, respectively; it is essential to note that efforts are made to test many different off-the-shelf differential amplifiers until those two proper differential amplifiers were found. Indeed, it is easy to test and deploy the off-the-shelf devices, and they are sufficiently good enough to show the feasibility of the detection method, but it can also be noticed that there is no control over the factors of the differential amplifiers in our implementation. There may be applications where no off-the-shelf differential amplifier is found, and as a result, it is essential to customize and manufacture a differential amplifier. For example, when the detection method needs to sense weak attacking signals, a larger gain with smaller noise is required so as to achieve a lower minimum detectable attack power; when the victim system is sensitive to high-frequency attacking signals, a larger maximum detectable frequency is required so as to handle those attacks. However, the tradeoff of doing so will lead to more expense on research and development.

4.5.4 Difficulty of Canceling Attacking Signals

An idea of mitigating the influence caused by attacks is generating an “anti-attack” signal to cancel out the attacking signal. The anti-attack signal and the attacking signal have the same frequency and amplitude, but they are 180 degrees out of

phase. When the anti-attack signal and the attacking signal meet, they destruct each other. This idea is similar to the sound noise cancellation technology that is used in headphones. However, it is hard to realize such a cancellation regarding the electromagnetic interference. In the air, an electromagnetic signal propagates around the light speed; in the circuit, the speed halves. In addition, it takes time for the actuator system to capture the attacking signal and then generate the anti-attack signal for the cancellation. This means that the anti-attack signal always lags behind the attacking signal. It is difficult to synchronize the anti-attack signal with the attacking signal unless the microcontroller can predict the attacking signal.

4.5.5 Difficulty of Drive Signal Injection

As mentioned previously, compared with the control signal injections, a drive signal injection may require much more power if the actuator is power-consuming. The power of a drive signal injection is estimated as follows. According to datasheets of an off-the-shelf motor, it needs a drive signal that is around 4.5 W; as for a microcontroller, such as an Arduino Uno microcontroller, it can output a control signal that is only 0.1 W. For simplicity, suppose that the attenuation on attacking signals is the same in those two injections. Then, the attacker needs to radiate at least $\frac{4.5\,\mathrm{W}}{0.1\,\mathrm{W}} = 45$ times more power to realize the drive signal injection than the control signal injection. This result implies that it is much more difficult and costly to conduct the drive signal injection than the control signal injection in practice.

Another piece of evidence to show that the drive signal injection is hard to achieve is to regard the injection as wireless power transmission [81]. In wireless power transmission techniques, scientists specifically designed both antennas of the transmitter and the receiver to achieve power transmission. Given the wire that works as a low-gain antenna in the actuator system, delivering enough power into the drive signal wire can be much more challenging.

4.6 Summary

In this chapter, a novel detection method that can detect electromagnetic signal injection attacks on actuator systems is proposed. This class of systems previously had to rely on physical security measures and signal decay, and had no meaningful security guarantees against a determined adversary. This detection system fills this critical gap and provides strong detection guarantees to any actuator system. The core idea of the detection method is straightforwad: any diffenrence caused by external attacks between two identical signals (the primary signal and the reference signal) indicates the attacks. This detection method provides provable guarantees against attacks, and can be tuned to any attack power and any amount of environmental noise. It has been shown that the detection method provides the actuator system with a strong security guarantee, and an attacker who attempts to effectively manipulate the actuator system will always be detected by the detection method. Despite this, the detection method requires only a few cheap off-the-shelf electronic components and does not add any significant weight to the system it protects. This is important in many contexts, such as aviation and implantable medical devices. Moreover, the implementation of the detection method on a speaker system and a motor control system proves its generality for different actuator systems, as well as its effectiveness and robustness in a practical setting.

5

Message Injection into Differential Signaling Systems

Contents

5.1 System Model and Adversary Model 82
5.1.1 System Model 82
5.1.2 Adversary Model 86
5.2 Bit Injection Attack 87
5.2.1 Bypassing Subtractor 88
5.2.2 Bit Detected Incorrectly in Receiver 89
5.3 Analysis of Success Rate 90
5.3.1 Parameterization 91
5.3.2 Success Rate of Bit Injection 92
5.3.3 Success Rate of Message Injection 95
5.4 Experiments 96
5.4.1 Testbed 96
5.4.2 Subtractor 97
5.4.3 Receiver 101
5.5 Message Injection into CAN 106
5.5.1 CAN Basics 106
5.5.2 Message Injection 108
5.6 Discussion 110
5.6.1 Gaining Knowledge 110
5.6.2 Restricted Attack Power and Distance 111
5.6.3 Future Countermeasures 111
5.7 Summary 111

The “antenna-like behavior” of wires has motivated the proliferation of studies

abusing adversarial electromagnetic signals to inject malicious information into wired communications, as introduced Section 2. Several researchers recommended using *differential signaling* to resist the attacks [14, 32], as it can reject external noise by looking at the difference between two complementary signals, supposing the noise impacts both signals equally. This chapter will show that because of circuits' asymmetry and nonlinearity, differential signaling cannot provide sufficient protection when faced with electromagnetic signal injection attacks. This allows an attacker, who has no physical access to a victim system, to conduct electromagnetic signal injection attacks to control a victim system. Since so many popular protocols such as USB, Ethernet and HDMI are based on differential signaling, and these protocols serve in countless safety- and security-critical applications (e.g., automotive, aviation, robotics), such attacks immediately put these applications at risk.

Specifically, the impact of the electromagnetic signals is to cause receiving circuits to detect incorrect bits [82]. Imagine a scenario of using a wired USB keyboard. An attacker can use electromagnetic signals to interfere with the USB cable and injects malicious bits (by extension, arbitrary user inputs); as a result, she can deceive the victim system into acting as she wishes. Such attacks on keyboards in critical infrastructures could cause catastrophic consequences, such as wrong prices in stock exchanges, incorrect traffic signals on railways, and overdoses of medicines for patients in hospitals. Similar attacks can also happen to many other applications that use the differential signaling technique.

It is essential to mention that in preliminary experiments, it is not difficult to use electromagnetic signal injection attacks (which use a similar attack setup in Section 5.5) to cause bit errors and stop data transmission of branded USB mice, USB keyboards, and routers. Such experimental results show that deployed attenuation methods, i.e., RF shielding materials and common-mode choke filtering, are not sufficient enough to block the electromagnetic signal injection attacks.

This pioneering work will systematically and experimentally investigate how and why the attacks can manipulate transmitted information in differential signaling. Despite various application scenarios or communication protocols, differential

signaling receivers work in largely the same way, which makes it possible to use a single attack strategy to attack a large number of well-known communication protocols and applications. The contributions of this chapter are as follows:

- It abstracts and parameterizes a generalized system model from practical differential signaling receivers so as to capture the characteristics of different systems by tuning the parameters; it also defines an adversary model, clarifying an attacker's objectives and capabilities, as well as her limitations in practice. (Section 5.1)
- It details how an electromagnetic signal injection attack can bypass the differential signaling technique and how it makes the receiving circuits detect intended bits. (Section 5.2)
- It also analyzes success rates of injections, and discusses critical factors that determine a high success rate. (Section 5.3).
- It further demonstrates the attacks on different chips to verify the attack principles (Section 5.4), and it successfully shows how to inject an arbitrary message into a CAN bus at a distance (Section 5.5).

5.1 System Model and Adversary Model

This section abstracts and parameterizes a system model of differential signaling from practical circuits. This model allows us to capture the characteristics of different circuits by tuning the parameters. Next, an adversary model is defined, which explains an attacker's capabilities and limitations. The system- and the adversary models together form a flexible tool to describe the attacks.

5.1.1 System Model

In differential signaling, information is transmitted in a pair of signals that are the same but have opposite polarities, and each in its own conductors/wires. These two wires are identical and put close to each other, e.g., twisted cables, on the one hand,

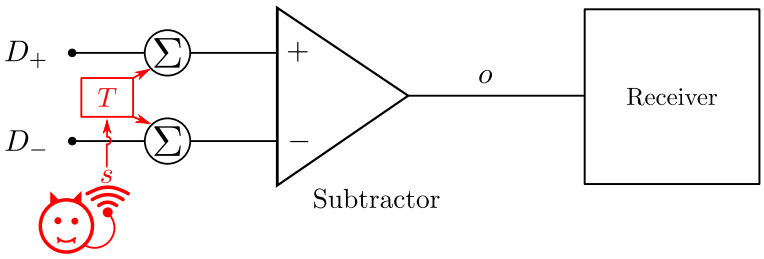


Figure 5.1: The system model consists of a pair of complementary signals (D_+ and D_-), a subtractor, and a receiver. The difference between the signals is extracted by the subtractor, and next, the difference is sent to the receiver. When an attack happens, the injected signals are superimposed with the complementary signals.

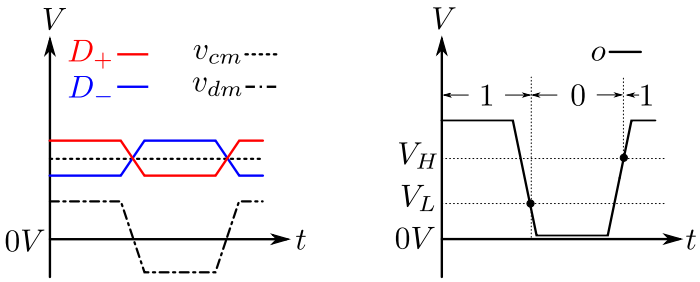
minimizing the electromagnetic radiation from the pair of wires, and on the other hand, making it resilient to electromagnetic interference (EMI) from outer sources to an extent [32, 83]. Because of such a setup, the external EMI will impact the wires equally. When the pair of signals are received, the information is extracted from the difference between these two signals. In this way, the equal impacts caused by the external EMI cancel out each other, leaving the intended information intact.

To obtain the difference, circuits that can "subtract" one signal from the other are used. Circuits with such a subtraction function are defined as a "Subtractor", and a block diagram of the subtractor is shown in Figure 5.1. The subtractor has two inputs, each receiving a signal of the differential pair. The subtractor calculates the difference and then sends it to the following circuits for further processing.

After receiving the subtractor output signal, an essential step is to convert its analog voltage levels into a sequence of bits. This is because the circuits that process information are usually digital (e.g., microprocessors) that only handle logic 1 and logic 0. After obtaining the bits, other functions or tasks, such as decoding, error checking, authentication, etc., can be further executed. A "Receiver" is defined, as presented in Figure 5.1, to incorporate all functions of the circuits that handle the subtractor's output signal.

Parameterizing Subtractor

The subtractor's two input signals are denoted as $D_+(t)$ and $D_-(t)$, and the output signal as $o(t)$. To simplify the notations, time t is omitted hereafter. In order to



(a) Subtractor **(b)** Receiver

Figure 5.2: (a) The subtractor's two input signals D_+ and D_- can be represented by their differential mode v_{dm} and common mode v_{cm}. (b) The receiver compares o with two thresholds to determine the logic levels.

explicitly show the information that is carried by the two input signals, these two input signals can also be rewritten by their differential mode and common mode. The differential mode is defined as the difference between two signals, and it is denoted as $v_{dm} = D_+ - D_-$, and note that v_{dm} represents the transmitted information; the common mode is defined as the average of two signals, and it is denoted as $v_{cm} = \frac{D_+ + D_-}{2}$. Such a relationship between the two input signals and their modes is illustrated in Figure 5.2a. Note that v_{cm} is a non-zero constant in almost all protocols, and thus, it is assumed that it is non-zero hereafter unless stated otherwise.

In practice, the subtractor is essentially a differential amplifier, and it is reasonable and prevalent to model its output signal as a sum of the amplified differential mode and the amplified common mode [84]. The gains for the differential mode and the common mode are denoted as G_{dm} and G_{cm}, respectively; note that the gains are functions of frequencies. The amplified terms are expressed as $G_{dm} \cdot v_{dm} + G_{cm} \cdot v_{cm}$.

In addition, there also exist distortion and noise that contaminate the output signal. The distortion originates from nonlinear properties of electronic components (e.g., transistors) that make up the subtractor [84], and a function $F(v_{dm}, v_{cm})$ is defined to incorporate the impacts of the distortion phenomenon on the input signals. The noise is modeled as additive Gaussian white noise, denoting it as n. Thus, the subtractor output is expressed as:

$$o = G_{dm} \cdot v_{dm} + G_{cm} \cdot v_{cm} + F(v_{dm}, v_{cm}) + n \tag{5.1}$$

In Equation 5.1, the first term $G_{dm} \cdot v_{dm}$ explicitly carries the transmitted information, i.e., v_{dm}. It needs to be emphasized that every subtractor has a finite operational band, within which it is designed to function properly. Inside this operational band, the differential-mode gain remains constant, and as such, the subtractor can guarantee a consistent output while handling input signals at different bit rates. The common-mode gain is so small that it makes typical attenuation of $70\,\text{dB} - 120\,\text{dB}$ to the common mode of the inputs [85], thus making $G_{cm} \cdot v_{dm}$ nearly zero. The distortion of the subtractor is also well maintained, and thus the impact of $F(v_{dm}, v_{cm})$ is negligible. Therefore, the subtractor is sufficiently good enough at rejecting the impacts of the common mode within the operational band. However, this no longer holds out of the operational band, and the reasons will be detailed in Section 5.2.

Parameterizing Receiver

The primary function of the receiver is to convert analog voltages into bits, as mentioned previously. It is a common way in practical circuits that the logic levels are determined by comparing analog voltage levels with two pre-determined thresholds [86]. The reason for using two thresholds instead of a single one is that the difference between two thresholds can prevent the noise from causing wrongly detected logic levels. These two thresholds are denoted as V_H and V_L, and $V_H > V_L$. The detection rule is straightforward: as depicted in Figure 5.2b, when $o \geq V_H$, a logic 1 will be detected; when $o \leq V_L$, a logic 0 will be detected. Specifically, when o is between these two thresholds (e.g., the noise causes the voltages to fluctuate into this region), the detected bit will retain its value. Note that the receiver detects bits periodically.

In addition, it is also essential to cover the circuits before the logic level detection, as an attacker needs to exploit these circuits to achieve a wrongly detected bit. The attack principles will be detailed in Section 5.2.2, but here, a model is abstracted, and the functions of the circuits are explained.

