Diss. ETH No. 19644

Physical-layer Identification of Wireless Devices

A dissertation submitted to
ETH Zurich

for the degree of
Doctor of Sciences

presented by

Boris Danev

MSc en informatique EPFL
born June 28, 1979
citizen of Bulgaria

accepted on the recommendation of

Prof. Dr. Srdjan Čapkun, examiner
Prof. Dr. David Basin, co-examiner
Prof. Dr. Wayne Burleson, co-examiner
Prof. Dr. Refik Molva, co-examiner

2011

Abstract

Wireless technologies are becoming increasingly present and important in our daily lives. They are being incorporated in more and more applications such as identity documents, payment systems, intelligent homes, environmental monitoring, supply chains, medical devices. Certain critical issues in the security and privacy of these applications relate to the identification of devices.

There are two major ways to identify devices in the network. The first one, mostly used in today's networks, relies on what devices hold (e.g., unique identification numbers, cryptographic keys). The second one, which is the focus of this thesis, consists of extracting unique characteristics which are inherent to the device and can be observed.

In this thesis, we study the feasibility of uniquely identifying wireless devices using physical characteristics of their analog radio circuitry. These characteristics are the result of hardware impairments introduced during the manufacturing process. We focus on those features that appear in the transmitted radio signals and are therefore measurable during the physical-layer device communication. We propose techniques that enable the accurate identification of several types of wireless devices, analyze the underlying assumptions and clarify the implications on the security and privacy of wireless applications.

In the introductory part of this thesis, we provide a real-world example that illustrates one problem with authenticating devices by what they hold. We realize a practical attack on car access control systems and discuss the potential of device identification to complement traditional authentication and prevent this and other device identity attacks.

Secondly, we study the problem of identifying same-model-same-manufacturer active and passive wireless devices using physical-layer characteristics. We consider low-power wireless transceivers and passive RFID transponders. We explore timing, modulation and spectral properties of the radio signals and show that wireless devices can be accurately identified under certain assumptions.

Finally, we evaluate the resilience of physical-layer device identification methods to impersonation. We show that physical-layer identification is vulnerable to certain types of impersonation attacks. We also provide a classification of attacks and discuss the implications of the use of physical-layer device identification in applications such as intrusion detection, device cloning detection and device privacy protection.

Zusammenfassung

Drahtlose Technologien werden zunehmend allgegenwärtig und wichtig in unserem täglichen Leben. Sie werden in immer mehr Anwendungen eingesetzt, wie zum Beispiel in Identitätskarten, Zahlungsystemen, intelligenten Häusern, bei der Überwachung, in Lieferketten und medizinischen Geräten. Für die Sicherheit der Anwendungen spielt auch die eindeutige Identifizierung dieser Geräte eine wichtige Rolle.

Es gibt zwei Wege um Geräte im Netzwerk zu identifizieren. Erstens können Geräte aufgrund von Informationen wie MAC-Adressen oder kryptographischen Schlüsseln identifiziert werden. Zweitens, und dies ist der Fokus dieser Arbeit, können Geräte aufgrund von einzigartigen Charakteristika identifiziert werden, welche bei jedem Gerät beobachtet werden können.

In dieser Arbeit erforschen wir die Durchführbarkeit der eindeutigen Identifizierung von drahtlosen Geräten durch physikalische Charakteristika ihrer analogen Funkschaltkreise. Diese Charakteristika resultieren aus Abweichungen im Produktionsprozess. Wir konzentrieren uns auf Charakteristika, welche im übertragenen Signal auftauchen und dadurch auf der physikalischen Ebene beim Empfänger messbar sind. Wir stellen Techniken vor, welche die genaue Identifizierung mehrerer Klassen von drahtlosen Geräten ermöglichen, analysieren die zugrundeliegenden Annahmen und erläutern die Folgen für die Sicherheit der drahtlosen Anwendungen.

In der Einführung dieser Arbeit stellen wir ein Beispiel aus der Praxis vor, welches das Problem der ausschliesslich kryptographischen Identifizierung verdeutlicht. Wir erläutern unseren Versuchsaufbau, der Angriffe auf Fahrzeugschliesssysteme ermöglicht und, diskutieren das Potential der Identifizierung von Geräten auf der physikalischen Ebene als Ergänzung zu traditionellen Authentifizierungssystemen, um diesen und ähnliche Angriffe zu verhindern.

Zweitens erforschen wir das Problem der Identifizierung von aktiven und passiven drahtlosen Geräten aus der gleichen Serie eines Herstellers, aufgrund von physikalischen Charakteristika. Wir betrachten Charakteristika mit Bezug auf Zeit, Modulation und Spektraleigenschaften der Funksignale und zeigen, dass drahtlose Geräte unter bestimmten Voraussetzungen eindeutig identifiziert werden können.

Schliesslich werten wir die Widerstandsfähigkeit unserer Identifizierung auf der physikalischen Ebene gegen Imitationsangriffe aus. Wir

zeigen, dass die physikalische Identifizierung anfällig gegenüber einigen Arten von Imitationsangriffen ist. Wir klassifizieren diese Angriffe und diskutieren die Folgen für die Nutzung der Identifizierung auf der physikalischen Ebene für Anwendungen wie die Erkennung von Eindringlingen, Gerätekopien und Methoden zum Schutz der Privatsphäre.

Résumé

Les technologies sans fil deviennent de plus en plus présentes et importantes dans la vie quotidienne. Elles sont incorporées dans de nombreuses applications telles que documents d'identité, systèmes de paiements, maisons intelligentes, surveillance, chaînes d'approvisionnement, équipements médicaux. La sécurité de ces applications est fortement liée à l'identification de ces équipements radio.

Il existe deux moyens pour identifier les équipements radio dans le réseau. Le premier, le plus couramment utilisé, est basé sur ce que l'équipement en question possède (des numéros d'identification uniques, clés cryptographiques). Le deuxième consiste à extraire des caractéristiques uniques, intrinsèques à l'équipement et qui peuvent être mesurées.

Dans cette thèse, nous étudions la faisabilité d'identifier d'une manière unique les équipements radio en utilisant des caractéristiques physiques de leurs circuits intégrés. Ces caractéristiques sont dues à des imperfections de fabrication de leurs composants. Nous nous concentrons sur les caractéristiques qui apparaissent dans les signaux radio transmis et en conséquence peuvent être mesurées pendant la transmission au niveau physique. Nous proposons et analysons des méthodes pour identifier plusieurs types d'équipements de manière précise, et nous expliquons les conséquences possibles pour la sécurité et la protection de la vie privée de leurs utilisateurs.

Dans la première partie de la thèse, nous présentons un exemple de la vie réelle qui montre un problème important d'authentification des équipements basés sur des clés cryptographiques. Nous effectuons des attaques contre des systèmes de contrôle d'accès de véhicules, et nous discutons le potentiel de l'identification basée sur les caractéristiques physiques pour prévenir ce type d'attaques, ainsi que pour d'autres attaques liées à l'identité de l'équipement.

Deuxièmement, nous explorons le problème d'identification d'équipements radio du même fabriquant et du même modèle en utilisant des caractéristiques extraites de la communication au niveau physique. Nous considérons des émetteurs-récepteurs radios actifs et des transpondeurs RFID passifs. Nous étudions des propriétés de temps, de modulation et les propriétés spectrales des signaux émis par l'équipement. Nous montrons que les équipements radio peuvent être identifiés de manière précise sous certaines conditions.

Finalement, nous évaluons les possibilités de compromettre les méthodes d'identification d'équipements. Nous montrons que certaines attaques basées sur l'imitation des signaux sont possibles et efficaces. Nous classifions l'ensemble de ces attaques et discutons les conséquences de l'utilisation de ces techniques d'identification d'équipements radio sur la détection d'intrusion, la détection de clones et la protection de la vie privée.

Acknowledgments

I express deep gratitude and appreciation to my advisor Prof. Srdjan Capkun for his support and guidance during my doctoral degree. His enthusiasm, scientific believes and constructive feedback kept me going forward. I am also thankful to the co-examiners Prof. David Basin, Prof. Wayne Burleson and Prof. Refik Molva for accepting to evaluate this thesis and provide comments and suggestions.

I would like to thank my colleagues Dr. Aurelien Francillon, Thomas Heydt-Benjamin, Ramya Jayaram Masti, Heinrich Luecken and Davide Zanetti who were committed co-authors to most of the publications that led to this thesis. It was a pleasure to work with them. My gratitude also goes to Hansruedi Benedickter from the Laboratory for Electromagnetic Fields and Microwave Electronics at ETH for his wise suggestions and assistance in the radio design.

I also thank my other colleagues Ghassan O. Karame, Christina Poepper, Kasper Bonne Rasmussen, Mario Strasser and Nils Ole Tippenhauer for the enjoyable work environment and challenging discussions. Additional thanks should go to Ghassan O. Karame, Christina Poepper and Nils Ole Tippenhauer for the great projects and supervised student works done together. I also thank our group assistant Barbara Geiser for her help and celerity in assisting the projects.

Furthermore, my thanks also go to the Wireless Communications Group at ETH, especially Heinrich Luecken, Dr. Marc Kuhn and Prof. Armin Wittneben for their feedback and assistance. I also thank the Computer Science Department for being always collaborative and friendly.

Last but not least, this thesis would not have been possible without the every day support, patience and love of my wife, Yordanka. Thank you so much for being there during the ups and downs in my studies. I also thank my parents for their encouragements during the years.

Contents

List of Figures

List of Tables

Part I

Motivation

Chapter 1

Introduction

In today's digital world, increasingly many consumer and business applications are carried out by wireless technologies. A number of wireless access systems have been established (e.g., personal mobile communications) and others are emerging in new areas such as identity documents, payment systems, intelligent homes, environmental monitoring, supply chains, medical devices.

The wireless technologies and applications need to provide sufficient level of security and privacy to users in order to be deployed in their full potential. A critical part of any security and privacy mechanisms is related to the device identity and the ability of identifying devices. In particular, the device identity serves as a building block of authentication protocols which ensure that only authorized devices are allowed to use system resources and perform transactions. The authentication task becomes especially challenging when it must function in adversarial settings, i.e., under device identity spoofing, identity compromise, replication and cloning. In many emerging scenarios, device identities would also have to be protected in order to avoid malicious actions such as unauthorized device tracking [1].

Wireless devices are traditionally identified by unique information that they hold such as a public identification number and/or cryptographic private key. A prominent example of unique identification number is the Media Access Control (MAC). While intended to be a permanent and globally unique identification for a device, in practice it is possible to modify it. Wireless applications, therefore, cannot safely rely on such device identities [2]. For security applications, devices

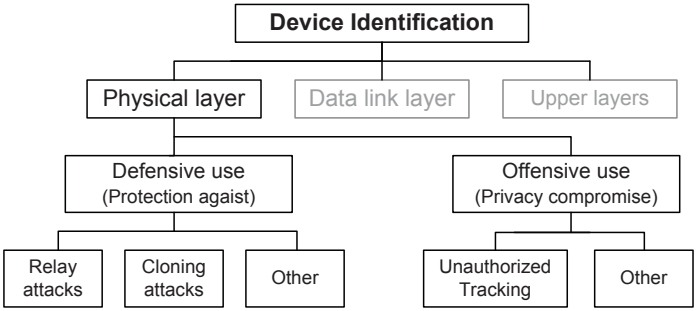


Figure 1.1: Physical-layer device identification presents defensive and offensive uses. As a defensive mechanism, it can be used for protection against relay, cloning and other identity-based attacks. As an offensive mechanism, it aims at compromising the device privacy for purposes such as unauthorized tracking.

would typically embed a cryptographic private key that can neither be read nor copied, but its existence in the device can be proven. Such keys are the basis for secure device authentication and secret key establishment [3].

A common feature of identification numbers and cryptographic keys as device identity is that they relate to what the device holds. This presents several security threats. First, authentication methods based on what devices carry could be vulnerable to relay attacks. We demonstrate a real-world relay attack in the following chapter. Second, if a cryptographic key is compromised, secure device authentication would be disabled. Furthermore, the key can be replicated to other devices, i.e., create device clones and disturb network operations and services. Device cloning that resulted in significant financial losses was the GSM SIM card cloning [4].

Besides by what they hold, devices can be identified by what they are, i.e., by some unique characteristics that they exhibit and that can be observed. Examples include characteristics related to device components such as operating system, drivers, clocks, radio circuitry. Analyzing these components for identifiable information is commonly referred to as fingerprinting, since the goal is to create fingerprints similar to their biometric counterparts [5]. A special class of characteristics inherent to a device are the hardware related impairments. These physical artifacts are caused by physical limitations and imprecision in the man-

ufacturing process. If the hardware imperfections are unique to a device and can be easily measured, they can be used to create a fingerprint of the device that cannot be easily modified or compromised.

In this thesis, we investigate hardware characteristics for identification of wireless devices that manifest themselves in the transmitted radio signals. We refer to it as physical-layer device identification. More formally, physical-layer device identification is the process of fingerprinting the analog circuitry of a device (or a class of devices) by analyzing its communication at the physical layer. It is possible due to analog radio circuit impairments introduced in the manufacturing process that are measurable during wireless communication.

Physical-layer device identification presents both defensive and offensive uses (Figure 1.1). As a defensive mechanism, it can provide an additional layer of security against a number of threats such as relay attacks, device cloning attacks, cryptographic key compromise. As an offensive mechanism, it can be used by an attacker to extract fingerprints of devices which allow device (user) tracking without its (his) prior consent. This may become a major hurdle in devising solutions that preserve the identity privacy of devices in the network [6].

The main goal of this thesis is to understand whether physical-layer device identification is feasible and accurate for same-model-same-manufacturer devices, analyze the underlying conditions and assumptions and clarify the defensive and offensive implications on the security and privacy of wireless networks.

Contributions

- We demonstrate real-world relay attacks on automobile access control systems. Our attacks enable vehicle access and drive even though the secure authentication is not compromised. This strongly motivates the study of device identification using characteristics inherent to the device. If it is feasible and secure, it would prevent relay attacks and other identity threats.

- We revisit transient-based identification of wireless transceivers and show that specific spectral features in the turn-on signal transient enable accurate identification of same-model-same-manufacturer transceivers. We also find that the transient contains channel-specific artifacts that cannot be easily removed. This property

exposes the limitation of transient-based approaches for device identification in dynamic environments.

- We design and implement novel methods for physical-layer identification of HF and UHF RFID devices. Our methods consist of purpose-built readers for precise signal acquisition and a set of time domain and spectral techniques for fingerprint extraction. Our experimental results demonstrate that same-model-same-manufacturer RFIDs can be identified in a controlled setup.

- We elaborate on HF RFID device identification with signal acquisition and feature extraction optimizations. We devise a method that allows identification across different setups. Our results show that HF RFID identification is not only feasible, but also accurate and stable over time and across setups. This is an insightful result as it confirms that identifiable information is readily available in the RFID circuit for anti-cloning protection.

- Our investigation on longer range UHF RFID devices demonstrates the feasibility of tracking RFIDs independently of their location using the physical layer. This result proves the location and/or identity privacy mechanisms on the logical layer are not sufficient to guarantee privacy.

- We analyze the resilience of several physical-layer device identification methods to attacks. More precisely, we design and implement a set of impersonation attacks. We show that such attacks on device identification are feasible and realistic. One conclusion is that these methods cannot be safely used in a number of application scenarios, where their use has been suggested.

- We contextualize the implications of physical-layer device identification on the security and privacy of several applications.

Thesis Outline

We devote the remainder of Part I to a real-world example of relay attacks on modern automobile entry and start systems. Our attacks allow access to the vehicle and enable vehicle driving even though the secure device authentication mechanism based on a shared cryptographic key is not compromised. This security problem motivates the study of

device identification using physical-layer characteristics as a defensive mechanism against relay and other identity related attacks.

In Part II, we investigate the problem of identifying same-model-same-manufacturer active wireless transceivers and passive RFID devices using physical-layer characteristics.

In the case of wireless transceivers, we consider IEEE 802.15.4 sensor nodes and show that the transient of the transmitted signal from the device contains enough distinguishable information to enable accurate device identification from short and long distances. We further explore the effects of distance, antenna polarization, voltage and temperature and analyze how these parameters affect the accuracy. Our findings expose the limitations of transient-based techniques adoption in mobile networks and dynamic environments.

For passively powered wireless devices, we focus on HF and UHF RFID. These are incorporated in a number of applications such as electronic passports, contactless identity and payment cards, supply chain systems. We explore timing, modulation and spectral features extracted from device communication to in- and out-of-specification reader requests. Our results indicate the presence of accurate and stable physical-layer HF RFID device fingerprints. These fingerprints can also be practical in the detection of cloned and/or counterfeit HF RFID-enabled identity documents. Our study on UHF RFID tags demonstrates the existence of timing characteristics that enable tracking of users carrying multiple tags by a network of readers.

In Part III, we evaluate the resilience of several physical-layer device identification approaches to impersonation attacks. Our findings show that physical-layer device identities are vulnerable to certain types of impersonation attacks. We also provide a classification of attacks and discuss the implications of our results on physical-layer device identification in application scenarios such as intrusion detection, device cloning detection and device privacy protection.

The related work is presented and compared to the work in this thesis in Part IV. Finally, we conclude the thesis and discuss remaining issues and possible future directions.

Chapter 2

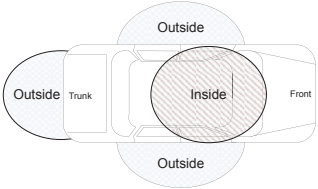
Example: Relay Attacks on Car Entry Systems

Device identities typically consist of unique identifiers and cryptographic private keys stored in the device. Authentication protocols leverage on these to securely authenticate devices. We demonstrate that authentication protocols based on what devices hold can be vulnerable to relay attacks if no protection measures are put in place.

Our real-world example focuses on a set of modern automobile entry and start systems. These systems allow to open and start a car without owner interaction and have been developed by a number of manufacturers. We build two efficient and inexpensive attack realizations, wired and wireless physical-layer relays, that allow an attacker to enter and start the car by quickly relaying messages between the car and the key. The secure authentication and encryption based on shared keys between the car and key cannot prevent our attack. Given the generality of the relay attack and the number of evaluated systems, it is likely that all car entry and start systems based on similar designs are vulnerable to the same attack.

Besides demonstrating relay attacks on PKES systems, we further analyze the characteristics of these systems and discuss the results. We discuss the fundamental issues with these systems and show the importance of studying physical-layer device identification as a possible solution.

(a) A PKES Key and its backup physical key.



(b) Car LF coverage.

Figure 2.1: Backup key and LF coverage regions.

2.1 Passive Keyless Entry and Start

Passive keyless entry and start systems first appeared in [7]. The authors proposed a system that automatically unlocks the vehicle when the user carrying the key approaches the vehicle and locks the vehicle when the user moves away. The system is referred to as 'Passive' as it does not require any action from the user. The communication between the key and car is characterized by a magnetically coupled radio frequency signal. In this system, the car concludes that the key is in the physical proximity when it is within the allowed communication range.

Current PKES car keys use an LF technology that provides short range communication (within 1-2 m in active and a few centimeters in passive mode) and an UHF transceiver for longer range communication (from 10 to 100 m). The LF channel is used to detect if the key is within regions *Inside* and *Outside* of the car. Figure 2.1(b) shows the areas in proximity of the car that are detected in order to allow a safe and convenient use of the PKES system. The regions are as follows.

- Remote distance to the car (typically up to 100 m). Car open and close is allowed by pushing a button on the key fob.
- Outside the car, but approximately 1 - 2 m from the door handle. Car open and close is allowed by using the door handle.
- Inside the car. Engine start and car drive is allowed.

The PKES protocols vary depending on the manufacturer. Typically two modes of operation are supported, namely *normal* and *backup* modes. The normal mode relies on a working battery, while the backup

Table 2.1: PKES access control summary

Key position	Authorization	Medium used	
		Car ⇒ Key	Key ⇒ Car
Normal mode: when the internal battery is present			
Remote	Active open/close	-	UHF
Outside	Passive open/close	LF	UHF
Inside	Passive start	LF	UHF
Backup mode: when the internal battery is exhausted			
Remote	Open/close	Not available	
Outside	Open/close	With physical key	
Inside	Start	LF	LF

mode operates without battery (e.g., when the battery is depleted). Table 2.1 summarizes the access control modes.

Figure 2.2 shows two example protocols of car opening in a normal mode. The car sends beacons on the LF channel either periodically or when the door handle is operated. These beacons could be either short wake-up messages or challenge messages that contain the car identifier. When the key detects the signal on the LF channel, it wakes up the micro-controller, demodulates the signal and interprets it. After computing a response to the challenge, the key replies on the UHF channel. This response is received and verified by the car. In the case of a valid response the car unlocks the doors. Subsequently, in order to start the car engine, the key must be present inside the car (region *Inside* in Figure 2.1(b)). The key then receives different types of messages that when replied will inform the car that the correct key is within the car itself. The car will then allow starting the engine. It should be noted that in normal mode the LF channel is only used to communicate from the car to the key.

In backup mode, the user is still able to open and start his car. The manufacturers usually embed a backup physical key within the key fob to open the car doors. These are shown in Figure 2.1(a). In order to start the engine the system uses the passive LF capabilities of the key. Given the very short communication range as discussed before, the user is required to place the key in the close proximity of some predefined location in the car (e.g., the Start button).

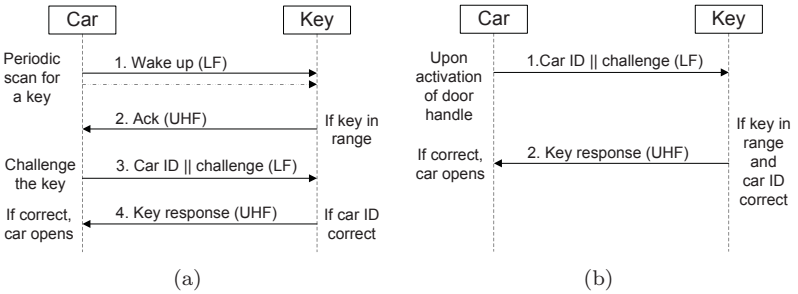


Figure 2.2: Examples of Passive Keyless Entry and Start system protocols. a) In a typical realization, the car periodically probes the channel for the presence of the key with short beacons. If the key is in range, a challenge-response protocol between the car and key follows to grant or deny access. b) In a second realization, upon activation of the door handle, the car directly sends a challenge that contains the car identifier. If the key is in range, it directly responds to that challenge.

2.2 Relay Attack on Smart Key Systems

In this section we first describe generic relay attacks, and then we present the attacks that we have implemented and tested on PKES systems of several cars from different manufacturers. In our experiments, we relay the LF communication between the car and the key. The relay of the UHF communication (from the key to the car) was not needed since this communication is 'long' range (approx. 100 m) and is not used for proximity detection. However, similar relay attacks can also be mounted on the UHF communication if a longer relay than 100 m is required.

2.2.1 Relay Attacks

The relay attack is a well-known attack against communication systems [8]. Realizations have been demonstrated on credit card transactions [9] and between nodes in wireless sensor networks, known as a wormhole attack [10]. An example of relay attack on HF RFID is available in [11]. It consisted of demodulating the signal, transmitting it as digital information using RF and then modulating it near the victim device. That relay adds 15–20 us of delay which may be detected by a

suitable key/car pair.

In this work, we design and implement a physical-layer relay attack. Our attack does not demodulate and modulate the signal, i.e., only introduces the delays typical for analog RF components. It is completely transparent to most security protocols designed to provide authentication or secrecy of the messages. Even if a PKES system uses strong cryptography (e.g., AES, RSA), it would still be vulnerable to our proposed relay attack.

It should be noted that many relay attacks previously presented are modulating and demodulating the signal. An obvious advantage of such attacks is that they can be performed with commercial off-the-shelf (COTS) hardware. The same setup can also be used to perform replay or message forging. However, this approach has several drawbacks. First, modulation and demodulation significantly increase the response time of the attack; this extra time may be used to detect the relay. Second, such a realization is dependent on the modulation and encoding of the signal, which makes the relay specific to some systems. Both drawbacks are avoided in our design and implementation of the relay attack.

2.2.2 Relay Over-Cable Attack

In order to perform this attack, we used a relay (Figure 2.3) composed of two loop antennas connected together with a cable that relays the LF signal between those two antennas. An optional amplifier can be placed in the middle to increase the signal power. When the loop antenna is presented close to the door handle, it captures the car beacon signal as a local magnetic field. This field excites the first antenna of the relay, which creates by induction an alternating signal at the output of the antenna. This electric signal is then transmitted over the coaxial cable and reaches the second antenna via an optional amplifier. The need for an amplifier depends on several parameters such as the antenna quality, cable length, signal strength. When the relayed signal reaches the second antenna of the cable it creates a current in the antenna which in turn generates a magnetic field in the proximity of the second antenna. Finally, this magnetic field excites the antenna of the key which demodulates this signal and recovers the original message from the car. In all evaluated PKES systems, this is sufficient to trigger the key sending an *open* or *start* authorization message over the UHF channel. The message sent by the key will depend on what

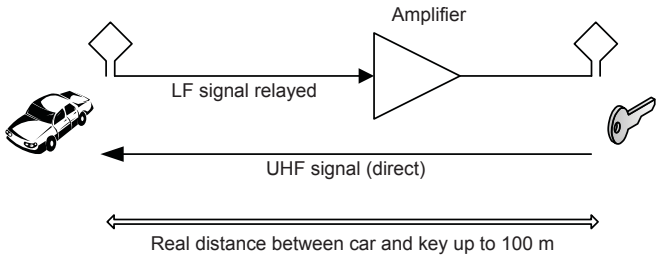


Figure 2.3: The relay with antennas, cables and an (optional) amplifier.

was originally sent by the car. The car will send *open* command to the key from the outside antennas and the *start* command form the inside antennas. Therefore, the attacker (e.g., car thief) first needs to present the relaying antenna in front of the door handle such that the key will send the open signal. Once the door is unlocked, the attacker brings the relaying antenna inside the car and after he pushes the brakes pedal or the start engine button the car will send the *start* message to the key. In both cases the key answers on UHF and the action (open or start) is performed.

2.2.3 Relay Over-The-Air Attack

Relaying over a cable might be inconvenient or raise suspicion. For example, the presence of walls or doors could prevent it. We therefore design and realize a physical-layer relay attack over the air. Our attack relays the LF signals from the car over a purpose-built RF link with minimal delays. The link is composed of two parts, the *emitter* and the *receiver*. The *emitter* captures the LF signal and up-converts it to 2.5 GHz. The obtained 2.5 GHz signal is then amplified and transmitted over the air. The *receiver* part of the link receives this signal and down-converts it to obtain the original LF signal. This LF signal is then amplified again and sent to a loop LF antenna which reproduces the signal that was emitted by the car in its integrity. The procedure for opening and starting the engine of the car remains the same as explained in the previous section.

Using the concept of analog up and down conversion allows the attacker to relay the LF signal further away from the car, while keeping the size, power consumption and price of the attack low. We note that it

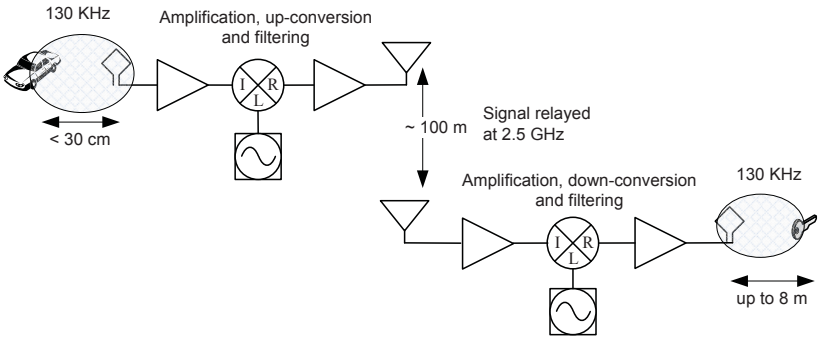


Figure 2.4: Simplified view of the attack relaying LF (130 KHz) signals over the air by upconversion and downconversion. The relay is realized in analog to limit processing time.

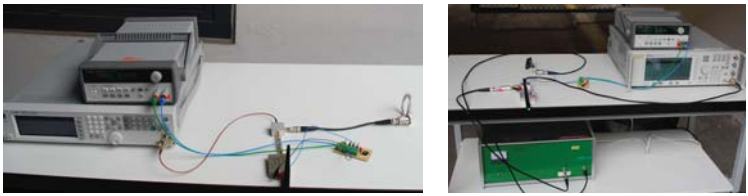
could be possible to transmit in LF over longer distances. However this would require large antennas and a significant amount of power [12].

2.2.4 Experimental Relays Results

Some measurement results on the delay versus distance are reported in Table 2.2 for both relay attacks.

In the cable LF relay, the delay is primarily introduced by the wave propagation speed in solid coaxial cables which is approximately 66% of that speed in the air. The delay of our amplifier is of the order of a few nanoseconds. In the wireless LF relay, our measurements show a delay of approximately 15 - 20 ns in both emitter and receiver circuitries, the remaining delay being due to the distance between the antennas, i.e., approximately 100 ns for 30 m. Therefore for larger distances, using the over-the-air relay should be preferred in order to keep the delay as low as possible. In order to compute the total delay of the relay attack, i.e., including both the LF and UHF links, we should add the UHF car-key communication which assumes wave propagation with the speed of light and will only depend on the distance.

Figure 2.5(a) shows the part of the wireless relay that receives messages from the car. Signals are received using the white loop antenna (right in the picture). This antenna must be positioned near to the car emitting antennas, for example at the door handle or the start button (Figure 2.6) in order to obtain a good signal from the car. This signal



(a) Car side (b) Key side

Figure 2.5: Experimental relay over-the-air realization.

Table 2.2: Relay setup distance and delay. The measured delays are for the LF channel only. The UHF channel delay is based on direct car-key communication and assumes wave propagation at the speed of light. The latter should be added to obtain the total relay delay.

Attack	Distance (m)	Delay (ns)	Comments
Cable	30	160 (±20)	Open and start works reliably
	60[1]	350 (±20)	Amplification may be required
Air	30[2]	120 (±20)	Reliable open, engine start works

[1] With an amplifier between two 30 m cables.
[2] Tested distance. Longer distances can be achieved.

is amplified, up-converted and retransmitted at 2.5 GHz with a dipole antenna (black in front of the picture).

Figure 2.5(b) shows the receiver side of the over-the-air relay which should be placed in the proximity of the key. The dipole antenna receives the relayed 2.5 GHz signal, and a down conversion setup extract the original car signal which is then relayed to the key using a loop antenna. While the setup on those pictures is made of experimental equipment, it can easily be reduced to two small and portable devices.

2.3 Evaluation on Different Car Models

Both above presented setups were initially tested on a few different car models. To further evaluate the generality of the attack we tested 10 cars (including one after-market PKES) on which we ran several

(a) Loop antenna placed next to the door handle.

(b) Starting the engine using the relay.

Figure 2.6: The relay attack in practice: (a) opening the door with the relay. (b) starting the car with the relay, in the foreground the attacker with the loop antenna starts the car, in the background the table (about 10 meters away) with the receiver side (Figure 2.5(b)) of the wireless relay and the key. Emitter side (Figure 2.5(a)) of the wireless relay is not shown on this picture.

experiments. The cars were either rented on purpose or the experiments were performed with the agreement of the car owners. In one case, a car manufacturer representative proposed us to evaluate the attack on a car he made available to us. In another case, a car owner, who recently had a similar car stolen asked us to evaluate his second car's PKES. The aftermarket PKES system was bought and analyzed for the purpose of our experiments for about 200 USD. Finding other car models for testing was not always easy. In some cases, we were able to rent cars or found volunteers through personal relationships. The tested cars models cover a wide range of types and price as follows: 2 models in SUV class, 4 executive or luxury class (>50K) cars, 1 minivan and 2 cars in the compact class (<30K). We had two different models for only two of the tested manufacturers. During the evaluation of the 10 different PKES systems, we observed that all of them differed in their implementation. We also noticed that even if they relied on the same general idea and similar chips the overall system behaved differently for each model [1]. The differences were found in timings, modulation and protocol details (e.g., number of exchanged messages,

[1] This was also the case for the models from the same manufacturers.

message length). Only the aftermarket system was obviously not using any secure authentication mechanisms.

