Author:
Paschou, Chrys

Title:
Physical Layer Security for Securing Resource-Constrained Networks

Physical Layer Security for Securing Resource-Constrained Networks

By

CHRYSANTHI PASCHOU

Department of Electrical & Electronic Engineering
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

JANUARY 2023

Count: 41450 words

“Simplicity is the ultimate sophistication”

Leonardo da Vinci

...however,

“Simplicity is not a simple thing”

Charles Spencer Chaplin

Abstract

Recent developments in society are trending towards the use of large networks comprising a large number of small devices. Traditional cryptographic techniques are no longer an attractive solution due to excessive memory and power requirements. Physical Layer Security (PLS) promises to be a good fit for forthcoming networks due to its potential for relatively low computational complexity. However, being a relatively new field, there are challenges to be addressed before PLS is integrated into real-world systems. After establishing the underlying theory in information theory and wireless communications, this thesis covers the background of PLS, identifies challenges, and suggests novel solutions for confidentiality, and authentication attacks in short-range systems. All suggested methods for security consider one end of the legitimate communication link to be severely constrained in terms of computational capabilities. The work presented after the literature review (chapter 4) challenges one of the most common assumptions made in key-based PLS, namely, the half-wavelength channel decorrelation assumption. It proves that such an assumption brings secrecy vulnerabilities which are brought to attention and quantified. The results motivate the definition of secure distance which facilitates the quantification of secrecy performance in terms of spatial channel correlation. Observations on the outcomes of chapter 4 motivate the solutions provided in the next two chapters. Chapter 5 presents a method that exploits base-station cooperation for the secure transmission of small data such as keys. The method addresses the main challenge of keyless PLS which is the requirement of a positive secrecy gap. Although spatial channel correlation is treated as an impairment for key generation purposes in chapter 4, chapter 6 demonstrates how it can be used to our benefit for protecting against authentication attacks in short-range systems via two novel methods. The first method targets relay attacks and replay attacks - the two most common impersonation attacks in short-range communications. The second method aims to verify the distance of devices that communicate through backscattering modulation and are equipped with little or no local power. The final research chapter finds a solution to the problem of a high reconciliation cost in key-based PLS. More precisely, chapter 7 presents a key agreement protocol which, different to existing key agreement protocols, achieves an arbitrarily low key disagreement rate without increasing the computational complexity. The practicality of the suggested key agreement protocol is successfully tested on a series of typical Internet-of-Things (IoT) boards.

Thesis Publications

The outcomes of the aforementioned contributions have resulted in five peer-reviewed papers and one patent. Although these are all co-authored works, I was the first author for each publication and led the research. Work dated 2023 is expected to be published by the Summer of 2023.

- C. Paschou, O. Johnson, A. Doufexi, Z. Zhu, and W.H. Chin. “Increasing the secrecy gap in quasi-static Rayleigh channels with secret splitting.” In 2020 IEEE Globecom Workshops (GC Wkshps, pp. 1-7. IEEE, 2020. [chapter 5]
- C. Paschou, O. Johnson, Z. Zhu, and A. Doufexi. “A Lightweight Protocol for Validating Proximity in UHF RFID Systems.” In 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), pp. 1-7. IEEE, 2021. [chapter 6]
- C. Paschou, O. Johnson, Z. Zhu, and A. Doufexi.“Re-Defining Secure Distance for CSI-based Key Generation Protocols.” In 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring), pp. 1-6. IEEE, 2022. [chapter 4]
- C.Paschou, Z. Zhu, M. Sandell. “Preventing replay/relay attacks in keyless entry systems.” Patent 54322US, Oblon, McClelland, Maier & Neustadt, L.L.P., 2022. [chapter 6]
- C. Paschou, O. Johnson, Z. Zhu, and A. Doufexi. “Physical Layer Protection Against Relay/Replay Attacks for Short-Range Systems.” In 2023 IEEE Wireless Communications and Networking Conference (IEEE WCNC 2023), pp. 1-6. IEEE, 2023.[chapter 6]
- C. Paschou, F. Raimondo, M. Gugala, D. McEwan, J. Pope, G. Oikonomou. “CRICKET: A Practical Physical Layer Key Agreement Protocol for IoT Networks.” In 2023 IEEE International Conference of Communications (IEEE ICC 2023), pp. 1-7. IEEE, 2023. [chapter 7]