When a signal enters the receiver, it first goes through an electrostatic discharge (ESD) circuit, which is commonly used to protect the input pins of any electronic device from overvoltages. A block diagram of the ESD circuit is presented in Figure 5.3: it clamps the negative overvoltages to a minimum allowed voltage (e.g., GND), and the positive overvoltages to a maximum allowed voltage (e.g., V_{DD}) [87]. After that, a buffer circuit follows. It is used to get rid of an impedance mismatch between the previous stage and the receiver, and more precisely, it provides isolation and prevents undesired interaction from the previous stage [88]. The buffer circuit is essentially built up with transistors, which work like switches, and its function is abstracted in a way as illustrated in Figure 5.3: its input signal controls the switches, generating an output signal to reproduce its input signal. In this way, the buffer circuit transfers its input signal to the logic level detection. In all, the circuits before the logic level detection are modeled as a combination of the ESD circuit and the buffer circuit.

5.1.2 Adversary Model

An attacker's objective is to inject a message with a length of L bits into a victim system. The attacker has no physical access to the victim system, implying that she cannot modify its circuits, nor can she tap wires to inject attacking signals into it. Because of no physical access, it is rather difficult to know which bit is transmitted in the wires. However, the attacker's ability to guess the bit is not limited, and this will be further explained and discussed (regarding her guess ability) in detail in Section 5.3. The attacker can wirelessly inject the attacking signals into the victim system by radiating electromagnetic waves, and she can tune her attacking signals, regarding their frequencies, power, etc. The attacker knows the period that the receiver detects a bit. In each bit injection, the duration of the attacking signal is set the same as the period, and hence, the attack can always interfere with the receiver when it detects a bit.

As mentioned previously, an electromagnetic signal injection is a complicated process in practice. A transfer function T is defined to explain any changes to

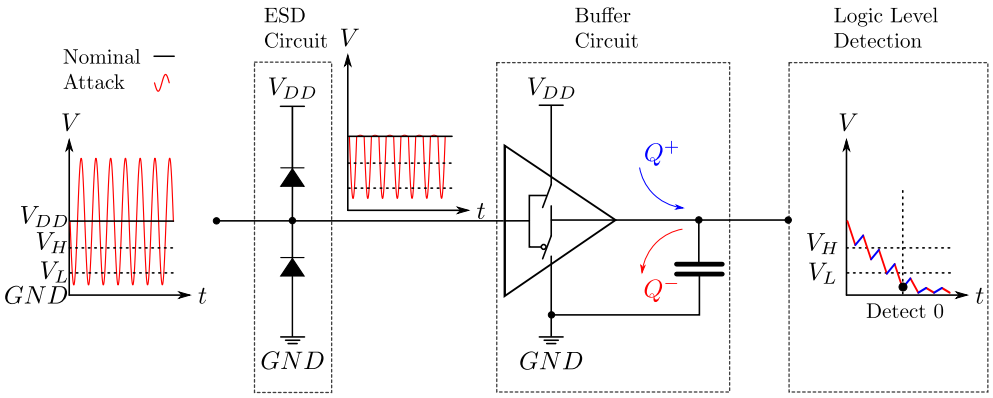


Figure 5.3: This diagram shows a model of circuits before logic level detection in the receiver. The charging and the discharging of the output parasitic capacitance are asymmetrical, and a net charge $Q^+ - Q^-$ causes the output voltage changes when an attack happens.

the attacking signal $s(t)$ due to the injection process, e.g., frequency selectivity, attenuation, spreading, etc. To simplify the notation, time t is omitted, and thus the injected signal is denoted as $T(s)$. Note that there could exist multiple injection places in the victim system. However, only when the injection impacts the signals that carry the information will the attacker be able to manipulate the bits, which are recognized by the receiver. Therefore, it is equivalent to modeling the pair of wires as the injection point, as shown in Figure 5.1. In fact, the differential signaling technique is usually deployed between two ends that are far from each other, and the pair of wires are effective antennas capturing the attacking signals in practice.

5.2 Bit Injection Attack

This section details how an attack can make the victim system detect an incorrect bit. Note that as mentioned in Chapter 2, since previous research has thoroughly studied the signal injection process, it is not further detailed here. However, this section focuses on explaining why injected signals can bypass the subtractor, and how the receiver responds to the bypassed injected signals.

5.2.1 Bypassing Subtractor

The injected signals are superimposed with the two input signals of the subtractor, leading to $D_+ + T(s)$ and $D_- + T(s)$. According to the definition of the differential mode and the common mode in Section 5.1.1, the differential mode will not be affected because the injected signals cancel each other out; however, an extra term $T(s)$ that is the average of the identical injected signals is added to the common mode. As explained previously, the common mode of the original input signals has almost no impact on the subtractor output signal. Thus, under the interference of the attack, it is equivalent to approximate the common mode to be $T(s)$, or $v_{cm} = T(s)$. Substitute it into Equation 5.1, and obtain an expression of a malicious subtractor output signal o' as follows:

$$o' = G_{dm} \cdot v_{dm} + G_{cm} \cdot T(s) + F(v_{dm}, T(s)) + n \tag{5.2}$$

Note that the terms $G_{cm} \cdot T(s) + F(v_{dm}, T(s))$ are malicious changes that are caused by the injected signal $T(s)$, and these terms explicitly represent the bypassed injected signal.

It is essential to point out that the subtractor's common-mode rejection ability is finite, and it is ascribed to two widely accepted reasons. First, the two inputs of the subtractor are not perfectly symmetrical in practice. This results in that common-mode variations in the inputs are converted to differential-mode variations in the output, which is also known as "common-mode to differential-mode conversion" [84, 89–92]. Second, nonlinearities of the subtractor lead to extra unexpected variations in the output [93, 94]. Especially beyond the operational band of the subtractor, on the one hand, the subtractor's common-mode rejection ability deteriorates dramatically because the common-mode to differential-mode conversion becomes more and more significant at higher frequencies [84], and this indicates that $G_{cm} \cdot T(s)$ becomes larger. On the other hand, the nonlinear phenomenon of the subtractor becomes stronger [94], leading to much more distortion, or larger $F(v_{dm}, T(s))$. The evidence above indicates that if the injected signals are out of the operational band, their impacts on the common mode are much more easily converted into

additional malicious voltage changes in the subtractor's output. In this way, the injected signals bypass the subtractor.

It is essential for the attacker to know the waveform of the bypassed injected signal so that she can have a controllable impact on the next stage, i.e., the receiver. However, since the subtractor is not initially designed for use beyond the operational band, it is not easy to precisely predict the waveform. To tackle this challenge, a determined attacker can get a replica of the subtractor and experimentally find the relationship between the output signal and the input signals of the subtractor, and as such, the attacker can estimate the bypassed injected signal. A demonstration is shown it in Section 5.4.2.

5.2.2 Bit Detected Incorrectly in Receiver

Recalling in Section 5.1.1, it explained that the receiver determines a bit by comparing the subtractor output signal with two thresholds. In order to make the receiver detect an incorrect bit, the bypassed injected signal must make the nominal voltage level (which is represented by $G_{dm} \cdot v_{dm}$) of the subtractor's output signal cross the threshold that determines the opposite bit. Since detecting an incorrect bit is literally flipping a bit, these two synonyms will be used interchangeably hereafter. To make the explanation concise, the nominal voltage is assumed to be at V_{DD}, which is above V_H, and it means that the receiver is supposed to receive 1 if no attack presents.

The oscillation of the bypassed injected signal either pushes the voltage level towards or away from V_L, as shown in Figure 5.3. However, only when the malicious voltage change causes the voltage level to move towards V_L will the receiver wrongly detect a bit. Luckily, the ESD circuit guarantees the direction of the malicious voltage change. This is because the positive part of the bypassed injected signal exceeds the maximum allowed voltage, leading to rectification by the ESD circuit, but the negative part remains, and thus the voltage level moves toward V_L, as shown in the middle of Figure 5.3,.

After being rectified, the malicious voltage change continues propagating through the buffer circuit of the receiver. Note that the frequency of such a malicious signal is further beyond the operational band of the receiver. The fast oscillation of the malicious voltage change will make the switches close and open periodically, thus charging and discharging the parasitic capacitance of its output periodically. However, the charging process and the discharging process are asymmetrical, leading to a quick accumulation of net charge across the output capacitance, and then, it holds still [95, 96]. If a bit is read when the voltage level crosses V_L, 0 is recognized.

In this way, the bypassed injected signal successfully makes the receiver detect 0. In a similar vein, when the nominal voltage level is below V_L (and the receiver expects 0), the bypassed injected signal also makes the receiver detect 1. Researchers pointed out that the impacts of the malicious voltage change are equivalent to adding a constant DC offset to the input signal of the receiver [95, 96]. The magnitude of this equivalent DC offset depends on the frequency and amplitude of the malicious voltage change, as well as the specific circuits that are impacted [95]. This implies that an attacker can successfully flip a bit by properly choosing the frequency and the power of her attacking signals. However, it is also not easy to predict the receiving circuits' responses out of their operational bands, and hence, it will be difficult to figure out a formula to calculate the effective frequency and power of the attacking signals. Still, a determined attacker can experimentally obtain such attacking signals by sweeping the frequency and power to find ranges of effective attacking signals. This will be demonstrated in Section 5.4.3.

5.3 Analysis of Success Rate

After knowing the principles of a bit injection attack, its success rate is analyzed. When an attacker intends to inject a bit, it is essential to consider which bit is being transmitted in the wire: if the transmitted bit is not what the attacker wants, she needs to emit an attacking signal so as to flip the bit; otherwise, she does not need to emit any attacking signal, leaving the bit unchanged. However, it is not an easy job to know which bit is transmitted in practice, but some methods will be

discussed in Section 5.6. Still, the attacker can make use of her knowledge about the victim system to make a guess of the bit, and her guess will further dictate her actions. To make a successful injection, on the one hand, it is essential to have a correct guess; on the other hand, the attacking signal can effectively flip a bit. In this section, first, the attacker's guess ability is parameterized, as well as the effectiveness of her attacking signals, and next, the success rate will be analyzed.

5.3.1 Parameterization

The bit that is transmitted in the wire is denoted as A, and the attacker's guess as G. A parameter g is used to quantify the attacker's knowledge about A, and $g \in [0, 1]$. Note that $g = \frac{1}{2}$ means the attacker knows nothing about the bit, and there is an equivalent chance that the attacker will guess 1 or 0. Furthermore, $g > \frac{1}{2}$ means the attacker knows information that indicates the bit could be 1. A larger g means that the attacker knows more information, and thus, it is more likely to guess 1; when $g = 1$, the attacker is sure that the bit is 1. Conversely, $g < \frac{1}{2}$ means the attacker knows information that indicates the bit could be 0, and a smaller g also implies knowing more information, and thus, it is more likely to guess 0; when $g = 0$, it means the attacker is sure that the bit is 0. G is thus modeled to follow a Bernoulli distribution with the parameter g, where G takes 1 with a probability of g and 0 with a probability of $1 - g$.

The performance of an attacking signal is quantified by two parameters: u represents the probability of flipping 1 to 0, and v represents the probability of flipping 0 to 1, and $u, v \in [0, 1]$. For a certain victim system, each attacking signal corresponds to a pair of u and v, and all u, v pairs together characterize this specific victim system's responses to attacks. *Feasible pairs* are defined to incorporate all these pairs. In practice, u and v can be measured experimentally, and the measurements and the characterization will be shown in Section 5.4.3. In addition, here are two special pairs that need to be paid attention to. The first pair is $u = 0$ and $v = 1$. Since $1 - u = 1$ means that a logic 1 is always kept unchanged and $v = 1$ means that a logic 0 will always be flipped successfully, the corresponding attacking

signal will force any bit to 1. Conversely, a pair of $u = 1$ and $v = 0$ corresponds to an attacking signal that can force any bit to 0. With these two ideal pairs, the attacker can inject any bit successfully without any guess all the time. Unfortunately, they are not always attainable in practice and this will be shown it in Section 5.4.3.

5.3.2 Success Rate of Bit Injection

Let's begin by considering that the attacker intends to inject a single 1. There are four combinations of A and G, and the attacker's actions and the success rate for each combination are as follows:

- If $A = 1$ and $G = 1$, the attacker makes a correct guess, and since she intends to inject 1, she will not radiate any attacking signal. The success rate is 1, which can be written as $A \cdot G$.

- If $A = 0$ and $G = 1$, the attacker wrongly thinks that the bit is already 1 so that she will not radiate any attacking signal, meaning that she will never flip the bit. Hence, the success rate is 0.

- If $A = 1$ and $G = 0$, the attacker wrongly thinks that the bit is 0 and she will radiate an attacking signal. However, the attacking signal needs to keep the bit unchanged such that it is still 1. Thus, the success rate is $1 - u$, which can be written as $A \cdot (1 - G) \cdot (1 - u)$.

- If $A = 0$ and $G = 0$, the attacker's guess is correct, and the attacker will radiate an attacking signal to flip the bit. The success rate of flipping 0 is v, which can also be written as $(1 - A) \cdot (1 - G) \cdot v$.

The success rate of injecting 1 is denoted as P_1, and it can be expressed as a combination of these cases:

$$P_1 = \begin{cases} G + (1 - G) \cdot (1 - u), & \text{if } A = 1 \\ (1 - G) \cdot v, & \text{if } A = 0 \end{cases}$$

Suppose the attacker intends to inject a single 0, a similar way can be used to reach an expression of the success rate of injecting 0, which is denoted as P_0 and expressed as:

$$P_0 = \begin{cases} G \cdot u, & \text{if } A = 1 \\ (1 - G) + G \cdot (1 - v), & \text{if } A = 0 \end{cases}$$

The injection of 1 will be focused on hereafter, as the injection of 0 is a symmetrical process and the explanation is similar.

Impact of g

To investigate the impact of g, let's begin from the expectation of P_1, which can be easily derived and expressed as:

$$E(P_1) = \begin{cases} u \cdot g + 1 - u, & \text{if } A = 1 \\ -v \cdot g + v, & \text{if } A = 0 \end{cases}$$

Essentially, the larger $E(P_1)$ is, the better. Since the impact of g is being discussed, it is reasonable to assume that u and v are non-zero here; otherwise, g vanishes in $E(P_1)$.