When possible, on each car we measured the distances for the relay, the maximum acceptable delay and the key response time and spread.

2.3.1 Distance Measurements

In order to validate the feasibility of the attack in practice, we tested several distances for the cable relay. This allows to evaluate the possible attack setup, a longer relay distance over the cable will allow the thief to act when the car owner is relatively far from his car, reducing chances of detection. We further measured the distance form the relaying antenna to the key, a longer distance will make the attack easier (e.g., avoid suspicion from the user).

The cable relay was performed with off-the-shelf coaxial cables. We built two 30 m cables that we combined for the 60 m relay tests. We used a set of antennas, two home made antennas, and a large antenna [2] for an improved antenna-key range. We performed the attacks with these antennas both with and without amplification. If the LF signal near the car was weak we used a 10 mW low-noise amplifier to increase the signal power. To further improve key to antenna range we used a power amplifier with a nominal power of 2 to 5 W.

The results of these experiments are shown in Table 2.3. The relays over the 3 cable lengths were always successful when we were able to test them. In most of the cases the signal received on the collecting antenna was strong enough to perform the relay over the cable without any amplification.

However, without amplification at the key-side relay antenna, the key could only be excited from a few centimeters up to 2 m. With a power amplifier, we were able to achieve a range between 2 and 8 m, (with the key fob in the person's pocket which corresponds to the typical key placement). We note that the distance achieved between the relay antenna and the key depends on the strength of the collected signal from the car side and the sensitivity of the key. Finally, the values reported here show that the attack is practical as the key can be activated up to 8 meters away from the antenna and the distance from the key to the car can be extended up to 60 meters. It is likely that using more powerful amplifiers would further increase these distances.

[2] Antenna size 1.0 x 0.5 m Texas Instruments RI-ANT-G04E

Table 2.3: Experimental results on distance. Legend: '✓' relay works without amplification, 'A' with amplification, '-' not tested

Car	Relay cable				Key to antenna distance [m]			
	30 m		60 m		No Ampli.		With Ampli.	
	open	go	open	go	open	go	open	go
Model 1	✓	✓	✓	✓	2	0.4	-	-
Model 2	A	A	A	A	0.1	0.1	2.4	2.4
Model 3	✓	✓	✓	✓	-	-	-	-
Model 4	✓	✓	-	-	-	-	-	-
Model 5	✓	✓	✓	✓	2.5	1.5	6	5.5
Model 6	A	A	A	A	0.6	0.2	3.5	3.5
Model 7	A	A	-	-	0.1	0.1	6	6
Model 8	✓	A	-	-	1.5	0.2	4	3.5
Model 9	✓	✓	✓	✓	2.4	2.4	8	8
Model 10	✓	✓	-	-	-	-	-	-

2.3.2 Maximum Acceptable Delay

In order to measure the maximum theoretical distance of a physical-layer relay, we computed for each tested PKES system the maximum acceptable delay by relaying LF messages with a variable delay. For this purpose we used a USRP1 from Ettus Research [13] with LFRX and LFTX boards. This allowed us to receive and send messages at 135 KHz. However, we found that the minimal processing delay achievable by this software radio platform (SDR) was between 10 and 20 ms. This proved to be too slow on all but one PKES we tested.

The delay in a software defined radio device is mainly due to buffering and sending data over the USB to (resp. from) the computer for processing and the software processing. To reduce this delay we modified the USRP FPGA to bypass the RX (resp. TX) buffers and the communications with the computer. With this modification and appropriate configuration of the USRP the digitized signals were directly relayed by the FPGA from the receiving path to the transmitting path. We experimentally measured the resulting minimal delay to be 4 us. To insert an additional, tunable, delay we added a FIFO between the RX and TX path. Changing FIFO memory and the decimation rate allowed us to accurately test delays between 4 us and 8 ms. For larger delays we had to use an unmodified USRP with a tunable delay in

Table 2.4: Maximum delay, key response time and spread

Car	Max. Delay	Key Response	
		Time	Spread
Model 1	500 us	1782 us (±8)	21 us
Model 2	5 ms	11376 us (±15)	47 us
Model 4	500 us	-	-
Model 5	1 ms	5002 us (±4)	11 us
Model 6	10-20 ms	23582 us (±196)	413 us
Model 7	620 us	1777 us (±12)	25 us
Model 8	620 us	437 us (±70)	162 us
Model 9	2 ms	1148 us (±243)	436 us
Model 10	35 us	2177 us (±8)	12 us

software. This, however, reduced the delay precision.

Table 2.4 shows the measured maximum delays on the tested vehicles. Large delays allow to relay messages over longer distances. The maximum delays were measured to be within 35 us to tens of ms depending on the car model. This leads to a theoretical distance between 5 and 1500 km for the wireless physical-layer relay. With faster software defined radio platforms demodulation and modulation relays can also be envisioned [14].

2.3.3 Key Response Time and Spread

Other characteristics of the smart key that are relevant to the physical-layer relay performance are the key response time and spread. The key response time is the elapsed time between the moment when the challenge is sent by the car and the beginning of the response from the smart key. The key response time spread is the difference between the minimum and maximum key response times that we have observed. The computation of these two measures allows us to estimate (i) how much delay could the physical-layer relay attack exploit without any practical detection being possible (ii) what is the design decision behind the maximum acceptable delays allowed by the evaluated systems. We note that the numerical differences of these two measures between car models are due to the hardware used as well as the implementation of the secure protocols (e.g., message size, type of encryption).

In order to measure the key response time and spread, we recorded

the protocol message exchanges between the car and key with an oscilloscope using high sampling rate (from 20 to 50 MS/s depending on the PKES system). This allowed us to have a precise estimation (within tens of nanoseconds) of the start and end of transmitted messages. Table 2.4 summarizes the average key response time with its standard deviation and the key response time spread computed from 10 different message exchanges during car open.

The results show large differences between different car models. The key response standard deviations vary from 4 to 196 us, and the maximum spread - from 11 to 436 us. These values show that the current implementations exhibit large variance. That is, possible solutions that rely on measurements of the average key response time in order to detect the time delay introduced by our attack would be infeasible; even the smallest key response time spread of 11 us (Model 5) is already too large to be used for the detection of our attack. We recall that our 30 m wireless physical-layer relay requires approximately 120 ns in one direction (Table 2.2).

Moreover, we also observe that higher key response spread leads to higher acceptable delay. The manufacturers seem to fix the maximum acceptable delay at 20 to 50 times of the measured spread (except for Model 10). The reason is most likely to provide high reliability of the system as any smaller delays could occasionally make car owners being denied access to the car and/or authorization to drive.

2.4 Implications

Our realization of relay attack on PKES systems could be put into practice in numerous scenarios. In one scenario, the attackers can install their relay setup in an underground parking, placing a relay antenna close to the passage point (e.g., exit corridor). When the user leaves his car and exits the parking confident that his car is locked, a second attacker can place the second antenna to the door handle and wait until the car owner passes by the passage point. At that moment the key in his pocket will receive car signals and will send back an *open* command to the car. As this message is sent over UHF, it will reach the car even if the car is within several tens of meters. The car will therefore unlock. Once that the attacker has access to the car, the signals from within the car are relayed and the key will believe it is inside the car and emit the *allow start* message. The car can now be started and driven. Indeed,

most of the cars will detect the missing key, however for safety reasons, the car will not stop. Similarly, the car might detect a missing key for several other reasons including if the key battery is depleted. None of the evaluated cars stopped the engine if the connection with key was cut.

In a second scenario, the attacker can go with one relay antenna close to a window to activate a key left inside a closed building (e.g., on a table). This is possible when the antenna–key range is larger. In such case, if the car is parked close to the building, the attacker is able to open and start it without entering the building.

The described relay attack could not be easily traced unless the car keeps a log of recent entries and records exchanged signals (e.g., for later analysis). It will be difficult for the owner to prove that he is not the one that actually opened and used the car, as there are no physical traces of car entry. This can have legal implications for car owners in case that their cars or property from their cars are stolen due to this PKES vulnerability.

Our attack provides physical access to the interior of the car, and therefore can be used in conjunction with other attacks. For example, *rootkits* on car computers can be installed that allow an attacker to take control of the entire car [15].

2.5 Summary and Discussion

We showed that a set of modern car entry systems (PKES) are vulnerable to fast physical-layer relay attacks. While immediate or mid-term countermeasures could be envisioned [16], our attacks reveal two fundamental problems with proximity access control based on cryptographic shared keys for device authentication.

The first is related to physical proximity. Instead of verifying that the correct key is in its physical proximity, the car verifies if it can communicate with the key. In adversarial settings this cannot be taken as a proof of physical proximity. Possible solutions include distance-bounding protocols which guarantee that physical proximity is securely verified [17–23]. This means that the attacker cannot convince the car that the key is closer than it really is.

The second problem is related to the device identity used for secure device authentication. The authentication is performed based on what the devices (car and key) hold (a shared cryptographic key) and not on

what the devices really are. The key does not verify that the messages come indeed from the car. In relay attacks, they are always transmitted by a third device owned by the attacker.

Device authentication can potentially be enhanced using device identification techniques that verify what the device really is. This implies that one entity (e.g., key) stores a fingerprint of the other entity (e.g., car) and can measure this fingerprint from the communication with that entity. Given that the attacker's device would ideally not have the same fingerprint as the car, the key would not recognize it as being the car. It would therefore discard all retransmitted messages and the relay attack would be prevented.

In the reminder of this thesis, we present our research and results on extracting and analyzing fingerprints of wireless devices. More precisely, we focus on device identification based on analog (radio) circuitry properties based on that they exhibit and that can be observed at the physical communication layer.

Part II

Physical-layer Device Identification

Chapter 3

Device Identification Background

Physical-layer device identification is based on device fingerprints corresponding to unique characteristics extracted from the device's communication at the physical layer. Here we present its functionality in the broader concept of a physical-layer device identification system. We introduce the building blocks of such systems and the key concepts in their design and performance evaluation. These serve as a basis for the proposed systems in the remaining parts of this thesis.

3.1 System Overview

A typical physical-layer device identification system involves three entities as shown in Figure 3.1: a wireless device, a device identification subsystem, and an application using the identification. The goals is to identify (or verify the identity of) devices or their affiliation classes based on device characteristics which are observable from the radio communication. That is, physical-layer device identification systems acquire, process, store, and compare signals transmitted from devices.

Such an identification system can be viewed as a pattern recognition system. It acquires signals from devices, also referred to as *identification signals*, extracts identification-relevant information from the those signals, also referred to as *fingerprints*, compares the extracted

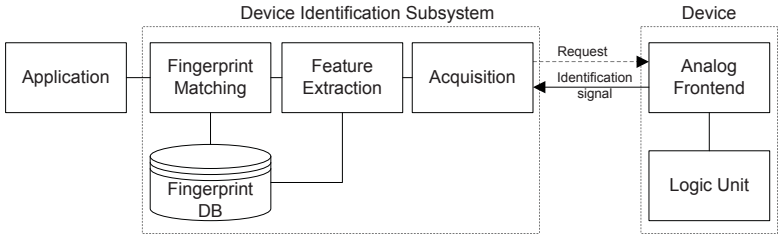


Figure 3.1: Building blocks of physical-layer identification system.

fingerprints with already enrolled device fingerprints and provides the relevant results to the application subsystem.

Typically, we distinguish two functional phases in the identification system: enrollment and identification. During enrollment, signals are captured from each device or a class of devices considered by the application. Fingerprints obtained from feature extraction are stored in a database typically linked with a device or class identifier. During identification, identification signals from the device (class of devices) are acquired, fingerprints are extracted and compared with the enrolled reference fingerprints. We consider the following modes of operation: (i) Device identity verification: verify that a device identity matches its claimed identity or class (1:1 comparison) (ii) Device classification: classify a device with unknown identity (but in the enrollment database) to the most likely device or class of devices in the enrollment database (1:N comparisons).

In this thesis, we focused on device identity verification as this mode of operation is most suitable for security applications such as intrusion detection, device authentication and relay detection. Device classification may also be a valid mode of operation in certain close-world scenarios. These scenarios are of very limited use. Therefore, we provide results on device classification only to compare with related work.

3.2 System Entities

In the following, we briefly discuss the main physical-layer device identification system entities, processes and evaluation criteria.

3.2.1 Device

Physical-layer device identification is based on fingerprinting the analog circuitry of devices by observing their radio communication. Consequently, any device that uses radio communication may be subject to physical-layer identification. Wireless devices are typically composed of antennas, analog front-ends, digital back-ends and logic units with different levels of complexity. For example, active wireless transceivers (e.g., 802.15.4 sensor nodes) are significantly more complex than passive RFID transponders.

The feasibility of uniquely identifying wireless devices among other devices or classes of devices depends on the hardware impairments introduced during manufacturing of the analog circuitry and its components (e.g., resistors, capacitors, transistors). Understanding the sources of variability within a given device or class of devices has relevant implications on the design of the device identification system.

3.2.2 Identification Signals

Wireless devices communicate via radio signals by sending data according to specifications and protocols. Given that physical-layer identification extracts unique characteristics from the radio signals during communication, we define *identification signals* as the signals that are collected for the purpose of identification.

Identification signals can be either device transmissions during normal operation or specific transmissions for the purpose of identification. Therefore, identification signals can be either passively acquired by the system (e.g., by recording device transmissions) or actively triggered (e.g., by sending a request to the device and recording its response). This procedure may be repeated a sufficient number of times such that the system is able to extract the device characteristics needed for identification.

3.2.3 Features and Device Fingerprints

The characteristics extracted from the identification signal for the purpose of identification are referred to as features. They could be located in specific data or non-data parts of the identification signal. They often relate to well-understood signal characteristics defined or not in device specifications. These specifications device typically include various

characteristics and error tolerances (e.g., amplitude, frequency). Some of them are used for quality control and FCC certification. In addition, features can be extracted from signals without a-priory knowledge of a specific signal characteristic. They can also be further enhanced by statistical analysis in order to improve their discriminant capabilities.

The purpose of the features is to form device fingerprints, also referred to as device physical-layer identities. A device fingerprint is a set of features or a composition of features. Depending on the application, various fingerprint properties can be required such as universality (every device must have the features), uniqueness (no two devices should have the same fingerprints), permanence (fingerprints should be robust, e.g., invariant to distance and location, stable over a specified time interval). Other application-specific requirements may be related to acquisition and data-dependency of the device fingerprints (Section 12).

3.3 System Processes

The physical-layer identification system (Figure 3.1) has to acquire identification signals (acquisition), extract features and form fingerprints (feature extraction), and compare fingerprints (fingerprint matching). The system may either passively collect identification signals or it may actively request devices to send back specific responses.

3.3.1 Acquisition

The acquisition process ensures capturing, digitalizing and storing the identification signals. It should neither influence nor degrade (e.g., by adding noise) the signals needed for identification, i.e., it should preserve and bring into the digital domain the unique signal characteristics which the identification relies on.

Acquisition can be passive and/or active. In passive acquisition, the system acquires identification signals without interacting with the devices, e.g., identification signals can simply relate to data packets sent by devices during normal communication with other devices. In active acquisition, the system acquires the identification signals after requesting the devices to transmit them. Besides the advantages of obtaining identification signals "on demand", active acquisition may request identification signals that are not part of the communication specification.

3.3.2 Feature Extraction

The feature extraction process deals with extracting features from identification signals that are used to distinguish devices or classes of devices. Typically, it implements functions that directly relate the identification signal to the features. For example, when considering features like modulation errors, the feature extraction process is a demodulator with functions to compute these errors. Feature extraction can also be based on time domain and/or spectral transformations of the identification signal to create new distinguishable features. These transformations could be further enhanced statistical analysis and/or feature selection techniques [24].

3.3.3 Fingerprint Matching

Fingerprint (feature) matching compares newly extracted device fingerprints with reference fingerprints enrolled in the system database. Depending on the application, it can provide an yes/no answer whether a device fingerprint matches a chosen reference fingerprint (identity verification) or a list of devices that the device fingerprint most likely originated from (identification). The choice of the matching algorithm depends on the extracted features and the application requirements.

3.4 Performance Evaluation

Identification systems are typically evaluated in terms of accuracy, robustness, computational speed, exception handling, cost and security.

3.4.1 Accuracy

A critical performance factor is often considered to be the accuracy, i.e., how precise is the system during identification. Here we detail our methodology and metrics to measure the accuracy in device identity verification. For completeness, we also discuss device classification.

Device Identity Verification

We evaluate the accuracy of our system based on the methodology for threshold-based identity verification since this is the most widely accepted way for evaluating biometric systems [5]. We adopt Equal

Error Rate (EER) and Receiver Operating Characteristic (ROC) as performance metrics. Their definition and computation are discussed below. We note that the system accuracy often cannot be theoretically established, but only statistically estimated using test databases.

The EER and ROC are based on the errors that occur during hypothesis testing that establishes matching between two samples. The *null* hypothesis H_o states that the two samples match and the *alternative* hypothesis H_a - that the two samples do not match. In such a setting, there are two possible errors: False Match and False Non-Match. False Match means that the system decides H_o when H_a is true. In our system this is equivalent to a decision that a device's (claimed) identity is legitimate while in reality it is an imposter device. We refer to it as a False Accept. False Non-Match means that the system decides H_a when H_o is true. In our system, this is equivalent to a decision that a device's identity is not legitimate while in reality it is. We refer to it as a False Reject.

The False Accept Rate (FAR) and False Reject Rate (FRR) represent the frequencies at which the above errors occur. The FAR and FRR are closely related to each other in the Receiver Operating Characteristic (ROC). The ROC is a curve which allows to automatically compute FRR when the FAR is fixed at a desired level and vice versa [5]. The operating point in ROC, where FAR and FRR are equal, is called the Equal Error Rate (EER). The EER represents the most common measure of the accuracy of an identification system [25].

We estimate the ROC and EER as follows. We compute the similarity score between all reference and test fingerprints from all devices. We then separate these scores in two categories: genuine and imposter. The genuine category includes all scores from matching two fingerprints from the same device. The imposter category contains all scores from comparing two fingerprints from different devices. Given that each score represents the similarity between two fingerprints (identities), we compute the rate of falsely rejected and falsely accepted fingerprints using a threshold. The scores from the genuine category that are above this threshold indicate the number of false rejects or the FRR, while the scores from the imposter category that are below the threshold indicate the number of false accepts or the FAR. The EER is the error rate where both FAR and FRR are equal. The value of the threshold at the EER is our threshold T for an accept/reject decision.

For ROC presentation, we use Genuine Accept Rate (GAR = 1 - FRR) instead of FRR because GAR shows the rate of accepts of

legitimate identities for a given FAR (e.g., FAR = 1%).

Our methodology for EER and ROC validation is based on the method of cross-validation [26]. Cross-validation consists of partitioning the dataset into disjoint subsets, training the system and validating it on independent subsets. We perform multiple rounds of cross-validation using different partitions, and the presented estimates are averaged over these rounds.

Device Classification

Device classification has often been used in evaluating device fingerprinting techniques even though for security applications it is not the correct accuracy metric (see Chapter 9). The system operating in classification mode measures unknown devices, obtains their fingerprints and assigns them to one of the devices (or classes) in its available set of devices. The commonly adopted metric in that case is the classification error rate (CER) which is defined as the percentage of incorrectly assigned device fingerprints to their respective devices or classes. We use this metric in certain cases to compare the efficacy of our proposed techniques to related work. Our methodology for CER validation is also based on cross-validation.

3.4.2 Feature Stability

Another important characteristic of physical-layer device identification system is the robustness of the measured fingerprints. We also refer to it as feature stability. The identification system may need to operate in different scenarios and conditions, e.g., identify devices from different distances and location. Therefore, the performance evaluation should ideally include the stability of the fingerprints to (i) external environmental aspects that directly influence the identification signal propagation such as signal interference, multipath and distance/location and (ii) device-related aspects like internal temperature and voltage.

Understanding feature stability is crucial to understanding the possible applications of physical-layer identification.

3.4.3 Other Performance Factors

Other performance factors also include computational resources, system cost and exception handling. In physical-layer identification the com-

putational resources such as acquisition speed, memory consumption and system cost are directly related to the hardware being used. The higher the hardware quality, the higher the cost. Where necessary, we discuss these issues in the case of our proposed prototype systems. We also suggest alternative implementations and improvements to reduce the overall cost of the identification system.

Identification systems including physical-layer device identification systems usually need an exception handling procedure. Typically, Failure to Use, Failure to Enroll and Failure to Acquire events need to be appropriately handled [5]. In this thesis, we do not discuss exception handling solutions as they are orthogonal to the identification system. We invite the reader to consult [5] for related discussion.

Chapter 4

Transient-based Identification of Wireless Transceivers

Identification of wireless transceivers based on the characteristics of their radio transmissions can provide an additional layer of security in wireless networks. This layer can be used to detect and/or prevent intrusion, replication and relay attacks. In this chapter, we investigate the feasibility of uniquely identifying same model and manufacturer low-power IEEE 802.15.4 (CC2420) radio devices using discriminant information present in the turn-on transient part of radio packet transmission. We propose a system comprised of acquisition, feature extraction and matching procedures and show that it enables accurate device identification under certain assumptions. We further explore the performance of our approach in terms of distance, antenna polarization, voltage and temperature and analyze how these parameters affect the accuracy. The obtained results expose the limitations of using transient-based identification in dynamic (mobile) environments.

Last, but not least, we validate the applicability of our proposed transient-based identification system to other radio transceivers. We show that it achieves similar accuracy on a set of low-power CC1000 radio transceivers. Therefore, our approach is likely to be applicable to a wider range of modern transceivers.

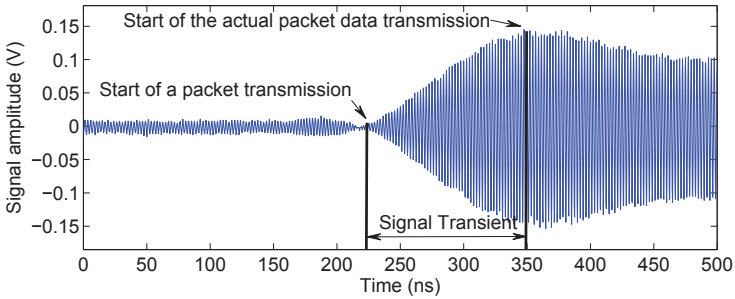


Figure 4.1: Turn-on transient at the start of each new packet transmission – IEEE 802.15.4 CC2420 radio transceivers.

4.1 System Overview

Our transient-based device identification system is an instance of the physical-layer device identification system in Chapter 3. We refer to it as transient-based because the physical-layer fingerprints are extracted from the turn-on transient in device packet transmissions. It is the part of the signal before data modulation where the signal amplitude gradually rises to a specified level (Figure 4.1). The unique properties of the transient are believed to originate in the analog circuitry which includes amplifiers, filters, mixers and transmitting antenna. Each of these components contains a number of passive (e.g., resistance) and active (e.g., capacitance, transistor) components which contribute to the behavior of the transient signal.

Our system consists of a single hardware acquisition setup with feature extraction and matching components implemented in software. The hardware setup acquires packet radio transmissions from devices and extracts the turn-on transient signal from each packet transmission. Feature extraction builds device fingerprints from a number of collected turn-on transients (identification signals). Fingerprint matching process verifies whether the measured fingerprints correspond to the stored fingerprints of the device during enrollment.

In terms of wireless transceiver devices, we considered 50 COTS Tmote Sky sensor nodes equipped with Chipcon CC2420 low-power radio transmitters. All devices were same model and manufacturer with a signature "4M 94V-0 H014-4787". Given that they were purchased

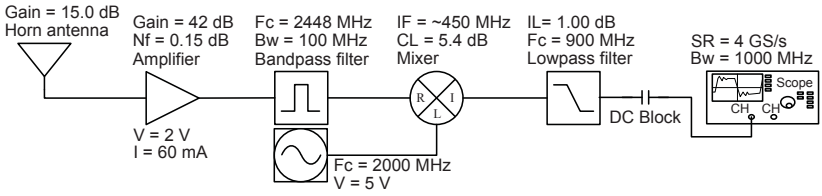


Figure 4.2: Hardware signal acquisition setup.

in two sets, we cannot fully assert that they were all produced at the same production line, even though such an assumption is highly plausible. We also validated our system on a number of Mica2 sensor nodes equipped with Chipcon CC1000 radios.

4.2 Signal Acquisition

4.2.1 Hardware Setup

Figure 4.2 shows the schematic of the hardware setup used to capture turn-on transient signals. The signals are acquired by a Standard Horn directional antenna and subsequently amplified by an ultra low-noise and low-power amplifier with a noise figure of 0.15 dB. Due to the low power of the sensor devices, it is critical to amplify the signal without losing its unique characteristics, as the signal-to-noise ratio degrades drastically within meters. An ultra low-noise and low-power amplifier proved to be the best choice among a number of amplifiers we tested. Figure 4.1 visualizes a turn-on transient signal acquired from 10 m.

We use a low insertion loss bandpass filter to eliminate radio frequencies outside the IEEE 802.15.4 band [27]. We then down-convert this amplified and filtered signal to an intermediate frequency of 450 MHz using a frequency mixer and a synthesizer. We down-convert the signal in order to acquire it with sufficient precision on a 1 GHz oscilloscope. If the transmitted 2.4 GHz signals are not down-converted, the oscilloscope significantly attenuates the received signals (25 dB less). Another solution is to use larger bandwidth oscilloscope. However, we did not have this type of equipment at our disposal.

Due to the frequency artifacts during conversion, we pass the intermediate frequency signal through a lowpass filter and a DC blocking capacitor. We then record it with a sampling rate of 4 GS/s.

Table 4.1: Acquired datasets.

	Goal	Distance	# Signals	# Devices	Total
1	Accur.	10 m	600	50	30000
2	Accur.	40 m	600	10	6000
3	Volt.	10 m	200	10	2000
4	Polar.	-	600	10	6000

4.2.2 Collected Data

During data collection, each device was positioned on the same tripod, previously fixed at a given distance from the fingerprinter's antenna. Polarizations of the sensor devices' antennas (all devices were equipped with standard on-board integrated antenna) and of the fingerprinter's antenna were aligned and perpendicular to the ground. The devices were run on 2 x 1.5V AA batteries (Dataset 1,2,4) and 2 x 1.2V AA batteries (Dataset 3). The experiments were made indoors (Dataset 1,3,4) and in an underground parking space (Dataset 2) for about 20 minutes with equally spaced packet transmissions in order to acquire a large number of signal samples for performance evaluation. The ambient temperature of the environment was varying between 18 and 23°C. The recorded datasets and main measurement parameters are summarized in Table 4.1.

4.2.3 Transient Extraction

From each acquired signal (one signal corresponds to one packet), we extracted its turn-on transient. It should be noted that in a regular transmission from the nodes, the transient is present in each transmitted packet. Each acquired signal trace lasted 500 ns, of which the transient consistently occupied approximately 125 ns for all devices in our population (Figure 4.1). Given the 4GS/s sampling rate of our oscilloscope, this corresponded to approximately 500 data points. We therefore defined the identification signal as the first 512 data points from the detected starting point of the transient determined according to the slope detection in [28].

We also note that we observed very similar and stable maximum signal amplitudes at the fingerprinter's antenna. The large majority of devices had a mean amplitude of approximately 145-7 mV with a standard deviation of about 4 mV (Dataset 1).

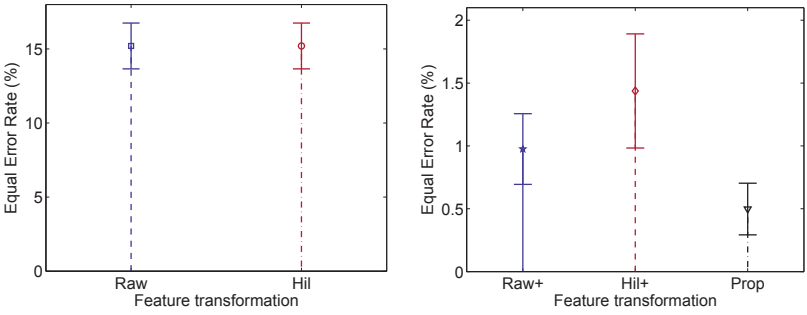


Figure 4.3: Accuracy after initial transformation (Dataset 1).

4.3 Feature Extraction

The goal of feature extraction is to obtain distinctive feature templates (fingerprints) from the identification signals. Our feature extraction process consists of two phases: (1) initial transformation and (2) feature extraction using statistical analysis. The initial transformation is chosen from a set of known transformations and is an input into a Linear Discriminant Analysis (LDA) for feature extraction [26].

In the initial transformation phase, we experimentally test a number of signal transformations to find initial features that capture most discriminant information. In the statistical analysis phase, we statistically reduce the dimensionality by discarding noisy dimensions using LDA projections. We note that LDA has been effectively applied to discriminate human biometrics [29, 30] and outperforms related methods when the training data is sufficiently large [31].

4.3.1 Initial Transformation

We considered the following initial transient transformations: *Raw* - the original identification signal (raw transient), no transformation; *Hil* - the envelope of the identification signal obtained by the Hilbert transform [32]; *Raw+* - the FFT spectra of the identification signal; *Hil+* - the FFT spectra of its envelope; *Prop* - relative differences between adjacent FFT spectra of the identification signal.

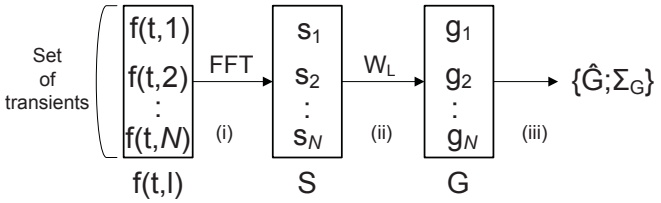


Figure 4.4: Feature extraction process.

We tested the use of these initial transformations in our identification system. The results using Dataset 1 are summarized in Figure 4.3 in terms of EER. They show that when using the original identification signals (*Raw*) or their envelopes (*Hil*), our system scores a high EER (15%) which translates into a low identification accuracy. This makes these two transformations unsuitable for further analysis. Using FFT spectra significantly decreases the error rate (*Raw+, Hil+, Prop*), with (*Prop*) scoring the lowest EER. We therefore chose the relative differences between adjacent FFT spectra (*Prop*) as the transformation for further feature extraction.

The above results were validated with 4-fold cross validation [26]. Three folds of Dataset 1 were used for training and the remaining one fold for testing. Each fold contained 150 identification signals per device. This resulted in a total of 300 genuine and 22050 imposter matchings per fold[1] to compute the EER.

4.3.2 Feature Extraction

In this section, we describe our feature extraction process based on the relative differences between adjacent FFT spectra as initial transformation.

For a given device, spectral Fisher-features are extracted from N identifications signals using a linear transformation derived from LDA. Figure 4.4 illustrates the process. First, we extract the transient part of the recorded signal l. We denote this part by $f(t, l)$, where $f(t, l)$ is

[1]Each fold contains 3 feature templates (fingerprints) per device. This results in 6 different matchings of fingerprints of the same sensor node (i.e., genuine matchings) and 441 different matchings of fingerprints from different sensor nodes matching(i.e., imposter matchings). This makes 300 genuine and 22050 imposter matchings for 50 devices.

the amplitude of the signal l at time t.

In Step (i), we apply a one-dimensional Fourier transformation on $f(t,l)$ to obtain $F(\omega,l)$:

$$F(\omega,l) \quad = \quad \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} f(t,l) \exp(-2\pi i \frac{t\omega}{M}) \tag{4.1}$$

where M is the length of transient and $0 \leq t \leq M-1$.
We then compute the relative difference between the adjacent spectra of the $|F(\omega,l)|$ denoted in a vector form as: $\vec{s_l}$ = [$|F(2,l)| - |F(1,l)|$ $|F(3,l)| - |F(2,l)|$ $\cdots$ $|F(M/2-1,l)| - |F(M/2-2,l)|$]^t where the DC component and redundant half of the spectrum are removed.