Acknowledgements

First and foremost I would like to thank my supervisors, Professor Oliver Johnson, and, Professor Angela Doufexi for their support and invaluable insight throughout the PhD. I would like to also thank my industrial mentors from Toshiba Research Ltd, Ziming Zhu and Woon Hau, and my colleagues Francesco, James, and George. Their general expertise and insightful contributions along the way have been invaluable and much appreciated. Special thanks also go to Dave McEwan for all the fruitful conversations and technical advice. The PhD journey would not be as colourful without my close friends Hannah, Malu, Anne, Laurence, Olivia, Roger, and Jonathan. Last but certainly not least, I would like to thank my brother, George, and my parents Konstantina and Dimitri, for their ever-present support and encouragement.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ....*Chrysanthi Paschou*....... DATE: ....*20/01/2023* ....

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

Figure **Page**

Acronyms

2D two-dimensional.

3D three-dimensional.

AES Advanced Encryption Standard.

AN Artificial Noise.

AoA Angle-of-Arrival.

ASK Amplitude Shift Keying.

AWGN Additive White Gaussian Noise.

BAN Body Area Network.

BCH Bose–Chaudhuri–Hocquenghem.

BLE Bluetooth Low Energy.

CHRYSP CHannel Reflection Yields Secure Proximity.

CRicKET Channel Reciprocity for KEy Transmission.

CSI Channel State Information.

DES Data Encryption Standard.

ECC Error Correction Coding.

EDC Error Detection Coding.

EM electromagnetic.

EPC Electronic Product Code.

FDD Frequency Division Duplexing.

i.i.d. independentally and identically distributed.

IoT Internet-of-Things.

IP Internet Protocol.

IT Information Theory.

KDR Key Disagreement Rate.

LDPC low-density-parity-check codes.

LFSR Linear Feedback Shift Register.

LoS Line-of-Sight.

MAC Media Access Control.

MIMO Multiple Input Multiple Output.

NFC Near Field Communication.

NIST National Institute of Standards and Technology.

OFDM Orthogonal Frequency-Division Multiplexing.

OSI Open Systems Interconnection.

p.d.f. probability density function.

PAN Personal Area Network.

PKC Public Key Cryptography.

PL Physical-Layer.

PLA Physical Layer Authentication.

PLKG physical layer key generation.

PLS Physical Layer Security.

PSK Phase Shift Keying.

QAM Quadrature Amplitude Modulation.

r.v. random variable.

RF Radio-Frequency.

RFID Radio Frequency Identification.

RIS Reconfigurable Intelligent Surfaces.

rms root mean square.

ROC Receiver Operator Characteristic.

RSS Received Signal Strength.

RSSI Received Signal Strength Indicator.

SDF Solo Distance Fraud.

SDF-RFID (against) Solo Distance Fraud for RFID.

SNR Signal-to-Noise-Ratio.

TDD Time Division Duplexing.

UHF Ultra-high frequency.

UWB Ultra-WideBand.

Notation

$(\cdot)^*$ Conjugate operator.

B_c Coherence bandwidth.

$J_0(\cdot)$ Zeroth-order Bessel function of the first kind.

$[\cdot]$ Set of positive integers comprising one, the enclosed integer, and all integers in between. E.g., $[3] = \{1, 2, 3\}$.

$\Leftrightarrow$ Logical equivalence. E.g. *statement A*$\Leftrightarrow$*statement B* means *statement A* is true if and only if *statement B* is true.

$\Rightarrow$ Material conditional. E.g. *statement A*$\Leftarrow$*statement B* means that if *statement A* is true, so does *statement B*.

$\forall$ Universal quantification. I.e., $\forall x$ means ‘for all possible values of variable x’.

$\hat{(\cdot)}$ Estimation of a statistical measure.

$\in$ Set membership. I.e., $x \in \mathscr{X}$ means that x is an element of the set $\mathscr{X}$. Notation $\notin$ indicates the opposite..

λ Wavelength (m).

$(\cdot)^H$ Hermitian transpose operator.

$\mathbb{C}$ Set of complex numbers.