If $A = 1$, $E(P_1)$ increases with g. According to the definition of g, a bigger g means knowing more information about $A = 1$, and thus it is more possible to make a correct guess. The importance of making a correct guess can be easily proved: if $A = 1$, P_1 is maximized when $G = A$. Thus, it can be concluded that if $A = 1$, the larger g is, the more possible that P_1 will be maximized, and the better. Similarly, if $A = 0$, $E(P_1)$ increases while decreasing g, and a smaller g means a higher chance of making a correct guess, and thus more possible to maximize P_1.

Regarding P_0, the analysis is similar and it is not further detailed here. Therefore, to make a correct guess to maximize the success rate, two points need to pay attention to: first, it is crucial that g is in a manner conforming with A, and second, it is always better to know more information about A.

Impact of u and v

As indicated by the equation of P_1, the larger $1-u$ and v are, the better. However, it needs to be emphasized that in a specific system, u and v are related in a certain way, and an example is shown in Figure 5.11. From the experiments with different chips in Section 5.4.3, it is observed that there is a trade-off between increasing $1-u$ and increasing v. Then, here comes the question: Which is the optimal pair?

Determining the optimal pair can be formulated into a multi-objective optimization problem, where $1-u$ and v are the objectives. The most extensively used method of solving such an optimization problem is called the weighted sum method [97, 98], where the two objectives are combined and converted into one scalar, composite objective function by assigning proper weights to them; note that the sum of the weights equals 1. Regarding the weights, g is selected as the weight for $1-u$, and $1-g$ as the weight for v, and the reasons are as follows.

Firstly, if the attacker has no knowledge about A (where $g=\frac{1}{2}$), it is equivalently important to "keep 1 unchanged" and "flip 0". Hence, it requires that the weights are equal, and $g=1-g=\frac{1}{2}$ meets the requirement. Secondly, if the attacker knows information indicating that the bit is 1 (where $g>\frac{1}{2}$), "keeping 1 unchanged" is more important, and hence, more weight for $1-u$ than v. Moreover, when more information is known, the importance of $1-u$ further increases, and so does the weight. Since $g>1-g$ and knowing more information also means that g increases, g can properly quantify the weight of $1-u$, and accordingly, $1-g$ quantifies the weight of v. Thirdly, if the attacker knows information indicating that the bit is 0, it is not difficult to deduce that g as the weight for $1-u$ and $1-g$ as the weight for v in a similar way, and the reason is not further detailed. Therefore, searching for the optimal pair of u and v of injecting 1 is solving the following problem:

$$\begin{aligned} \max_{(u,v)} \quad & g\cdot(1-u)+(1-g)\cdot v \\ \text{s.t.} \quad & (u,v)\in \text{feasible pairs} \end{aligned}$$

In a similar vein, concerning injecting 0, the larger u and $1 - v$ are, the better. Finding the optimal u and v of injecting 0 is solving the following problem:

$$\begin{aligned} \max_{(u,v)} \quad & g \cdot u + (1 - g) \cdot (1 - v) \\ \text{s.t.} \quad & (u, v) \in \text{feasible pairs} \end{aligned}$$

In Section 5.4.3, it will demonstrate how to use the method above to find the optimal pairs and then verify that the optimal pairs will do better than other pairs. Note that since the attacker has no access to the victim system, when she is preparing attacking signals, she needs to conduct experiments on a replica and use the methods above to find the optimal pairs.

Measuring Susceptibility

Although the attacker cannot access the victim system, a system designer of the victim system can do so. Thus, she can measure and obtain the optimal pairs, and then, use them to estimate the success rate. Note that the success rate is also a metric that sufficiently quantifies the susceptibility of the victim system: a higher success rate means that the victim system is more susceptible to an injection; conversely, a lower success rate means less susceptible. Thus, the system designers can use this analysis to quantitatively evaluate the security of their systems, and are able to change components or data modulation schemes to reduce adversarial success.

5.3.3 Success Rate of Message Injection

Recall that the attacker's objective is injecting L bits into the victim system, and the success rate of injecting a message will decrease exponentially with the message length. However, it needs to be pointed out Section 5.2.2 explained that an attack can cause voltage changes to accumulate quickly and then holds still. Therefore, suppose the attacker will perform identical attacks (i.e., the same attacking signal, the same intended injected bit) on a sequence of transmitted bits that are consecutive and identical, once the first bit injection is successful, the success rates of the following bit injections will increase. This is because the first

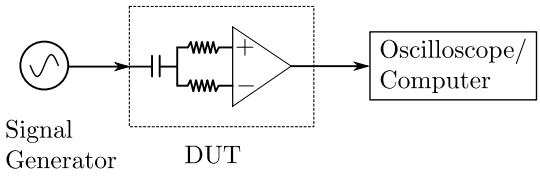


Figure 5.4: A testbed consists of a signal generator, a device under test (DUT), and an oscilloscope/computer.

successful bit injection attack gets rid of many uncertainties with respect to the guess, the effectiveness of the attacking signal, timing, etc. It is reasonable to approximate the success rate after the first successful bit injection to be 1 until the end of the consecutive injections. Such an approximation may overestimate the success rate of a message injection, as unpredictable responses in the victim system may still lead to a failure of a bit injection. An example of using such a method to estimate the success rate of a message injection will be shown in Section 5.5. It needs to be emphasized that system designers would rather overestimate the success rate than underestimate it because when they deploy measures to improve the security the nominal protection will make the victim system less susceptible in practice.

5.4 Experiments

Recalling in Section 5.2, it explains the principles of injecting bits. This section will demonstrate that injected signals can easily bypass the subtractor, and that the bypassed injected signals cause incorrectly detected bits in the receiver.

5.4.1 Testbed

To test different chips of the subtractor and the receiver, a testbed is built before the experiments. The testbed's functions are generating attacking signals and measuring responses of a device under test (DUT).

A setup of the testbed is shown in Figure 5.4. A signal generator produces an attacking signal and injects it into the DUT through a wire. Such a signal injection setup is also known as Direct Power Injection (DPI) methodology [59], and it is used because the injected frequency and power can be precisely controlled so that

the responses of the chips can be measured reliably. Moreover, an oscilloscope is used to capture and measure the waveforms of the injected signals and the DUT's output signals, and a computer is used to process and analyze the measurements.

5.4.2 Subtractor

Five different off-the-shelf subtractor chips are chosen, which are TJA1050, MCP2551, SN65HVD230, MX485, and SN751768P. They support CAN bus or RS485/422, and they are widely used in many critical applications such as automobiles, medical equipment, and industrial devices. The subtractor chip is configured in a way as shown in the DUT block in Figure 5.4: two same resistors are added to the input of the subtractor, and these two transistors are equivalent to the terminated resistors in practice that are required by the differential signaling standards. Note that the voltage difference between the two inputs is internally configured to keep unchanged.

Regarding the injected signal, it is coupled to the midpoint between the two resistors by a capacitor. Doing so is equivalent to injecting a common-mode interference into the subtractor. The injected signal is sinusoidal, and its frequency is swept from 10 kHz to 100 MHz, and its peak-to-peak voltage of the injected signal is set to be 1 V, 2 V, and 4 V. Note that other waveforms such as square and sawtooth are potentially effective, but due to the limit of the signal generator, it cannot produce high-frequency and powerful signals with these special waveforms, so sinusoidal signals are used in the experiments.

Impacts of Injected Frequency and Power

As explained in Section 5.1.1, when no attack happens, the subtractor's output signal remains consistent with the differential mode of its input signals. In the configuration above, since the voltage difference between the two inputs is constant, the subtractor's output signal is also constant. Note that the noise essentially exists, but it is too small to significantly disturb the output signal. When an injected signal applies, the subtractor's output signal will start oscillating, and such an oscillation represents the bypassed injected signal. Note that the bypassed

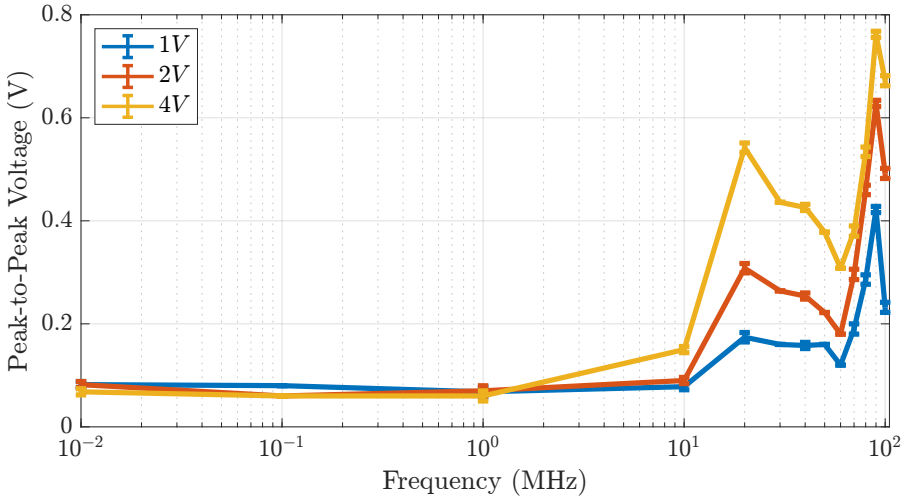


Figure 5.5: Test a subtractor chip TJA1050 with injected signals with frequencies ranging from 10 kHz to 100 MHz and peak-to-peak voltages from 1 V to 4 V. The y-axis represents the output of the subtractor.

injected signal is explained by $G_{cm} \cdot T(s) + F(v_{dm}, T(s))$ in Equation 5.2. Therefore, the peak-to-peak voltage of the subtractor's output signal is used to quantify the strength of the bypassed injected signal.

Taking a subtractor chip TJA1050 as an example, when there is no attack, the peak-to-peak voltage of the output signal is 0.06 V, which reflects the noise level. When an attacking signal is injected into the chip, the averaged peak-to-peak voltage and its standard deviation are shown in Figure 5.5. Between 10 kHz and 1 MHz, the output is as close as the noise level. This is because the common-mode interference is well handled within the operational band. However, when the frequency is increased above 10 MHz, the peak-to-peak voltage has an increasing trend along with the frequency. These results explicitly show that the subtractor's common-mode rejection ability deteriorates out of the operational band. In addition, two local maximums appear at 20 MHz and 90 MHz, as shown in Figure 5.5. This means that the injected signals at these two frequencies bypass this subtractor chip more efficiently than other frequencies. From the perspective of an attacker, she can take advantage of properly choosing the injected frequency to achieve a bypass using less attack power.

While increasing the injected power, the peak-to-peak voltage of the output

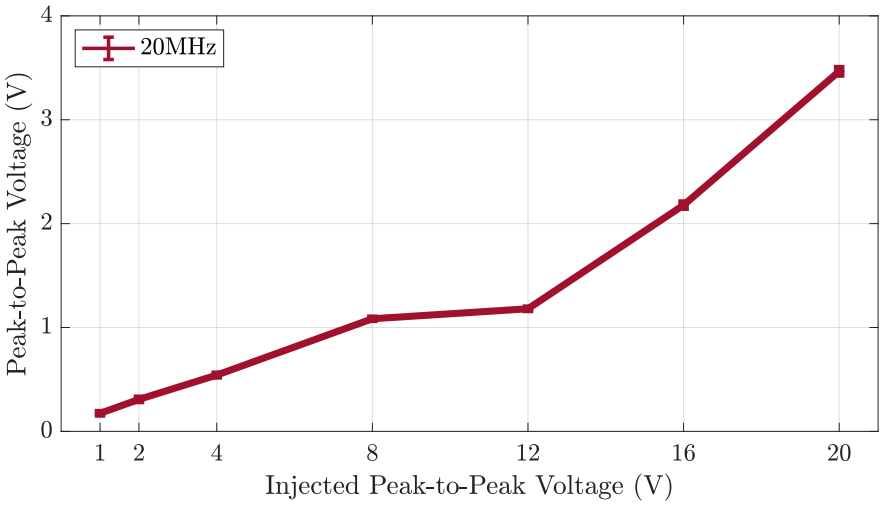


Figure 5.6: The strength of the bypassed injected signal increases while the injected power.

also increases, implying a stronger bypassed injected signal. However, as shown in Figure 5.5, with the injected power of 4 V, the highest peak-to-peak voltage of the bypassed injected signal is still below 1 V. To achieve a stronger bypassed injected signal in the subtractor output, an RF power amplifier is used to increase the injected signal up to 20 V at 20 MHz, which is an efficient frequency that the subtractor lets the injected signal bypass. The output of the subtractor is shown in Figure 5.6. The results indicate that with increasing the injected power, the strength of the bypassed injected signal also increases. Also, it can be observed that the strength of the bypassed injected signal is roughly proportional to the injected power, and this allows the attacker to estimate the strength of the bypassed injected signal.

Note that such a bypassing phenomenon does not only occur in the TJA1050 chip but also in many other subtractor chips. For example, in the other four chips, it is observed that the injected signal can always bypass them when the frequency is increased out of their operational bands; also, they all show that the higher the injected frequency/power is, the stronger the bypassed injected signal is.

Impacts of Noise and Distortion

As indicated by Equation 5.2, the distortion $F(v_{dm}, T(s))$ plus the noise n will make the bypassed injected signal differs from the injected signal regarding waveforms.

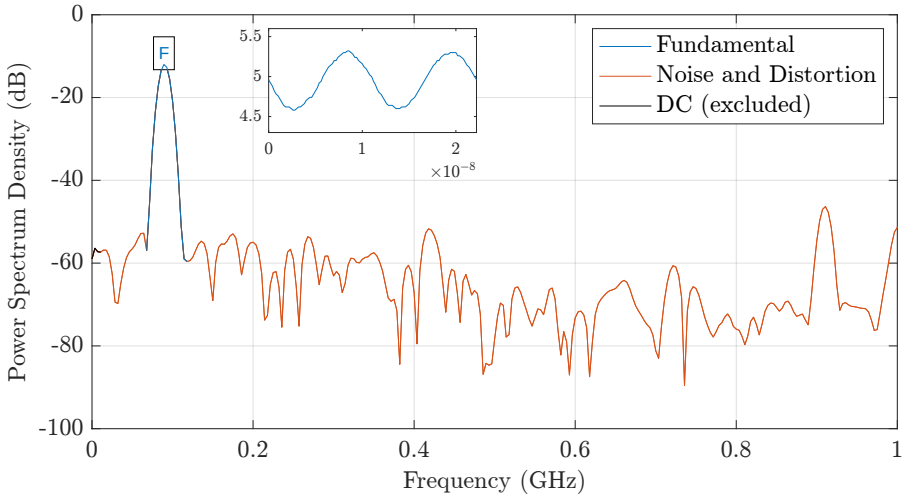


Figure 5.7: Set the injected frequency to be 90 MHz and the injected power to be 4 V. The frequency domain of the bypassed injected signal is presented. A time domain screenshot of the bypassed signal is in the floating window, where the x-axis is time (s) and the y-axis is voltage (V).