In Step (ii), a projected vector $\vec{g_l}$, also called a Fisher-feature, is extracted from the Fourier spectrum using an LDA matrix W_L:

$$\vec{g_l} = W_L^t \vec{s_l} \tag{4.2}$$

Based on the above description, the Fisher-feature extraction from N identification signals for a given sensor device is written as $G = W_L^t S$ where G is an array of g_l and S is a matrix S = [s_0 .. s_l .. s_N].

Finally in Step (iii), the feature template $\boldsymbol{h}$ used for matching (recognition) is computed:

$$\boldsymbol{h} = \{\hat{G}; \Sigma_G\} \tag{4.3}$$

where $\hat{G}$ denotes the mean vector of G and Σ_G denotes the covariance matrix of G.

The number of identification signals N used to build the feature template and the number of projected vectors in W_L (i.e., the Fisher subspace dimension) are experimentally determined.

4.3.3 Training and Feature Matching

The LDA matrix W_L is derived by a standard LDA procedure based on scatter matrices [26]. Here, W_L is the optimal Fisher discriminant projection given as the set of κ eigenvectors in matrix W that correspond to the κ-highest eigenvalues in the generalized eigenvalue problem: $S_b W = \Lambda S_w W$, where Λ is the eigenvalue matrix, S_w is the within-class scatter matrix showing the average scatter of sample features $\boldsymbol{h}$ from the same sensor device and S_b is the between-class scatter

representing the average scatter of sample features $\boldsymbol{h}$ from different sensor devices.

Mahalanobis distance is used to find the similarity between feature templates (fingerprints). The result of matching a reference $\boldsymbol{h}^R$ and a test $\boldsymbol{h}^T$ feature templates is a matching score, calculated as follows.

$$Matching\ score = \sqrt{(\boldsymbol{h}^T - \boldsymbol{h}^R)^t \Sigma_G^{-1} (\boldsymbol{h}^T - \boldsymbol{h}^R)} \tag{4.4}$$

Values of the matching score closer to 0 indicate a higher similairy.

It should be noted that the proposed feature extraction and matching method can be efficiently implemented in hardware as it uses only linear transformations for feature extraction and inter-vector distance matching to compute similarity. These operations have a low memory footprint and are computationally efficient.

4.4 Performance Evaluation

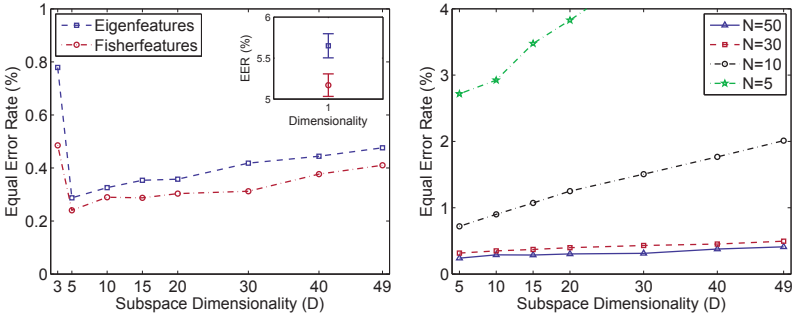
4.4.1 Accuracy analysis

In our evaluation, we first considered Dataset 1 which that contained identification signals from all 50 devices taken at distance of 10 meters. The number of identification signals used to build feature templates was fixed to $N = 50$. The results, validated by 4-fold cross validation, are presented in Figure 4.5(a). They show the accuracy (EER) of our system depending on the subspace dimensionality. The original feature dimensionality after the initial transformation was $D = 254$.

The results demonstrate a very small EER, which is, for dimensionality $D \geq 3$ between 0.0024 (0.24%) and 0.005 (0.5%). This means that our system correctly identifies the same-model-same-manufacturer transceivers with an accuracy higher than 99.5% (GAR at the EER operating point). We later show that the accuracy achieved in this set is also preserved for larger distances.

Figure 4.5(a) suggests that using the first 5 eigenvectors for projection is sufficient for accurate identification. This results in compact fingerprints[2]. EER degrades progressively for higher dimensional subspaces. This phenomenon is more pronounced when N decreases, in particular for $N < 30$ as shown in Figure 4.5(b).

[2]If each dimension is represented by a 4-byte floating-point number, the feature template $\boldsymbol{h}$ size is 20 (5x4) bytes for $\bar{G}$ plus 100 (5x5x4) bytes for Σ_G resulting in a total of 120 bytes.



(a) (b)

Figure 4.5: (a) Eigen- and Fisher-features accuracy for different subspace dimensionality. Dimensionality 1 is in the inner plot. (b) Fisher-features accuracy for different subspace dimensionality and identification signals N (Dataset 1).

Figure 4.5(a) also compares Eigen- vs. Fisher-feature extraction. Eigen-feature extraction is based on Principal Component Analysis (PCA). The validated EERs show that Fisher LDA is more effective for lower dimensional subspaces (1-3 eigenvectors). However, we cannot assert this with statistical confidence for higher dimensional subspaces.

In order to fully characterize the system accuracy, we plot the ROC for the 5-dimensional features versus the number of identification signals N (Figure 4.6). Reducing N degrades the Genuine Accept Rate for lower FAR (e.g., 0.01%). This is not readily visible in Figure 4.5(b) where the differences in EER for $N > 10$ are statistically insignificant (Table 4.2). The ROC analysis suggests that if an application is required to operate at low FAR ($< 0.1\%$), it must use more identification signals in order to achieve high GAR.

Table 4.2 summarizes the underlying data, namely the number of identification signals N, total genuine and imposter matchings performed, Accept/Reject threshold T (at EER point), EER and its confidence interval (CI).

4.4.2 Stability analysis

In the following analysis, we investigate the stability of our proposed technique in terms of distance, antenna polarization, voltage and tem-

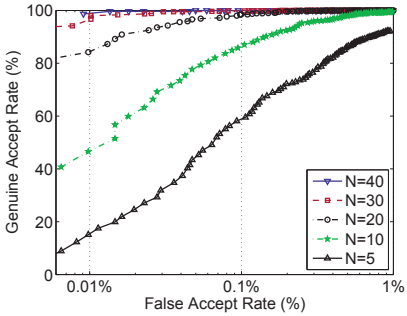


Figure 4.6: Receiver operating characteristic (ROC) for different number of identification signals N used to build the feature template (Dataset 1). The Fisher-feature subspace dimensionality is fixed to 5. See Table 4.2 for the underlying data.

Table 4.2: Summary of accuracy for Dataset 1.

N	Test matchings		Thr T	EER (%)	EER CI (%)	
	Genuine	Imposter			lower	upper
50	300	22050	3.01	0.24	0	0.49
40	300	22500	3.95	0.34	0.02	0.66
30	600	39200	3.87	0.32	0.07	0.56
20	1000	61250	4.10	0.34	0.21	0.47
10	1000	61250	6.74	0.72	0.62	0.82
5	1000	61250	16.04	2.72	2.38	3.06

perature. We also validate our system on other radio transceivers.

Distance

For distance evaluation, we performed measurements in the university parking, which allowed us to collect signals up to 40 m line-of-sight (LoS). We used the first 10 devices from our population (Dataset 2).

Table 4.3 compares the validated EERs for different N and distances of 10 and 40 m respectively. The system is trained separately for each distance. We did not observe statistically significant effects on the system accuracy. Our acquisition setup was successful in preserving the features in the transient signal.

In order to complete the analysis on the effect of distance on the

Table 4.3: EER at 10 and 40 meters (Dataset 1-2).

N	Test matchings		EER (%)		Valid.
	Genuine	Imposter	10m	40m	
50	60	810	0	0	4-fold
40	60	810	0	0	5-fold
30	120	1440	0	0	5-fold
20	200	2250	0.57	0.36	5-fold
10	200	2250	1.35	3.41	5-fold

recognition accuracy, we performed cross-matching between feature templates extracted from both distances. We registered a significant increase of EER = 0.38 (38.01%) for $N = 50$. This result shows that while the frequency information in the transient signal is unique within a given distance, it changes across different distances for the same antenna polarization.

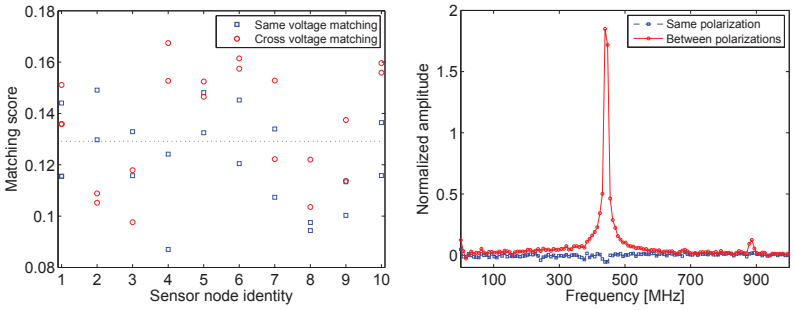
Polarization

In order to quantify the effect of antenna polarization, we collected identification signals under the same conditions as in Dataset 1, but with a changed polarization of the antenna on the device by 45° with respect to the fingerprinter antenna. We then matched the extracted feature templates to the reference feature templates in Dataset 1. This resulted in a impractical EER = 0.39 (39%).

As this result could have been influenced by the training procedure where only training data from one type of polarization was used, we collected transient data samples from 10 sensor nodes at 3 different antenna polarizations (Dataset 4). The identification accuracy did not improve, the reason being that varying the polarization changed the frequency information in the identification signal (Figure 4.7(b)). These changes could not be separated by a linear discriminant. We acknowledge that further work is needed to quantify how much change in polarization can be tolerated (e.g., small perturbations).

Voltage and Temperature

For voltage evaluation, we used 2x1.2V NiMH and 2x1.5V alkaline batteries which provided two different voltage levels of 2.4 and 3V respectively. Figure 4.7(a) shows the matching scores between fingerprints



(a) Voltage (b) Polarization

Figure 4.7: (a) Matching fingerprints acquired at the same voltage and between two different voltage levels provides similar values, within the genuine distribution. (b) Turn-on transients captured at two different antenna polarization present large spectral differences.

taken at the same voltage level (blue triangles) and between fingerprints taken at different voltage levels (red circles) for 10 devices. We do not observe a significant difference between genuine matching scores coming from the same and between voltage levels. The scores are close to 0 and within the boundary of the genuine score distribution (i.e., below $T = 3.01$). The EER for this set of 10 nodes remained 0.

This is an expected result given that the sensor nodes are equipped with a low-power micro-controller which requires 2.1 – 3.6 V for operation. It should be noted that such a result is not necessary true for high-power transmitters [33].

Our experiments did not show that surrounding temperature changes affect the accuracy. We point out however that the ambient temperature during our experiments did not vary substantially, the variance being approximately 5°C between the two environments used. We did not investigate extreme changes of temperature (e.g., intentional heating) and higher variance of the ambient temperature which usually occurs in outdoor environments.

4.4.3 Device classification and comparison

In application scenarios where the number of devices is known, CER can be used to evaluate the ability of a system to classify identifica-

Table 4.4: Average CER (Dataset 1).

N	# Samples	1-NN (%)	3-NN (%)	Valid.
50	300	0.07	0	4-fold
40	300	0.07	0	5-fold
30	600	0.25	0.07	5-fold
20	1000	0.97	0.45	5-fold
10	1000	3.71	2.43	5-fold

tion signals to their corresponding devices. Table 4.4 summarizes the average CER of our system on Dataset 1 using k-Nearest Neighbor classifiers. We provide these results for direct comparison with related work 9. It should be noted however that comparison by value could be misleading given the differences in device population (same vs. different manufacturers), radio hardware and experimental parameters.

We also applied our proposed features to the data collected by the authors in [28]. It consisted of 2000 transient data samples captured from 10 Mica2 devices equipped with CC1000 (433Mhz) radios from 15 cm distance. The transient part occupied approximately 100 ns (200 data points). Our system scored an EER=0.0167 (1.67%) on that data, showing that CC1000 radios can also be identified with high accuracy. It should be noted that this result can possibly be improved if the linear transformation W_L is trained for CC1000 radios. This was not possible due to the small size of the available data.

In terms of device classification, our system achieved a CER of 3.2% compared to the 30% reported in [28]. In that particular case, we can assert with certainty that our transient-based identification system significantly improves over related work.

4.5 Summary and Discussion

We investigated transient-based identification of low-power 802.15.4 transceivers (Chipcon CC2420). Our proposed system enabled accurate device identification as long as devices did not change their location and distance with respect to the signal acquisition antenna. It was also validated on other low-power transceivers (Chipcon CC1000) where it significantly improved previously reported results.

We also evaluated the performance of our techniques with respect to

distance, antenna polarization and voltage. We showed that large fixed distances and variable voltage preserve fingerprint properties, whereas varying distance and antenna polarization distort the fingerprints and cause significantly lower identification accuracy. These findings severely limit the usability of transient-based identification in mobile scenarios where devices change their locations.

The reason for the discussed behavior is most likely due to the presence of channel-specific characteristics in the turn-on transient introduced by different multi-path propagation, path loss and polarization. These create large differences in the frequency spectrum of the transient and prevent accurate fingerprint comparison. While channel estimation procedures could in theory be applied to estimate the channel and reduce its effect on the device communication, our efforts in that direction were not successful. More precise channel compensation is required in order to preserve the unique device features. It is an open research question under which type of wireless channels this may be possible.

Another issue with transient-based identification in general is the requirement on high-quality hardware. Transients are very short in time and acquisition requires fast analog-to-digital conversion (ADC). From practical considerations, it is important to investigate whether turn-on transients contain discriminant information at lower intermediate frequencies. This will significantly reduce the hardware cost.

Chapter 5

Physical-layer Identification of Passive RFID

RFID technology is deployed in a number of devices such as contactless identity cards, electronic passports, payment credit cards, consumer products. Many studies have addressed security and privacy issues related to RFID deployment on the logical layer. Examples include device authentication, cloning detection, unauthorized tracking, inventorying.

In this chapter, we investigate the security and privacy implications of RFID devices by looking at the properties of their physical communication layer. In particular, we focus on physical-layer identification of HF and UHF RFID devices. We present a hardware setup for RFID signal acquisition to in- and out-of-specification reader requests and a set of techniques which extract timing, modulation and spectral features from the acquired signals. We evaluate our system on HF RFID smart cards, UHF RFID tags and a set of electronic passports. We demonstrate that RFID exhibit physical-layer characteristics with different properties. Some of these can be effectively used to build accurate device fingerprints and therefore enable detection of cloned RFID devices. Others can be used in an offensive manner to perform unauthorized tracking of devices despite any logical layer protection mechanism.

5.1 System Overview

We consider a physical-layer device identification system that consists of a single acquisition setup and feature extraction and matching modules implemented in software. For RFID signal acquisition, we use a purpose-built reader to transmit in- and out-of-specification requests and record the corresponding RFID device responses, also referred to as identification signals. Given that HF and UHF RFID acquisition setups are similar, we only detail the signal acquisition setup in case of HF RFID. For details on UHF RFID, we invite the reader to consult [34].

Our identification system operates in two modes, namely RFID device identity verification and classification. The former is used to assess the ability to verify the identity of same model and manufacturer RFID devices. In case of identity documents, this could mean identifying documents from the same country, year and place of issuance. The latter is used to check the ability to associate RFID devices to predefined classes. In case of identity documents, classes may include the country that issued the document or the year of issuance.

For performance evaluation, we considered HF and UHF RFID device populations (Table 5.1). Our primary set of HF devices consisted of 50 JCOP NXP 4.1 smart cards [35] which contain ISO 14443 compliant NXP RFID transponders [36]. We chose these cards since they are popular for use in identity documents and access cards, and because they have been used by hackers to demonstrate cloning attacks against e-passports [37]. We also validated our techniques on 8 HF RFID electronic passports[1]. In case of UHF RFID, we used a primary set of 50 ALN9540 tags compliant with EPCglobal UHF Class 1 Generation 2 (EPC C1G2) standard [38]. We chose these particular tags as they are representative for industrial applications.

5.2 Signal Acquisition

In this section, we first describe our signal acquisition setup. We then detail the different types of experiments we performed and present the collected datasets from our population of devices.

[1]The small quantity of the electronic passports used in the experiments is due to the difficulty of finding people who are in possession of such passports and at the same time willing to allow experimentation on them.

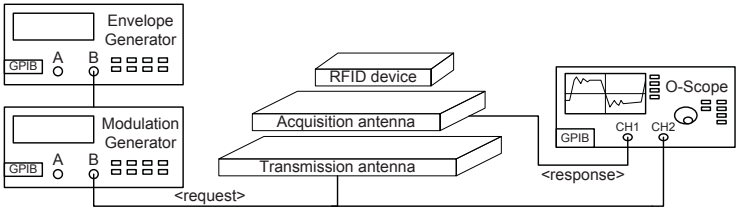


Figure 5.1: Signal acquisition setup. Envelope and modulation generators generate wake-up signals that initiate the response from the RFID transponder. This wake-up signal is transmitted by the transmitting antenna. The acquisition antenna captures both the wake-up signal and the response from the transponder. The signal from the acquisition antenna is then captured and recorded by the oscilloscope.

5.2.1 Hardware Setup

Figure 5.1 displays the hardware setup that we use to collect RF signals from the HF RFID devices. Our setup is essentially a purpose-built RFID reader that can operate within the standard-specified RFID communication [36] and out of specification. This enables a broader range of experiments. The setup consists of two signal generators used for envelope generation (envelope generator) and for signal modulation (modulation generator) as well as two transmission and acquisition antennas. The envelope generator is loaded with a waveform that represents the communication protocol wake-up command[2] required for initiating communication with the RFID. The envelope waveform is then sent to the modulation generator and is modulated according to the ISO/IEC 14443 protocol Type A or B, depending on the device being used. The modulated signal is then sent over the transmission antenna. Finally, the wake-up signal and the RFID response (identification signal) are received at the acquisition antenna and digitized with the oscilloscope. The separation of the envelope generation and modulation allows to independently vary baseband and RF characteristics in our experiments.

In order to collect identification signals, we make use of an antenna arrangement (Figure 5.2 (b) where the acquisition antenna is positioned between the transmission antenna and the RFID. A wooden platform holds the transmission and acquisition antennas in a fixed position to

[2]ISO/IEC 14443 for RFID communication defines two different communication protocols, Type A and B, which use different wake-up commands.

(a) (b)

Figure 5.2: (a) Transmission and acquisition antennas. (b) An electronic identity document being placed in the fingerprinting setup.

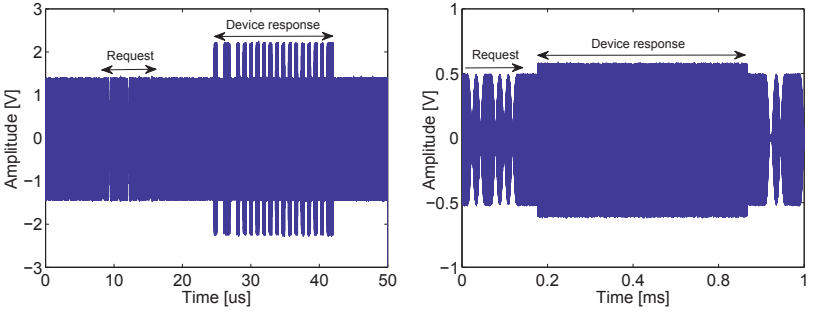
avoid changes in antenna position. The platform is separated from the desk by a non-conductive wooden cage. The transmission and acquisition antennas are both connected to an oscilloscope. We use the reader signal at the transmission antenna to trigger the acquisition at the oscilloscope. We note that device responses can also be observed at the transmission antenna. Given that our acquisition antenna had higher gain, we opted for the described setup to obtain better signal quality.

5.2.2 Performed Experiments

Using the proposed setup, we performed the following experiments.

Experiment 1: In this experiment we initiate communication with the RFID according to its specification. In case of HF RFID, the envelope generator generates Type A or B envelopes in baseband at the nominal bit rate of $F_b = 106$ kbit/s. The modulation generator modulates the baseband signal at the standard carrier frequency $F_c = 13.56$ MHz using 100% ASK for Type A and 10% ASK for Type B[3]. In UHF RFID, the envelope generator outputs a select command with phase-reversal amplitude shift keying (PR-ASK) baseband modulation according to [38]. The modulation generator up-converts the baseband signal to one of the standard carrier frequencies $F_c = 866.7$ MHz. For both technologies, the experiment consists of the following steps. A

[3]For 100% ASK modulation, we used pulse modulation as the built-in amplitude modulation (AM) in our generators could not reach the required precision.

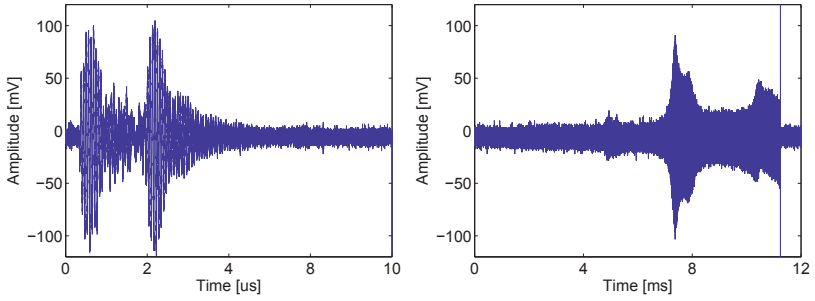


(a) HF RFID (b) UHF RFID

Figure 5.3: Examples of RFID reader request and response. (a) HF RFID communication according to ISO/IEC 14443 Type A ($F_c = 13.56$ MHz). (b) UHF RFID communication according to EPCglobal UHF Class 1 Gen. 2 ($F_c = 868$ MHz).

period of unmodulated carrier is transmitted to power the device and the oscilloscope begins recording the data at this instant. The carrier is then modulated according to the wake-up command specification. When the commands are no longer transmitted, an unmodulated period of carrier is maintained to get the device response. The carrier is turned off between each observation to ensure that the device reboots each time. Figure 5.3 shows examples of HF and UHF RFID responses at RF. This experiment allows to test if devices can be distinguished when they respond to standard reader requests.

Experiment 2: In this experiment, we challenge the RFID device with the same signals as the previous experiment, but using out-of-specification frequencies. For HF RFID, we varied the carrier frequency F_c from 12.96 MHz to 14.36 MHz with a step of 100 KHz. In case of UHF RFID, we specified different backscatter link frequency (BLF) through the TRcal parameter considering TRcal = 15, 17, 33, 83, 225, 250. We expect the variation in the RFID response to be higher when the device is performing out-of-specification communication. The reason behind this reasoning is that manufacturers mainly focus on optimizing the standard frequency ranges.



(a) (b)

Figure 5.4: HF RFID responses to out-of-specification signals. (a) Response to a 10-cycle burst signal of non-modulated 5 MHz carrier. (b) Response to a frequency linear sweep of carrier from 100 Hz to 15 MHz, 10 ms

Experiment 3: This set of experiments is only performed on HF RFID as it is specifically designed for close proximity. We send out-of-specification burst and frequency sweep signals and recorded the device response. The burst signal consists of 10 cycles non-modulated 5 MHz carrier (maximum allowed burst frequency). The frequency sweep signal consists of non-modulated carrier linear sweep from 100 Hz to 15 MHz. The duration of the sweep is fixed to the maximum allowed by our generator, 10 ms. Both signals were sent with the maximum allowed peak-to-peak amplitude of 10 V and power of 1 W. Figure 5.4 shows recorded device responses to the described burst and frequency sweep reader requests. One should note the different shape artifacts.

Using the out-of-specification signals allows to test the devices under extraordinary conditions. We expect to see variations between devices since each device antenna and charge pump are possibly unique. During power-up they may present a unique modification of the activating field. Testing many different frequencies provides details about the RF circuit resonance in each device.

5.2.3 Collected Data

Using the proposed hardware setup, we performed the above described experiments and collected device responses (identification signals) from all available sets of devices. For the privacy of our research subjects, we

Table 5.1: Population of electronic passports

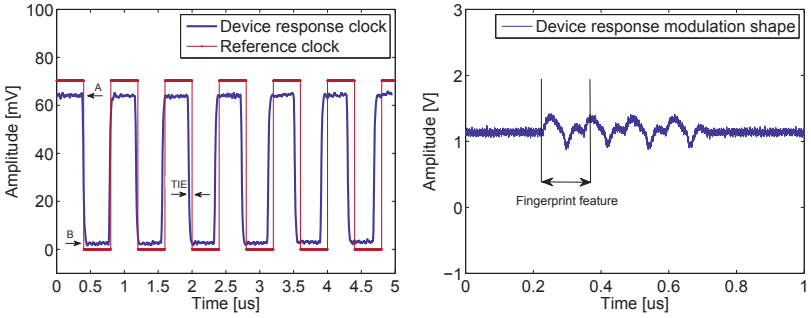
# Passports	Label	Country	Year	Place of Issue
2	ID1, ID2	C1	2006	P1
1	ID3	C1	2006	P2
1	ID4	C1	2006	P3
1	ID5	C1	2007	P4
1	ID6	C2	2008	P5
1	ID7	C3	2008	P6
1	ID8	C1	2008	P1

Table 5.2: Collected data (* identification signals per device per run)

Dataset	Model	# Dev.	Experiment	# Runs	Total*
1	E-passport	8	1,2,3	2	50
2	JCOP card	50	1,2,3	2	50
3	ALN9540	10	1,2	10	100
4	ALN9540	50	2	1	100

labeled the passports from ID1 to ID8. To further protect their privacy, we replaced the country and place of issuance with pseudonyms C1 to C3 and P1 to P6 respectively (Table 5.1).

Our data collection procedure for a single experiment run was as follows: We positioned the target RFID device on the experimental platform with all other devices being at an out-of-range distance from the activating field. We then placed a heavy non-conductive weight on top of the device in order to fix it firmly and horizontally on the platform. For each RFID device, we performed all experiments at fixed acquisition timing offsets and a sampling rate of 4 GS/s. We saved the recorded data on a disk for later analysis. For each device we performed at least two runs, completely removing and replacing the RFID on the experimental platform between runs. For UHF RFID, we also performed independent runs that varied the location of the tag with respect to the reader. For each experiment, we collected between 50 and 100 responses per device per run. Table 5.2 summarizes the collected datasets.



(a) (b)

Figure 5.5: Timing and modulation feature extraction. (a) Time interval error (TIE) in an example UHF RFID response. (b) Modulation feature in an example HF RFID response.

5.3 Feature Extraction and Matching

The goal of the feature extraction is to obtain device fingerprints which most effectively support the two initial objectives, namely classification and identification. In this section, we detail the extraction and matching procedures of timing, modulation and spectral features.

5.3.1 Timing Features

In RFID communication protocols, the reader typically initiates the communication by sending commands to the RFID device (tag). The device then responds to the reader with the requested information (e.g., identification number). Although similar on the logical layer, these protocols differ on the physical layer where they use different modulation schemes and/or include a number of stages.

The RFID standards specify the time within which the device needs to respond to commands issued by the reader as well at the duration of the response. These characteristics depend on the standard and specified time and frequency tolerances. There are different ways to measure these characteristics including looking at the number of cycles and cycle duration or data rate frequency. Given that we have acquired device responses with a high sampling rate, we had a precise time domain information. Therefore, we measured how far each active edge of

the clock varies from its ideal position in time. This deviation is often referred to as the time interval error (TIE). Figure 5.5 shows the cycles of an RFID response and TIE. Given that TIE increases linearly, we define its slope ∂_{TIE} as our timing feature; ∂_{TIE} is proportional to the data rate frequency.

In order to compute ∂_{TIE}, we have to accurately find the points of start and end of each clock cycle. We use a two-step threshold-based detection algorithm. In the first step, a rough-point detection of the start and end time points of each clock cycle is computed for device response. This step allows to quickly determine approximate points amenable to more precise point detection. In the second step, we look at the signal around the rough detected points and use an slope detection algorithm to accurately detect the start and end points of each clock cycle.

After computing the described points, we apply a standard linear least square fitting algorithm (LSF) to determine ∂_{TIE}. More precisely, we fit a line $y = a \cdot x + b$ to the set of cycle points $\{(x_i, y_i) : i \in \{1, ..., C\}\}$, by minimizing the least square error. Here C is the number of clock cycles used to fit the line, x_i is the index of the clock cycle, and y_i is the TIE at clock cycle i. The ∂_{TIE} is the fitted line coefficient a.

For each cycle i, we compute TIE_i with respect to the 10% of the cycle step height, i.e., at $0.1 \cdot (A_i - B_i) + B_i$, where B_i and A_i are respectively the average low-state amplitude and the average high-state amplitude of the response for cycle i (Figure 5.5).

It should be noted that the notions of TIE and ∂_{TIE} are close to the notion of clock offset and clock skew as in [6, 39]. The difference resides in the communication layer used for measurement. We measure TIE from the physical-layer signal, while in related work, the clock offset/skew are derived from timestamps available from upper-layer protocols (e.g., TCP). Such timestamp information is not available in RFID communication and therefore cannot be used.

5.3.2 Modulation Features

RFID communication defines a number of different data modulation mechanisms (e.g., on off keying, amplitude shift keying). A choice of features can therefore be the shape of the RFID modulated response at a given carrier frequency F_c. Figure 5.5 (b) shows the shape of the On-Off keying modulation from one of our smart cards. Below is the procedure to obtain the shape.

Let's denote the RFID response as $f(t, l)$, where $f(t, l)$ is the amplitude of the signal l at time t. One way to obtain the signal shape (envelope) is to use Hilbert transformation [40].

In Step (i), we apply Hilbert transform on $f(t, l)$ to obtain $H(t, l)$:

$$H(t, l) \quad = \quad \text{Hil}(f(t, l)) \tag{5.1}$$

where Hil is a function implementing the Hilbert transform [41].

In Step (ii), the starting point of the modulation in $H(t, l)$ is determined using slope detection algorithm. The end point is fixed to a predefined value (see Section 5.4) and then the modulation fingerprint is extracted.

Feature matching between a reference and a test fingerprints is performed using standardized Euclidean distance, where each coordinate in the sum of squares is inversely weighted by the variance of that coordinate [42].

5.3.3 Spectral Features

In this section, we describe the extraction and matching of spectral features from HF RFID device responses (identification signals) to a burst and a frequency sweep (Section 5.2.2).

Both frequency sweep and burst device responses were high dimensional: each sweep response contained 960000 samples (dimensions) and each burst – 40000. Our data contained many noisy dimensions and we could not determine which frequencies were effective for discrimination. We therefore used a statistical approach based on Principal Component Analysis (PCA) for high-dimensional data [26]. Dimensions that did not contribute to the total covariance were discarded. Given that the number of dimensions was very high, orders of magnitude higher than the number of data samples we could process, standard PCA could not be applied. In the following, we describe the spectral feature extraction and matching.

Extraction and Matching

For a given RFID device, spectral PCA features are extracted from N captured samples using a linear transformation derived from PCA for high-dimensional data. We denote a signal by $f(t, l)$, where $f(t, l)$ is the amplitude of the signal l at time t. The features are extracted in the following three steps:

In Step (i), we apply a one-dimensional Fourier transformation on $f(t,l)$ to obtain $F(\omega,l)$:

$$F(\omega,l) = \frac{1}{\sqrt{M}}\sum_{m=0}^{M-1} f(t,l)\exp(-2\pi i\frac{t\omega}{M}), \tag{5.2}$$

where M is the length of signal considered and $0 \leq t \leq M-1$ is time. We then remove the DC component in $F(\omega,l)$ and the redundant part of the spectrum; we denote the remaining part of the spectrum by $\vec{s_l}$. In Step (ii), a projected vector $\vec{g_l}$, also called a spectral feature, is extracted from the Fourier spectrum using a PCA matrix W_{PCA}:

$$\vec{g_l} = W_{PCA}^t \vec{s_l} \tag{5.3}$$

The feature extraction from N captured samples for a given RFID is then given by $G = W_{PCA}^t S$ where G is an array of $\vec{g_l}$ and S is a matrix $S = [\ \vec{s_0}\ ..\ \vec{s_l}\ ..\ \vec{s_N}\ ]$.
Finally, in Step (iii), the feature template (fingerprint) h used for matching is computed:

$$h = \{\hat{G}; \Sigma_G\} \tag{5.4}$$

where $\hat{G}$ denotes the mean vector of G and Σ_G denotes the covariance matrix of G. The number of captured samples N used to build the feature template and the number of projected vectors in W_{PCA} (i.e., the subspace dimension) are experimentally determined.