It is essential to know how much the bypassed injected signal is distorted because the bypassed injected signal will act on the receiver straight away and its waveform determines how the receiver responds.

To measure the impacts, a signal to noise-and-distortion (*SINAD*) ratio is used as a metric, which is calculated by the following equation:

$$SINAD = 10 \times \log_{10} \frac{P_s}{P_{n+d}} \tag{5.3}$$

where P_s is the power of the fundamental frequency of the signal, and P_{n+d} is the power of noise and distortion. The *SINAD* ratio is a widely used measure that quantifies the quality of a signal that is particularly degraded by the noise and distortion [99, 100]. The higher the *SINAD* ratio of a signal is, the better the signal quality is, and hence, less distortion in the signal. In Figure 5.7, a frequency domain of a bypassed injected signal is presented to show the difference between the fundamental frequency and the noise plus the distortion, and the *SINAD* is around 27 dB. In this figure, the distortion exists in the bypassed injected signal as harmonics, but they are too small to distort the bypassed injected signal significantly, which can also be observed from the time domain of the signal (please refer to a floating window at the top-left corner in the figure).

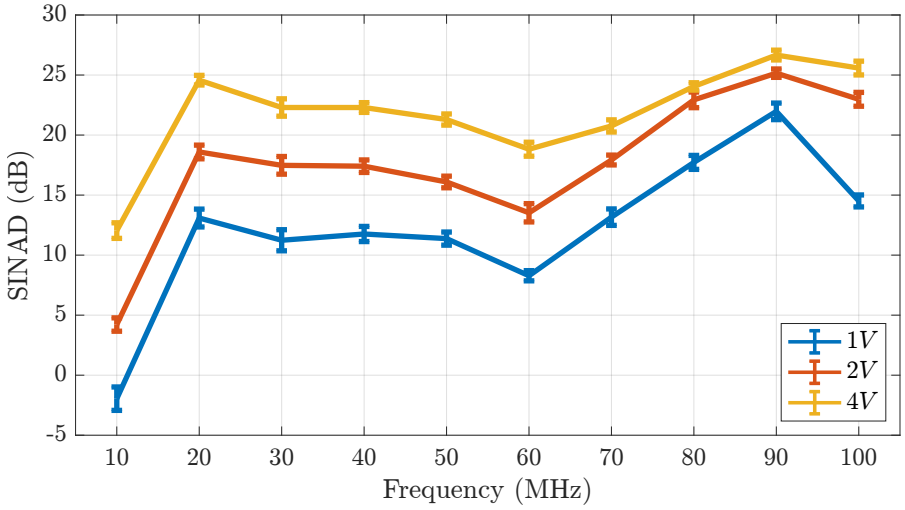


Figure 5.8: The signal to noise-and-distortion (*SINAD*) ratio quantifies the signal quality of the bypassed injected signal. Specifically, it measures how much the fundamental frequency component of the bypassed injected signal is stronger than the noise plus distortion. The larger the ratio, the less the signal is distorted.

Taking the TJA1050 chip as an example, in Figure 5.8, it shows the *SINAD* ratio when different injected signals apply. The ratio is low when the injected frequency and the injected power are small (e.g., 10 MHz and 1 V), and this is because only a tiny amount of injected signal can bypass the subtractor, as explained previously. While either increasing the injected power or the frequency, the ratio has an increasing trend. In addition, the *SINAD* ratio is at least 10 dB for most of the measurements. Such a result implies that this chip demonstrates weak distortion and noise, which do not need to be worried too much while modeling its output signal. However, it does not mean that every subtractor chip has such weak distortion and noise, and an attacker still needs to handle them carefully.

5.4.3 Receiver

In various systems, microcontrollers are usually the devices that realize the receiver functions: they detect the logical level of the input signals and then execute specific tasks according to the received information. Three different microcontroller chips are selected to test, which are nRF52833, ATMEGA328P, and ATSAM3X8E. The testbed that is shown in Figure 5.4 is used to study how the injected signal impacts bits that are recognized by the receiver chips. However, there are small modifications

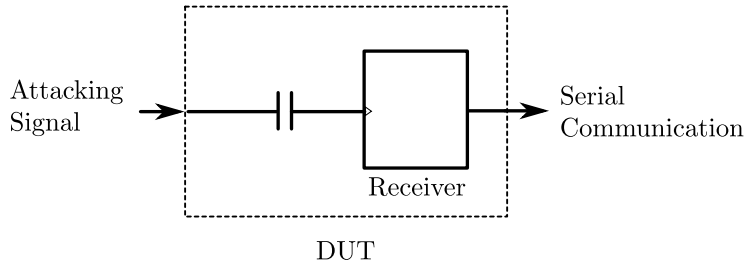


Figure 5.9: Change the DUT for a receiver: the DC level of the digital input pin is internally fixed at a certain level (V_{DD} or GND), and the receiver sends measurements to a computer by serial communication.

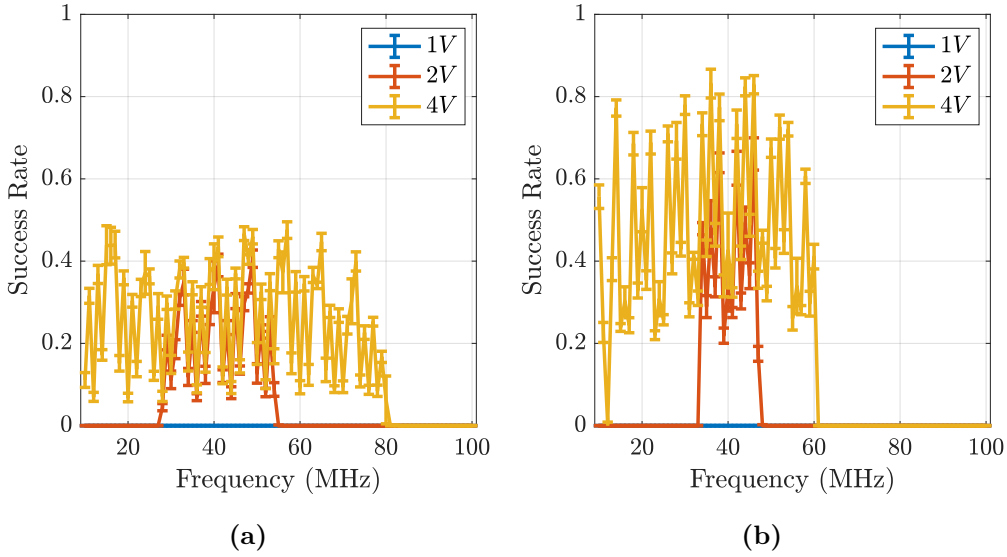


Figure 5.10: Success rates of bit injections in nRF52833. (a) Flip bits from 1 to 0. (b) Flip bits from 0 to 1.

in the DUT block, and they are shown in Figure 5.9. First, the input is changed to single-ended. Second, internally fix the input voltage level at a high (or low) voltage level, which corresponds to logic 1 (or 0). Third, because these chips support serial communication with the computer, the DUT directly sends recognized bits to the computer through a serial communication line.

The injected frequency is set from 10 MHz to 100 MHz with a step of 1 MHz. Note that since the subtractor chips have demonstrated that they can well remove the common-mode interference below 10 MHz, that frequency range is not further tested. The peak-to-peak voltage of the injected signal is set to be 1 V, 2 V, and 4 V. For each combination of the injected power and the injected frequency, 10

measurements are recorded; in each measurement, 256 bits are collected by the chip, and calculate the percentage of successfully flipped bits as the success rate; then, the mean and the standard deviation of the success rates are calculated and presented.

Taking an nRF52833 chip that works at $V_{DD} = 3\,\mathrm{V}$ as an example, it has $V_H = 2.1\,\mathrm{V}$ and $V_L = 0.9\,\mathrm{V}$ according to its datasheet [101]. Recalling in Section 5.1.1, V_H and V_L are two thresholds that are used to determine logic levels. To flip 1, the voltage change needs to be at least $3\,\mathrm{V} - 0.9\,\mathrm{V} = 2.1\,\mathrm{V}$; conversely, to flip 0, it needs to be at least $2.1\,\mathrm{V} - 0\,\mathrm{V} = 2.1\,\mathrm{V}$. The experimental results of flipping 1 are shown in Figure 5.10a, and the results of flipping 0 are shown in Figure 5.10b.

When the injected signal is 1 V, no bit flip is observed. This is because the injected signal is too weak to cause the voltage change beyond the threshold. When the injected signal is increased above 2 V, bit flips happen. Although the injected signal of 2 V is still weaker than the required threshold of 2.1 V, recall that as explained in Section 5.2.2 the voltage change can accumulate quickly and lead to a voltage change over the threshold ultimately, and consequently, the bit flips happen. When the injected power is increased to 4 V, the success rate becomes higher. Also, the frequency range where bit flip happens widens when the injected signal becomes much stronger. The results also imply that this chip is more susceptible in a frequency range that is centered at 40 MHz, and it is relatively easier to cause bit injections in this frequency range with less attack power.

In the other two chips, it is also observed that the success rates of bit injections are related to both the power and the frequency of the injected signal. The results show that the higher the power is, the higher the success rate is, and the wider the frequency range in which bit flips happen. Note that regarding the chip nRF52833 in Figure 5.10, the success rates show a periodic pattern in terms of the injected frequency: a peak appears every 2 MHz. Such a repeated pattern has nothing to do with the testing circuits outside the chip because the periodic pattern is not observed in other chips. It is speculated that some deterministic properties of the nRF52833 chip lead to this periodic pattern. However, it is trivial to figure out what

$$\max_{u,v} \quad g \cdot (1-u) + (1-g) \cdot v$$

g	Optimal (u, v)
$0.00 - 0.88$	$(0.09, 0.83)$
$0.89 - 0.90$	$(0.07, 0.69)$
$0.91 - 1.00$	send nothing

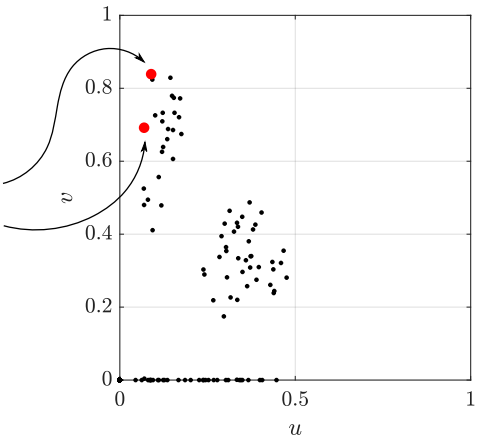


Figure 5.11: The pairs of u and v characterize the chip's responses to the attacks. Regarding injecting 1, the optimal pair with respect to g can be decided by solving the optimization problem.

these deterministic properties are because this periodic pattern only exists in this chip, and knowing the deterministic properties does not help attacks on other chips.

Characterizing Receiver's Response

Recalling in Section 5.3, two parameters u and v are used to characterize a victim device's responses to attacks. It is not difficult to find that the success rate of flipping 1 is u (see Figure 5.10a), and the success rate of flipping 0 is v (see Figure 5.10b). Thus, the feasible pairs of u and v are obtained, and they are plotted in Figure 5.11, which visualizes the chip's (nRF52833) responses to the attacks.

As mentioned previously $u = 0$ and $v = 1$ are an ideal pair, which represents an attacking signal that forces any bit to 1. The closer a pair is to it, the easier the injection of 1 will be. Similarly, $u = 1$ and $v = 0$ is the other ideal pair, which represents an attacking signal that forces any bit to 0. As shown in Figure 5.11, the feasible pairs' distribution is skewed to $u = 0$ and $v = 1$, meaning that it is much easier to inject 1 than 0 into this chip. Since injecting 1 and injecting 0 are symmetrical processes and the analysis will be similar, the following parts focus on injecting 1.

Recalling in Section 5.3.2, the method of determining the optimal pair of u and v is formulated. To determine the optimal pair regarding injecting 1, it is assumed

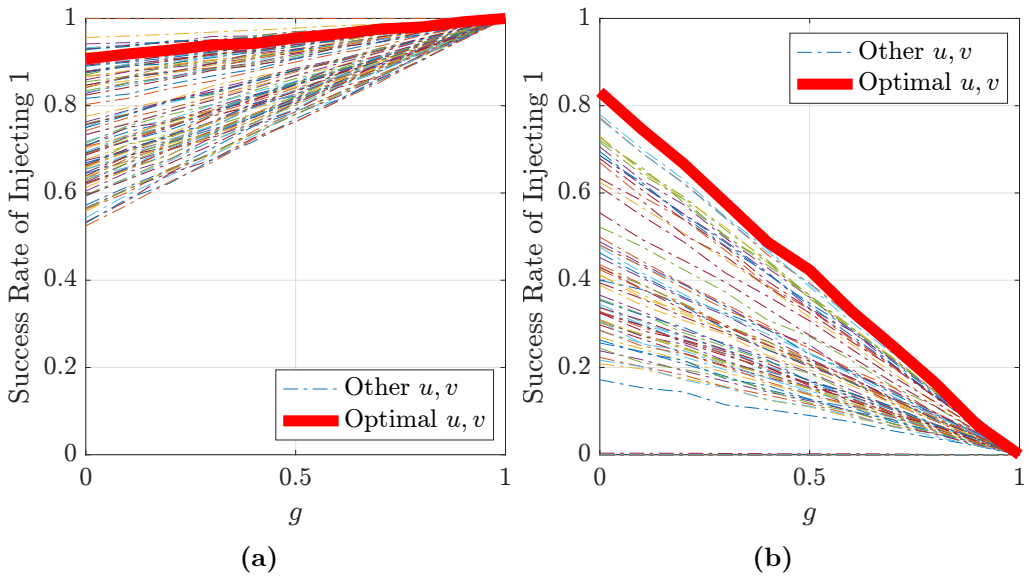


Figure 5.12: (a) If $A = 1$, the success rate of injecting 1 with using different pairs of u, v. (b) If $A = 0$, the success rate of injecting 1 with using different pairs of u, v.

that g is always correct, and the results are presented in Figure 5.11. When g is below 0.88, the optimal pair is $u = 0.09$ and $v = 0.83$. Such an attacking signal can successfully flip 0 with a probability of 0.83 and keep 1 unchanged with a probability of $1 - 0.09 = 0.91$. With further increasing g, as explained in Section 5.3.2, the attacker is becoming more and more sure that the bit is 1, and hence, u decreases to 0.07. When g is greater than 0.9, the solution indicates that the attacker will send nothing. Next, attacks are simulated to show how the optimal pair outperforms.