Mahalanobis distance is used to find the similarities between fingerprints[4]. The result of matching a reference h^R and a test h^T feature templates is a matching score, calculated as follows.

$$\begin{aligned} scr(h^R, h^T) = min(&\sqrt{(\hat{G}^T - \hat{G}^R)^t \Sigma_{G^R}^{-1} (\hat{G}^T - \hat{G}^R)}, \\ &\sqrt{(\hat{G}^T - \hat{G}^R)^t \Sigma_{G^T}^{-1} (\hat{G}^T - \hat{G}^R)}) \end{aligned} \tag{5.5}$$

Values of the matching score closer to 0 indicate a better match between the feature templates. The proposed matching uses the mean and covariance of both test and reference templates. It also ensures the symmetric property, that is $scr(h^R, h^T) = scr(h^T, h^R)$.

[4]We discovered that the feature templates are distributed in ellipsoidal manner and therefore use Mahalanobis distance that weights each projected feature according to the obtained eigenvalues.

Training PCA transformations

In order to compute the eigenvalues and corresponding eigenvectors of the high-dimensional data (the number of device responses $\ll$ the number of dimensions), we used the following lemma:

Lemma: For any $K \times D$ matrix W, mapping $x \rightarrow Wx$ is a one-to-one mapping that maps eigenvectors of W^TW onto those of WW^T.

W denotes a matrix containing K samples of dimensionality D. Using this lemma, we can first evaluate the covariance matrix in a lower space, find its eigenvectors and eigenvalues and then compute the high-dimensional eigenvectors in the original data space by normalized projection [26]. Based on this description, we compute the PCA matrix W_{PCA}=$[\vec{u_1}\vec{u_2}\ldots\vec{u_i}]$ by solving the eigenvector equation:

$$(\frac{1}{K}X^TX)(X^T\vec{v_i}) = \lambda_i(X^T\vec{v_i}) \tag{5.6}$$

where X is the training data matrix $K \times D$ and $\vec{v_i}$ are the eigenvectors of XX^T. We then compute the eigenvectors of our matrix $\vec{u_i}$ by normalizing:

$$\vec{u_i} = \frac{1}{\sqrt{K\lambda_i}}(X^T\vec{v_i}) \tag{5.7}$$

It should be noted that other techniques for dimensionality reduction could be employed (e.g., linear discriminant analysis, probabilistic PCA). Given the satisfactory results with this scheme, we did not consider other methods.

5.4 Performance Evaluation

We present our experimental evaluation starting with the spectral features which demonstrated capabilities of uniquely identifying RFID devices. We then discuss the capabilities of the timing and modulation features.

5.4.1 Spectral Features

For the evaluation of our spectral features, we considered the sets of HF RFID smart cards and e-passports (Dataset 1 and 2). We estimated the benchmark accuracy over a single run collected data and quantified the feature stability by considering all independent runs together.

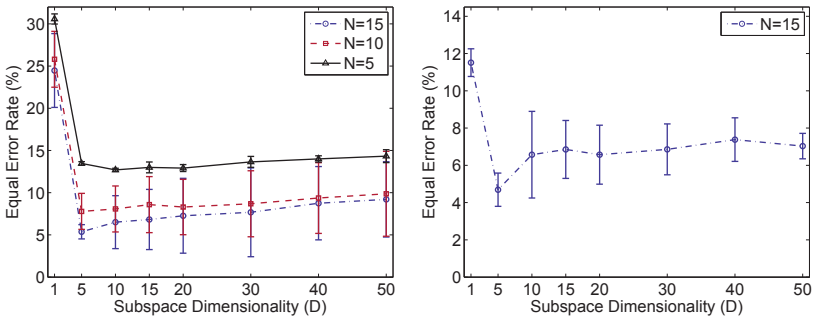


Figure 5.6: Spectral features identification accuracy for N identification signals used to built the fingerprint and feature dimensionality D (Dataset 2). (a) Burst spectral features (b) Sweep spectral features

We validated our results using cross-validation [26]. Given the collected 50 identification signals per device per run, we used 5-10 signals for training and the remaining 40-45 for testing. The exact number depended on the number of identification signals N used to build the fingerprint. The training and testing data were thus separated.

Feature accuracy

Burst feature accuracy. The results are presented in Figure 5.6 (a) for different number of identification signals N and subspace dimensionality. The dimension of the features before the projection is 19998. Our system reached an EER of 0.0537 (5.37%) for $N = 15$. This means that the considered RFID devices were correctly identified with an accuracy of approximately 95% (GAR at the EER operating point).

Sweep feature accuracy. For computational reasons, we did not consider the entire sweep identification signal. Instead, we extracted the spectral features from the part of the signal between 6 ms and 9 ms. This part contained the biggest shape changes as shown in Figure 5.4(b). This decision reduced allowed significantly faster feature extraction.

Figure 5.6 (b) shows the results for $N = 15$ and different subspace dimensionality. The dimension of the original features before projection is 49998. We obtained an EER of approximately 0.0469 (4.69%), when

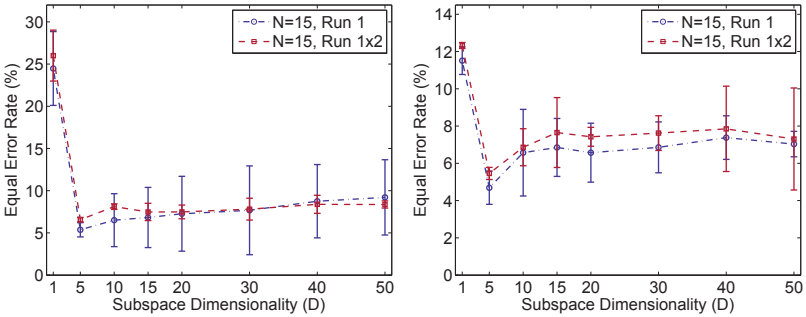


Figure 5.7: Feature stability using two independent runs (Dataset 2). (a) Burst spectral features (b) Sweep spectral features.

using the first 5 eigenvectors to project and store the feature template. The obtained accuracy is therefore similar to the one obtained with the burst features, i.e., our system correctly identifies the individual devices with an accuracy of approximately 95% (GAR at the EER point).

Using the collected data from the electronic passports (Dataset 1), we obtained an EER = 0 with both the burst and sweep spectral features. While this result validates the efficacy of the spectral features, we note that the passport data contains different classes of devices which facilitates the identification task.

Table 5.3 summarizes the underlying data, namely the number of responses N, total genuine and imposter matchings performed for EER computation[5], Accept/Reject threshold and EER.

Feature stability

In the previous sections we have analyzed the identification accuracy using burst and sweep spectral features within a single experiment run. This allows us to have a benchmark for estimating the stability of the

[5]The number of genuine and imposter matchings depends on the number of available fingerprints per device. For N=10, we are able to built 4 different fingerprints with the testing data within a run. This results in 6 different matchings of fingerprints from the same device (i.e., genuine matchings) and 392 different matchings of fingerprints from different devices (i.e., imposter matchings). For 50 devices, this makes 300 genuine and 19600 imposter matchings.

Table 5.3: Summary of accuracy for spectral features (Dataset 2, $D = 5$, 4-fold cross validation).

Type	Run	N	Test matchings		T	EER (%)
			Genuine	Imposter		
Burst	1	15	150	11025	1.88	5.37 (4.38;6.36)
	1	10	300	19600	2.91	7.79 (5.29;10.28)
	1x2	15	200	9800	2.64	6.57 (6.25;6.89)
Sweep	1	15	150	11025	1.68	4.69 (3.65;5.74)
	1x2	15	200	9800	1.93	5.46 (5.08;5.84)

features. In particular, we performed the following stability analysis:

1. Using the linear transformations W_{PCA} obtained in the first run, we selected 4 feature templates (2 from each run) and computed again the EER by considering only the cross matching scores of fingerprints from different runs[6]. The process was repeated 3 times with different feature templates from the two runs to validate the feature stability.

2. We trained the system over the first 20 devices and then used the obtained linear transformation to estimate the accuracy over the remaining 30 devices. This analysis tests the stability of the obtained linear transformations to discriminate independent transponder populations[7].

Table 5.4: Accuracy of spectral features for independent training and testing sets (Dataset 2, $D = 5$, 3-fold cross validation)

Type	Run	N	Test matchings		T	EER (%)
			Genuine	Imposter		
Burst	1x2	15	120	3480	2.78	7.33 (6.01;8.65)
Sweep	1x2	15	120	3480	2.03	5.75 (5.45;6.05)

Figure 5.7 compares the EER accuracy obtained with the first run (Run 1) and the accuracy obtained by mixing fingerprints of both runs

[6]This procedure is required in order to remove any possible bias from cross matching scores of fingerprints from the same run. We point out that this results in a reduced number of genuine and imposter matchings for the EER computation.

[7]The motivation behind this division (20 vs. 30) is that it gives sufficient number of samples for both training and testing.

(Run 1×2) for a fixed $N = 15$. Table 5.3 displays the confidence interval for subspace dimension of 5 eigenvectors. The obtained EERs do not show a statistically significant difference between the two experiments for both the burst and sweep features using 4-fold cross validation.

Table 5.4 summarizes the EER accuracy obtained using independent transponder sets for training and testing using the two runs in Dataset 2. The subspace dimensionality was fixed to $D = 5$ and the number of identification signals $N = 15$. Even if the testing population (30 devices) is smaller, we observe that both features perform similarly to the benchmark accuracy (Table 5.3).

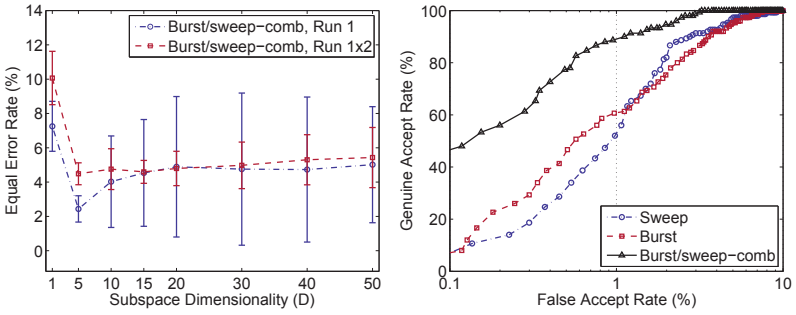
Combining sweep and burst features

Given that the identification accuracies of both burst and sweep spectral features are similar; in order to fully characterize the identity verification we computed the ROC curves for the burst and sweep features as shown in Figure 5.8(b). We notice that while the EERs are similar, the curves exhibit different accuracies at different FARs. In particular, for low FAR $\leq 1\%$ the sweep features show lower GAR.

The burst and sweep features discriminate the fingerprints in a different way, and therefore these features can be combined in order to further increase the accuracy. Such combinations are being researched in multi-modal biometrics [43] where different "modalities" (e.g., fingerprint and vein) are combined to increase the identification accuracy and bring more robustness to the identification process [43].

A number of integration strategies have been proposed based on decision rules [44], logistic functions to map output scores into a single overall score [45], etc. Figure 5.8 shows the EERs and ROC curves of feature combination by using the sum as an integration function. The overall matching score between a test and a reference template is the sum of the matching scores obtained separately for the burst and sweep features. Table 5.5 summarizes the results.

For the benchmark datasets (Run 1), we observe significant improvement of the accuracy reaching an EER = 2.43%. The improvement is also significant for all target FARs (e.g., 0.1%, 1%) as shown in Figure 5.8 (b). We also observe a statistically significant improvement on using fingerprints from both Run 1 and 2. The accuracy is slightly lower (EER=4.38%). These results motivate further research on feature modalities and novel integration strategies.



(a) (b)

Figure 5.8: (a) Identification accuracy by combining the sweep and burst features (b) Receiver operating characteristic (ROC) for burst and sweep spectral features and their combination (Dataset 2, $D = 5$ and $N = 15$). See Table 5.5 for the underlying data.

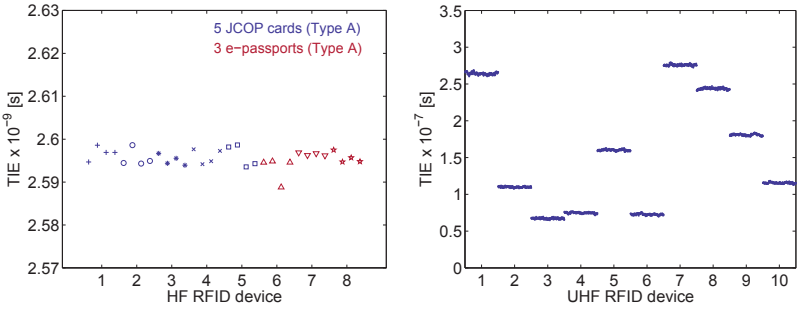
Table 5.5: Summary of accuracy when combining burst and sweep features (Dataset 2, $D = 5$, 4-fold cross validation).

Type	Run	N	Test matchings		T	EER (%)
			Genuine	Imposter		
B&S	1	15	150	11025	1.56	2.43 (1.54;3.33)
B&S	1x2	15	200	9800	2.18	4.38 (3.9;4.9)

Other performance factors

Our analysis showed that using the first 5 eigenvectors kept the system accuracy high. If 5 eigenvectors are used to store the device fingerprint, the proposed burst and sweep spectral features form compact fingerprints of RFID devices. More precisely, if each dimension is represented by a 4-byte floating-point number, the size of the device fingerprint $h = \{\hat{G}; \Sigma_G\}$ is 20 (5×4) bytes for $\hat{G}$ and 100 (5x5x4) bytes for the square covariance matrix Σ_G resulting in a total of 120 bytes.

In terms of acquisition and extraction efficiency, the burst spectral features are significantly more efficient for digital acquisition and extraction due to their lower dimensionality. More precisely, we measured an acquisition and extraction time of 2 s per burst vs. 26 s per sweep on a machine with 2.00 GHz CPU, 2 GB RAM running Linux Ubuntu.



(a) (b)

Figure 5.9: Timing feature for HF and UHF RFID. (a) HF RFID smart cards and e-passports of the same type exhibit stable, but indistinguishable time interval errors (TIE). (b) Same type UHF RFID tags present stable and distinguishable TIE over different configurations.

Even though all components of the feature extraction can be implemented in hardware, processing sweep signals would be significantly slower to due to higher sampling rates required.

5.4.2 Timing Features

Here we discuss our findings on the timing feature accuracy using HF and UHF RFID devices (Dataset 1,2,3). Figure 5.9(a) visualizes the time interval error (TIE) for a set of randomly selected 5 HF RFID smart cards and 3 e-passports (ISO 14443 Type A). For each device, 4 fingerprints are visualized at the nominal frequency $F_c = 13.56$ MHz. The results show that TIE is stable over time, but cannot be used to distinguish same type HF RFID devices. This is also true for all tested in- and out-of-specification frequencies. We could only observe timing differences between Type A and Type B HF RFID. Given that we had only one model/manufacturer of Type B HF RFID at our disposal, we cannot conclude that TIE would at least vary between manufacturers.

Figure 5.9(b) illustrates the first derivative of TIE (∂_{TIE}) for a set of 10 randomly selected UHF RFID tags for TRcal = 15. For each tag, 220 fingerprints collected using 11 different configurations are visualized. Our configurations consisted of 8 different locations up to 6 meters from the acquisition antenna as well as 3 cases where

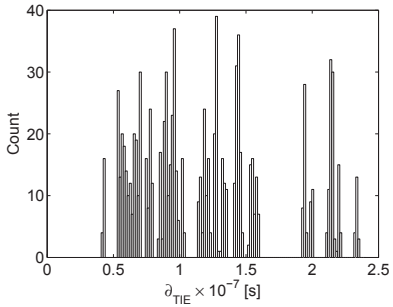


Figure 5.10: Empirical distribution of ∂_{TIE} based on measuring 50 UHF RFID (ALN9540) tags (Dataset 4). A total of 1000 fingerprints (20 per tag) are used to fill in the histogram bins. The bin width is fixed to two times the average standard deviation of ∂_{TIE}.

the transmitting power and/or device orientation were varied. ∂_{TIE} is not only stable, but also different for same manufacturer and type tags. This is primarily due to the allowed standard tolerances in the backscatter link frequency (BLF) [38].

In order to quantify the capability of ∂_{TIE} to distinguish UHF RFID tags, we estimated the entropy from the empirical distribution of ∂_{TIE} obtained from the entire set of 50 devices. The obtained result suggested that we could learn 5.84 bits of information about a given tag. The entropy measure depends only on the ∂_{TIE} variation which directly relates to the BLF tolerances. According to the empirical distribution in Figure 5.10, we observed a BLF tolerance of $\pm 14.01\%$ around a mean frequency of approx. 1400 KHz. If we consider the maximum allowed BLF tolerance in the standard [38], i.e., $\pm 22\%$ for BLF = 320 KHz, the maximum possible entropy would be 9.86 bits[8]. However, we could not observe such large BLF variations in our tested population of UHF RFID devices. We also observed similar distinguishable behavior on two smaller sets of 10 devices of two other manufacturers (Analog Devices (AD) and UPM). We acknowledge that more investigation is required to find the frequencies that allow highest entropy. These frequencies may also depend on the manufacturer.

[8]The maximum possible entropy is achieved when the probability distribution of ∂_{TIE} is assumed to be uniform [46].

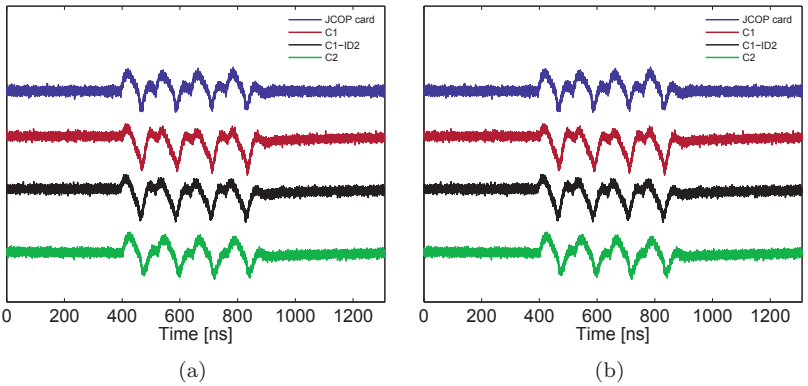


Figure 5.11: Modulation of responses of 4 different classes (C1),(C1-ID2),(C2),(JCOP). (a) First run (b) Second run. The modulation is stable across acquisition runs (Dataset 1,2).

5.4.3 Modulation Features

For the modulation feature evaluation, we considered e-passports and smart cards (Dataset 1 and 2). The passports ID1-4 and ID7-8 as well as the smart cards used Type A, whereas ID5-6 used Type B ISO 14443 communication. It is interesting to notice that within the same country (C1) we had documents with different types: ID1-4 used Type A and ID5 - Type B.

Our modulation features showed discriminant artifacts that differ from one device to another on out-of- specification carrier frequencies. Figure 5.11 shows the modulation shapes of 4 different classes of Type A protocol devices. These were recorded at an out-of-specification carrier frequency $F_c = 13.16$ MHz. Visual inspection shows that the modulation shapes were stable and different for the considered classes.

In order to quantify these observations more precisely, we considered classification with 3 classes (2 countries + JCOP cards) with all fingerprints from two different runs. The classification process was repeated 8 times with 8 different reference fingerprints per class for validation. The results showed an average classification error rate of 0%. In addition, after detailed inspection of the modulation features we discovered that ID2 from C1 differed significantly from the representatives of that class. We therefore formed a new classification scenario with 4 classes

and obtained an error rate of 0%. We should note that ID1 and ID2 were issued by the same country, in the same year and place of issue. However, the embedded RFID differed a lot. The modulation of ID1, ID3 and ID4 from C1 could not be further distinguished using the combination of modulation features and Euclidean matching.

Similar to Type A, the 2 Type B passports from two different countries (C1,C3) available in our population showed complete separability.

In summary, the modulation shapes at an out-of-specification carrier frequency present potential to classify different models (e.g., countries). They are quickly extractable and stable across different runs. We acknowledge that our data set is insufficient due to the difficulty of obtaining e-passports. We believe however that our investigation could stimulate future work with a larger set of e-passports.

5.5 Summary and Discussion

We investigated timing, modulation and spectral features for physical-layer identification of RFID. Our results demonstrate that RFID devices exhibit physical-layer characteristics that enable their identification in a controlled setup.

Spectral features extracted from the device responses to burst and linear frequency sweep signals enable device identification with an error rate (EER) as low as 5%. These features are also stable and can be combined in order to further improve the accuracy. The obtained results motivate the use of HF RFID physical-layer fingerprints in document cloning detection solutions. For related results on UHF RFID devices, we invite the reader to consult [34].

Timing features such as the time interval error (TIE) and modulation could be effective in distinguishing certain types of devices. In particular, TIE can be used to identify up to 2^6 UHF RFID tags independently of the tag location. While more investigation is required to consider practical issues (e.g., lower cost hardware) and assess more scenarios (e.g., mobility), given the allowed standard tolerances, UHF RFID devices leak distinguishable information that enables their (un)authorized tracking. This was not possible for the tested HF RFID devices due to higher manufacturing precision of their internal clocks.

Chapter 6

Towards Practical Identification of HF RFID Devices

In the previous chapter, we have demonstrated the feasibility of distinguishing RFID devices based on physical-layer fingerprints. In this chapter, we go further into practical issues of physical-layer HF RFID identification such as accuracy, stability and transferability.

We propose an improved signal acquisition and enhanced feature extraction and matching methods. This system enables significantly more accurate identification with an EER as low as 0.005 (0.5%). It also removes the requirement for statistical analysis. The extracted fingerprints are stable over multiple independent acquisitions during an extended period of time. We also propose a solution based on channel equalization that allows fingerprint verification across acquisition setups. This scenario is of practical importance when device fingerprints are acquired at one setup and verified on another one.

Our improvements strengthen the application of physical-layer HF RFID identification in the detection of cloned and/or counterfeit identity documents. In this scenario, presented documents are measured in a controlled setup, their fingerprint is then extracted and verified with the enrolled fingerprints of legitimate documents. We discuss this application in more details in Chapter 8.

6.1 Problem and System Overview

In this work, we focus on building accurate and stable physical-layer fingerprints of short range HF RFID-enabled devices (same manufacturer and type) for the purpose device identity verification (Chapter 3).

Our system is comprised of a fingerprinting hardware, acquisition (antenna) setup and feature (fingerprint) extraction and matching procedures. The block diagram of its components is illustrated in Figure 6.1. We use two acquisition setups to test the accuracy and stability. These are comprised of two independent sets of same manufacturer and type antennas. Due to cost, we could not replicate the fingerprinting hardware which would have provided two entirely independent systems. However, given the precision of our hardware in rendering the signals, its influence on the system performance should be negligible.

The feature extraction and matching procedures are implemented in software and performed offline. We consider extracting physical-layer fingerprints from device responses to high-energy bursts of sinusoidal carrier (Figure 6.3). Our decision stems from the fact that these fingerprints are suitable for identification, fast to acquire and computationally efficient to use in feature extraction and matching.

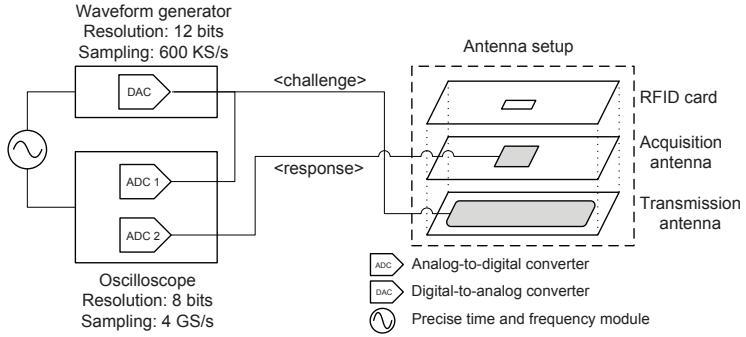
For the performance analysis, we considered the same set of 50 HF RFID smart cards (ISO 14443, 13.56 MHz) as in Chapter 5. Due to limited resources, we could not evaluate on larger sets. Nevertheless, the considered data population allows fair comparison with prior results and should be indicative for the accuracy and stability of the proposed techniques.

6.2 Signal Acquisition

In this section, we first describe our hardware and antenna setup. We then detail the different types of experiments we have performed and present the collected datasets from our population of RFID devices.

6.2.1 Hardware Setup

Figure 6.1 shows the block diagram of the fingerprinting hardware and acquisition setup in our system. The fingerprinting hardware consists of an arbitrary waveform generator [47] and oscilloscope [48]. A time and frequency reference module provides a precise and stable 10 MHz



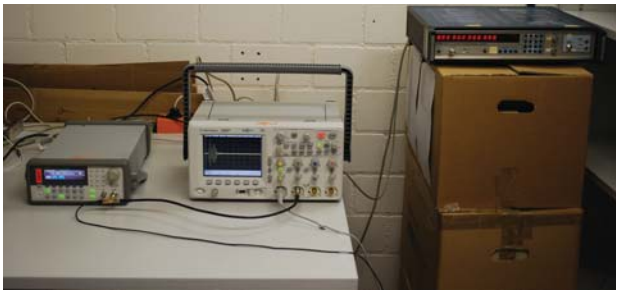
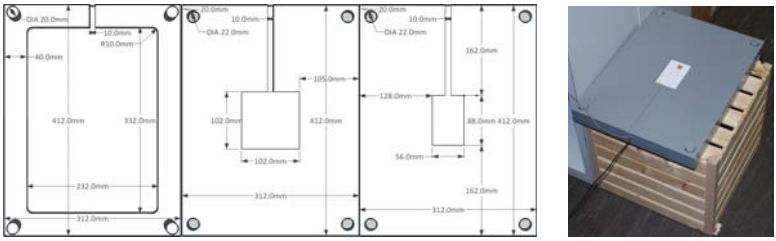


Figure 6.1: Hardware setup. The setup consists of an arbitrary waveform generator, oscilloscope and acquisition (antenna) setup. The arbitrary waveform generator and oscilloscope are connected to a common time and frequency reference module.

reference clock to both waveform generator and oscilloscope. This guarantees optimal precision and stability in rendering and acquiring the signals. The fingerprinting hardware is connected to the acquisition (antenna) setup (Figure 6.2). The latter consists of three PVC plates to hold the two antennas and device to be measured. The acquisition antenna[1] is positioned between the transmission pad antenna[2] and the device (smart card). The proposed setup guarantees a robust and precise structure in order to avoid antenna and card position fluctuations. The design also optimizes the signal-to-noise ratio (SNR) of the cap-

[1]Texas Instruments ANT 100x100MM 50 Ω 13.56 MHz
[2]Texas Instruments RFID ANT PAD 320x240MM 13.56 MHz



(a) Platform plates (b) Assembled setup

Figure 6.2: Antenna setup. The setup consists of a platform with three plates that hold two antennas and the device to be measured. The bottom plate holds a pad-style transmission antenna, the middle plate holds the acquisition antenna and the top plate provides a slot for placing devices (smart cards).

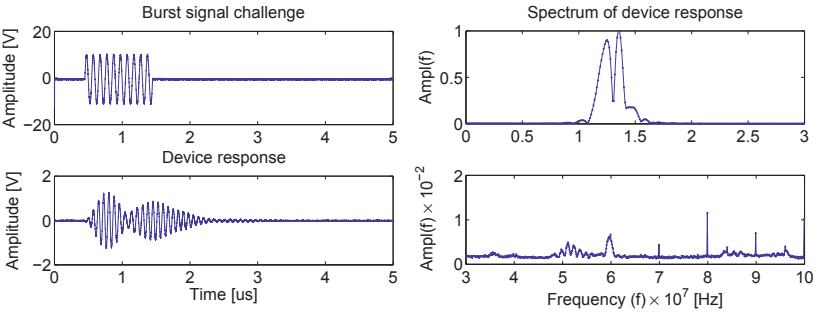
tured signals[3]. Appropriate low-noise cabling is also used to reduce the noise in the entire setup.

The hardware setup is used as follows. The purpose-build burst signal (challenge) is sent by the arbitrary waveform generator to the transmission antenna and the device response as observed by the acquisition antenna is recorded at the oscilloscope for later processing. The challenge signal is also sent in parallel to the second channel of the oscilloscope in order to provide an exact trigger for digitizing the device response. This design guarantees precise time alignment of all recorded responses. The entire process is controlled by a computer.

6.2.2 Performed Experiments and Collected Data

Our physical-layer analysis focused on collecting device responses to high-energy bursts of sinusoidal carrier (see Section 6.1 for motivation). We power the RFID smart card by a burst signal of RF energy. Our burst consists of 10 cycles non-modulated carrier at a center frequency F_c, output power of 1 W (30 dBm) and peak-to-peak amplitude of $V_{pp} = 10$ V. The center frequency is a parameter that we determine experimentally. Figure 6.3(a) shows the recorded response to an RF

[3]We note that the device response could also be observed at the transmission antenna. Our two antenna design is more suitable for identification as it increases the received SNR and provides separate transmission and reception channels.



(a) Signal challenge and response

(b) Spectrum of the response 0 - 100 MHz

Figure 6.3: Reader challenge and device response. (a) A non-modulated carrier sinusoidal signal burst of 10 cycles is sent to the smart card and its response is acquired simultaneously. (b) The spectrum of the response is mostly contained between 11 and 17 MHz. Additional artifacts could also be observed at higher frequencies.

burst signal challenge at $F_c = 12$ MHz with a signal-to-noise ratio of 21 dB. Since the device internal components (e.g., charging capacitor) and antenna characteristics are unique, we observe that during power-up each device exhibits a unique modification of the activating field.

Using our hardware setup, we collected device responses from 50 HF RFID smart cards (same model and manufacturer) at a rate of 2 responses/second. This rate was selected to allow enough time to record the data on a stable storage. Each received device response was sampled at 4 GS/s (maximum allowed with our oscilloscope) during a period of 10 μs. Our data collection procedure for a single collection "run" was as follows: We positioned the target RFID device on the antenna setup ensuring that all other devices remain out-of-range of the activating field. We then collected 50 device responses at a fixed acquisition timing offset and saved them on a disk for later analysis. For any subsequent collection run, we removed and replaced the RFID device on the acquisition setup. The hardware was switched off between runs collected on different days.

Table 6.1 summarizes the collected datasets and measurement parameters. Dataset 1 contains device responses from 10 different smart cards. For each card, we collected 50 responses per burst frequency of

Table 6.1: Collected data from 50 HF RFID smart cards.

Dataset	Setup	# Dev.	Burst Freq.	# runs	# signals
1	I	10	9-15 MHz	1	3500
2	I	50	12 MHz	2	5000
3	II	50	12 MHz	2	5000
4	I	10	12 MHz	7	3500

9 to 15 MHz with a step of 1 MHz. This dataset is used for parameter selection (e.g., frequency). Dataset 2 and 3 contain device responses from the entire set of 50 smart cards. For each card, we collected 100 responses at a burst frequency of 12 MHz in two separate runs; that is the card was removed and repositioned on the setup across each run. The experiment was performed on the two acquisition setups, referred to as setup I and II in the rest of the paper. This dataset is used to evaluate the accuracy and stability of our proposed techniques within a single setup (Section 6.4.1) and across setups (Section 6.4.3). Dataset 4 is composed of device responses from 10 smart cards acquired in a period of approximately 60 days (7 different runs). The dataset is used to evaluate the stability of our features to multiple independent acquisitions (Section 6.4.2).

6.3 Feature Extraction and Selection

In this section, we describe the extraction, selection and matching of features (fingerprints) from the collected data (Section 6.2.2).