Simulation and Success Rate

First, attacks with the optimal pair are simulated. The transmitted bit A is set to either 1 or 0, and g ranges from 0 to 1. The simulated success rates of each g are averaged, and the results are presented in Figure 5.12. In Figure 5.12a, when $A = 1$, the success rate increases with g; in Figure 5.12b, when $A = 0$, the success rate decreases with g. The simulation results match with the model of $E(P_1)$ in Section 5.3.2, and in addition, the importance of having a g that is in a manner conforming with A is also explicitly shown.

Next, the simulation is repeated with other pairs of u and v, and compare them with the optimal pair, and the results are shown in Figure 5.12. In Figure 5.12a,

when $A = 1$, some pairs outperform the optimal pairs, but these pairs are those that have small u and small v: they are good at keeping 1 unchanged, but they cannot flip 0 effectively. Therefore, as shown in Figure 5.12b, when $A = 0$, the optimal pair outstrips others.

To decide whether the optimal pair outperforms any other pair significantly, multiple t-tests are conducted. Since the success rate has a linear relationship with g as shown in both Figure 5.12a and Figure 5.12b, the averaged success rate at $g = \frac{1}{2}$ is used as a metric to represent the attack performance. Note that the simulation is repeated 100 times for each pair, thus 100 samples for each pair. Next, t-tests are conducted to test against the alternative hypothesis that the optimal pair has a higher averaged success rate, or namely, outperforms the other pair. The significance level is set to 0.05, which is conventionally accepted as the threshold. These tests show that they reject the null hypothesis, except the pair of $u = 0.092$ and $v = 0.82$. It is not surprising because it is the pair that is close to the optimal pair of $u = 0.09$ and $v = 0.83$, as shown in Figure 5.11.

5.5 Message Injection into CAN

A Controller Area Network (CAN) is a protocol that is devised to allow many devices to communicate with each other on a two-wire bus, and it is now deployed in many different applications from medical instruments to automotive. The CAN is a broadcast type of bus, and any device, also known as a node, can freely send/receive data. This feature makes it possible for an attacker to broadcast whatever she wants on a CAN bus. In this section, it first briefly introduces the basics of the CAN, and then it demonstrates how to inject an arbitrary message into the CAN.

5.5.1 CAN Basics

A basic structure of the CAN is shown in Figure 5.13. In a node, a transceiver is an interface between the wires and the microcontroller, and its function is to convert the differential signals into a single signal that the microcontroller can use while receiving data, or the other way around while transmitting data. The

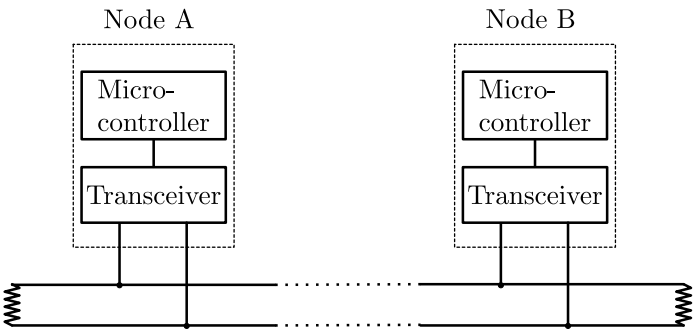


Figure 5.13: In the CAN system, nodes are connected to the same bus, where two wires are terminated by resistors.

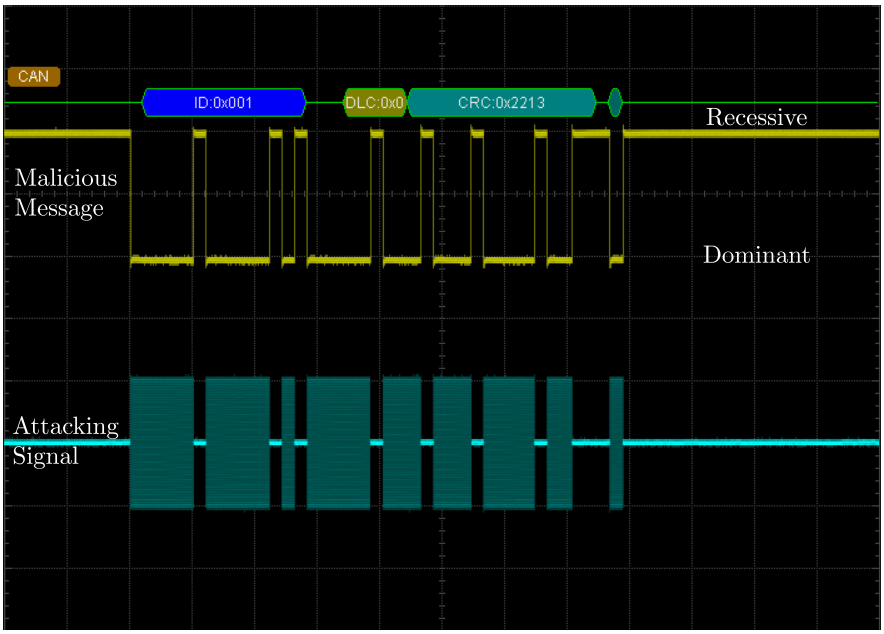


Figure 5.14: The attacker wants to inject a malicious message, and she generates an attacking signal according to the malicious message.

microcontroller handles signals on a software level, including identifying the type of the data, error checks, bus arbitration, etc.

On the physical level, when the voltage levels of the differential signals are the same, a recessive state (1) is defined; otherwise, a dominant state (0). Note that when no message is broadcast, the CAN system always remains at the recessive state.

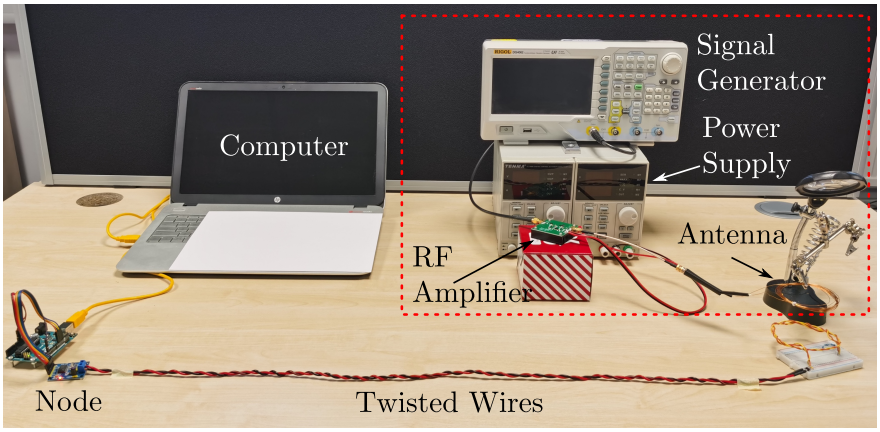


Figure 5.15: A practical setup of message injection attack on a CAN bus. The devices in the red rectangle form an attacker's setup

5.5.2 Message Injection

It is not difficult to find that such a CAN system matches the system model: the two wires in the CAN system correspond to the two input wires in the system model; the transceiver is the subtractor; the microcontroller is the receiver. Thus, it is possible for an attacker to use the bit injection attack to inject arbitrary messages into the CAN system. The attack is detailed as follows.

Assume that the attacker has $g = 1$, i.e., she knows the line is always at a recessive state. With an arbitrary message that the attacker wants to inject, the first step is to convert it into a sequence of bits according to the rules of the CAN protocol. Based on the bits, an attacking signal is generated. For example, the attacker wants to inject a malicious message that is shown in Figure 5.14, which contains an identity (ID) field with a value of 0x001, a data length (DLC) field with a value of 0x0, and a cyclical redundancy check (CRC) field with a value of 0x2213. Note that this malicious message is just an example, and the attacker can craft any valid message as she wishes. Since the line is always at a recessive state, the attacker only needs to radiate electromagnetic interference when dominant bits need to be injected. Therefore, the attacker can craft an attacking signal as shown in Figure 5.14, where the electromagnetic interference corresponds to all 29 dominant bits in this malicious message. When such an attacking signal is injected into the wires, it first bypasses the transceiver, and then the bypassed

injected signal further forces the microcontroller to receive dominant bits. The microcontroller will check the received message; if no error is detected, it will ultimately recognize the malicious message.

Commercially off-the-shelf electronic devices are used to build a CAN bus system. As shown in Figure 5.15, a node is connected to one end of two twisted wires. In the node, a TJA1050 is used as the transceiver, and an ATMEGA328P that is integrated with an MCP2515 CAN controller is used as the microcontroller. This node is programmed to always listen to the wires. Moreover, the node is connected to a computer through serial communications so that the received message can be recorded and shown on the computer. As for the attacking signal, a signal generator is connected to an RF power amplifier, and the amplified signal is radiated by a coil antenna. In order to inject the attacking signal into the wires effectively and efficiently, the coil antenna is put at around 5 cm above the wires. Note that this is limited by both local RF equipment regulations and the gain of RF amplifier, but a determined attacker will not be regulated by laws, and she can also increase her attack power by extra cost, thus conducting the attack at a farther distance. The frequency of the electromagnetic waves is set to be 22 MHz and the amplitude to be 20 V, which has the highest u that is around 0.74 according to preliminary experiments before the attack. Then, the message injection attack is conducted, and consequently, 3 malicious messages will be successfully recognized every 1000 attacks in 2 seconds, and the success rate is 0.003.

Such a success rate matches expectations. Since there are 29 bits to be injected in this message injection, if each bit injection is regarded as independent, the expected success rate will be $0.74^{29} \approx 0.0002$. However, as explained in Section 5.3.3, once the first bit injection of several consecutive injections is successful, the success rate for the following injections will be higher, and the success rate can be approximated to 1. As shown in Figure 5.14, to inject this message, 9 groups of consecutive flips from 1 to 0 are needed, and hence, the expected success rate is around $0.74^9 \approx 0.06$. A practical result should lie between 0.0002 and 0.06, and thus, it is reasonable to obtain a success rate of 0.003 in practice. Note that this success rate is for injecting

a complete well-formed CAN message. It might not be high, but it still allows an attacker to inject a full message every 2 seconds. As discussed in Section 5.6.2, the experiments were conducted at a fairly low power level so as to not radiate too much power on unlicensed frequencies. With that in mind, it is a pretty good result.

It is essential to emphasize that the setup uses commercial receivers and hardware that are similar to those in automobiles, and the experimental results sufficiently reflect the feasibility of the signal injection attacks on CAN bus. Moreover, provided that it would be hard to guarantee the safety of other drivers if we did the experiment at high power outside, we chose to let the laboratory experiments be enough.

5.6 Discussion

In this section, it discusses methods of gaining knowledge of transmitted bits, power restrictions in experiments, and future countermeasures.

5.6.1 Gaining Knowledge

Knowing transmitted bits indeed gives an attacker advantage in achieving high success rates of injections. There are multiple methods of obtaining information about the bits. For example, a recent work [102] showed that whatever the state of a CAN bus is, causing two bit errors that are separated by a fixed number of clock cycles can force the bus into an idle state. In addition to actively interfering with the victim system, the attacker can also use existing information about the victim system to figure out what is transmitted. For example, a preamble of a packet is usually predetermined and published in the protocol, and it is relatively easy to know the transmitted bits in the preamble. Despite the fact that the payload or checksum may be hard to guess, the attacker can also use a magnetic field probe to listen for electromagnetic leakage from the wires, and the attacker can obtain the bits by analyzing and processing the leakage, which essentially carries information about the bits [32].

5.6.2 Restricted Attack Power and Distance

A well-known trade-off between distance and power (free space path loss) indicates that a close attacker with low power is equivalent to a faraway attacker with high power. Since it is limited legally in how much power that can be emitted on frequencies within which the author does not have a license, the author has chosen only to stay close and go as high as necessary to show that our method works. Indeed, more power will extend attack distance, but this chapter focuses on the novelty of injecting into differential pairs rather than characterize the distance/power relationship.

5.6.3 Future Countermeasures

Although differential signaling performs well while rejecting electromagnetic noise, this work shows that it is incapable of preventing electromagnetic signal injection attacks. Countermeasures are suggested from two aspects: detection and mitigation. Detection allows a victim device to notice attacks, and mitigation aims to mitigate adverse impacts caused by the attacks. Despite the fact that abundant detection/mitigation methods have been proposed, they are designed for specific applications or they need to fit certain system models [14–16, 19, 20, 34, 35, 49, 59, 61, 65, 80, 103–108]. Similar ideas may apply to differential signaling, but millions of devices are not yet protected, and electromagnetic signal injection attacks still pose threats to them. Hence, future work is encouraged to fill this critical gap. Note that deploying countermeasures requires extra software/hardware, which will challenge many practical constraints of systems, such as budget caps, size/weight requirements, and computational resources. It is essential for system designers to weigh up between countermeasures' performance and the constraints/costs.

5.7 Summary

Despite the fact that differential signaling was proposed for making communication cables more immune to external interference, in this work, it shows that

electromagnetic signal injection attacks can inject arbitrary information into a differential signaling system. Because of the input asymmetry and nonlinearities of the subtractor, the rejection ability of the differential signaling technique is not sufficiently good for high-frequency signals to prevent attackers from successfully injecting adversarial signals. Moreover, in the receiver, the ESD circuit's rectification plus the buffer circuit's net charge accumulation results in high-frequency signals ultimately being incorrectly detected as either 0's or 1's depending on the frequency. The experiments have demonstrated the attack principles, and how to properly choose the frequency and the power of the attacking signal, in order to achieve successful injection. It analyzes the success rate of injection of more complicated bitstrings, taking into account any knowledge that the attacker might have about the existing data transmissions in the cable. It shows how this knowledge and the choice of attacking signals will affect the success rate. This analysis can also be used defensively by system designers who want to evaluate the security of their own systems, and are able to change the components and data modulation scheme to minimize adversarial success. Finally, this chapter demonstrates arbitrary message injection into a CAN bus, allowing an attacker to dictate the actions of the victim system.