6.3.1 Basic Feature Extraction

For a given RFID card, we first proceed in extracting our basic feature from N acquired responses (samples). We denote a sample by $f(t,l)$, where $f(t,l)$ is the normalized amplitude of the signal l at time t. We apply one-dimensional Fourier transformation on $f(t,l)$ to obtain $F(\omega,l)$:

$$F(\omega,l) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} f(t,l) \exp(-2\pi i \frac{t\omega}{M}), \tag{6.1}$$

where M is the length of signal considered and $0 \leq t \leq M-1$ is time. We then remove from $F(\omega, l)$ the DC component and redundant part of the spectrum; we denote the remaining part of the spectrum by $\vec{s_l}$.

The extracted basic features $\vec{s_l}$ are in a high dimensional subspace (20000 dimensions). This requires effective dimensionality reduction to remove noisy dimensions (e.g., spectra from other transmissions) and dimensions that do not contribute to discriminating the devices. We describe a new way of effective dimensionality reduction based on filtering. We also refine the dimensionality reduction by statistical analysis in previous work to allow using the new features.

6.3.2 Feature Selection by Filtering

A projected vector $\vec{b_l}$, also called filtered features, is extracted from the basic features $\vec{s_l}$ using a bandpass filter transformation W_{BPF}:

$$\vec{b_l} = W_{BPF}^t \vec{s_l}. \tag{6.2}$$

The feature extraction from N captured responses for a given RFID device is then given by $G = W_F^t S$ where G is an array of $\vec{b_l}$ and S is a matrix $S = [\ \vec{s_0}\ ..\ \vec{s_l}\ ..\ \vec{s_N}\ ]$.

In our implementation, we used a Chebyshev I bandpass filter design with filter order of 100. The passband frequencies (Fp1 and Fp2) are experimentally determined. Using this technique the dimensionality of filtered features could be drastically reduced to only tens of dimensions depending on Fp1 and Fp2.

6.3.3 Feature Selection by Statistical Analysis

In this section, we detail the feature selection by statistical analysis. Our goal is similar to the above, but instead of using filtering, we use principal component analysis (PCA) for effective dimensionality reduction[4]. The main idea behind PCA is find a compact feature subspace that contains most of the total covariance in the data. This statistical step can either be applied directly to the basic features or to the filtered features as follows.

A projected vector $\vec{p_l}$, also called statistical features, is extracted from the basic $\vec{s_l}$ or filtered $\vec{b_l}$ features using a previously obtained PCA transformation W_{PCA}:

[4]We note that other alternative dimensionality reduction techniques could also be effective (e.g., Discriminant Analysis).

$$\vec{p_l} = \begin{cases} W_{PCA}^t \vec{s_l} & \text{if initial feature is } \vec{s_l} \\ W_{PCA}^t \vec{b_l} & \text{if initial feature is } \vec{g_l} \end{cases} \tag{6.3}$$

The feature extraction from N captured responses for a given RFID device is then given by $G = [\ \vec{p_0}\ ..\ \vec{p_l}\ ..\ \vec{p_N}\ ]$. For computation of the eigenvalues and corresponding eigenvectors of the PCA transformation we used the lemma and procedures described in Section 5.3.3.

6.3.4 Feature Matching

Finally, the feature template (fingerprint) h consists of two components computed from G:

$$h = \{\hat{G}; \Sigma_G\}, \tag{6.4}$$

where $\hat{G}$ denotes the mean vector of G and Σ_G denotes the covariance matrix of G. It is important to note that we computed the covariance Σ_G by shrinkage estimation [49] because standard covariance algorithms proved to be unstable when the number of dimensions was higher. An additional advantage of the shrinkage method is that it also yields a positive definite and well conditioned covariance. We used the implementation provided in the R tool [50].

Mahalanobis distance is used to find the similarity between a reference h^R and test h^T fingerprints.

$$\begin{aligned} scr(h^R, h^T) = min(&\sqrt{(\hat{G}^T - \hat{G}^R)^t \Sigma_{G^R}^{-1} (\hat{G}^T - \hat{G}^R)}, \\ &\sqrt{(\hat{G}^T - \hat{G}^R)^t \Sigma_{G^T}^{-1} (\hat{G}^T - \hat{G}^R)}) \end{aligned} \tag{6.5}$$

Values of the matching score closer to 0 indicate a better match between the feature templates. The proposed matching uses the mean and covariance of both test and reference templates. Thus, it also ensures the symmetric property $scr(h^R, h^T) = scr(h^T, h^R)$.

6.3.5 Channel Equalization

Here, we present a simple channel equalization technique that allows to preserve fingerprint quality in cases where different sets of antennas create significant channel differences and distort the extracted fingerprints. We note that channel equalization is a common procedure in wireless communications that aim at reducing amplitude, frequency and phase distortion introduced by the channel [51]. The procedure is as follows.

Before any set of measurements on a given antenna setup, we first collect a number of device responses (50) and compute the channel frequency response. In order to equalize the channels between two setups, we compute an equalization vector $\vec{w} = [w_1 w_f \cdots w_F]$ that contains weights for each frequency $f \in [Fp1, Fp2]$ such that the sum of mean square errors between the normalized amplitudes of two frequency responses $\vec{x}, \vec{y}$, $\sum_{f=0}^{F}(y_f - w_f x_f)^2$ is minimal; F is the dimensionality of the frequency response. Before matching the test fingerprints extracted from one setup to reference fingerprints extracted from another, we apply the compensation w for each test fingerprint. Therefore, the test fingerprint becomes $\{\hat{wG}; \Sigma_{wG}\}$.

We note that in our channel estimation, we measured each setup without and with an RFID smart card on it. We discovered that the plastic material of the smart cards slightly shifts the channel frequency response with the same constant factor for our entire set of devices. We therefore need to adjust the channel equalization weights measured for each setup without card with that constant factor.

6.4 Performance Evaluation

We evaluate the accuracy of our system based on the metrics and methodology from Chapter 3. We perform multiple rounds of cross-validation using different partitions, and the presented estimates are averaged over these rounds. As we have collected two or more runs in each dataset, we used one run to train the system and extract reference fingerprints from each device. Within that run, we divided the responses per device in five disjoint sets of 10 responses. In each cross-validation round, we used 10 responses per device for training and the remaining up to 40 to build reference fingerprints. Depending on the number of responses N to build a fingerprint, we had 2 to 4 different reference fingerprints per device (e.g., 4 for $N = 10$, 2 for $N = 15$,20). The remaining run(s) in our datasets were used to build independent test fingerprints. All test fingerprints were then matched to all reference fingerprints and the metrics computed as discussed above.

6.4.1 Accuracy Analysis and Comparison

In this section, we present the results of our analysis on the accuracy of the our proposed system on two separate acquisition setups. We first

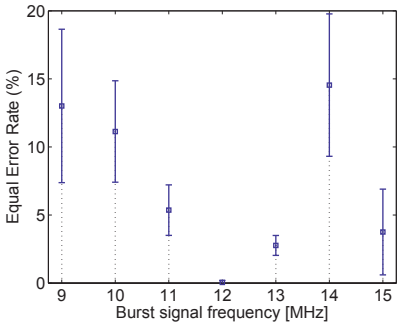


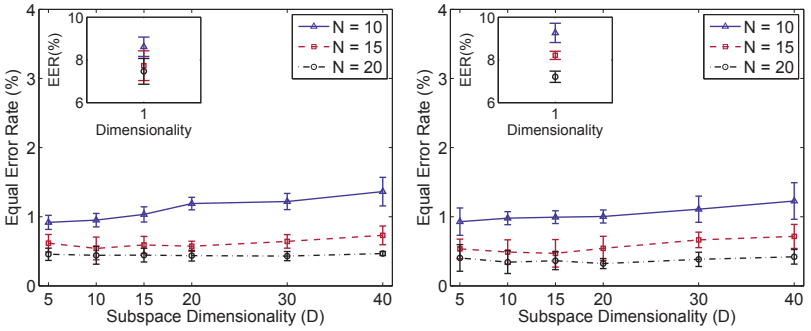
Figure 6.4: Accuracy in EER for different frequencies of the burst signal challenge (Dataset 1). The number of responses used to build the fingerprint was fixed to $N = 10$. We observe that the discriminant capabilities of our features depend on the chosen frequency. For further analysis, we selected burst frequency of 12 MHz.

experimentally determine the burst signal frequency on a smaller set of devices (Dataset 1). We then use the entire set of devices (Dataset 2 and 3) to evaluate the system parameters and obtain reliable estimates of the identification accuracy.

One of the critical parameters for our identification system is to determine (experimentally) the appropriate frequency of the burst signal. Given that our antenna setup was tuned for HF RFID communication at 13.56 MHz, we could only record device responses to burst frequencies between 9 and 15 MHz. Statistical analysis was directly applied to the extracted basic features as it was difficult to perform filtering without knowing where the discriminant features reside. We observed that burst frequency of 12 MHz yields the lowest EER for our antenna set configuration (Figure 6.4). We therefore fixed it to this value for all our subsequent measurements and analysis. We later show that the second antenna setup behaves in the same way. It should be noted that the suitable burst frequency is likely to depend on the resonance and tuning of the chosen antenna set[5]. Further investigation with specific antenna measurements is needed in order to quantify the exact causes of this behavior.

After having determined the burst frequency, we proceeded to an-

[5]We have also tested another different set of antennas, and the lowest EER was reached for both 10 and 12 MHz.



(a) Setup I (b) Setup II

Figure 6.5: Accuracy in EER for different number of responses used to build the fingerprint (N) and subspace dimensionality (Dataset 2 and 3). The EER for $D = 40$ corresponds to the system accuracy after feature selection by filtering. Lower dimensional subspaces are obtained by an additional PCA. The results show that feature selection by filtering already achieves high accuracy. Statistical analysis could additionally be used to effectively reduce the feature dimensionality (i.e., fingerprint size).

alyzing the features and estimating the identification accuracy of our system. In particular, we considered the number of device responses used to build each fingerprint (N) and the dimensionality (D) of the feature subspace.

For all collected device responses, we observed that the features that have contributed the most to discriminating device fingerprints were contained between 11 and 15 MHz. This is also corresponding to the most significant part of the device response spectrum. Therefore, we applied feature selection by filtering (Section 6.3.2) with parameters $Fp1 = 11$ MHz and $Fp2 = 15$ MHz. This procedure reduced the features (fingerprint) dimensionality to $D = 40$. In our figures/hfrfid2, unless specified otherwise, the accuracy at $D = 40$ represents the features (fingerprints) as selected by filtering only. An additional feature selection by PCA was applied on those features in order to display the accuracy in lower dimensional feature subspaces ($D < 40$).

Figure 6.5 shows the obtained EERs varying the number of responses N and principal components (referred to as subspace dimen-

sionality[6]) for our two acquisition setups (Setup I and II). The obtained results demonstrate that both setups exhibit similar EERs of approximately 0.5%. The EER improves significantly by using higher N and reaches stable estimates for $N \geq 15$. The additional statistical analysis based on PCA confirms that using the first 5 eigenvectors ($D = 5$) to project and store the fingerprint keeps the accuracy the same while reducing the fingerprint size. More importantly, with our filtered features only ($D = 40$), the accuracy is similar. This shows that there is no need for additional statistical analysis if the memory requirements are met. Table 6.2 shows the fingerprint size in bytes depending on how many dimensions are used. Typically, if each dimension is represented by a 4-byte floating-point number, the size of the corresponding feature template $h = \{\hat{G}; \Sigma_G\}$ is $D \times 4$ bytes for $\hat{G}$ and $D \times D \times 4$ bytes for the square covariance matrix Σ_G where D is the dimensionality.

Table 6.2: Fingerprint size (in bytes) from subspace dimensionality

	Subspace dimensionality (D)					
	5	10	15	20	30	40
Fingerprint size	120	440	960	1680	3720	6560

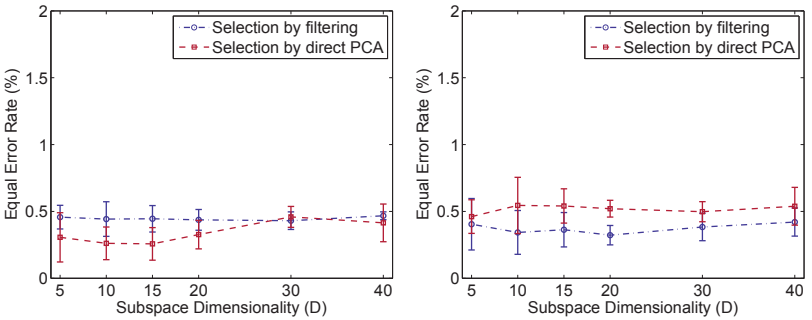
Feature selection by filtering vs. by direct statistical analysis

We compare the accuracy of our proposed feature selection by filtering to the feature selection by direct dimensionality reduction from the original high dimensional data space. Figure 6.6 shows the EERs for $N = 20$ using Datasets 2 and 3. We observe that our filtering method is at least as accurate. In addition, it presents the following advantages: (i) Statistical training is not required in order to perform identification (ii) Feature extraction is more efficient as it does not require additional linear transformations (e.g., W_{pca}) (iii) The bandpass filter operation is directly implementable in the analog domain with a filter and the features can be immediately acquired by a spectrum analyzer.

6.4.2 Stability Analysis

In the previous sections we have analyzed and compared the identification accuracy using two runs and two acquisition setups. This allowed

[6]The number of principal components is directly related to the fingerprint size. The higher the number of principal components, the bigger the fingerprint size.



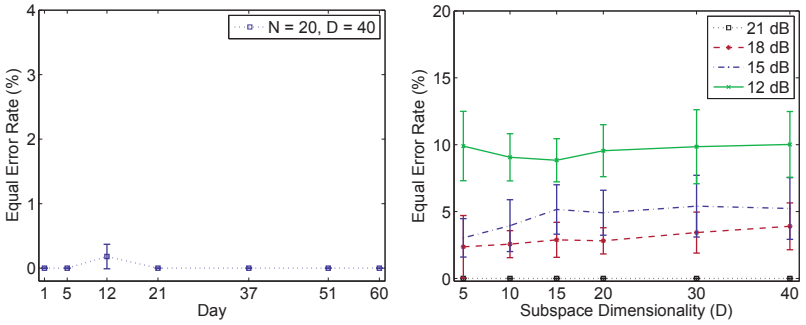
(a) Setup I (b) Setup II

Figure 6.6: Accuracy in EER using feature selection by filtering only and by direct statistical analysis (PCA), i.e., no prior filtering (Dataset 2 and 3). The number of responses used to build the fingerprint was fixed to $N = 20$. The results for the two setups show a comparable accuracy of both feature selection approaches.

us to have a benchmark of the system accuracy. Here, we further explore various factors that could influence the accuracy of our proposed features within a single acquisition setup. Given the specificity of the different setup case, we detail it in the next section. Here, we considered the following scenarios:

Multiple runs: The purpose of this analysis is to evaluate the longer term stability of our features. We collected multiple acquisitions from 10 devices during a period of 60 days. The setup was placed in a air-conditioned room for the entire duration of the experiment. We acquired a total of 7 runs in different days/time within the measurement period. The fingerprints from each device in these runs were matched to the reference fingerprints and the EER is reported. The hardware equipment was switched off after each run.

SNR: We evaluate the effect of the signal-to-noise ratio (SNR) on the identification accuracy of our system. For that purpose, we gradually decreased the SNR of all responses in our set of 10 devices. We then performed fingerprint matching between fingerprints extracted at a decreased SNR and reported the EER.

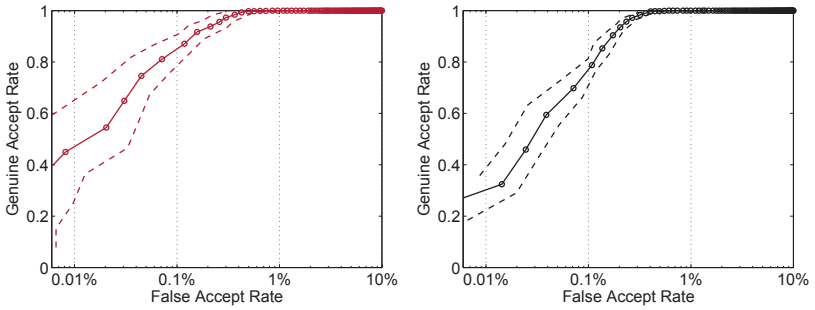


(a) Multiple runs (b) SNR

Figure 6.7: Longer term stability and effect of signal-to-noise ratio (SNR). (a) Accuracy in EER using multiple acquisition runs over a period of 60 days, $N = 20$ and $D = 40$ (Dataset 4). The results demonstrate the stability of our fingerprints, i.e., EER is stable. (b) Accuracy in EER using fingerprints with reduced SNR. Significant influence on the accuracy is only observed for SNR $\leq$ 18 dB and $N \leq 10$.

Figure 6.7(a) shows the accuracy for multiple runs over a period of 60 days. Our system accurately reproduces the test fingerprints of the considered RFID devices, and the fingerprints remained stable during the period of measurement. We also observed an interesting behavior related to the signal acquisition. Even if our time base had a frequency stability of 10^{-7} Hz, the clock frequency seemed to stabilize at a slightly different value (within 0.1 Hz) for some runs. Therefore, the fingerprints acquired in these runs exhibited minor constant offsets that we had to compensate using our channel equalization procedure. While these differences were small, for high accuracy applications, they should be compensated. An alternative approach would be to use a more precise time base. There exists time bases that provide a frequency stability of approximately 10^{-12} Hz with a very low phase noise [52].

The original collected data had an SNR of approximately 21 dB. At that SNR for the 10 devices, we reached an EER of 0% for both acquisition setups. We only started observing a significant decrease of the accuracy when $N \leq 10$ and SNR $\leq$ 18 dB. Therefore, Figure 6.7(b) shows the accuracy at a reduced SNR for $N = 10$. Given that the SNR is directly related to the output power of our arbitrary waveform



(a) Setup I (b) Setup II

Figure 6.8: Receiver operating characteristic (ROC).

generator, the analysis shows that that output power could be reduced two times to 500 mW and still provide high accuracy. This maybe required for certain regulatory considerations. In summary, not only the center frequency, but also the SNR is an important parameter that needs to be carefully adjusted. We note that lower SNR could probably be compensated by using a larger number of responsess to build the device fingerprint. However, this would increase the acquisition and computational time.

Last but not least, it should be noted that our experiments were performed in a controlled environment where there were no large deviations in the ambient temperature. Physical-layer properties are likely to be influenced by large temperature changes. Due to lack of special environment for testing this factor and also the fact that temperature could be well controlled, we did not consider it in our analysis.

We complete the analysis by providing the Receiver Operating Characteristic (ROC) of our identification system (Figure 6.8). The ROC characterizes the accuracy with respect to the operating False Accept Rate (FAR). The two setups exhibit very similar behavior. If the FAR is fixed to a value higher than 0.5% our system will correctly identify 100% of the devices. However if an application is required to operate at much lower FAR points (e.g., 0.01%), the accuracy would be affected. The ROC therefore clarifies that our identification system is currently not suitable for applications that require operation with very low probability of false acceptance. Table 6.3 summarizes the details about the number of genuine and imposter matching scores used in our estimate

Table 6.3: Summary of ROC and EER computation settings (5-fold cross validation)

Set	N	Test matchings		Threshold	EER (%)
		Genuine	Imposter	T	
I	10	500	24900	13.03	0.92 (0.82;1.02)
I	15	300	14900	8.30	0.62 (0.49;0.75)
I	20	200	9800	6.39	0.46 (0.37;0.55)
II	10	500	24900	12.87	0.93 (0.73;1.13)
II	15	300	14900	8.56	0.53 (0.39;0.67)
II	20	200	9800	6.12	0.40 (0.21;0.59)

computations, the operating threshold at EER, the average EER and number of cross validations.

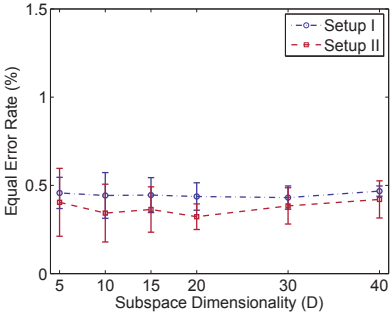
6.4.3 Different Setups Case

In the previous sections, we have shown that our physical-layer fingerprint extraction and matching is accurate within single acquisition setups, demonstrated on two such realizations. We recall that the two antenna setups were composed of same model and manufacturer Tx and Rx antennas. While this property of both antenna setups guaranteed us high and reproducible accuracy, it did not allow us to achieve the desired accuracy across setups.

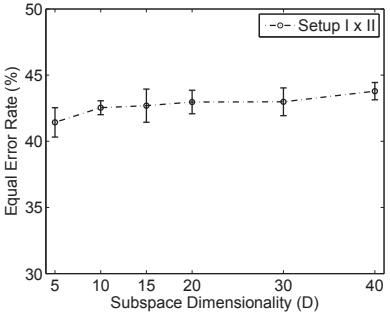
Figure 6.9 presents the accuracy of the two antenna setups for $N = 20$ separately (a) and when the fingerprints from setup II are considered as test fingerprints and matched to the reference fingerprints extracted from setup I (b). The EER is significantly higher reaching rather impractical error rates of 40%.

We discovered that the reason for this poor performance is to do with the significant differences in the wireless channel incurred by the two set of antennas. Figure 6.9 (c) shows the differences in similarity between two consecutive channel measurements with one setup and between setups. We observe large variations in the frequency response from 11 to 15 MHz, the bandwidth at which our fingerprints are extracted. In terms of distance similarity (Euclidian distance), there was more than an order of magnitude difference between the two antenna setup frequency responses.

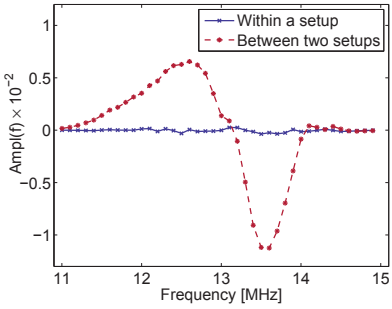
Given the above observations, we applied channel equalization (Sec-



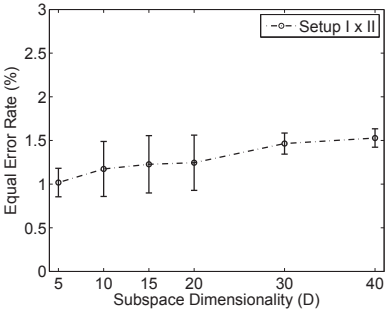
(a) Separate setups



(b) Cross-setup fingerprint matching



(c) Channel difference between setups



(d) Cross-setup fingerprint matching after channel equalization

Figure 6.9: The effect of channel differences on the system accuracy. (a) EER for setup I and II separately. (b) EER after matching fingerprints between setup I and II (c) Channel frequency response within a single setup and between setups (d) EER after equalizing the channel and matching fingerprints between setup I and II. The results show that the two setups exhibit different channel properties which prevent direct cross-setup fingerprint comparison. Channel equalization is required to verify the fingerprints acquired across setups in order to achieve high accuracy.

tion 6.3.5) to compensate the channel effects. Our channel equalization consisted of measuring the frequency response difference between setups, compensating the test fingerprints in one setup with the appropriate channel coefficients and matching those fingerprints to the reference fingerprints in the other setup. We note that the methodology remained the same, except for the additional pre-processing (channel equalization) step for the test fingerprints.

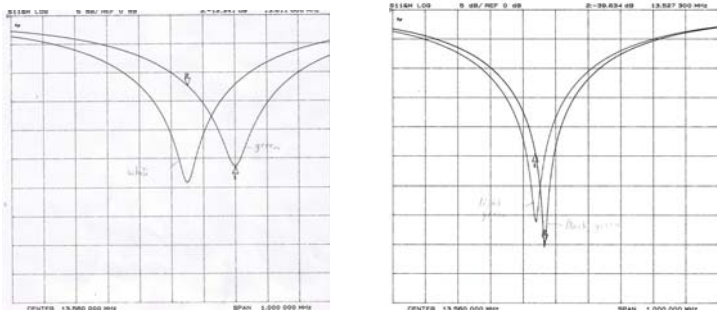
Figure 6.9 (d) shows the EER after the channel equalization of approximately 1% for the lower dimensions. These error rates are very close to the nominal accuracy of our fingerprints on a single setup. They clearly demonstrate that the RFID-enabled devices do have distinctive fingerprints that are independent of the acquisition setup.

On antenna properties. We further investigated the differences between our two sets of antennas by measuring the antenna reflection coefficients (S11 parameter). These are likely to being among the factors that could affect our proposed features[7]. Our Tx and Rx antennas were made of a magnetic loop with capacitors to excite a resonance where $f_r = 1/(2\pi\sqrt{LC})$. L is the inductance of the magnetic loop and C the total capacitance of the combination of capacitors.

For the Tx antennas, assuming a linear frequency span of 1 MHz, the error between both antennas was about 150 kHz/13.56 MHz = 1.1%. For the Rx antennas, similarly assuming a linear frequency span of 1 MHz, the error between both measurements is about 30 kHz/13.56 MHz = 0.2%. Given that standard SMD capacitors usually have a tolerance of approximately 5%, it is considered rather common that the resonance frequencies differ. In fact, 1.1% and 0.2% is already much better than 5% which shows that the manufacturer might have already tried to find the best possible combination of capacitors.

While these measurements and related findings cannot fully explain the differences that we have observed in the two antenna setups, they show that it is difficult to achieve identical antenna characteristics by standard manufacturing and components. The same point would be valid in the case if one wants to produce similar RFID-enabled devices. These observations strengthen the need to better understand the roots of identifiability (Section 12).

[7]Many other antenna characteristics can also be measured. However, the measurement procedures require special environment and hardware. It is out of the scope of this work.



(a) 2 Tx Antennas (b) 2 Rx Antennas

Figure 6.10: Antenna reflection measurements. (a) Set of transmission antennas (b) Set of receiving antennas. The middle line in the graphs is at the center frequency of 13.56 MHz. Each square represents 1 MHz along the X-axis and 5 dB on the Y-axis. The measurements show that even same model and manufacturer antennas exhibit stable and distinguishable properties. For the Tx antennas, the deviation in reflection coefficients is more significant than for the Rx antennas.

6.4.4 Possible Performance Optimizations

In our system, device responses are acquired at a fixed rate of 2 responses per second. This rate was chosen to provide enough time for the system to acquire the data on a computer (Matlab). As a consequence, the raw data to build a typical device fingerprint takes between 5 and 10 seconds for $N \in [10; 20]$. Further pre-processing is required to extract the features which is in the order of milliseconds if PCA is used for feature selection. In the case of feature selection by filtering, additional 5 s are required per response. The times are measured on a machine with 2.00 GHz CPU, 2 GB RAM running Linux Ubuntu. This shows that digital filtering is an expensive operation on high-dimensional data. However, this step can be handled with analog filters before acquisition at the oscilloscope. Moreover, it is even possible to directly acquire our spectral features by using a spectrum analyzer as an acquisition hardware. The combination of analog filtering and spectrum analyzer would allow to directly obtain the fingerprints without passing by the time domain and digital processing. Given the conclusion that no statistical analysis is required, our fingerprints can be directly matched to

stored fingerprint templates. This optimization together with higher acquisition rate would enable device identification within only a few seconds, i.e., practical in real-world deployments.

Last but not least, it should be noted that all the components of the feature extraction can also be implemented efficiently in hardware for further performance improvements.

6.5 Summary

We investigated practical issues in physical-layer identification of HF RFID devices (smart cards). We proposed an improved hardware setup and enhanced feature extraction techniques that enable significantly more accurate device identification than related work. We further showed that our techniques provide physical-layer fingerprints that are stable over longer periods of time. Their identification quality can be preserved across different acquisition setups if channel equalization is applied. Our results significantly strengthen the use of physical-layer identification in anti-cloning solutions.

Part III

Security Analysis and Implications

Chapter 7

Attacks on Physical-layer Identification

In the previous chapters, we proposed a set of techniques for physical-layer identification of active and passive wireless devices. Here, we consider the transient-based identification in Chapter 4 and modulation-based identification [53] and analyze their resilience to impersonation attacks. We chose those techniques as they represent two different classes of physical-layer identification and have been shown to provide the highest identification accuracy for active wireless transceivers.

More precisely, we investigate impersonation attacks by feature replay, signal replay and hill-climbing strategies. In feature replay, we modify radio signals to match the targeted identification features, while in signal replay we capture and replay radio signals in RF. In hill-climbing, we modify the transmitted signals by varying the polarization of the radio waves.

We study the required hardware and conditions that make physical-layer identification vulnerable to impersonation attacks and show that the considered techniques are subject to impersonation; however, transient-based techniques are more difficult to reproduce due to wireless channel and antenna artifacts. We assess the feasibility of performing impersonation attacks by extensive measurements as well as simulations using collected data from wireless devices.

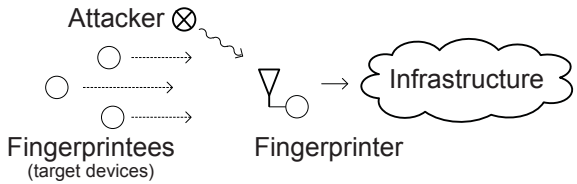


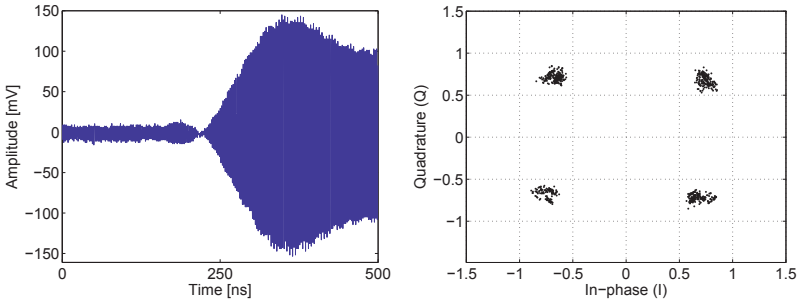
Figure 7.1: Our system consists of a wireless network with a number of wireless devices (fingerprintees) and a fingerprinting device (fingerprinter). We assume that in the system initialization phase the fingerprints of the devices are registered with the fingerprinter. The fingerprints are extracted from the packets sent by the devices and verified by the fingerprinter. The goal of the attacker is to impersonate a target device by generating packets that contain the fingerprints of that device.

7.1 System and Attacker Model

We consider the following setting: a wireless network is deployed in an area $\mathcal{A}$. The network consists of N wireless devices and a fingerprinting device. A physical-layer device identification system is used in the network as described in Chapter 3. During the initialization phase, the fingerprinting device (e.g., wireless access point) extracts a physical-layer fingerprint of each wireless device in its network and stores it in a back-end database. During network operation, the fingerprinter records each packet radio transmission of wireless devices, extracts their fingerprints (according to the specified fingerprinting methodology) and verifies if the extracted fingerprints match one of the reference fingerprints in the back-end database.

We focused on two instances of physical-layer identification systems, one based on the transient technique in Chapter 4 and a second one following the modulation-based technique proposed in [53]. Figure 7.2 visualizes the parts of the transmitted signals used for identification in both techniques. The approaches have been demonstrated to provide accurate identification of active wireless transceivers from the same model and manufacturer.

The attacker's goal is to break the physical-layer identification system in the network which operates at a fixed application specific threshold T. The threshold serves as an Accept/Reject decision boundary for determining if a given fingerprint is genuine (belonging to the set of legitimate devices) or if it is an imposter (belonging to an intruder de-



(a) Transient technique (b) Modulation technique

Figure 7.2: (a) Transient-based techniques extract unique features for device identification from the radio signal transient shape at the start of each new packet transmission. (b) Modulation-based techniques extract frequency and constellation symbol imperfections (i.e., modulation errors).

vice) (see Chapter 3 for more details). The purpose of the attacker is therefore to create impersonating signals whose identification features fall in the accept region of the identification system.