6
Related Work

Contents

6.1 Detection 114
6.1.1 Methods for Sensor Systems 114
6.1.2 Methods for Actuator Systems 115
6.1.3 Other Anomaly Detectors 117
6.2 Attenuation 117
6.2.1 Shielding 117
6.2.2 Miniaturization 118
6.2.3 Filtering and Mitigation 118
6.2.4 Robust Hardware 119

This chapter will discuss related work on how to protect embedded systems from electromagnetic signal injection attacks, and it will focus on two aspects: detection and attenuation. Detection helps embedded systems notice attacks so as to take further measures to mitigate adverse impacts caused by the attacks. Attenuation aims to reduce the power of injected signals to a low level at which the injected signals hardly cause practical impacts to circuits. Please note that detection methods and attenuation methods can be applied together, making the victim system more immune to the attacks than only applying one of them.

6.1 Detection

Recall that Chapter 3 proposes a generalized detection method for sensor systems, and Chapter 4 for actuator systems. In addition, many studies also proposed various detection methods for these two systems as follows.

6.1.1 Methods for Sensor Systems

Detection essentially requires the capture of the attacking signals first. The intuitive idea is to add a specific channel to monitor the attacking signals. Kune et al. [14] investigated using extra antennas or conductors to capture and measure attacking signals. If the measured power is beyond a pre-determined threshold, an attack is confirmed, and the measurements can be then used by their adaptive filtering mechanism to mitigate the attacks, as mentioned previously. Tu et al. [106] proposed adding a dummy sensor for detection and correction. The dummy sensor "shares the same vulnerabilities with the (normal) sensor but is not sensing the legitimate signal". When an attack happens, the dummy sensor's readings can indicate the attack. Further, the readings can also be used to mitigate the attack's impact in normal sensor measurements. It can be noticed that these detection methods require redundant channels to handle the attacking signals. However, it is not an easy job to craft such channels. For example, as per Tu et al. [106], the extra channel needs to respond to the attacking signal in the same way as the original channel, and tiny *mismatch* can weaken the security; indeed, it will cause the extra cost to produce two identical channels, increasing the complexity of manufacture and expense of deployment. Regarding the detection method for the sensor systems in Chapter 3, it does not add an extra channel to monitor the attacking signals, meaning less complexity and more lightweight. Also, the detection method is essentially designed regardless of the types of the sensors, making it more flexible to be deployed.

In other specific devices that interact with users closely, it is possible to utilize users' reactions or behaviors while using these devices to identify the existence of attacks on the sensors. In a cardiac implantable electrical device (CIED), Kune et al. [14] proposed to immediately send a pace pulse to cardiac tissue just after

the cardiac tissue contracts. Since the cardiac tissue already contracts, the pace pulse won't cause any contract. Based on this fact, if the sensors of CIED find a contract after sending a pace pulse, it indicates that there exists an attack. However, since the extra pace pulses are sent to the heart tissues, more health care studies are needed to figure out whether this method is potentially harmful or not. For electromagnetic signal injection attacks on smartphone screens, Wang et al., [16] proposed a detection algorithm that utilizes the touching interval between pressing and lifting fingers to identify whether an attack exists. Since these detection methods count on the interactions between systems and users, it is essential to learn and model such interactions first so as to detect anomalies as attacks. However, the detection method in Chapter 3 avoids such customization.

Shoukry et al. [61] proposed a detection method named PyCRA that is similar to the detection method for sensor systems in Chapter 3. Since PyCRA has been thoroughly discussed in Chapter 3, it is not repeated here. Fang et al. [107] proposed adding unique noise (fingerprints) to sensor measurements and using machine learning techniques to detect the attacks. Several works mentioned that multiple built-in sensors of a device could react to variations of the electromagnetic environment, and the characteristics could be exploited to detect abnormal electromagnetic activities [15, 108]. Such a detection approach is also known as sensor fusion, which has been widely studied to detect signal injections that use other types of attacking signals such as ultrasonics and lasers [49, 59].

These detection methods work well for the sensors because the computational capabilities of the receiver (microcontroller) make authentication possible. However, it is not easy to apply similar ideas to the actuator systems because the receiver (actuator) lacks computational capabilities to authenticate its input signals.

6.1.2 Methods for Actuator Systems

As mentioned in Section 1.1, little research studied the electromagnetic signal injection attacks on actuator systems, and fewer studies on protection. Muniraj and Farhood [80] proposed that reliable sensor measurements can be used to indicate

whether actuators are under attack. In unmanned aircraft systems, they proposed to artificially cause minor disturbances to the actuators at a random time and use sensors to capture the disturbances; unexpected disturbances imply attacks. Note that this method trades off the stability of the whole system against its security. The same authors proposed another detection method that casts the actuator attack detection problem as an unknown input estimation problem and uses a two-stage extended Kalman filter to estimate actuator attacks from sensor measurements, requiring additional computational power. Moreover, the authors also proposed a method that adds randomness to control signals to improve the resilience of the actuator against malicious attacks.

It is not difficult to find that these defenses are devised for a specific application, i.e., unmanned aircraft vehicles. It will be challenging to apply similar ideas to other applications such as smart locks and insulin pumps, as they work in distinct ways, e.g., they do not have abundant sensors to interact with the environment. Moreover, it is essential to point out again that these detection methods require "reliable sensor measurements", meaning that the sensors must be properly protected from the attacking signals. In fact, the sensors in the systems still lack protection, which will corrupt the security of these detection methods and their usability. Therefore, it is encouraged to deploy the detection methods for the sensor systems to guarantee trustworthy sensor readings. Compared with these detection methods, the detection approach that is proposed for actuator systems in Chapter 4 outstrips. First, regarding flexibility, it is designed for different actuator systems regardless of their types, making it quick to be deployed in different applications. Second, recalling that the detection approach relies on the difference between the primary signal and the reference signal, rather than any specific models or sensor measurements, and as such not only reduces the complexity but also retains the reliability. Please note that my detection approach requires minor modifications of circuits, but the other detection methods are implemented in software without touching the hardware.

6.1.3 Other Anomaly Detectors

Researchers developed standalone detection systems that capture electromagnetic waves by dedicated antennas and then use intricate circuits to process the captured signals for detection [103–105]. Attacks are detected if abnormal electromagnetic signals or activities appear. However, it is essential to point out that the detection circuits are cumbersome and complicated, making them difficult to be integrated with applications where size and weight are critical, e.g., implantable medical devices. Tu et al. [19] proposed leveraging the superheterodyne technique to create an anomaly detector to check whether sensor measurements carry malicious frequency components. Moreover, in a cryptographic integrated circuit, Fujimoto et al. [109] proposed a detection method against the attacking signal by monitoring the built-in voltage variation of the power supply using the on-chip voltmeter.

6.2 Attenuation

Four distinct strategies, i.e., shielding, miniaturization, filtering and mitigation, and robust hardware, are discussed here. Note that shielding and miniaturization aim to attenuate attacking signals before they are injected into victim circuits, while the other two strategies handle the attacks after the injections.

6.2.1 Shielding

Better isolation from the external world can make a system more immune to attacking signals. Wrapping components with proper RF shielding materials is a common method to attenuate attacking signals [14–16, 19, 22, 23, 26, 34, 37, 51, 64]. For example, Kune et al. showed a 40 dB attenuation of the injected signal into a webcam. It is essential to emphasize that the shielding materials provide finite attenuation [21], and a powerful attacker may still breach the protection by increasing her attack power. Although adding thicker shielding materials can increase the attenuation level, it will still challenge the device's weight and size, especially for applications such as implantable medical devices and aviation.

Moreover, some unavoidable holes in the shielding can lead to degradation of the attenuation: for example, apertures in a shielded enclosure for ventilation or optical displays [110], seams on the shielding [111], and cable penetrations [21]. Conductive shielding materials are used to eliminate coupled electric and magnetic fields or lone electric fields [13, 26, 111]. However, for a lone magnetic field that can also induce adversarial signals into cabling, magnetic shielding with high permeability should be employed [112]. Selvaraj et al. [26] pointed out that magnetic shielding is usually not considered due to its weight and cost [113].

Regarding traces in a printed circuit board (PCB), researchers [27, 32] suggested that via-fenced striplines, where the vias and ground planes behave as a solid conductor enclosing (i.e., shielding), can also eliminate attacking signals by a finite amount (approximately 15 dB [114]).

6.2.2 Miniaturization

Making circuits smaller essentially pushes the resonant frequency up to a higher value. From the attackers' perspective, in order to achieve effective and efficient injections, they need to increase the attack frequency. A higher attack frequency also means more advanced signal generators, more powerful amplifiers, and more complex antenna designs, which directly increase the complexity of conducting attacks. Dayanıklı [32] analyzed and showed that minimizing the lengths of signal traces and PCB thickness will also make the attack injection more challenging. For example, halving the PCB traces can force the attacker to increase the attack power by 6 dB so as to cause the same effects. In short, miniaturizing circuits can raise the bar for the attackers.

6.2.3 Filtering and Mitigation

Filtering is another prevalent solution to mitigate attacking signals; specifically, it eliminates unwanted frequency bands. For example, low-pass filters (LPFs) can significantly attenuate out-of-band attacking signals [14, 19, 26, 34, 37, 51]. However, in-band attacking signals can still pass through the LPFs. Researchers

also pointed out that the parasitics in surface mount components can convert the LPF into a band-stop filter, allowing out-of-band attacking signals to pass [115]; hence, careful design and thorough tests are necessary. Besides, EMI filters are specifically designed to suppress electromagnetic noise, but it is not always feasible due to their size and weight.

In addition to the conventional filtering strategies, novel filtering methods have been proposed in the last years. Kune et al. [14] proposed to deploy an adaptive filtering mechanism [116] that makes use of knowledge about ambient electromagnetic emissions to attenuate the interference in sensor measurements. Crovetti and Musolino [117] also proposed a novel way to suppress the EMI-induced errors in the sensor measurements. Specifically, in the filtering stage, they crafted an extra channel, which the EMI will impact in an opposite way from the original channel; after digitization, the EMI-induced error thus can be easily compensated by a weighted sum of data from two channels.

Furthermore, some researchers recommended using differential rather than single-end comparator to attenuate the attacking signals in a finite frequency band, thereby raising the bar for attackers [14, 32, 106]. However, Tu et al. [19] pointed out that the input asymmetries may lead to insufficient mitigation [19]. Indeed, Chapter 5 explicitly shows that such asymmetries allow an attacker to bypass the differential signaling and inject arbitrary messages.

6.2.4 Robust Hardware

Attackers can exploit the hardware imperfections, such as asymmetry and non-linearities of victim circuits, to realize malicious controls. In response to such imperfections, abundant EMI-robust circuits have been proposed in the literature. For example, Maekettos and Moore suggested using carefully balanced transistors to reduce the asymmetries in ring oscillators [23]. For another example, regarding operational amplifier (OPAMP), Crovetti and Musolino [117] summarized existing methods focus on filtering [118–125], source-buffering [126, 127], cancellation by compensating asymmetries [119, 128–130], and compensation of nonlinearities [119,

122, 124, 131–134]. Despite the novelties, the authors emphasized that the EMI immunity provided by these solutions can be easily impaired by device mismatch, which thus must be carefully treated and avoided in the deployment.

7 Conclusion

Starting from the threat of electromagnetic signal injection attacks on embedded systems, this thesis focuses on three crucial signals: sensor measurements, actuator control signals, and differential signals. Although attacks on sensor measurements have been thoroughly studied and attacks on actuator signals are being investigated, generalized detection methods that can protect variously different devices from these attacks were absent until the novel works in this thesis fill the gap. Furthermore, this thesis also presents a pioneering work that systematically and experimentally shows the feasibility of arbitrary message injections into differential signaling, and such a state-of-the-art attack immediately poses threat to many protocols that derive their electromagnetic noise immunity from differential signaling. In short, this thesis brings not only profound insights into security issues of the embedded system concerning electromagnetic signal injection attacks but also novel solutions.

Chapter 3 focuses on electromagnetic signal injections into sensor measurement and fills the gap of a lightweight and generalized detection method that fit any sensors, including active sensors, powered-, and non-powered passive sensors. The novelty of the detection method is secretly encoding sensor power, forcing an attacker to correctly guess a secret to avoid detection. The detection method provides a provable security guarantee, where the possibility of a successful attack without being detected is negligible. It is essential to highlight that this detection

method can effectively mitigate negative impacts from the attack, as the detection will immediately notify the embedded system to stop processing malicious sensor measurements. Also, this detection method requires small hardware and software modification, allowing quick deployment in practice, which is shown by implementing the approach in two practical systems (a microphone system and a temperature sensor system). Further, the experimental results exhibit high true positive rates and low false positive rates in detecting attacks, demonstrating the approach's effectiveness and robustness.

Although electromagnetic signal injection attacks on actuator signals are not misused as frequently as sensor measurements, it is also essential to develop proper defenses to prevent the attacks before they become prevalent. In Chapter 4, a novel detection method that fits various actuator systems is presented. Its detection principle is simple: any difference caused by attacks between two identical signals, which are a primary signal and a reference signal, indicates the attacks. This detection method provides provable security guarantees, and any attacks that effectively impact the actuators will always be detected. Also, this detection method allows system designers to dynamically tune a detection threshold to handle any attack power and any amount of external environmental noise. Besides, this method is lightweight, and its implementation only needs several inexpensive off-the-shelf electronic components. Its implementation on practical systems such as a speaker system and a motor control system demonstrates its generality for different actuator systems, as well as its effectiveness and robustness.

Chapter 5 digs into hardware imperfections of differential signaling that allow an attacker to inject any bits, and further arbitrary messages, into communications between embedded systems. This chapter details the principles that an attacker can exploit to achieve the injections, and it also provides a systematic way to analyze the success rates of the attacks. Note that the success rate is also a helpful metric that lets the system designer evaluate the security against the attacks. Extensive experiments explicitly show circuits' responses to the adversary-injected signals. A case study demonstrates the feasibility of bypassing differential signaling and

injecting malicious control commands into the CAN bus, which is commonly used in domestic appliances, medical devices, and automobiles. It is essential to point out that this work indeed shows how to inject malicious messages into differential signaling, but more importantly, this work aims to draw attention to security issues and motivate the development of defenses.