Definition 1 *We say that an impersonation attack is successful with a probability p if the matching score between fingerprints of a device targeted for impersonation (D) and that of the attacker (A) is below the application specific threshold T with probability p.*

Given that the considered modulation-based identification technique was evaluated in related work using device classification, we adapt the above definition in the case of classification. We note that in standard classification, there is no notion of rejection based on threshold, i.e., the classifier assigns an unknown device fingerprint to the device that has the highest similarity in the entire set of devices. Therefore, Definition 1 should be modified as follows.

Definition 2 *We say that an impersonation attack is successful with a probability p if the classification process assigns the fingerprints of attacker (A) to the class of fingerprints of the device targeted for impersonation (D) with probability p.*

We consider the following three impersonation methods and related assumptions:

- *Impersonation by Feature Replay*: In this attack, we modify the radio signal characteristics of an attacker device to closely match all or part of the features used to identify the device targeted for impersonation. We assume that the attacker knows the features used by the identification system and the exact feature extraction, matching and decision making processes.

- *Impersonation by Signal Replay*: In this attack, we record signals from a device targeted for impersonation and retransmit those signals without modification at RF with high-end arbitrary waveform generators. We do not assume any knowledge of the features used for identification.

- *Impersonation by Hill-Climbing*: In this attack, we vary the antenna polarization of an attacker device during transmission in order to find a polarization degree that closely match the features of a device targeted for impersonation. We do not assume any knowledge of the features used for identification.

For all impersonation methods, the attacker is in possession of all necessary hardware equipment to measure and reproduce radio communication signals at any location. He can also build a second fingerprinting device for emulating the entire identification process. The attacker does not have access to the true reference fingerprints captured by the fingerprinter F and the only feedback he can get from F is an Accept/Reject response. However, in some application scenarios the attacker might have access to the location of the fingerprinter in order to collect the signals from it.

As an instance of the above system and attacker models, we considered a network with 3 wireless devices (Universal Software Radio Peripheral - USRP [13]) and the fingerprinting device is a high-end Agilent Digital Signal Analyzer (DSA) [54]. The attacker is in possession of two devices for the proposed impersonation attacks: a 4-th USRP device and a high-end 20 GS/s arbitrary waveform generator (Tektronix AWG 7000B [55]). These two types of devices allow evaluating an attacker with different strengths: low-cost USRP versus high-quality, but costly signal generator.

7.2 Impersonation by Feature Replay

In this section, we present an impersonation attack on the modulation-based identification proposed in [53]. We first provide background on the identification technique and then detail the attack design, implementation and test scenarios.

7.2.1 Modulation-based Identification

Modulation-based identification was proposed in [53] as an alternative to transient-based techniques to uniquely identify same model and manufacturer wireless devices. This class of techniques focuses on extracting unique features from the modulated signal. More precisely, the authors in [53] extracted five distinctive signal properties of IEEE 802.11b modulated signals, namely the Frame frequency offset (F1), Frame SYNC correlation (F2), Frame I/Q origin offset (F3), Frame magnitude error (F4) and Frame phase error (F5). These five features together formed a fingerprint of the wireless device, subsequently used for device identification. They were extracted from each packet frame by means of a high-end vector signal analyzer at 70 MHz intermediate frequency (IF) for high precision. The accuracy of the fingerprints for device identification was tested with a k-NN classifier with L1 distance similarity and an SVM classifier with maximum-margin separation [26]. The experimental results from over 100 IEEE 802.11 Network Interface Cards (NIC) demonstrated an identification (classification) accuracy of over 99%.

7.2.2 Attack Design

In this attack, we use the capabilities of a USRP with the GNU radio software library [56] to modify parameters in the radio transmission of individual 802.11 packets. In particular, we find that a combination of digital and analog techniques can be applied to modify F1, F3, F4 and F5 detailed below. The basic ideas are summarized in Figure 7.3.

Frame frequency offset (F1) is the most discriminative feature [53] in the considered modulation-based technique. It represents the difference (offset) between the carrier frequency of the fingerprintee and the fingerprinter. In order to pretend being a given device with respect to F1, we need to adjust the carrier frequency of our attacking device to the carrier frequency of the targeted for impersonation device.

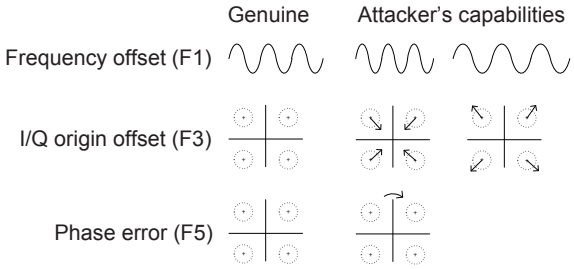


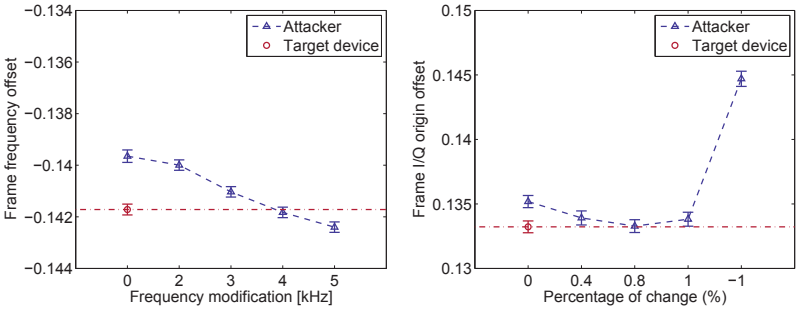
Figure 7.3: Attacks on modulation-based identification. We are able to modify the signal frequency offset (F1) by changing its carrier frequency in the analog domain, the I/Q origin offset (F3), magnitude (F4) and phase (F5) errors by modifying its original constellation in the digital domain.

We achieved this by using the analog circuit of the USRP which allows arbitrary changes of the carrier frequency with the precision of 0.01 Hz.

Frame SYNC correlation (F2) is the second most discriminative feature. It measures the modulation quality of the frame synchronization preamble by normalized cross-correlation with the ideal synchronization sequence. We found that this feature is difficult to modify in a deterministic way. We later demonstrate that it is not necessary to modify this feature in order to impersonate a targeted device with high accuracy. We also show that an attack including impersonation of this feature improves the impersonation accuracy (Section 7.3).

Frame I/Q origin offset (F3) is the third most discriminative feature in the modulation-based identification. It shows the distance of the ideal I/Q plane centered at (0,0) and the average of all measured I/Q values (symbols in an I/Q constellation) within a packet frame. The Frame I/Q origin offset is usually specific to a given transceiver under the assumption that the analog circuit is provided with the ideal fixed constellation symbols (e.g., $\pm 0.707 \pm 0.707i$ in a Gray-coded constellation). The latter are generated digitally in the digital signal processing (DSP) module of the radio transceiver. In our attack, we digitally shrink or expand the ideal constellation symbols' position in order to change the Frame I/Q origin offset.

Frame Magnitude (F4) and *Phase* (F5) errors are the least discriminative features in the modulation-based identification. The frame magnitude error is the average difference in the scalar magnitude between all ideal and measured I/Q symbol values, while the frame phase



(a) Frequency offset (b) I/Q offset

Figure 7.4: Experimental results on incremental modification of the frame frequency offset (F1) and frame I/Q origin offset (F3). The results show that we can deterministically change feature values of one device in order to match those of a targeted device.

error is the average difference in phase (i.e., angle in degrees) between the ideal and all measured I/Q symbol values in the frame. We modify these values in the digital domain by shrinking/expanding the I/Q symbols in order to impersonate these features.

It is important to note that the digital modifications of F3, F4 and F5 must take into consideration the analog circuit deviations that occur in processing the signal from the D/A converter to the antenna and compensate them. In addition, any modifications must also not go beyond the standard tolerances of the impersonated technology [57].

In Figure 7.4 we show some experimental results from deterministically decreasing the features F1 and F3 of the attacker's device to the values exhibited by the target device (Device 2). In particular, the frame frequency offset is closely equalized at $f = f_C + 4.7$ kHz where f_C is the original carrier frequency of the attacker's device. The Frame I/Q origin offset exhibited by the target device was closely equalized by shrinking the attacker's QPSK constellation points by a factor of 0.7%.

7.2.3 Measurement Setup and Attack Procedure

For the purpose of performing and evaluating the attack, we used four USRPs (3 genuine devices and 1 attacker device). For close matching of the signals used in [53], we developed an 802.11-style QPSK digital

baseband modulator. The frame is constructed according to the IEEE 802.11 specification [57] with a preamble (used for coarse frequency offset estimation), followed by a longer preamble for fine frequency offset and channel estimation and the actual data payload. The frequency estimation algorithms were implemented according to [58] which are well established algorithms for that purpose. It should be noted that more sophisticated algorithms will only improve the computation of the errors. The data payload was modulated using QPSK modulation [40]. All packet frames contained the same content transmitted at a data rate of 1Mb/s.

The design of the fingerprinter is shown in Figure 7.5. Each signal was captured with a standard 2 dB dipole antenna and subsequently amplified by an ultra low-noise and low-power amplifier (NF=0.15 dB) and filtered by a low insertion loss bandpass filter to eliminate radio frequencies outside the industrial, scientific and medical (ISM) band. The received signal was digitized by an Agilent Digital Signal Analyzer [54] and processed by our 802.11-style QPSK digital demodulator for feature extraction. Feature matching and classification was performed offline with Matlab. The genuine devices were positioned at fixed locations to the fingerprinter's antenna. We note that for the modulation-based features the distance should not have an effect on the classification accuracy as outlined in [53].

We started the impersonation attack by modifying the carrier frequency in order to reach the one of the targeted genuine device. We determined the carrier frequency of the targeted device by analyzing the power spectrum density of the radio transmission. Subsequently, we adjusted the frame I/Q origin offset, magnitude and phase of the attacking device by digitally modifying its ideal QPSK constellation symbols (Figure 7.3) to closely reproduce the feature values of the targeted device after the entire analog processing at the attacking device. Here, we chose to measure the targeted device communication, compute the corresponding features and then adjust them appropriately. There is a second possible approach that consists of launching a hill-climbing attack [5] by repeatedly sending signals with modified features until they are identified as the targeted device.

7.2.4 Attack Evaluation Results

In this evaluation, we used the capabilities of a software-defined radio for feature replay and followed the design described in Section 7.2. For

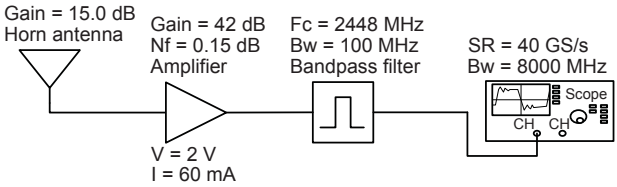


Figure 7.5: Fingerprinter hardware setup.

data collection, feature extraction and matching, we followed the procedures in [53]. We briefly summarize them: we used 80 valid frames [1] per genuine device and computed the corresponding F1, F2, F3, F4 and F5 features. A device reference fingerprint was built from a total of 20 frames and the remaining 60 frames were used to build testing fingerprints. All presented results were validated using 4-fold cross validation [26]. The similarity score between reference and testing device fingerprints was computed with L1 distance as proposed in [53].

For evaluation with respect to Definition 1, we had to fix the application specific threshold T. We chose to set T to the threshold of the EER operation point which is the mostly used threshold for evaluation [5, 25]. In our particular case, EER = 0% and the corresponding $T_{EER} = 0.05$. It should be noted that if one would like to have a realistic estimate of the EER and corresponding T, a much larger amount of devices must be considered [5]. Therefore, the above results, should only be used to assess the attacker's ability to go below the system's operating point T.

To visualize the impersonation attack performance, we computed the genuine, imposter and attacker scores in all folds and show them in the form of histograms. The genuine matching scores were computed by matching the testing frames from the devices to their respective reference fingerprints. The imposter matching scores were computed in the same way, but using the reference fingerprints of the other devices. The attacker scores were computed by matching the impersonating (attacker) frames to the reference fingerprint of the targeted device. We used an average of 5 frames to compute the overall matching score. This is consistent with [53] where it was shown that averaging over more than 4 frames is needed to achieve the highest accuracy.

Figure 7.6 shows the matching scores of the impersonating (at-

[1] We consider as valid the frames that comply with the standard [57].

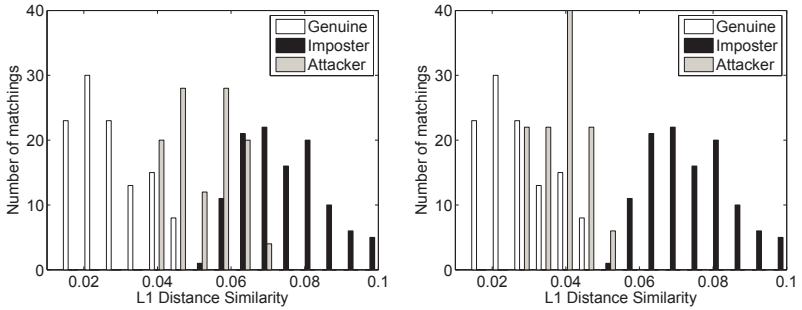


Figure 7.6: Modulation-based identification: genuine, imposter and attacker matching score histograms. (a) Impersonation attack by feature replay of F1 and F3. (b) Impersonation attack by feature replay of F1, F3 and F5. The device fingerprints were computed by averaging the features over 5 packet frames. The application specific operating threshold was fixed to $T = 0.05$.

tacker's) frames against the target device (Device 2). If we reproduce only F1 and F3 features, the impersonating frames will be rejected by the system in approx. 60% of the cases according to Definition 1 with $T = T_{EER}$. This is shown in Figure 7.6a. If we lower the operating point, the system can reject 80% of the impersonating frames while only slightly increasing its FRR. On the other hand, if we reproduce F1, F3 and F5 features, we successfully place 98% of the impersonating frames below T_{EER}, i.e., the impersonation success rate is 98% (Figure 7.6b).

It should be noted that if the system can tolerate some false rejects, it can reduce the attack success rate, however annihilating the attack without significantly increasing the FRR cannot be achieved (e.g., at $T = 0.025$ the system will reject all impersonating frames, but also 50% of its genuine frames).

In Figure 7.6b, we also observe that the attacker matching scores are still shifted towards the imposter histogram scores. This is due to the fact that our attack did not modify the F2 and F4 characteristics of the attacking device. We found that F2 was hard to change digitally and F4 could not be independently modified without influencing F3 due to computational dependence. Therefore, we chose to modify the most discriminative of the two, F3. We now show that the impersonation attack by signal replay sufficiently preserves all the features and places

Table 7.1: Classification success rates on genuine devices

1-NN	3-NN	5-NN	SVM
87.65%	97.78%	100%	100%

all impersonating frames in the genuine matching score space.

Impersonation in classification

We considered the k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) classifiers trained and executed as in the related work [53]. For complete compliance with [53], in the k-NN [2] classifier half of the training frames (10) were discarded from the reference device fingerprint by removing the frames whose features deviated the most from the overall mean. No frames were removed from the testing set. The similarity measure was L1 distance.

The classification success rates using k-NN and SVM classifiers for distinguishing the 3 genuine devices are shown in Table 7.1. Both k-NN and SVM classifiers successfully classify the fingerprints of the genuine devices. Inline with [53], the k-NN classifier requires averaging over a number of frames ($k \geq 4$) to reach its highest accuracy. In our case, a success rate of 100% was reached for $k = 5$.

After tuning our attacking software-defined radio device in order to match the feature F1, F3 and F5 of the target device (Device 2) as well as possible (see Figure 7.4), we injected the attacker's collected frames in the k-NN and SVM classifiers by replacing all Device 3 frames and computed again the classification success rates.

The results in Table 7.2 show the success rate of classifying genuine frames and impersonating frames with feature replay of F1 and F3. The impersonation attack success rate is 62% for the 5-NN classifier, while for the SVM classifier it tops 100%. On the other hand, if the attacker performs a feature replay with F1, F3 and F5, it will impersonate both classifiers in 100% of the cases (Table 7.3). An impersonation attack by signal replay also succeeds in 100% inline with the previous results.

It should be noted that the above results on classification are highly dependent on the number of classes (devices) and the separability be-

[2] We complied to the definition of parameter k and notation k-NN in [53]. We note that these definitions are different from the commonly accepted ones in pattern recognition [26].

Table 7.2: Genuine and attacker classification success rate on Device 2 by feature replay of F1 & F3

	1-NN	3-NN	5-NN	SVM
Input	Device 2	Device 2	Device 2	Device 2
Device 2	73.33%	98.33%	100%	100%
Attacker	50%	50%	62.33%	100%

Table 7.3: Genuine and attacker classification success rate on Device 2 by feature replay of F1, F3 & F5

	1-NN	3-NN	5-NN	SVM
Input	Device 2	Device 2	Device 2	Device 2
Device 2	73.33%	98.33%	100%	100%
Attacker	63.33%	98.33%	100%	100%

tween different device fingerprints. It is interesting to observe that a system with highly discriminative classifier such as SVM was easier to impersonate ($p = 100\%$ with 2 reproduced features). In our case, this is due to the fact that SVM builds large decision boundaries well separating the three devices. Therefore, few modifications of the features towards the features of one of the 3 devices make the impersonating frames cross the decision boundary of that device. However, if the number of classes is larger, this might not be sufficient and more impersonated features would be required. This finding suggests that if the attacker can modify only some of the features of an identification technique, a good strategy would consist of identifying a device in the network that differs the most from all other devices and try impersonating that device. We also point out that a general problem of standard classification is that without a rejection criterion, the attacker would be always assigned to one of the genuine devices.

7.3 Impersonation by Signal Replay

In this section, we demonstrate a device impersonation attack by radio signal replay on modulation and transient-based identification. As opposed to the previous attack, we do not modify the signal characteristics, but retransmit the entire radio packet frame in its integrity at the RF frequency. For the impersonation attacks, we considered the

same modulation-based identification technique (Section 7.2.1) and the transient-based technique described in Chapter 4.

7.3.1 Attack Design

In this attack, we use the capabilities of the 20 GS/s arbitrary waveform generator Tektronix AWG 7000 Series [55]. Due to its fast digital to analog converter, this generator can output any 802.11 signals directly at the required radio frequency of 2.4 GHz. Unlike in the previous attack, where the attacker tries to match as close as possible the features of a device targeted for impersonation, in this attack, we captured the signals of the target device at the RF frequency and replayed them without any modification. This attack is more powerful than feature replay attacks since it does not require knowledge of the features that are extracted by the fingerprinter. It simply requires that the attacker records the transmissions of the targeted device.

A more sophisticated attack based on signal replay would be to produce crafted signals by replaying parts of the message. In the case of modulation-based identification, the attacker can replay the preamble part of the message to reproduce F1 and F2 and craft its own payload. Furthermore, the attacker can also craft his own payload and at the same time reproduce all F3, F4 and F5 features. This is due to the fact that he has full control over the features in the digital domain and relies on the arbitrary waveform generator to directly output the crafted signal in RF thanks to the 20GS/s digital-to-analog (D/A) converter.

In transient-based identification only the transient part of the signal is used for identification. Therefore, the attacker can create a message with the transient part in its integrity concatenated with the actual payload. In this case, the replay attack becomes an impersonation attack. We point out however that such an attack can only be mounted with a high-end arbitrary waveform generator which has the available bandwidth to output the crafted transient signals.

7.3.2 Measurement Setup

To evaluate the impersonation attack by signal replay, we built an experimental setup in a lab environment. The setup consisted of two tripods: the first was used to hold the device to be impersonated; the second holds two identical 2 dB dipole antennas, connected to the fingerprinter and the attacker respectively. Both antennas were fixed on

the platform separated by a distance of 30 cm in order to avoid near-field effects, but still get a high signal-to-noise ratio (SNR). The design of the fingerprinter was the same as shown in Figure 7.5 with an additional implementation of the transient-based feature extraction and matching procedures that were proposed in [59].

We first collected frames from the targeted device. Subsequently, we replayed the recorded frames to the fingerprinter and evaluated the attack performance.

7.3.3 Attack Evaluation Results

In this evaluation, we used our high-end 20 GS/s arbitrary waveform generator to retransmit device packet frames in their integrity at RF. Following the procedures in Section 7.3, we collected 20 frames from the target device (Device 2) at the attacker's position. Subsequently, we retransmitted those frames towards the fingerprinter twice, resulting in 40 impersonating frames. It should be noted that the device signals were captured at RF = 2.4 GHz and sampling rate of 20 GS/s in order to preserve as much as possible the radio signal (e.g., no downconversion to intermediate frequency). The genuine and replayed matching score histograms are shown in Figure 7.7.

We observe that all the genuine score bins are filled with the scores resulted from matching with the impersonating (replayed) frames. The results demonstrate that signal replay at RF is a powerful attack that makes the impersonating (attacker's) frames very difficult to distinguish from the genuine device frames.

As in the previous section, we used the high-end arbitrary waveform generator to retransmit transient signals. We implemented the transient-based identification technique in [59] and followed the proposed procedure in Section 7.3. We collected transient signals from 3 Tmote Sky sensor nodes in order to fully match the conditions in [59]. We present our results for replaying these signals both using a cable and air interface to better assess the limitations of our attack.

Figure 7.8 shows the genuine and imposter histograms from matching transient-based features from the original devices captured with our setup as well as the histograms of matching original and replayed transients by arbitrary waveform generator over a cable and over the air. The results clearly show that the replayed signals over the cable closely match the original signals. This is an important result as it shows that the arbitrary waveform generator can retransmit transient signals with

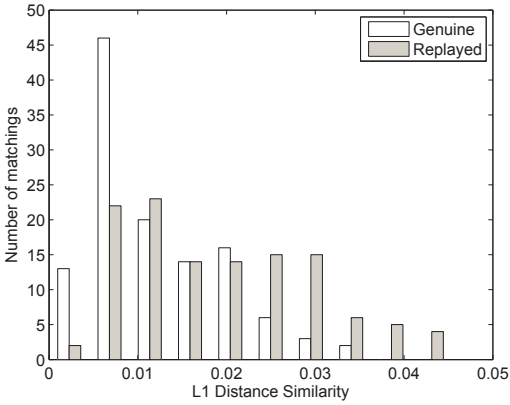


Figure 7.7: Modulation-based identification: genuine and replayed (attacker) matching score histograms obtained by signal replay at RF with a 20 GS/s arbitrary waveform generator. The two histograms are overlapping making it very difficult to distinguish a genuine device from the attacking device.

high accuracy.

On the other hand, replaying the same signals over the air altered the signals, so that the replayed signals were recognized as imposter signals and the impersonation attack failed. We further investigated the issue and discovered that in addition to the device fingerprint in the transient-based features, there is also the presence of the wireless channel characteristics. In order to confirm the channel effect on the transient, we simulated a frequency selective channel to estimate the degree of modification of the original transient signals under channel changes. The results showed that different channels modify the transient features and the system rejects all attacker's replayed transient signals at the threshold $T = 3.01$ [59].

Impersonation of transient-based spectral features [59] is inherently more difficult due to channel and antenna effects on the transient part of the signal as shown in our analysis. While our high-end signal generator can accurately reproduce it as well over a cable (i.e., fixed channel), replaying over the air from a different location is not likely to be successful to impersonate a device. However, we could impersonate the targeted device from its location. There are two possible scenarios that

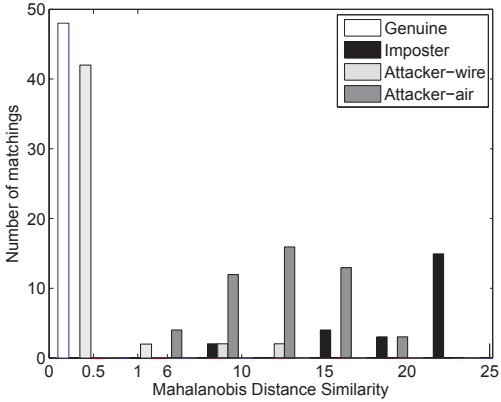


Figure 7.8: Transient-based identification: genuine and imposter matching scores are from the genuine fingerprinting system; attacker matching scores with fingerprints of the device targeted for impersonation over a wire and over the air. The attacker transient signals over cable are indistinguishable from those of the genuine device.

could achieve this depending on the attacker model. In the first scenario, if we are allowed to measure the transient signal of the targeted device before actual transmission through the antenna (e.g., capture the device and measure over a cable with an oscilloscope), we can then replay it with the arbitrary waveform generator from the same location. In the second scenario, a possible compromise of the fingerprinter would reveal the transient signal received at the fingerprinter. Subsequently, we need to estimate the wireless channel response between the targeted device location and the fingerprinter, compensate the transient signal accordingly and replay it with the arbitrary waveform generator. This second scenario can also be applied if the attacker is allowed to collect frames at the location of the fingerprinter.

In summary, the modulation-based features and L1 distance similarity measure proposed in [53] are vulnerable to impersonation attacks by feature and signal replay. Impersonation by signal replay at RF makes the impersonating (attacker) frames almost indistinguishable from the genuine frames of the targeted for impersonation device.

Table 7.4: Hill-climbing attack on device ID = 9.

N	50	20	10	5
Hill-attack distance	42.74	38.12	35.89	21.61
Threshold	3.01	4.10	6.74	16.04

7.4 Impersonation by Hill-Climbing

In this section, we analyze the resilience of transient-based features to a hill-climbing attack by varying antenna polarization and show that impersonation would be possible if a small number of signals is used for feature extraction.

7.4.1 Attack Design and Results

A hill-climbing attack is a well-known attack on biometric recognition systems [5]. This attack consists of repeatedly submitting data to an algorithm with slight modifications. Only modifications that preserve or improve the matching score are kept in the process. Eventually, a score that exceeds the identification system operating threshold T might be achieved. This results in successful impersonation without providing the genuine biometric.

To perform the attack, we would ideally need a specialized device that is able to create transient signals (similar to the ones generated by the sensor nodes) and at the same time allow for introducing variations in it. In order to closely much these requirements, we decided to use 3 additional sensor nodes that are not part of the population of 50 sensor nodes used so far. In order to create variations in the shapes, we mounted external antennas on the sensor nodes. We then manually changed the radio wave propagation by rotating the nodes' antenna in order to find a polarization that impersonates a sensor node from the targeted network.

We collected 50 transient data samples from 7 different polarizations of the antennas of the 3 sensor nodes. We then supplied these transient data samples to identification system. Figure 7.9 displays the matching scores obtained during the attack in a 3D representation for $N = 5$. For clarity reasons, all scores that exceed 100 are not displayed.

The identification procedure becomes more vulnerable to the impersonation attack when N decreases. In particular, the matching scores

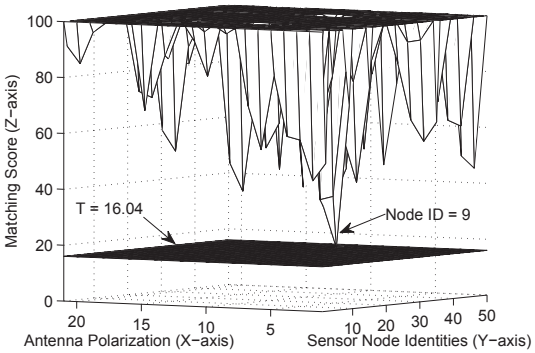


Figure 7.9: Hill-climbing attack scores. The X-axis contains the 21 (3 sensor nodes x 7 antenna polarizations) attacking features; the Y-axis shows the reference features of the 50 sensor nodes targeted for impersonation; the Z-axis is the matching score obtained between each attacking and reference features. The thick surface is the Accept/Reject threshold (T=16.04).

against one of the sensor nodes (ID = 9) for $N = 5$ were consistently very close to the Accept/Reject threshold $T = 16.04$ (Table 7.4). Device impersonation is possible for $N \leq 5$. A real system needs to consider acquiring $N > 5$ transient identification signals in order to build the fingerprint to ensure protection against this type of impersonation.

7.5 Summary and Discussion

We investigated the feasibility of performing impersonation attacks on certain physical-layer identification techniques. We designed and implemented impersonation attacks by feature replay, signal replay and hill-climbing on modulation and transient-based identification. Our results show that modulation-based features can be impersonated with high accuracy by simply modifying and replaying them. Transient-based features can also be reproduced using a high-end arbitrary waveform generator over a wire, but they are hard to record by an external attacker since they are channel- and antenna-dependent. Therefore, actual replay of transient features over the air is likely to succeed only from the location of the device targeted for impersonation. In the alternative case, where the attacker could only launch impersonation attacks from other locations, he would need to accurately estimate and com-

pensate appropriately all channel properties present in the transient features. Given that feedback signaling is typically needed for channel state information (CSI) estimation [60], it would be difficult to launch deterministic impersonation attacks without receiver cooperation. We note however that further work is required to better quantify the exact influence of the channel and antenna on the transient features.

Given that the characteristics of transient-based features are related to the features used in RFID identification in Chapters 5 and 6, i.e., they are both frequency based, we conjecture that replay attacks with high-end arbitrary waveform generators are also feasible in that case.

Chapter 8

Implications on Selected Applications

In this chapter, we discuss the implications of physical-layer device identification on the security and privacy of wireless networks. These implications stem from the main findings and results in this thesis and are presented in the context of selected applications. We first provide a classification of possible attacks on physical-layer identification. We then describe the requirements of each application and analyze the feasibility and security aspects of physical-layer device identification.

8.1 Classification of Attacks

We distinguish between attacks on the identification system that aim at subverting the decision of an application (e.g., grant or not grant access) and attacks on the anonymity of wireless devices that aim at identifying them disregarding their will to be identified. We do not discuss attacks that could be performed by an attacker who controls internal system components. The latter are classical to identification systems and have been already extensively discussed [5].

We assume a Dolev-Yao style attacker [61]. The attacker has the ability of observing, capturing, modifying, composing, and (re)playing identification signals transmitted by authorized devices. To observe and capture identification signals, the attacker needs to have access to the identification area or may directly acquire identification signals

from the target device. This implies temporal possession and possibly knowledge of the challenges used to acquire the identification signals. The attacker can arbitrarily modify and compose identification signals. To (re)play identification signals, the attacker is allowed access to the identification area. We discuss restrictions of this strong attacker in the context of each application.

8.1.1 Signal replay attack

In a signal replay attack, the attacker's goal is to observe analog identification signals of a target device, capture them in a digital form (digital sampling), and then transmit (replay) these signals towards the identification system by some appropriate equipment. The attacker does not modify the captured identification signals, i.e., the analog signal and the data payload are preserved. This attack is similar to message replay in the Dolev-Yao model and is illustrated in Figure 8.1. The difference resides in the level of replayed information. The message replay attack preserves the bits of information within the message (e.g., 01101), while a signal replay aims at preserving the digital sampling of the signal. Message replay is subsumed by signal replay. It should be noted that replaying digital signal samples cannot be as exact as replaying information bits. This is due to inherent randomness in hardware components and the wireless medium. Improvement of replay accuracy can be achieved by high-end hardware and controlled wireless medium.

The signal replay attack does not assume attacker's knowledge of the feature extraction and matching procedures used by the identification system. However, knowledge on how to observe, capture, and submit identification signals to the system is required.

Given the requirement of digitizing analog signals and subsequent analog conversion for radio transmission, the attacker needs appropriate devices for capturing and rendering the identification signals. Some knowledge on the features used (e.g., baseband or RF) would narrow the choice of device. The range of devices includes low-cost hardware [13], high-end signal analyzers [54], and arbitrary waveform generators [55].

It should be noted that analog identification signals can also be relayed without being previously stored in a digital form. A few components such as amplifiers and antennas are required to perform the task. An example of such relay attack is our physical-layer relay in Chapter 2.

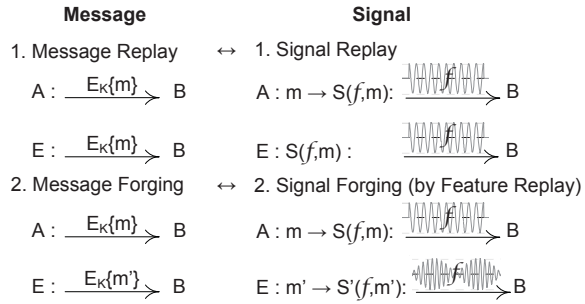


Figure 8.1: Comparison of message and signal attacks. A and B are genuine parties. E is a Dolev-Yao attacker that can observe, capture, modify, and compose messages or signals. A signal replay attack aims at preserving the digital sampling S(f,m) of the signal carrying message m and features f as opposed to a message replay attack which preserves the bits of information within the message. In signal replay, the features do not need to be known to the attacker. In a feature replay attack, the attacker needs to know the features f and then can compose signals that reproduce f. The difficulty for the attacker resides in composing the signal S'(f,m') that preserves f.

8.1.2 Feature replay attack

Unlike signal replay attacks, where the goal of the attack is to reproduce the captured identification signals in their integrity, the feature replay attack creates, modifies or composes identification signals that reproduce only the features of the identification system. The analog representation of the forged signals may be different, but the features should be the same (similar) as illustrated in Figure 8.1.

The feature replay attack is comparable to message forging in the Dolev-Yao model where the attacker can arbitrarily modify and compose messages. The difference is that forging involves analog and/or digital signal samples and data payload (information bits) as opposed to information bits only in the Dolev-Yao model. The difficulty of this attack resides in the ability to compose signals while preserving the identification features. In order to impersonate a device, the attacker needs to know the features that the identification system extracts to identify a device and needs to be able to forge signals while preserving their distinctive features. This corresponds to attacks on message authentication, where the attacker typically needs to know the secret key in order to create an authentic message or has to be able to modify/forge

existing messages such that they appear authentic to the receiver.

The feature replay attacks could be launched in a number of ways. Similar to a signal replay attack, special devices such as arbitrary waveform generators could be used to produce the modified or composed signals. The attack could also be launched by finding a device that exhibits similar features to a targeted device, which is then used during the identification procedure. This scenario is relevant in applications where a large set of possibly same-model-same-manufacturer devices could be obtained by the attacker.

A third way for launching feature replay attacks is to replicate the entire circuitry of the targeted device or at least the components influencing the identification features. This is probably the hardest way as it assumes precise knowledge of the hardware component(s) affecting/causing the features. It is unclear if that is feasible in practice.

8.1.3 Other attacks

For completeness, we should mention the coercion attacks. In this attack, an attacker needs to come into (temporal) possession of an authorized device, and use it during the identification procedure. While straightforward, such an attack has relevant implications in certain applications.

8.2 Intrusion Detection in Wireless Networks

As a first application of physical-layer identification, we consider its most widely discussed use as an intrusion detection mechanism in wireless networks. Several scenarios can be envisioned covering wide and local area networks.

In a first scenario, a primary layer of access control by a cryptographic mechanism that authenticates the devices with the network is deployed to prevent unauthorized devices (users) of using the resources of the network (e.g., WLAN). Physical-layer identification is deployed in the network as a second layer of access control to defend against authentication break-in (e.g., cryptographic key compromise). An attacker who succeeds to compromise the authentication keys will not be able to gain access to the network with her own device unless he is able to subvert the physical-layer identification.

Additional benefits of physical-layer identification in such settings are (i) detection of multiple identification numbers used by a device (e.g., MAC addresses in WLAN) (ii) detection of an identification number or a cryptographic key used by multiple devices.

In a second scenario, physical-layer device identification can be used to detect relay attacks. The primary layer of access control is again based on cryptographic authentication. However, the attacker does not aim at compromising the authentication, but forwards legitimate packets received at one point of the network to another point that is usually multiple hops. Relaying legitimate network packets can be used to influence network operations [62, 63] or to gain access as demonstrated in Chapter 2. Given that the attacker is only able to perform relay attacks with his own wireless devices, his activity would be detected as the physical-layer identities would be different from the legitimate ones. Once an intrusion attempt is witnessed, the network can take appropriate measures to raise an alarm or ignore the affected packets.

System requirements and analysis. The discussed scenarios pose specific requirements to the physical-layer identification system. In the case of static network configurations, physical-layer device fingerprints need to be temporally stable and occasional external interferences have to be removed by appropriate exception handling. In the mobile case, the devices would typically communicate between each other and with the infrastructure from random locations under different wireless channel conditions. Mobility implies that the physical-layer device fingerprints have to be resilient to distance, location and channel randomness.

Our results on transient-based identification (Chapter 4) suggest that access control in static with quasi-static channels could be provided with high accuracy. However, the fingerprints extracted from the transient signal do not only contain device specific information, but also wireless channel characteristics which vary depending on the location and distance. These fingerprint properties introduce severe restrictions on the usability of transient-based identification (e.g., authentication must be performed only from a particular location). Other physical-layer identification techniques such as [53] could be more appropriate in mobile scenarios.

Security requirements and analysis. In terms of security requirements, the identification system must be resilient to remote impersonation attacks. In particular, given the distance and uncontrolled nature

of the considered wireless networks, attacks by signal and feature replay are of a particular concern.

Our results show that the investigated transient and modulation-based physical-layer identification approaches are vulnerable to impersonation attacks and cannot be safely used to detect intrusion. With the appropriate equipment, the attacker is able to reproduce physical-layer identification features (fingerprints) with high accuracy. Certain physical-layer techniques can also be impersonated with relatively low-cost devices such as software-defined radios [13]. Furthermore, the less location sensitive a physical-layer fingerprint is, the easier it is for the attacker to impersonate it from any location.

Our results further motivate the investigation of techniques that can detect impersonation. These techniques have to either make sure that the signals or features are not known to the attacker such that he cannot replay them or have to detect from the replayed signals that they have been replayed. Whether such impersonation detection is feasible, is an open question that motivates future work.

8.3 Document Cloning Detection

We already witness the inclusion of wireless technologies, more precisely RFID in identity documents [36], electronic passports [64] and payment cards [65]. RFID allows storage of data (e.g., photos, private keys) and is intended to authenticate document holders and communicate data to authorized systems in a secure way.

Despite a number of protection measures, it has been shown that confidentiality of passport data can be compromised [66] and data stored on the RFID chip can be successfully extracted and cloned on other RFID-enabled devices [37, 67–69] even if defense mechanisms specified by the standard [64] are in place.

Our results from Chapters 5 and 6 strongly support the use of physical-layer RFID identification to protect against document cloning in two different settings. In the first setting, the fingerprints are measured before RFID deployment and are stored in a back-end database, indexed with the unique document identifier. When the authenticity of the document with identifier ID is verified, the fingerprint of the document transponder is measured and then compared with the corresponding transponder fingerprint of document ID stored in the database. In the second setting, the physical-layer fingerprints are measured before

their deployment, but are stored in the RFID memory, digitally signed by the document-issuing authority and protected from unauthorized remote access. When the document authenticity is validated, the binding between the document ID and the fingerprint stored on the RFID is ensured through cryptographic verification of the authority's signature. If the signature is valid, the stored fingerprint is compared to the measured fingerprint of the document. The main advantage in this use case is that the document authenticity can be verified offline. The main drawback is that the fingerprint is stored on the chip and requires appropriate memory resources and access protection.

System requirements and analysis. The requirements on the properties of the physical-layer fingerprints are significantly different compared to the intrusion detection scenario. Given that the anti-cloning verification must be achievable in multiple locations (e.g., country border controls), special purpose-built devices need to be devised. This relaxes the requirement on the fingerprints to be robust to environmental factors and channel effects as these can be controlled in the purpose-built measurement setup. However, the setup should be of high quality in order to preserve the fingerprint from undesirable distortions. Additionally, the fingerprints have to be compact enough to fit in the RFID chip memory. The cloning detection accuracy will depend on the system error rates.

Our RFID identification techniques meet the above requirements in case of identity documents and electronic passports. The proposed techniques can distinguish RFID-enabled devices with high accuracy, are stable over extended periods of time and can be verified on different measurement setups. Furthermore, the fingerprints are compact enough to be stored on the current chips. More precisely, the standard [64] provides space for such storage in files EF.DG[3-14], which are left for additional biometric and future use; RFID device fingerprints can be stored in those files. Our proposal does not require the storage of a new public key or maintenance of a separate public-key infrastructure, since the integrity of the fingerprints, stored in EF.DG[3-14] will be protected by the existing passive authentication mechanisms implemented in current e-passports.

Security requirements and analysis. For launching signal and feature replay attacks, the attacker has to obtain the fingerprint of the RFID in the original document. In order to extract a fingerprint he

needs to fully control the target document (hold it in possession) for long enough time to complete the extraction. Using the methods from our study, it would be hard, if not infeasible, for the attacker to extract the same fingerprints remotely (e.g., from 1 meter away). In our experiments, such remote feature extraction process resulted in an EER of approximately 50%. We assume that this is due to the change in wireless channel conditions (e.g., antenna orientation, multipath, noise).

After obtaining the original fingerprint, the attacker can try to launch impersonation attacks, producing or finding a device with similar properties. Impersonation attacks by signal replay and feature replay require a generator device that has similar external appearance as the one that is being cloned. In case of e-passports or identity smart cards, replay attacks require introducing a special device. It is not clear if and how this is feasible.

Building (producing) a cloned device is currently considered a hard task because of the complexity. Although manufacturing process variation influences the RFID micro-controller, it is likely that the main source of detectable variation lies in the RFID radio circuitry and antenna. Understanding how each component in the circuitry contributes to the device fingerprint or at least which components contribute the most is still an open research problem.

A more realistic strategy would be to find an RFID chip and antenna configuration that exhibit similar fingerprints to these of the target for cloning document. Such a task also requires knowledge on the feature extraction process. To realize this attack, the attacker needs to test a given quantity of RFID from the same manufacturer and model. The number of devices that have to be tested depends on the system error rates.

8.4 Device Privacy Protection

In the above applications, we presented the implications on defensive uses of physical-layer device identification. We now consider an offensive scenario where physical-layer identification aims at compromising device privacy. In such a scenario, legitimate devices in a wireless network communicate in a way that preserves the user (device) identity from being leaked to active or passive attackers. Such mechanisms typically aim at hiding unique identifier information in packet transmissions and reveal it only to the intended recipients. They have been

studied in the case of anonymous routing [70, 71] in wireless ad-hoc networks, location privacy in wireless local area networks (WLAN) [72] and recently has attracted lots of attention in the RFID community.

The deployment of RFID raises various privacy concerns for users, in particular clandestine tracking and inventorying [1]. Given that RFID is already present in identity documents, payment credit cards [] and is likely to replace barcodes in consumer products, a person carrying several RFID could be vulnerable to clandestine tracking. Given the large reading ranges of certain RFID (e.g., UHF RFID tags), users carrying them can be profiled, identified and tracked using a network of RFID readers without their prior consent. Several solutions that guarantee protection against clandestine tracking have been proposed [73]. They typically exploit RFID identification number pseudonymity by means of cryptographic mechanisms [74–78] at the logical protocol layer.

Our results on RFID physical-layer identification show that RFID leaks distinguishable information at the physical layer independently of any logical layer protocol. In case of UHF RFID tags, intended for consumer products [79], distinguishable information can be extracted independently of the location and distance to the reader up to 6 meters.

Therefore, using our techniques, people can be profiled based on physical-layer properties of a set of tags and then tracked with high accuracy. More precisely, each tag in our tested set of tag models provides approximately 6 bits of distinguishable information which corresponds to uniquely identifying up to $n = 2^6$ tags. As a consequence, a set of k tags can be identified among $C(n+k-1,k) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$ sets. This means that for a set of $k = 5$ tags (same model and manufacturer), there exist $6 \cdot 10^6$ unique combinations which correspond to approximately 22.5 bits of distinguishable information. It should be noted that while these numbers would depend on tag manufacturers, they should be comparable for other models/manufacturers given that all manufacturers should comply with the link frequency tolerances in the standard (Chapter 5).

In summary, our findings demonstrate the feasibility of tracking a set of RFID tags using the physical layer. While further work is required to evaluate certain practical issues (e.g., using low-cost and portable hardware), it is clear that RFID leaks information and privacy-preserving protocols at the logical layer cannot guarantee tag pseudonymity and untraceability. This problem adds to the multiple-tag RFID security and privacy issues [80].

8.5 Summary

We discussed the main results of this thesis in the context of intrusion detection, cloning detection and privacy protection. Physical-layer techniques can provide accurate identification of wireless devices under certain conditions. However, they cannot be directly used in certain real-life scenarios due to either robustness problems that force severe restrictions on the usability or security issues that prevent the safe use of identification in adversarial settings. Examples of such scenarios include access control and intrusion detection in wireless networks.

Document cloning detection scenarios that rely on a controlled setup are suitable for physical-layer identification. In case of RFID-enabled documents/smart cards, physical-layer fingerprints are accurate, robust and stable over longer periods of time.

Finally, physical-layer identification can also be used to compromise device identity privacy despite any logical-layer protection mechanisms. Certain classes of RFID exhibit physical-layer properties that enable device tracking with high accuracy.

Part IV

Related Work

Chapter 9

Device Identification Techniques

Device identification also referred to as device fingerprinting covers a broad spectrum of techniques spanning the physical, data link, transport and application layers. In this chapter, we describe the related work in the area of wireless device identification using hardware-related characteristics. We also compare the existing techniques to our proposed methods. Table 9.1 provides a comparison of selected approaches. Finally, we briefly discuss other device identification approaches.

9.1 Wireless Transceiver Identification

Identification of radio signals gained interest in the early development of radar systems during the World War II [81, 82]. In a number of battlefield scenarios it became critical to distinguish own from enemy radars. This was achieved by visually comparing oscilloscope photos of received signals to previously measured profiles [81]. Such approaches gradually became impractical due to the increasing number of transmitters and more consistency in the manufacturing process.

In mid and late 90s a number of research works appeared in the open literature to detect illegally operated radio VHF FM transmitters. Subsequently, research efforts continued on radio transceivers for wireless personal and local area networks. Depending on the signal part used for identification, fingerprinting techniques can be catego-

rized in transient-based, modulation-based and other approaches. For each category, we discuss the works in a chronological order.

Transient-based approaches use the turn-on/off transient of a radio signal for device identification and could be traced back to the early 90s. These approaches require transient detection and separation before feature extraction and matching. The detection and separation of the turn-on transient depend on the channel noise and device hardware and have been shown to be critical to most systems [83, 84]. Toonstra and Kinsner [85,86] introduced wavelet analysis to characterize the turn-on transients of 7 VHF FM transmitters from 4 different manufacturers. Device fingerprints were composed of wavelet spectra extracted from signal transients captured at the FM discriminator circuit. The extracted fingerprints were classified by means of a genetic algorithm (neural network) without any error. Gaussian noise was added to the original transients in order to simulate typical field conditions. Hippenstiel and Payal [87] also explored wavelet analysis by filter banks in order to characterize the turn-on transients of 4 different VHF FM transmitters. They showed that Euclidean distance could be used to accurately classify extracted device fingerprints to each manufacturer. Choe et al. [88] presented an automated device identification system based on wavelet and multi-resolution analysis of turn-on transient signals and provided results on classification of 3 different transmitters.

Ellis and Serinken [89] studied the properties of turn-on transients exhibited by VHF FM transmitters. They discussed properties of universality, uniqueness, and consistency in 28 VHF FM device profiles characterized by the amplitude and phase of the transients. By visual inspection, the authors showed that there were consistent similarities between device profiles within the same manufacturer and model and device profiles from different models that could not be visually distinguished. Moreover, some devices did not exhibit stable transient profiles during normal operation. The authors suggested that further research is needed to quantify environmental factors (e.g., doppler shift, fading, temperature). Following these recommendations, Tekbas et al. [33, 90] tested 10 VHF FM transmitters under ambient temperature, voltage, and noise level changes. The device fingerprints were composed of transient amplitude and phase features obtained from the signal complex envelope. A probabilistic neural network (PNN) was used for classifying the fingerprints. The experimental results showed that the system had to be trained over a wide temperature and opera-

tional supply-voltage ranges in order to achieve low classification error of 5%. Classification accuracy of low-SNR transients could be improved by estimating the SNR and modifying its level in the training phase [90].

Transient-based approaches were also investigated in modern wireless local and personal area networks (WLAN/WPAN), primarily for intrusion detection and access control. Hall et al. [91–93] focused on Bluetooth and IEEE 802.11 transceivers. The authors captured the transient signals of packet transmissions from close proximity (10 cm) with a spectrum analyzer. They extracted the amplitude, phase, in-phase, quadrature, power, and DWT coefficients and combined them in device fingerprints. Classification results on 30 IEEE 802.11 transceivers composed of different models from 6 different manufacturers [92,94] showed error rates from 0 to 14% depending on the model and manufacturer. The average classification error rate was 8%. The same technique was also applied to a set of 10 Bluetooth transceivers and showed similar classification error rates [93]. The authors also introduced dynamic profiles, i.e., each device fingerprint was updated after some amount of time, in order to compensate internal temperature effects in the considered devices. These works used a limited set of same model and manufacturer devices. The experiments were performed from close proximity. No indication on the feasibility of device identification in practical scenarios was provided.

Ureten et al. [32] proposed extracting the envelope of the instantaneous amplitude of IEEE 802.11 transient signals for device classification. The authors classified signals captured at close proximity from 8 different manufacturers using a probabilistic neural network. The classification error rates fluctuated between 2 and 4% depending on the fingerprint size. One weakness of this approach is that it works only on devices from different models. An attacker could easily compromise such a system by using a device from the same manufacturer.

In the discussed works, signal transients were captured at close proximity to the fingerprinting antenna, approximately 10 to 20 cm. The classification error rates were primarily estimated from a set of different model/manufacturer devices; only a few devices possibly had identical hardware. Physical-layer identification of same-model-same-manufacturer devices was considered by Rasmussen et al. [28]. The authors explored similar characteristics as prior works, namely transient duration, amplitude variance, number of peaks of the carrier signal, difference between normalized mean and maximum value of the transient power, and the first DWT coefficient. Experimental results

on 10 UHF (Mica2/CC1000) sensor devices with identical radio hardware showed a classification error rate of 30% from close proximity. This result clearly demonstrated that more investigation was required to evaluate the feasibility of accurately identifying identical devices.

None of the aforementioned works considered the feature (fingerprint) stability with respect to acquisition distance, antenna polarization and location. They did not study the resilience of physical-layer device identification to attacks.

In Chapter 4 ([59]) we revisited wireless transceiver identification based on turn-on transients in order to address the above issues. In particular, we considered a large set of same-model-same-manufacturer transceivers and stability analysis. We showed that it is possible to identify device with high accuracy using the frequency information within the transient signal. However, frequency information within the transient changes with distance, polarization and location. This and the security analysis of transient-based identification in Chapter 7 exposed for the first time the limitations of transient-based approaches for device identification in a number of security applications where their use has been suggested. These are detailed in Chapter 8.

Modulation-based approaches to device identification focused on extracting unique features from the modulated part of the signal. Brik et al. [53] used five distinctive signal properties of modulated signals, namely the frequency error, SYNC correlation, I/Q origin offset, and magnitude and phase errors as features for physical-layer device identification. The latter were extracted from IEEE 802.11b packet frames, previously captured using a high-end vector signal analyzer. Device fingerprints were built using all five features. Fingerprint classification was performed with k-NN and SVM classifiers specifically tuned for the purpose. The system was tested on 138 identical 802.11b NICs and achieved a classification error rate of 3% and 0.34% for k-NN and SVM classifiers respectively. The signals were acquired at distances from 3 to 15 m. Preliminary experimentation on varying locations of 3 devices showed that the extracted fingerprints were stable to location changes. However, no convincing information on the underlying conditions was provided. Attacks were not considered as well.

Candore et al. [95] further analyzed the accuracy of modulation-based features by exploring classifier combination methods. They authors focused on weighted voting and maximum likelihood strategies. Experimental results on a set of experimental radios showed an average

device detection rate of 88% and false alarms of 12%. The accuracy could be improved with increasing the number of frames used to build the device fingerprint.

Other approaches to physical-layer identification of wireless transceivers considered more regions in the transmitted signals. In particular, a number of works have appeared extracting identification features from non-transient signal parts such as near-transient, packet preamble.

Suski et al. [96] proposed using the baseband power spectrum density of the packet preamble to uniquely identify wireless devices. Device fingerprints were created by measuring the power spectrum density of the preamble of an IEEE 802.11a (OFDM) packet transmission. Fingerprints comparison was performed by spectral correlation. The authors evaluated the accuracy of their approach on 3 devices and achieved an average classification error rate of 20% for packet frames with SNR greater than 6 dB. Klein et al. [97] further explored IEEE 802.11a (OFDM) device identification by applying complex wavelet transformations and multiple discriminant analysis (MDA). The classification performance of their technique was evaluated on 4 same model Cisco wireless transceivers. The experimental results showed improvements in terms of SNR of approximately 8 dB for a classification error rate of 20%. Varying SNR and burst detection error were also considered in this work.

Reising et al. [98] used the near-transient and midamble regions of GSM-GMSK burst signals to classify 4 mobile phones from four different manufacturers. The authors observed that that the classification error using the midamble is significantly higher than using transients. Various factors were identified as potential areas of future work on identification of GMSK signals. In a subsequent study [99], the same authors demonstrated that the near-transient RF fingerprinting is suitable for GSM signal fingerprinting consistent with their prior work on 802.11a (OFDM) benchmark accuracy [97].

Jana et al. [39] proposed an identification technique based on clock skews in order to protect against unauthorized access points (APs) in a wireless local area network. The AP fingerprint consisted of its clock skew as measured by the client station. This technique has been previously shown to be effective in wired networks [6]. The authors showed that they could distinguish between different APs and therefore detect an intruder AP with high accuracy. The possibility to compute the clock skew relied on time-stamps available during AP association.

We point out that most of the above discussed works considered standard classifiers and classification error rate as a performance metric. While such a metric is appropriate for applications with well-known type and number of classes (e.g., [100]), it is not suitable for applications such as intrusion detection, device authentication, wormhole detection due to: 1) In intrusion-related applications, the number of classes (i.e., devices) is unlimited. 2) Standard classifiers will classify test signals coming from a device that does not belong to the classes of devices to one of these classes.

9.2 Passive Transponder Identification

Passive Radio Frequency Identification (RFID) are being incorporated in contactless identity cards [36], electronic passports [64], payment credit cards [65] and products in supply chain systems [101]. Due to the security and privacy threats related to RFID deployment, a number of works have addressed RFID authentication, key management and privacy-preserving deployment, among others [1, 102–105]. Although the literature contains many investigations of RFID security and privacy on the logical level, the security implications of the RFID physical communication layer have remained largely unexplored.

Passive RFID device identification using the physical layer has been recently considered [106–110]. Periaswamy et al. [106] addressed fingerprinting of UHF RFID tags. The authors proposed a method to enable ownership transfer of UHF RFID tags using the minimum power response of tags as a physical-layer fingerprint. The authors used a small set of 8 tags from 2 models and showed visual evidence that UHF tags can be distinguished. In a more detailed study [109], the authors evaluated the ability of the minimum power response to uniquely identify large sets of UHF tags. An experimental evaluation on two manufacturers (50 tags per manufacturer) demonstrated an average identification success rate of 94.4% (with false accept rate of 0.1%) and 90.7% (with false accept rate of 0.2%). While these results are particular to the physical behavior of UHF RFID backscatter communication, they confirm the main findings in Chapters 5 and 6 ([111,112]), namely the ability to accurately identify same-model-same-manufacturer RFID devices. Even though, the authors did not consider any feature stability experiments, the proposed technique is unlikely to be able to remotely identify tags from any distance and location due to: (i) the minimum

power response provides UHF tag's energy-harvesting information and it is indicated at a specified frequency and distance [113]. This implies that it varies with the distance. (ii) Experiments in reflective environments have demonstrated significant variations in the tag minimum power response [114].

The closest work to our UHF RFID in Chapter 5 ([34]) is [110]. In an independent investigation, the authors experimentally showed that UHF tag responses exhibit stable and consistent duration within same-model-same-manufacturer tags. This is very similar to our time interval error (TIE) feature for distinguishing UHF RFID tags. Experimental results on 30 tags from 3 major manufacturers (10 tags per manufacturer) showed a classification error of 2%, 4% and 60% for the respective manufacturers. A limitation of that work compared to ours is that the authors did not attempt to evaluate their findings on larger sets, neither they tried to theoretically explain and compute the entropy of the tag response duration as a device identification feature. Nevertheless, we can consider these results as an important confirmation that distinguishing UHF tags using timing characteristics is possible for other major manufacturers then the ones considered in this thesis.

Romero et al. [107] investigated electromagnetic characteristics and showed that different models of HF RFID cards were identifiable. Their method consisted of observing certain frequencies in the transient and frequency response of the device. The authors showed visually that the fingerprints of devices from 4 different models form clearly separable clusters in the feature space.

Following our work [111], Romero et al. improved their previously proposed technique. Their study [108] demonstrated that precise measurements of the unloaded resonance frequency and quality factor of HF RFID cards also allow identification of different card models as well as identification of individual cards of the same model. The results from combining resonance information together with measurements of the energy at the carrier harmonics during transmission enabled identification with an error rate as low as 4%. These rates were obtained on a set of 4 models with 10 devices per model.

The aforementioned works on HF RFID device identification are the closest to our Chapters 5 and 6 ([111, 112]). However, they did not consider neither feature stability analysis, nor the transferability of the device fingerprints between acquisition setups. Moreover, our optimized identification features achieved lower error rates (an Equal Error Rate of 0.5%).

RFID uniqueness for cloning protection can also be achieved with physical unclonable functions (PUFs) [115–117], certificates of authenticity [118, 119] and watermarking schemes [120]. PUF-enabled RFIDs contain a special circuit that maps input challenges to output responses using a function (PUF) determined by the inherent variations of that circuit. The difficulty of controlling these variations prevents an adversary from duplicating the PUF-enabled chips given some assumptions on its capabilities. The main limitation of PUF-based identification is that it requires PUF-enabled devices. However, it presents the advantage of relying on "controlled" variability as opposed to unintentionally introduced manufacturing variability that physical-layer device identification exploits. Recently, it has been shown that several PUF-constructions can be broken by numerical modeling attacks [121].

Lakafosis et al. [119, 122] realized RF physical objects to be included in RFID tags as certificates of authenticity (CoA) in order protect against counterfeiting [118]. The physical object consisted of a structure of copper wires which exhibits unique RF effects in the high-frequency range 5 - 6 GHz. The authors also implemented a purpose-built reader to extract the objects and provided performance and security evaluation. This type of designs is similar to watermarking schemes which aim at providing unforgeable properties in the device hardware [120, 123, 124]. One limitation is that watermarks typically require specialized procedures to be verified.

For completeness, we should also mention that identification of HF RFID-enabled identity documents has also been attempted on the logical layer. Richter et al. [125] reported on the possibility of detecting the country that issued a given passport by looking at the bytes that an e-passport RFID sends as a reply in response to some carefully chosen commands from the reader. This technique enabled classification of RFID chips used in electronic passports. Our technique differs from that proposal as it enables not only classification, but also identification of individual passports. Moreover, the technique proposed in [125] cannot be used for cloning detection since the attacker can modify the responses of a tag on the logical level.

9.3 Other Identification Approaches

This thesis considered identification of wireless devices based on imperfections in their analog circuitry that can be measured during radio

Table 9.1: Summary of selected physical-layer identification techniques

Approach	Signal	Features	Evaluation Data			Evaluated Factors	Method[1]	Error rate
			Device Type	#	Origin[1]			
Toonstra et al. [85]	Transient	Wavelets	Analog VHF	7	D1	-	C	0%
Serinken et al. [89]	Transient	Amplitude, phase	Analog VHF	28	D1	fixed distance	VI	n/a
Tekbas et al. [33]	Transient	Amplitude, phase	Analog VHF	10	D1	wide temp. range, voltage and SNR	C	5%
Hall et al. [91]	Transient	Amplitude, phase, DWT	IEEE 802.11	14	D2	close proximity and temp.	C	8%
Hall et al. [93]	Transient	Amplitude, phase, DWT	Bluetooth	10	D2	close proximity	C	7%
Ureten et al. [32]	Transient	Amplitude envelope	IEEE 802.11	8	D2	close proximity	C	2%
Rasmussen et al. [28]	Transient	Duration, amplitude, DWT	UHF sensor	10	D3	close proximity	C	30%
Brik et al. [53]	Data	Modulation errors	IEEE 802.11	138	D3	varied distance and location	C	0.34%
Jana et al. [39]	Data	Clock skew	IEEE 802.11	5	D1	virtual AP, temp. and NTP sync.	C	0%
Suski et al. [96]	Preamble	Spectrum	IEEE 802.11	3	D3	close proximity, SNR	C	13%
Romero et al. [108]	Data	Resonance	HF RFID	40	D1	close proximity	ID	4%
Periaswamy et al. [110]	Data	Duration	UHF RFID	30	D1	close proximity	C	2-4%

[1] D1: Devices from different manufacturers and some of the same model; D2: Devices from different manufacturers and models; D3: Devices from the same manufacturer and model (identical)
[2] C: Classification; VI: Visual Inspection; ID: Identification (identity verification)

communication. Physical fingerprints for device identification can also be extracted from the internals of the device circuitry. Examples include measuring the MOSFET threshold voltages [126], threshold voltage mismatch in NOR cells [127], and the power-up of the SRAM [128] for RFID identification. An advantage of these techniques is that they can possibly be applied to any hardware. The drawback, however, is the requirement of special access to the device circuitry as opposed to the techniques developed in this work.

We note that other physical properties of the wireless communication could be used for security applications such as access control and location distinction. In particular, a number of works have explored the physical properties of the wireless channel for device authentication [129, 130] and device location distinction [131]. While the channel characteristics are believed to be unique within a given location due to specific multipath effects, these characteristics are not inherent to the device and therefore cannot be used for device identification unless the device is bound to its location.

We conclude this section by mentioning that the present work covers the physical-layer subset of techniques on device identification. In general, device identification spans all communication layers of the OSI architecture [132] and a variety of network devices. Device identification (fingerprinting) has been explored on both wired and wireless devices at the link, transport, and application layers, among many others [6, 125, 133–139].

Chapter 10

Security Related Work

In this chapter, we focus on security aspects of device identification. In particular, we include the literature on security analysis of device identification and relay attacks.

10.1 Attacks on Device Identification

The large majority of works have focused on exploring feature extraction and matching techniques for physical-layer device identification. In parallel to our investigation of attacks, Edman and Yener [140] developed impersonation attacks on modulation-based identification [53]. They showed that low-cost software-defined radios [13] could be used to reproduce modulation features and impersonate a targeted device with a success rate of 50–75%. To the best of our knowledge this was the only related work considering impersonation attacks on existent physical-layer device identification techniques.

Our contributions in Chapter 7 ([59, 141]) differ in a number of aspects. First, we developed a broader range of impersonation attacks, namely feature-replay, signal-replay and hill-climbing attacks. We tested these attacks on both modulation and transient-based identification using both software-defined radios (SDR) and high-end arbitrary waveform generators. The authors in [140] studied only impersonation of modulation-based identification using SDR. Second, the work in [140] achieved lower impersonation rates of 55-75%. This was likely due to the following differences in our feature- and signal-replay attack

semantics, measurement setup and fingerprint extraction: (i) We used 8 GHz oscilloscope to measure the signal imperfections with high precision. (ii) Our modulation feature extraction and matching followed [53], while in [140] some of the most discriminative features (F2) were not computed. This modified the original modulation identification design. (iii) We used a high-end arbitrary waveform generator for signal replay at RF as opposed to an SDR in [140]. Finally, we evaluated the effectiveness of our proposed attacks using both threshold-based identification and two classification methods (k-NN and SVM); [140] considered only SVM classification.

10.2 Relay attacks

Relay attacks are well-known attacks in wireless systems. They have been demonstrated in other scenarios, e.g., in [142] as mafia-fraud attacks, in [10] as wormhole attacks. Similarly, the relationship between secure communication and physical neighborhood notions has been previously studied in [143–145].

The closest work to our investigation on car key systems can be found in [146, 147]. The authors perform security analysis of keyless entry systems including relay attacks. While the performed analysis identifies the relay problem, the authors did not provide neither hardware design, nor practical implementation of the attack. Finally, no adequate countermeasures were proposed. Some practical attacks on PKES systems have been recently reported [148]. However, no detailed information was available to verify wether the attack was real.