Although related research on the electromagnetic signal injection attacks and corresponding defenses is thriving, it is rare to see broad commercialization of new defenses, such as those state-of-the-art detection methods. There are no doubt many challenges from development to deployment. This process indeed requires more cross-discipline collaboration between multiple parties such as academia, industry, and legislature. Nonetheless, it is still good timing to act from now on because electromagnetic signal injection attacks have not been widely abused. Moreover, based on the current detection methods, it would be better to consider post-detection methods. For example, after detecting the attacks, the embedded system can somehow mitigate the attacks' negative impacts, log what is happening, and inform the system administrator of the attacks. However, it is not an easy job, as different embedded systems execute distinct tasks, and how to handle the attacks after detection may also be application-specific. Future work can focus on generalizing post-detection methods and developing tools that can analyze the security of products under the threat of electromagnetic signal injection attacks. They will help system designers to know how secure their products are in specific electromagnetic conditions and what they could do to improve the security further.

References

[1] Steve Heath. *Embedded Systems Design*. Elsevier, 2002, pp. 1–8.

[2] Edward A Lee and Sanjit Arunkumar Seshia. *Introduction to Embedded Systems - A Cyber-Physical Systems Approach*. Vol. 499. LeeSeshia.org, 2011, pp. 2–6.

[3] Shoei Nashimoto et al. "Sensor CON-Fusion: Defeating Kalman filter in signal injection attack". In: *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. 2018, pp. 511–524.

[4] Yunmok Son et al. "Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors". In: *24th USENIX Security Symposium (USENIX Security 15)*. 2015, pp. 881–896.

[5] Timothy Trippel et al. "WALNUT: Waging Doubt on the Integrity of MEMS Accelerometers with Acoustic Injection Attacks". In: *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2017, pp. 3–18.

[6] Yazhou Tu et al. "Injected and Delivered: Fabricating Implicit Control over Actuation Systems by Spoofing Inertial Sensors". In: *27th USENIX Security Symposium (USENIX Security 18)*. 2018, pp. 1545–1562.

[7] Chen Yan, Wenyuan Xu, and Jianhao Liu. "Can You Trust Autonomous Vehicles: Contactless Attacks against Sensors of Self-driving Vehicle". In: *Def Con* 24.8 (2016), p. 109.

[8] Yazhou Tu, Sara Rampazzi, and Xiali Hei. "Towards Adversarial Control Loops in Sensor Attacks: A Case Study to Control the Kinematics and Actuation of Embedded Systems". In: *arXiv preprint arXiv:2203.07670* (2022).

[9] Youngseok Park et al. "This Ain't Your Dose: Sensor Spoofing Attack on Medical Infusion Pump". In: *10th USENIX workshop on offensive technologies (WOOT 16)*. 2016.

[10] Hocheol Shin et al. "Illusion and Dazzle: Adversarial Optical Channel Exploits against Lidars for Automotive Applications". In: *International Conference on Cryptographic Hardware and Embedded Systems*. Springer. 2017, pp. 445–467.

[11] Yasser Shoukry et al. "Non-Invasive Spoofing Attacks for Anti-Lock Braking Systems". In: *International Conference on Cryptographic Hardware and Embedded Systems*. Springer. 2013, pp. 55–72.

[12] Perry F Wilson. "Radiation Patterns of Unintentional Antennas: Estimates, Simulations, and Measurements". In: *2010 Asia-Pacific International Symposium on Electromagnetic Compatibility*. IEEE. 2010, pp. 985–989.

[13] Clayton R Paul. *Introduction to Electromagnetic Compatibility*. Vol. 184. John Wiley & Sons, 2006.

[14] Denis Foo Kune et al. “Ghost Talk: Mitigating EMI Signal Injection Attacks Against Analog Sensors”. In: *Security and Privacy (S&P), 2013 IEEE Symposium on*. IEEE. 2013, pp. 145–159.

[15] Chaouki Kasmi and Jose Lopes-Esteves. “IEMI Threats for Information Security: Remote Command Injection on Modern Smartphones”. In: *IEEE Transactions on Electromagnetic Compatibility* 57.6 (2015), pp. 1752–1755.

[16] Kai Wang et al. “GhostTouch: Targeted Attacks on Touchscreens without Physical Touch”. In: *31st USENIX Security Symposium (USENIX Security 22)*. Boston, MA: USENIX Association, 2022. URL: `https://www.usenix.org/conference/usenixsecurity22/presentation/wang-kai`.

[17] Haoqi Shan et al. “Invisible Finger: Practical Electromagnetic Interference Attack on Touchscreen-based Electronic Devices”. In: *2022 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, May 2022, pp. 1548–1548. URL: `https://doi.ieeecomputersociety.org/10.1109/SP46214.2022.00119`.

[18] Yan Jiang et al. “WIGHT: Wired Ghost Touch Attack on Capacitive Touchscreens”. In: *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society. 2022, pp. 1537–1537.

[19] Yazhou Tu et al. “Trick or Heat? Manipulating Critical Temperature-based Control Systems Using Rectification Attacks”. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 2019, pp. 2301–2315.

[20] Sebastian Köhler, Richard Baker, and Ivan Martinovic. “Signal Injection Attacks against CCD Image Sensors”. In: *ACM ASIA Conference on Computer and Communications Security*. Association for Computer Machinery, 2022.

[21] Frederick M Tesche, Michel Ianoz, and Torbjörn Karlsson. *EMC Analysis Methods and Computational Models*. John Wiley & Sons, 1996.

[22] Kasper Bonne Rasmussen et al. “Proximity-based Access Control for Implantable Medical Devices”. In: *Proceedings of the 16th ACM conference on Computer and communications security*. 2009, pp. 410–419.

[23] A Theodore Markettos and Simon W Moore. “The Frequency Injection Attack on Ring-Oscillator-Based True Random Number Generators”. In: *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer. 2009, pp. 317–331.

[24] Robert A Stevenson and Dick H Ni. *EMI Filter for Human Implantable Heart Defibrillators and Pacemakers*. US Patent 5,751,539. May 1998.

[25] Richard Lee Ozenbaugh and Timothy M Pullen. *EMI filter design*. CRC press, 2017.

[26] Jayaprakash Selvaraj et al. “Electromagnetic Induction Attacks Against Embedded Systems”. In: *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. ACM. 2018, pp. 499–510.

[27] Gökçen Y Dayanıklı et al. “Electromagnetic Sensor and Actuator Attacks on Power Converters for Electric Vehicles”. In: *IEEE Workshop on the Internet of Safe Things*. IEEE. 2020.

[28] Paul Horowitz, Winfield Hill, and Ian Robinson. *The Art of Electronics*. Vol. 3. Cambridge university press Cambridge, 2015, p. 473.

[29] Michał Maćkowski. "The Influence of Electromagnetic Disturbances on Data Transmission in USB Standard". In: *International Conference on Computer Networks*. Springer. 2009, pp. 95–102.

[30] Ke-Jie Li et al. "Statistical Inference of Serial Communication Errors Caused by Repetitive Electromagnetic Disturbances". In: *IEEE Transactions on Electromagnetic Compatibility* 62.4 (2019), pp. 1160–1168.

[31] Fei Ren et al. "Effects of Electromagnetic Interference on Control Area Network Performance". In: *2007 IEEE Region 5 Technical Conference*. IEEE. 2007, pp. 199–204.

[32] Gökçen Y Dayanıklı. "Electromagnetic Interference Attacks on Cyber-Physical Systems: Theory, Demonstration, and Defense". PhD thesis. Virginia Tech, 2021.

[33] Sebastian Köhler et al. "Brokenwire : Wireless Disruption of CCS Electric Vehicle Charging". In: *Network and Distributed Systems Security (NDSS) Symposium*. 2023.

[34] Youqian Zhang and Kasper Rasmussen. "Detection of Electromagnetic Interference Attacks on Sensor Systems". In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2020, pp. 203–216.

[35] Youqian Zhang and Kasper Rasmussen. "Detection of Electromagnetic Signal Injection Attacks on Actuator Systems". In: *25th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2022)*. ACM. 2022.

[36] Youqian Zhang and Kasper Rasmussen. "Electromagnetic Signal Injection Attacks on Differential Signaling". In: *arXiv preprint arXiv:2208.00343* (2022).

[37] Ilias Giechaskiel, Youqian Zhang, and Kasper Rasmussen. "A Framework for Evaluating Security in the Presence of Signal Injection Attacks". In: *European Symposium on Research in Computer Security*. Springer. 2019, pp. 512–532.

[38] Alan Cheng et al. "Effects of Surgical and Endoscopic Electrocautery on Modern-day Permanent Pacemaker and Implantable Cardioverter-defibrillator Systems". In: *Pacing and Clinical Electrophysiology* 31.3 (2008), pp. 344–350.

[39] John Loewy, Amanda Loewy, and Edward J Kendall. "Reconsideration of Pacemakers and MR Imaging". In: *Radiographics* 24.5 (2004), pp. 1257–1267.

[40] Ariel Roguin et al. "Magnetic Resonance Imaging in Individuals with Cardiovascular Implantable Electronic Devices". In: *Europace* 10.3 (2008), pp. 336–346.

[41] Yakup Bayram et al. "High Power EMI on Digital Circuits within Automotive Structures". In: *IEEE International Symposium on Electromagnetic Compatibility, Portland, OR*. 2006, pp. 507–512.

[42] Yu-Ichi Hayashi et al. "Analysis of Electromagnetic Information Leakage from Cryptographic Devices with Different Physical Structures". In: *IEEE Transactions on Electromagnetic Compatibility* 55.3 (2013), pp. 571–580.

[43] Yu-ichi Hayashi et al. "Information Leakage from Cryptographic Hardware via Common-Mode Current". In: *2010 IEEE International Symposium on Electromagnetic Compatibility (EMC)*. IEEE. 2010, pp. 109–114.

[44] Yuichi Hayashi et al. “A Threat for Tablet PCs in Public Space: Remote Visualization of Screen Images using EM Emanation”. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 2014, pp. 954–965.

[45] William A Radasky, Carl E Baum, and Manuem W Wik. “Introduction to the Special Issue on High-Power Electromagnetics (HPEM) and Intentional Electromagnetic Interference (IEMI)”. In: *IEEE Transactions on Electromagnetic Compatibility* 46.3 (2004), pp. 314–321.

[46] Daniel Nitsch et al. “Susceptibility of Some Electronic Equipment to HPEM Threats”. In: *IEEE transactions on Electromagnetic Compatibility* 46.3 (2004), pp. 380–389.

[47] Frank Sabath and Heyno Garbe. “Risk Potential of Radiated HPEM Environments”. In: *Electromagnetic Compatibility, 2009. EMC 2009. IEEE International Symposium on*. IEEE. 2009, pp. 226–231.

[48] F Brauer, F Sabath, and JL Ter Haseborg. “Susceptibility of IT Network Systems to Interferences by HPEM”. In: *Electromagnetic Compatibility, 2009. EMC 2009. IEEE International Symposium on*. IEEE. 2009, pp. 237–242.

[49] Chen Yan et al. “SoK: A Minimalist Approach to Formalizing Analog Sensor Security”. In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2020, pp. 233–248.

[50] Harald T Friis. “A Note on A Simple Transmission Formula”. In: *Proceedings of the IRE* 34.5 (1946), pp. 254–256.

[51] Saki Osuka et al. “EM Information Security Threats against RO-based TRNGs: The Frequency Injection Attack Based on IEMI and EM Information Leakage”. In: *IEEE Transactions on Electromagnetic Compatibility* 61.4 (2018), pp. 1122–1128.

[52] David Samyde et al. “On A New Way to Read Data from Memory”. In: *First International IEEE Security in Storage Workshop, 2002. Proceedings.* IEEE. 2002, pp. 65–69.

[53] J-J Quisquater. “Eddy Current for Magnetic Analysis with Active Sensor”. In: *Proceedings of Esmart, 2002* (2002), pp. 185–194.

[54] Jörn-Marc Schmidt and Michael Hutter. *Optical and EM Fault-Attacks on CRT-based RSA: Concrete Results*. Citeseer, 2007.

[55] Amine Dehbaoui et al. “Electromagnetic Transient Faults Injection on A Hardware and A Software Implementations of AES”. In: *2012 Workshop on Fault Diagnosis and Tolerance in Cryptography*. IEEE. 2012, pp. 7–15.

[56] Nicolas Moro et al. “Electromagnetic Fault Injection: Towards A Fault Model on A 32-bit Microcontroller”. In: *2013 Workshop on Fault Diagnosis and Tolerance in Cryptography*. IEEE. 2013, pp. 77–88.

[57] Sébastien Ordas et al. “Evidence of A Larger EM-Induced Fault Model”. In: *International Conference on Smart Card Research and Advanced Applications*. Springer. 2014, pp. 245–259.

[58] Marco Leone and Hermann L Singer. “On the Coupling of An External Electromagnetic Field to A Printed Circuit Board Trace”. In: *IEEE Transactions on Electromagnetic Compatibility* 41.4 (1999), pp. 418–424.

[59] Ilias Giechaskiel and Kasper Bonne Rasmussen. "Taxonomy and Challenges of Out-of-Band Signal Injection Attacks and Defenses". In: *IEEE Communications Surveys & Tutorials (COMST)* 22.1 (2020), pp. 645–670.

[60] *Manchester Coding Basics*. Atmel Corporation. 2325 Orchard Parkway, San Jose, CA 95131, USA, Sept. 2009.

[61] Yasser Shoukry et al. "PyCRA: Physical Challenge-response Authentication for Active Sensors under Spoofing Attacks". In: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 2015, pp. 1004–1015.

[62] Philip Sedgwick. "Pearson's Correlation Coefficient". In: *BMJ* 345 (2012), e4483.

[63] Jacob Benesty, Jingdong Chen, and Yiteng Huang. "On the Importance of the Pearson Correlation Coefficient in Noise Reduction". In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.4 (2008), pp. 757–765.

[64] Hocheol Shin et al. "Sampling Race: Bypassing Timing-based Analog Active Sensor Spoofing Detection on Analog-digital Systems". In: *10th {USENIX} Workshop on Offensive Technologies ({WOOT} 16)*. 2016.

[65] Henri Ruotsalainen, Albert Treytl, and Thilo Sauter. "Watermarking Based Sensor Attack Detection in Home Automation Systems". In: *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE. 2021, pp. 1–8.

[66] J Lopes Esteves and C Kasmi. "Remote and Silent Voice Command Injection on a Smartphone through Conducted IEMI: Threats of Smart IEMI for Information Security". In: *Wireless Security Lab, French Network and Information Security Agency (ANSSI), Tech. Rep* (2018).

[67] Constantine A Balanis. *Antenna Theory: Analysis and Design*. John Wiley & Sons, 2016.

[68] Behzad Razavi. *Design of analog CMOS integrated circuits*. McGraw-Hill Education, 2005, pp. 100–126.

[69] Chunyu Wu et al. "Characterization of the RFI Rectification Behavior of Instrumentation Amplifiers". In: *2018 IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI & PI)*. IEEE. 2018, pp. 156–160.

[70] Robert E Richardson. "Modeling of Low-level Rectification RFI in Bipolar Circuitry". In: *IEEE Transactions on electromagnetic Compatibility* (1979), pp. 307–311.

[71] Marle L Forcier and Robert E Richardson. "Microwave-rectification RFI Response in Field-effect Transistors". In: *IEEE Transactions on Electromagnetic Compatibility* (1979), pp. 312–315.

[72] Curtis E Larson and James M Roe. "A Modified Ebers-Moll Transistor Model for RF-interference Analysis". In: *IEEE Transactions on Electromagnetic Compatibility* (1979), pp. 283–290.

[73] Charles Kitchin and Lew Counts. *A Designer's Guide to Instrumentation Amplifiers*. Analog Devices Norwood, MA, 2006.

[74] Ron Mancini. "Op Amps for Everyone: Design Reference". In: Newnes, 2003, pp. 189–191.

[75] Analog Devices. *RFI Rectification Concepts*. 2009. URL: `https://www.analog.com/media/en/training-seminars/tutorials/MT-096.pdf`.

[76] Franco Fiori. "Susceptibility of Smart Power ICs to Radio Frequency Interference". eng. In: *IEEE Transactions on Power Electronics* 29.6 (2014), pp. 2787–2797.

[77] Texas Instruments. *AN-1698 A Specification for EMI Hardened Operational Amplifiers*. `https://www.ti.com/lit/an/snoa497b/snoa497b.pdf`. 2013.

[78] Calogero Bona and Franco Fiori. "A New Filtering Technique that Makes Power Transistors Immune to EMI". In: *IEEE Transactions on Power Electronics* 26.10 (2010), pp. 2946–2955.

[79] Calogero Bona and Franco Fiori. "EMIs-Inducted Failures in MOS Power Transistors". In: *2009 International Conference on Electromagnetics in Advanced Applications*. Sept. 2009, pp. 564–567.

[80] Devaprakash Muniraj and Mazen Farhood. "Detection and Mitigation of Actuator Attacks on Small Unmanned Aircraft Systems". In: *Control Engineering Practice* 83 (2019), pp. 188–202.

[81] Naoki Shinohara. *Wireless Power Transfer via Radiowaves*. Wiley Online Library, 2014.

[82] Frank Sabath. "Classification of Electromagnetic Effects at System Level". In: *Ultra-Wideband, Short Pulse Electromagnetics 9*. Springer, 2010, pp. 325–333.

[83] David Barnett, David Groth, and Jim McBee. "Cabling: the Complete Guide to Network Wiring". In: John Wiley & Sons, 2006, pp. 50–60.

[84] Behzad Razavi. "Design of Analog CMOS Integrated Circuits". In: Tata McGraw-Hill Education, 2002, pp. 100–133.

[85] Analog Devices. *Op Amp Common-Mode Rejection Ratio (CMRR)*. Tech. rep. Analog Devices, 2009.

[86] David Harris and Sarah L Harris. "Digital Design and Computer Architecture". In: Morgan Kaufmann, 2010, pp. 2–52.

[87] Jean-Michel Redouté and Michiel Steyaert. "EMC of Analog Integrated Circuits". In: Springer Science & Business Media, 2009, pp. 72–82.

[88] Bruce Carter and Thomas R Brown. "Handbook of Operational Amplifier Applications". In: Texas Instruments Dallas, TX, 2001, pp. 15–16.

[89] G Meyer-Brotz and A Kley. "The Common-mode Rejection of Transistor Differential Amplifiers". In: *IEEE Transactions on Circuit Theory* 13.2 (1966), pp. 171–175.

[90] Richard Jaeger. "Common-mode Rejection Limitations of Differential Amplifiers". In: *IEEE Journal of Solid-State Circuits* 11.3 (1976), pp. 411–417.

[91] MING-GUANG YI. "Common-mode Rejection Ratio of Differential Amplifiers". In: *IEEE Journal of Solid-State Circuits* 15.2 (1980), pp. 214–221.

[92] Gianluca Giustolisi, Giuseppe Palmisano, and Gaetano Palumbo. "CMRR Frequency Response of CMOS Operational Transconductance Amplifiers". In: *IEEE Transactions on instrumentation and Measurement* 49.1 (2000), pp. 137–143.

[93] Paolo Stefano Crovetti and Franco Fiori. "Finite Common-mode Rejection in Fully Differential Operational Amplifiers". In: *Electronics Letters* 42.11 (2006), pp. 615–617.

[94] Paolo S Crovetti. "Finite Common-mode Rejection in Fully Differential Nonlinear Circuits". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 58.8 (2011), pp. 507–511.

[95] Franco Fiori. "Susceptibility of CMOS Voltage Comparators to Radio Frequency Interference". In: *IEEE transactions on electromagnetic compatibility* 54.2 (2011), pp. 434–442.

[96] P Crovetti and F Fiori. "IC Digital Input Highly Immune to EMI". In: *2013 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE. 2013, pp. 1500–1503.

[97] R Timothy Marler and Jasbir S Arora. "The Weighted Sum Method for Multi-objective Optimization: New Insights". In: *Structural and multidisciplinary optimization* 41.6 (2010), pp. 853–862.

[98] Jurgen Branke et al. "Multiobjective Optimization: Interactive and Evolutionary Approaches". In: vol. 5252. Springer Science & Business Media, 2008, pp. 11–13.

[99] Hank Zumbahlen. "Linear Circuit Design Handbook". In: Elsevier-Newnes, 2008, pp. 448–449.

[100] Walt Kester. "Understand SINAD, ENOB, SNR, THD, THD+ N, and SFDR So You Don't Get Lost in the Noise Floor". In: *MT-003 Tutorial* (2009).

[101] Nordic Semiconductor. *nRF52833 Objective Product Specification*. 2019. URL: `https://docs.rs-online.com/f99e/A700000006639409.pdf`.

[102] Matthew Rogers and Kasper Rasmussen. "Silently Disabling ECUs and Enabling Blind Attacks on the CAN Bus". In: *Embedded Security in Cars (escar)*. Nov. 2022.

[103] Christian Adami et al. "HPM Detector System with Frequency Identification". In: *2014 International Symposium on Electromagnetic Compatibility (EMC Europe)*. IEEE. 2014, pp. 140–145.

[104] Christian Adami et al. "HPM Detection System for Mobile and Stationary Use". In: *EMC Europe 2011 York*. IEEE. 2011, pp. 1–6.

[105] JF Dawson et al. "A Cost-efficient System for Detecting An Intentional Electromagnetic Interference (IEMI) attack". In: *2014 International Symposium on Electromagnetic Compatibility*. IEEE. 2014, pp. 1252–1256.

[106] Yazhou Tu et al. "Transduction Shield: A Low-Complexity Method to Detect and Correct the Effects of EMI Injection Attacks on Sensors". In: *ACM ASIA Conference on Computer and Communications Security*. 2021.

[107] Kai Fang et al. "Detection of Weak Electromagnetic Interference Attacks Based on Fingerprint in IIoT Systems". In: *Future Generation Computer Systems* 126 (2022), pp. 295–304.

[108] Chaouki Kasmi and Jose Lopes-Esteves. "Automated Analysis of the Effects Induced by Radio-frequency Pulses on Embedded Systems for EMC Functional Safety". In: *2015 1st URSI Atlantic Radio Science Conference (URSI AT-RASC)*. IEEE. 2015, pp. 1–1.

[109] Daisuke Fujimoto et al. "Detection of IEMI Fault Injection Using Voltage Monitor Constructed with Fully Digital Circuit". In: *2018 IEEE International Symposium on Electromagnetic Compatibility and 2018 IEEE Asia-Pacific Symposium on Electromagnetic Compatibility (EMC/APEMC)*. IEEE. 2018, pp. 753–755.

[110] Mats G Backstrom and Karl Gunnar Lovstrand. "Susceptibility of Electronic Systems to High-Power Microwaves: Summary of Test Experience". In: *IEEE Transactions on Electromagnetic Compatibility* 46.3 (2004), pp. 396–403.

[111] Henry W Ott. *Electromagnetic Compatibility Engineering*. John Wiley & Sons, 2011.

[112] Richard B Schulz. "ELF and VLF Shielding Effectiveness of High-Permeability Materials". In: *IEEE Transactions on Electromagnetic Compatibility* 1 (1968), pp. 95–100.

[113] Alan Rich. "Shielding and Guarding". In: *Analog Dialogue* 17.1 (1983), pp. 8–13.

[114] George E Ponchak et al. "The Use of Metal Filled Via Holes for Improving Isolation in LTCC RF and Wireless Multichip Packages". In: *IEEE Transactions on Advanced Packaging* 23.1 (2000), pp. 88–99.

[115] Ryan Hurley. *Design Considerations for ESD/EMI Filters: II Low Pass Filters for Audio Filter Applications*. ON Semiconductor. 2007.

[116] John G Proakis. "Digital Signal Processing: Principles Algorithms and Applications". In: Pearson Education India, 2001, pp. 500–519.

[117] Paolo Crovetti and Francesco Musolino. "Digital Suppression of EMI-Induced Errors in a Baseband Acquisition Front-End including Off-the-Shelf, EMI-Sensitive Operational Amplifiers". In: *Electronics* 10.17 (2021), p. 2096.

[118] Andrea Lavarda et al. "On the Robustness of CMOS-chopped Operational Amplifiers to Conducted Electromagnetic Interferences". In: *IEEE Transactions on Electromagnetic Compatibility* 60.2 (2017), pp. 478–486.

[119] Simone Becchetti et al. "A Comprehensive Comparison of EMI Immunity in CMOS Amplifier Topologies". In: *Electronics* 8.10 (2019), p. 1181.

[120] Subrahmanyam Boyapati, Jean-Michel Redouté, and Maryam Shojaei Baghini. "Modeling and Design of EMI-Immune OpAmps in 0.18-μ m CMOS Technology". In: *IEEE Transactions on Electromagnetic Compatibility* 58.5 (2016), pp. 1609–1616.

[121] Franco Fiori. "On the Susceptibility of Chopper Operational Amplifiers to EMI". In: *IEEE Transactions on Electromagnetic Compatibility* 58.4 (2016), pp. 1000–1006.

[122] Anna Richelli, Gilbert Matig-a, and Jean-Michel Redoute. "Design of a Folded Cascode Opamp with Increased Immunity to Conducted Electromagnetic Interference in 0.18 μm CMOS". In: *Microelectronics Reliability* 55.3-4 (2015), pp. 654–661.

[123] Anna Richelli. “Increasing EMI Immunity in Novel Low-voltage CMOS OpAmps”. In: *IEEE Transactions on Electromagnetic Compatibility* 54.4 (2012), pp. 947–950.

[124] Anna Richelli. “CMOS OpAmp Resisting to Large Electromagnetic Interferences”. In: *IEEE Transactions on Electromagnetic Compatibility* 52.4 (2010), pp. 1062–1065.

[125] Cedric Walravens et al. “Efficient Reduction of Electromagnetic Interference Effects in Operational Amplifiers”. In: *Electronics Letters* 43.2 (2007), p. 1.

[126] Jagapathi Gundla, Subrahmanyam Boyapati, and Vijaya Sankara Rao Pasupureddi. “Compact CMOS Miller Opamp with High EMI-Immunity”. In: *IEEE Transactions on Electromagnetic Compatibility* 62.6 (2020), pp. 2394–2400.

[127] Jingjing Yu, Ahmed Amer, and Edgar Sanchez-Sinencio. “Electromagnetic Interference Resisting Operational Amplifier”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 61.7 (2014), pp. 1917–1927.

[128] Anjan Kumar Pudi NS, Jean-Michel Redouté, and Maryam Shojaei Baghini. “A Generic EMI-Immune Technique for Differential Amplifiers with Single-Ended Output”. In: *IEEE Transactions on Electromagnetic Compatibility* 60.4 (2017), pp. 958–964.

[129] Anna Richelli, Simon Kennedy, and Jean-Michel Redouté. “An EMI-Resistant Common-Mode Cancelation Differential Input Stage in UMC 180 nm CMOS”. In: *IEEE Transactions on Electromagnetic Compatibility* 59.6 (2017), pp. 2049–2051.

[130] Jean-Michel Redouté and Anna Richelli. “A Methodological Approach to EMI Resistant Analog Integrated Circuit Design”. In: *IEEE Electromagnetic Compatibility Magazine* 4.2 (2015), pp. 92–100.

[131] Franco Fiori and Paolo S Crovetti. “Nonlinear Effects of Radio-Frequency Interference in Operational Amplifiers”. In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 49.3 (2002), pp. 367–372.

[132] Subrahmanyam Boyapati, Jean-Michel Redouté, and Maryam Shojaei Baghini. “Modeling and Design of an EMI-Immune Source-Buffered Miller OpAmp in 0.18 μm CMOS Technology”. In: *2017 International Symposium on Electromagnetic Compatibility-EMC EUROPE*. IEEE. 2017, pp. 1–5.

[133] Subrahmanyam Boyapati, Jean-Michel Redoute, and Maryam Shojaei Baghini. “Design of A Novel Highly EMI-Immune CMOS Miller OpAmp Considering Channel Length Modulation”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 64.10 (2017), pp. 2679–2690.

[134] Paolo S Crovetti. “Operational Amplifier Immune to EMI with No Baseband Performance Degradation”. In: *Electronics letters* 46.3 (2010), p. 207.