In terms of PKES system details, major electronic parts suppliers provide components for their realization [149–152], those components are then used by various car manufacturers. Although variations exist in the protocols and cryptographic blocks (Keeloq in [152], TI DST in [150], AES in [149]), all manufacturers provide systems based on the same combined LF/UHF radio technology. We also note that certain remote key entry systems have been shown to have weaknesses caused by short keys and weak encryption algorithms [153–156]. Our attack is independent of cryptography mechanisms, so solving such issues will not provide protection against physical-layer relay attacks.

Part V

Closing Remarks

Chapter 11

Conclusion

The successful application of wireless technologies to new areas depends on the deployment of appropriate security and privacy measures. One critical issue related to security and privacy is device identification. It presents both defensive and offensive perspective. As a defensive mechanism (e.g., authenticating devices), it must function in the face of device identity spoofing, device cloning, cryptographic key compromise, relay attacks. As an offensive mechanism (e.g., breaking user's privacy), it tries to gain information about user's identity and location without his prior consent. In this thesis, we investigated device identification from both perspectives using the physical layer.

In the introductory part, we demonstrated that certain automobile entry and start systems are vulnerable to relay attacks. The secure car/key authentication was not sufficient to prevent unauthorized vehicle access and drive. One of the fundamental problems was that the system had no means to recognize that the car credentials were relayed to the key by another device. This problem shows that it is important to identify devices based on inherent physical characteristics in order to prevent relay and other identity-based attacks.

Secondly, we investigated the feasibility and related assumptions of identifying wireless devices using the physical-layer. We looked at radio hardware imperfections introduced during the manufacturing process, which appear in radio transmissions. We considered active wireless transceivers and passive RFID transponders. For wireless transceivers, we showed that the transient part of a packet transmission provided unique characteristics for identification of same-model-same-

manufacturer transceivers with high accuracy. We clarified the underlying conditions allowing for such high accuracy. Our findings suggest that the transient properties are modified when the device changes distance and/or location, and therefore cannot be directly used for identification in mobile scenarios.

In the case of passive RFID devices, we explored timing, modulation and spectral features for device identification using in- and out-of-specification reader requests. Spectral features enabled device identification with low error rates in a controlled setup. Device fingerprints based on such features proved to be stable over time. They could also be extracted on one acquisition setup and verified on another if channel equalization was applied. Our results strongly support the use of physical-layer RFID identification in document cloning detection (e.g., e-passports), where the presented document is measured, its fingerprint is extracted and then compared to a previously enrolled fingerprint of the legitimate document.

Timing and modulation features showed effectiveness in distinguishing certain classes of RFID. In particular, time interval errors can be used to distinguish UHF RFID tags of the same model and manufacturer irrespective of the tag location to the reader. If a user carries a number of these tags, his location can be followed by a network of readers. This type of tracking cannot be easily prevented unless new cross-layer (physical and logical) mechanisms are devised.

Finally, we evaluated the resilience of selected physical-layer device identification techniques to impersonation attacks. We designed and implemented attacks by feature replay, signal replay and hill-climbing strategies. Our evaluation showed that physical-layer identification was vulnerable to impersonation. We clarified the underlying assumptions. Our findings suggest that these techniques cannot be safely used in a number of security scenarios where their use has been suggested. A prominent example are intrusion detection scenarios where unauthorized access is detected by device identification.

Chapter 12

Future Work

This thesis made a further step into understanding physical-layer device identification, its assumptions and implications on the security and privacy of wireless devices and networks. However, there are several research areas that remain to be investigated as follows.

Sources of identification

The exact components that make devices uniquely identifiable remain unclear. While their exact determination is a difficult task as it requires low-level analog circuit modeling, simulations and measurements, such analysis can provide (i) means to improve feature extraction for optimal accuracy (ii) means to perform a detailed security analysis (iii) insights on whether more precise manufacturing is able to prevent physical-layer device identification. We note that our feature extraction and matching methods are based on observations of the device behavior and experimentation. Understanding the source of variability in wireless devices would allow modeling the device circuitry and possibly draw theoretical bounds about the accuracy, entropy and stability of physical-layer fingerprints.

While the above proposed analysis may look difficult in the case of wireless transceivers given their analog complexity, passive RFID devices present relatively simple circuits that are a good starting point for investigation.

We note that more precise manufacturing and quality control may

minimize the hardware imperfections. It remains an open research problem whether this is practical and if all device imperfections could technically be removed.

Remote identification

The feasibility or non-feasibility of accurately identifying wireless devices remotely at the physical-layer under different channel conditions (e.g., due to mobility or dynamic environment) is an open challenge. Although this thesis demonstrated that certain classes of devices could be distinguished remotely, the possibility to do the same on wireless devices in general and with high accuracy remains an open issues. Future insights on that matter are likely to have important implications on the design of location and identity privacy-preserving protocols.

A possible starting point in that direction would be to look at precise wireless channel estimation and compensation procedures. Such procedures could possibly preserve the inherent device characteristics from the random effects of the channel. Under which type of wireless channels this may be possible is an interesting area of research.

Security aspects

In the security context, the difficulty of impersonating or building or finding wireless devices that would exhibit similar physical-layer identities needs to be better understood and quantified. While we have demonstrated that high-end arbitrary waveform generators have sufficient capabilities to accurately reproduce physical-layer signals, their use is unrealistic in several scenarios (e.g., device cloning detection). Protection again impersonation and replay attacks could be further studied by looking at data-dependent physical-layer characteristics. Such characteristics, if they exist and are unique, would enable physical-layer device identification that vary with the data sent by the device. This proposed direction goes towards the concept of physical unclonable functions (PUFs). In PUF solutions, embedded hardware-based functions produce outputs that depend on the data input.

Publications

The work presented in this thesis is based on the following publications co-authored during my doctoral studies at ETH Zurich.

1. B. Danev, S. Capkun, R. Jayaram Masti, T. S. Heydt-Benjamin, Towards Practical Identification of HF RFID Devices, Under submission

2. B. Danev, D. Zanetti, S. Capkun, On Physical-layer Identification of Wireless Devices, ACM Computing Surveys, 2011

3. A. Francillon, B. Danev, S. Capkun, Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars, Proc. 18th Network and Distributed System Security Symposium (NDSS), 2011

4. D. Zanetti, B. Danev, S. Capkun, Physical-layer Identification of UHF RFID Tags, Proc. 16th ACM Annual International Conference on Mobile Computing and Networking (MOBICOM), 2010

5. B. Danev, H. Luecken, S. Capkun and K. Defrawy, Attacks on Physical-layer Identification, Proc. 3rd Conference on Wireless Network Security (WISEC), 2010

6. B. Danev, T. S. Heydt-Benjamin, S. Capkun, Physical-layer Identification of RFID Devices, Proc. 18th USENIX Security Symposium, 2009

7. B. Danev, S. Capkun, Transient-based Identification of Wireless Sensor Nodes, Proc. 8th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2009

The following list contains the other publications co-authored during my studies.

8. B. Danev, R. Jayaram Masti, G. Karame and S. Capkun, Towards Secure VM-vTPM Migration Protocols, Under submission

9. G. Karame, B. Danev, C. Banwart and S. Capkun, On the Security of Network Measurements based on Packet-Pair Dispersions, Under submission

10. C. Poepper, N. Tippenhauer, B. Danev, S. Capkun, Investigation of Signal and Message Manipulations on the Wireless Channel, Proc. 16th European Symposium on Research in Computer Security (ESORICS), 2011

11. M. Strasser, B. Danev, S. Capkun, Detection of Reactive Jammers, ACM Transactions on Sensor Networks (TOSN), 2010

Bibliography

[1] A. Juels, "RFID security and privacy: A research survey," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, 2006.

[2] J. Bellardo and S. Savage, "802.11 denial-of-service attacks: real vulnerabilities and practical solutions," in *Proceedings of the 12th conference on USENIX Security Symposium - Volume 12*. Berkeley, CA, USA: USENIX Association, 2003, pp. 2–2.

[3] C. Boyd and A. Mathuria, *Protocols for Authentication and Key Establishment*. Springer, 2003.

[4] I. Goldberg and M. Briceno, *GSM Cloning*, 1998, http://www.isaac.cs.berkeley.edu/isaac/gsm.html.

[5] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, *Guide to Biometrics*. Springer, 2003.

[6] T. Kohno, A. Broido, and K. Claffy, "Remote physical device fingerprinting," *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 2, 2005.

[7] T. Waraksa, K. Fraley, R. Kiefer, D. Douglas, and L. Gilbert, "Passive keyless entry system," US patent 4942393, 1990.

[8] G. P. Hancke, K. Mayes, and K. Markantonakis, "Confidence in smart token proximity: Relay attacks revisited," *Computers & Security*, vol. 28, no. 7, pp. 615–627, 2009.

[9] S. Drimer and S. J. Murdoch, "Keep your enemies close: distance bounding against smartcard relay attacks," in *Proceedings of 16th*

USENIX Security Symposium. Berkeley, CA, USA: USENIX Association, 2007.

[10] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Wormhole attacks in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 370–380, 2006.

[11] G. Hancke, "Practical attacks on proximity identification systems (short paper)," in *Proc. of the 27th IEEE Symposium on Security and Privacy*, 2006.

[12] P. Dodd, *The low frequency experimenter's handbook*. Herts : Radio Society of Great Britain, 2000, iSBN : 1-872309-65-8.

[13] M. Ettus, "Universal software radio peripheral (USRP)," Ettus Research LLC, http://www.ettus.com/.

[14] K. Tan, J. Zhang, J. Fang, H. Liu, Y. Ye, S. Wang, Y. Zhang, H. Wu, W. Wang, and G. M. Voelker, "Sora: high performance software radio using general purpose multi-core processors," in *Proc. of the 6th USENIX symposium on Networked Systems Design and Implementation (NSDI)*. Berkeley, USA: USENIX Association, 2009, pp. 75–90.

[15] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile." in *Proc. of the 31st IEEE Symposium on Security and Privacy*, May 2010.

[16] A. Francillon, B. Danev, and S. Capkun, "Relay attacks on passive keyless entry and start systems in modern cars," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2011.

[17] S. Brands and D. Chaum, "Distance-bounding protocols," in *EUROCRYPT '93*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1994, pp. 344–359.

[18] G. P. Hancke and M. G. Kuhn, "An RFID distance bounding protocol," in *Proc. International ICST Conference on Security and Privacy in Communications Networks (SecureComm)*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 67–73.

[19] J. Munilla, A. Ortiz, and A. Peinado, "Distance bounding protocols with void-challenges for RFID," Proc. Workshop on RFID Security (RFIDSec), Ecrypt, Graz, Austria, July 2006.

[20] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *WiSe '03: Proceedings of the 2nd ACM workshop on Wireless security.* New York, NY, USA: ACM, 2003.

[21] N. O. Tippenhauer and S. Capkun, "ID-based secure distance bounding and localization," in *In Proceedings of European Symposium on Research in Computer Security (ESORICS)*, 2009.

[22] M. Kuhn, H. Luecken, and N. O. Tippenhauer, "UWB impulse radio based distance bounding," in *Proc. of the Workshop on Positioning, Navigation and Communication (WPNC)*, 2010.

[23] K. B. Rasmussen and S. Capkun, "Realization of RF distance bounding," in *Proc. of the 19th USENIX Security Symposium*, 2010.

[24] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[25] "Fingeprint verification competitions (FVC)." [Online]. Available: http://bias.csr.unibo.it/fvc2006/

[26] C. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.

[27] *IEEE Standard 802.15.4-2006: Wireless MAC and PHY Specifications for Low-Rate WPANs*, IEEE Standards Association, 2006.

[28] K. Rassmussen and S. Capkun, "Implications of radio fingerprinting on the security of sensor networks," in *Proc. International ICST Conference on Security and Privacy in Communication Networks (SecureComm)*, 2007.

[29] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1996.

[30] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Proc. Conference on Automatic Face and Gesture Recognition*, 1998, pp. 336–341.

[31] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, no. 2, pp. 228–233, 2001.

[32] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Canadian Journal of Electrical and Computer Engineering*, vol. 32, no. 1, Winter 2007.

[33] O. Tekbas, N. Serinken, and O. Ureten, "An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions," *Canadian Journal of Electrical and Computer Engineering*, vol. 29, no. 3, 2004.

[34] D. Zanetti, B. Danev, and S. Capkun, "Physical-layer identification of UHF RFID tags," in *Proc. of the 16th ACM Conference on Mobile Computing and Networking (MOBICOM)*. ACM, 2010.

[35] *JCOP - The IBM GlobalPlatform JavaCard implementation*, IBM, 2002, ftp://ftp.software.ibm.com/software/pervasive/info/JCOP_Family.pdf.

[36] *ISO/IEC 14443 Standard*, ISO/IEC, 2008.

[37] J. van Beek, "ePassports reloaded," in *Black Hat Briefings USA*, 2008.

[38] EPCglobal, "UHF Class 1 Gen 2 Standard v. 1.2.0," Standard, 2008.

[39] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," in *Proc. ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2008, pp. 104–115.

[40] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*, 2nd ed. Prentice-Hall Signal Processing Series, 1998.

[41] S. Marple, "Computing the discrete-time analytic signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, 1999.

[42] B. Manly, *Multivariate Statistical Methods: A Primer*. Chapman & Hall, 2004.

[43] A. Ross and A. Jain, "Multimodal biometrics: An overview," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2004.

[44] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, 1998.

[45] A. Jain, S. Prabhakar, and S. Chen, "Combining multiple matchers for a high security fingerprint verification system," in *Pattern Recognition Letters*, 1999.

[46] S. Guiasu and A. Shenitzer, "The principle of maximum entropy," *The Mathematical Intelligencer*, vol. 7, pp. 42–48, 1985.

[47] *Function/Arbitrary Waveform Generator 33250A*, Agilent, 2007, http://www.home.agilent.com/agilent.

[48] *Agilent InfiniiVision 6104A*, Agilent, 2007, http://www.home.agilent.com/.

[49] J. Schaefer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 32, 2005.

[50] J. Schäfer, R. Opgen-Rhein, and K. Strimmer, "Efficient estimation of covariance and (partial) correlation," The Comprehensive R Archive Network, 2010, http://strimmerlab.org/software/corpcor/.

[51] B. Sklar, *Digital communications: fundamentals and applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2001.

[52] *GPS Timing and Frequency Standards*, Quartzlock, 2010, http://www.quartzlock.com/downloads/datasheets/E8-Y_4pp.pdf.

[53] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2008.

[54] *Digital Signal Analyzer (DSA) 90804A*, Agilent, 2008, http://www.home.agilent.com/.

[55] *Arbitrary Waveform Generator 7000 Series*, Tektronix, 2008, http://www.tek.com/products/signal_sources/awg7000/.

[56] E. Blossom, "GNU software radio," GNU Software Radio, http://www.gnu.org/software/gnuradio/.

[57] *IEEE Standard 802.11b-1999: Wireless LAN MAC and PHY Specifications*, IEEE Standards Association, 1999.

[58] T. Schmidl and D. Cox, "Robust frequency and timing synchronization for ofdm," *IEEE Transactions on Communications*, vol. 45, no. 12, 1997.

[59] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *Proc. ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN)*, 2009.

[60] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[61] D. Dolev and A. C. Yao, "On the security of public key protocols," *IEEE Transactions on Information Theory*, vol. 2, no. 29, pp. 198–208, 1983.

[62] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," in *Proc. IEEE Workshop on Sensor Network Protocols and Applications*, 2003.

[63] Y. Hu, A. Perrig, and D. Johnson, "Packet leashes: A defense against wormhole attacks in wireless networks," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, 2003.

[64] ICAO, *Machine Readable Travel Documents (ICAO Document 9303)*, 2006, http://www.icao.int/.

[65] MasterCard, *MasterCard PayPass M/Chip Application Note*, 2009, http://www.paypass.com/documentation.html.

[66] S. Boggan, "Cracked it!" 2006, http://www.guardian.co.uk/technology/2006/nov/17/news.homeaffairs/.

[67] K. Zetter, "Hackers clone e-passports," Online publication, 2006, last access: 21.11.2009, http://www.wired.com/science/discoveries/news/2006/08/71521/.

[68] L. Grunwald, "New attack to RFID-systems and their middleware and backends," in *Black Hat Briefings USA*, 2006.

[69] M. Witteman, "Attacks on digital passports," in *What The Hack*, 2005.

[70] J. Kong and X. Hong, "ANODR: anonymous on demand routing with untraceable routes for mobile ad-hoc networks," in *Proc. ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc)*, 2003, pp. 291–302.

[71] B. Zhu, Z. Wan, M. Kankanhalli, F. Bao, and R. Deng, "Anonymous secure routing in mobile ad-hoc networks," in *Proc. IEEE International Conference on Local Computer Networks*, 2004, pp. 102–108.

[72] M. Gruteser and D. Grunwald, "Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis," *Mobile Networks and Applications*, vol. 10, no. 3, pp. 315–325, 2005.

[73] S. Spiekermann and S. Evdokimov, "Privacy enhancing technologies for RFID - A critical investigation of state of the art research," in *Proc. IEEE Privacy and Security*, 2009.

[74] T. Dimitriou, "A lightweight RFID protocol to protect against traceability and cloning attacks," in *Proc. International ICST Conference on Security and Privacy in Communication Networks (SecureComm)*, 2005.

[75] C. Berbain, O. Billet, J. Etrog, and H. Gilbert, "An efficient forward private RFID protocol," in *Proc. ACM Conference on Computer and Communications Security*, 2009, pp. 43–53.

[76] D. N. Duc, J. Park, H. Lee, and K. Kim, "Enhancing security of EPCglobal Gen-2 RFID tag against traceability and cloning," in *Proc. Symposium on Cryptography and Information Security (SCIS)*, 2006.

[77] Y. K. Lee, L. Batina, D. Singelée, and I. Verbauwhede, "Low-cost untraceable authentication protocols for RFID," in *Proc. ACM Conference on Wireless Network Security*, 2010.

[78] E.-O. Blass, A. Kurmus, R. Molva, G. Noubir, and A. Shikfa, "The Ff-family of protocols for RFID-privacy and authentication," *"IEEE Transactions on Dependable and Secure Computing"*, August 2010.

[79] EPCglobal, "Architecture framework v. 1.2. standard," 2007, http://www.epcglobalinc.org/standards/architecture/architecture_1_2-framework-20070910.pdf.

[80] E.-O. Blass and R. Molva, "New directions in RFID security," in *IFIP "Advances in Information and Communication Technology", Vol. 309*, 2009.

[81] D. Margerum, *Pinpointing Location Of Hostile Radars.* Microwaves, 1969.

[82] R. Jones, *Most Secret War: British Scientific Intelligence 1939-1945.* Hamish Hamilton, 1978.

[83] D. Shaw and W. Kinsner, "Multifractal modeling of radio transmitter transients for classification," in *Proc. IEEE Conference on Communications, Power and Computing*, 1997.

[84] O. Ureten and N. Serinken, "Detection of radio transmitter turn-on transients," in *Electronic Letters*, vol. 35, 2007, pp. 1996–1997.

[85] J. Toonstra and W. Kinsner, "Transient analysis and genetic algorithms for classification," in *Proc. IEEE Conference on Communications, Power, and Computing (WESCANEX)*, 1995.

[86] ——, "A radio transmitter fingerprinting system ODO-1," in *Proc. Canadian Conference on Electrical and Computer Engineering*, 1996.

[87] R. Hippenstiel and Y. Payal, "Wavelet based transmitter identification," in *Proc. International Symposium on Signal Processing and Its Applications (ISSPA)*, 1996.

[88] H. Choe, C. Poole, A. Yu, and H. Szu, "Novel identification of intercepted signals for unknown radio transmitters," in *Proc. SPIE*, vol. 2491, 1995, pp. 504–517.

[89] K. Ellis and N. Serinken, "Characteristics of radio transmitter fingerprints," *Radio Science*, vol. 36, pp. 585–597, 2001.

[90] O. Tekbas, O. Ureten, and N. Serinken, "Improvement of transmitter identification system for low SNR transients," in *Electronic Letters*, vol. 40, 2004.

[91] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *Proc. Communications, Internet, and Information Technology (CIIT)*, 2004.

[92] ——, "Radio frequency fingerprinting for intrusion detection in wireless networks," *Submission to IEEE TDSC (Electronic Manuscript)*, 2005.

[93] ——, "Detecting rogue devices in Bluetooth networks using radio frequency fingerprinting," in *Proc. IASTED Conference on Communications and Computer Networks (CCN)*, 2006.

[94] J. Hall, "Detection of rogue devices in wireless networks," Ph.D. dissertation, Carleton University, 2006.

[95] A. Candore, O. Kocabas, and F. Koushanfar, "Robust stable radiometric fingerprinting for wireless devices," *IEEE Intl. Workshop on Hardware-Oriented Security and Trust*, pp. 43–49, 2009.

[96] W. Suski, M. Temple, M. Mendenhall, and R. Mills, "Using spectral fingerprints to improve wireless network security," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2008.

[97] R. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance network security," *Jour. of Communications and Networks, Special Issue: Secure Wireless Networking*, vol. 11, no. 6, pp. 544–555, Dec 2009.

[98] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improved wireless security for GMSK-based devices using RF fingerprints," *Intl. Jour. of Electronic Security and Digital Forensics (IJESDF)*, vol. 3, no. 1, pp. 41–59, Apr 2010.

[99] ——, "Improving intra-cellular security using air monitoring with RF fingerprints," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, 2010, pp. 1–6.

[100] B. Wang, S. Omatu, and T. Abe, "Identification of the defective transmission devices using the wavelet transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 696–710, 2005.

[101] *The EPCglobal Architecture Framework v. 1.3*, EPCglobal, 2009.

[102] I. Vajda and L. Buttyán, "Lightweight authentication protocols for low-cost RFID tags," in *Proc. Workshop on Security in Ubiquitous Computing*, 2003.

[103] G. Avoine and P. Oechslin, "RFID traceability: A multilayer problem," in *Proc. Financial Cryptography*, ser. LNCS, vol. 3570, 2005, pp. 125–140.

[104] A. Juels, R. Pappu, and B. Parno, "Unidirectional key distribution across time and space with applications to RFID security," in *Proc. USENIX Security Symposium*, 2008, pp. 75–90.

[105] G. Avoine, "RFID security and privacy lounge," http://www.avoine.net/rfid/index.html.

[106] S. C. G. Periaswamy, D. Thompson, and J. Di, "Ownership transfer of RFID tags based on electronic fingerprint," in *Proc. International Conference on Security and Management*, 2008.

[107] H. P. Romero, K. A. Remley, D. F. Williams, and C.-M. Wang, "Electromagnetic measurements for counterfeit detection of radio frequency identification cards," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 5, pp. 1383–1387, 2009.

[108] H. P. Romero, K. A. Remley, D. F. Williams, C.-M. Wang, and T. X. Brown, "Identifying RF identification cards from measurements of resonance and carrier harmonics," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 7, 2010.

[109] S. Chinnappa Gounder Periaswamy, D. Thompson, and J. Di, "Fingerprinting RFID tags," *Dependable and Secure Computing, IEEE Transactions on*, vol. PP, 2010.

[110] S. C. G. Periaswamy, D. R. Thompson, and H. P. Romero, "Fingerprinting radio frequency identification tags using timing characteristics," in *Proc. Workshop on RFID Security (RFIDSec, Asia)*, 2010.

[111] B. Danev, T. S. Heydt-Benjamin, and S. Capkun, "Physical-layer identification of RFID devices," in *Proc. USENIX Security Symposium*, 2009.

[112] B. Danev, S. Capkun, R. Jayaram Masti, and T. S. Heydt-Benjamin, "Towards practical identification of HF RFID devices," *submitted to ACM Transactions on Information and System Security (TISSEC)*, 2011.

[113] L. Sydanheimol, J. Nummela, L. Ukkonen, J. McVay, A. Hoorfar, and M. Kivikoski, "Characterization of passive UHF RFID tag performance," *IEEE Antennas and Propagation Magazine*, vol. 50, no. 3, pp. 207–212, 2008.

[114] M. Nikkari, T. Bjorninen, L. Sydanheimo, L. Ukkonen, A. Elsherbeni, F. Yang, and M. Kivikoski, "Performance of a passive UHF RFID tag in reflective environment," in *Proc. IEEE Antennas and Propagation Society International Symposium*, 2008.

[115] B. Gassend, D. Lim, D. Clarke, S. Devadas, and M. van Dijk, "Identification and authentication of integrated circuits," *Concurrency and Computation: Practice and Experience*, vol. 16, no. 11, pp. 1077–1098, Sep 2004.

[116] P. Tuyls and L. Batina, "Rfid-tags for anti-counterfeiting," in *Topics in Cryptology - CT-RSA 2006, Vol. 3860 of LNCS*, 2006, pp. 115–131.

[117] S. Devadas, E. Suh, S. Paral, R. Sowell, T. Ziola, and V. Khandelwal, "Design and implementation of PUF-based "unclonable" RFID ICs for anti-counterfeiting and security applications," *Proc. IEEE International Conference on RFID*, pp. 58–64, 2008.

[118] G. Dejean and D. Kirovski, "Rf-dna: Radio-frequency certificates of authenticity," in *Proc. 9th international workshop on Cryptographic Hardware and Embedded Systems (CHES)*, 2007, pp. 346–363.

[119] V. Lakafosis, A. Traille, H. Lee, E. Gebara, M. Tentzeris, G. DeJean, and D. Kirovski, "RF fingerprinting physical objects for anticounterfeiting applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 59, no. 2, pp. 504–514, Feb 2011.

[120] A. T. Abdel-hamid, S. Tahar, and E. M. Aboulhamid, "Ip watermarking techniques: Survey and comparison," in *IEEE International Workshop on System-on-Chip for Real-Time Applications*, 2003.

[121] U. Rührmair, F. Sehnke, J. Sölter, G. Dror, S. Devadas, and J. Schmidhuber, "Modeling attacks on physical unclonable functions," in *Proc. ACM Computer and Communications Security Conference (CCS)*, 2010.

[122] V. Lakafosis, A. Traille, H. Lee, G. Orecchini, E. Gebara, M. Tentzeris, G. DeJean, and D. Kirovski, "An RFID system with enhanced hardware-enabled authentication and anticounterfeiting capabilities," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, 2010, pp. 840–843.

[123] G. Becker, M. Kasper, A. Moradi, and C. Paar, "Side-channel based watermarks for integrated circuits," in *IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, 2010, pp. 30–35.

[124] F. Koushanfar and Y. Alkabani, "Provably secure obfuscation of diverse watermarks for sequential circuits," in *IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, 2010, pp. 42–47.

[125] H. Richter, W. Mostowski, and E. Poll, "Fingerprinting passports," in *NLUUG Spring Conference on Security*, 2008.

[126] K. Lofstrom, W. Daasch, and D. Taylor, "IC identification circuit using device mismatch," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, 2000, pp. 372–373.

[127] Y. Su, J. Holleman, and B. Otis, "A 1.6pj/bit 96%-stable chip ID generating circuit using process variation," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, 2007.

[128] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Transactions of Computers*, vol. 58, no. 9, pp. 1198–1210, 2009.

[129] D. B. Faria and D. R. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," in *Proc. ACM Workshop on Wireless Security (WiSe)*, 2006.

[130] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in the ether: Using the physical layer for wireless authentication," in *Proc. IEEE International Conference on Communications (ICC)*, 2007.

[131] N. Patwari and S. Kasera, "Robust location distinction using temporal link signatures," in *Proc. ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2007.

[132] *ITU-T Recommendation X.200: Information technology - Open Systems Interconnection - Basic Reference Model: The basic model*, ITU, 1994.

[133] Nmap Security Scanner, http://www.insecure.org/nmap/2004.

[134] Xprobe, http://www.sys-security.com/.

[135] R. Gerdes, T. Daniels, M. Mina, and S. Russell, "Device identification via analog signal fingerprinting: A matched filter approach," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2006.

[136] S. J. Murdoch, "Hot or not: revealing hidden services by their clock skew," in *Proc. ACM Conference on Computer and Communications Security (CCS)*, 2006, pp. 27–36.

[137] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. Randwyk, and D. Sicker, "Passive data link layer 802.11 wireless device driver fingerprinting," in *Proc. USENIX Security Symposium*, 2006.

[138] D. Loh, C. Cho, C. Tan, and R. Lee, "Identifying unique devices through wireless fingerprinting," in *Proc. ACM Workshop on Wireless Security (WiSe)*, 2008.

[139] S. Bratus, C. Cornelius, D. Peebles, and D. Kotz, "Active behavioral fingerprinting of wireless devices," in *Proc. ACM Conference on Wireless Network Security (WiSec)*, 2008.

[140] M. Edman and B. Yener, "Active attacks against modulation-based radiometric identification," Rensselaer Institute of Technology, Technical report 09-02, August 2009.

[141] B. Danev, H. Luecken, S. Capkun, and K. Defrawy, "Attacks on physical-layer identification," in *Proc. of the 3th ACM Conference on Wireless Network Security (WiSec)*. ACM, 2010, pp. 89–98.

[142] Y. Desmedt, C. Goutier, and S. Bengio, "Special uses and abuses of the Fiat-Shamir passport protocol." in *CRYPTO*, 1987, pp. 21–39.

[143] P. Papadimitratos, M. Poturalski, P. Schaller, P. Lafourcade, D. Basin, S. Capkun, and J.-P. Hubaux, "Secure neighborhood discovery: A fundamental element for mobile ad hoc networking," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 132–139, February 2008.

[144] M. Poturalski, P. Papadimitratos, and J.-P. Hubaux, "Towards provable secure neighbor discovery in wireless networks," in *Proc. of the 6th ACM workshop on formal methods in security engineering*, 2008.

[145] P. Schaller, B. Schmidt, D. Basin, and S. Capkun, "Modeling and verifying physical properties of security protocols for wireless networks," in *22nd IEEE Computer Security Foundations Symposium (CSF)*, 2009, pp. 109–123.

[146] A. Alrabady and S. Mahmud, "Some attacks against vehicles' passive entry security systems and their solutions," *Vehicular Technology, IEEE Transactions on*, vol. 52, no. 2, pp. 431 – 439, March 2003.

[147] ——, "Analysis of attacks against the security of keyless-entry systems for vehicles and suggestions for improved designs," *IEEE*

Transactions on Vehicular Technology, vol. 54, no. 1, pp. 41–50, January 2005.

[148] Bender Enterprises Inc, http://vintrack.com/SIU.html.

[149] P. Lepek and P. Hartanto, "RF design considerations for passive entry systems," Atmel automotive compilation, 2009, http://www.atmel.com/dyn/resources/prod_documents/article_passive_entry_s.pdf.

[150] Texas Instruments, "Car access system: Car access solutions from Texas Instruments," http://focus.ti.com/docs/solution/folders/print/528.html.

[151] NXP Semiconductors, "Passive keyless entry systems," http://www.nxp.com/applications/automotive/vehicle_access/rke/.

[152] Microchip Technology Inc., "Passive keyless entry (PKE) reference design, user manual," ww1.microchip.com/downloads/en/DeviceDoc/DS-21986A.pdf.

[153] S. Indesteege, N. Keller, O. Dunkelman, E. Biham, and B. Preneel, "A practical attack on KeeLoq," in *Proc. of the 27th Annual Eurocrypt Conference*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 1–18.

[154] C. Paar, T. Eisenbarth, M. Kasper, T. Kasper, and A. Moradi, "KeeLoq and side-channel analysis-evolution of an attack," *Fault Diagnosis and Tolerance in Cryptography, Workshop on*, vol. 0, pp. 65–69, 2009.

[155] N. T. Courtois, G. V. Bard, and D. Wagner, "Algebraic and slide attacks on KeeLoq," in *15th Intl. Workshop on Fast Software Encryption (FSE)*. Springer-Verlag, 2008, pp. 97–115.

[156] S. C. Bono, M. Green, A. Stubblefield, A. Juels, A. D. Rubin, and M. Szydlo, "Security analysis of a cryptographically-enabled RFID device," in *Proc. of the 14th USENIX Security Symposium*. Berkeley, USA: USENIX Association, 2005.