$\mathbf{0}_N$ The all zero matrix of size $N \times N$.

$\mathbf{I}_N$ Unit matrix of size $N \times N$.

$\mathbf{x} \sim \mathscr{CN}(\mathbf{0}, I)$ Variable $\mathbf{x}$ is a standard complex random variable vector.

$\mathscr{CN}(\mathbf{0}, I)$ The all zero matrix of size $N \times N$.

ρ Spatial Channel correlation.

$\tau_{\mathbf{rms}}$ Root mean square delay spread.

$\mathbf{dist}(\cdot,\cdot)$ The number of places where the enclosed vectors disagree.

$\vec{x}$ Uplink transmission vector.

$\{a_i\}$ Abbreviation for the set $\{a_1, a_2, \ldots, a_n\}$.

c The speed of light in vacuum ($\approx 299\,792\,458\,\mathrm{m/s}$).

f Frequency (Hz).

f_c Carrier frequency (Hz).

$x \sim \mathscr{CN}(0,1)$ Variable x is a standard complex random variable. I.e., x follows the circularly symmetric complex distribution.

$I(x;y)$ The mutual information between variables x and y.

$\Gamma(k,\theta)$ Gamma distribution with shape parameter k and scale parameter θ.

$\Gamma_{\mathbf{inc}}(\cdot,\cdot)$ Upper incomplete Gamma function defined as $\Gamma s, x := \int_x^\infty t^{s-1}e^{-t}\mathrm{dt}$.

$\gamma_{\mathbf{inc}}(\cdot,\cdot)$ Lower incomplete Gamma function defined as $\Gamma s, x := \int_x^0 t^{s-1}e^{-t}\mathrm{dt}$.

$\mathscr{CN}(0,1)$ Circularly-symmetric complex distribution. Also referred to as the complex normal distribution. A special case of $\mathscr{CN}(\mu,\sigma^2)$.

$\mathscr{CN}(\mu,\sigma^2)$ Complex normal distribution of mean μ and variance σ^2.

$\mathrm{sinc}(\cdot)$ The unnormalised sinc function defined as $\mathrm{sinc}(x) = \frac{\sin x}{x}$ when $x \neq 0$, and $\mathrm{sinc}(0) = 1$. Also known as the zeroth-order spherical Bessel function of the first kind.

$|\cdot|$ Magnitude or determinant depending on the enclosed mathematical object. When $x \in \mathbb{C}$, $|x| = \sqrt{\mathrm{Re}(x)^2 + \mathrm{Im}(x)^2}$, whereas $|\mathbf{X}|$ denotes determinant of matrix $\mathbf{X}$.

$||\cdot||$ The Frobenious norm of the enclosed matrix..

Bold lower-case letters denote vectors. E.g. $\mathbf{y}$ has the form of $(y_1, y_2, \ldots, y_n)$, whereas $\mathbf{X}$ is a vector of random variables: $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

Bold upper-case letters multi-dimensional random variables or matrices. Interpretations are clarified within the text..

Upper-case letters represent random variables or matrices, with interpretations clarified in the text. Some, followed by parentheses ($H(\cdot,\cdot)$) are reserved for metrics and determined separately.

Chapter 1

Introduction

1.1 Motivation

In an era marked by the rapid proliferation of the fifth generation of wireless networks (5G), where hundreds of new devices connect to wireless networks worldwide every second, the need for security, privacy, and confidentiality has never been more pressing. Many of these devices have severe constraints on memory and energy due to their small size or low cost. Traditional cryptographic techniques implemented in the upper layers are no longer a good fit due to being computationally complex and the research community is seeking alternative ways of providing security. Physical Layer Security (PLS) has gained much attention due to its potential in relatively low computational complexity. This thesis covers the journey of PLS, identifies challenges and suggests novel solutions for confidentiality as well as authentication in short-range systems.

1.1.1 Limitations of Traditional Ways of Security

The traditional way of securing communications is performed in the upper layers of the Open Systems Interconnection (OSI) protocol stack by means of Public Key Cryptography (PKC) and symmetric cryptography [2]. Symmetric cryptography facilitates confidentiality, whereas PKC can be used for confidentiality and authentication purposes.

In a way, PKC transforms a public, and therefore, insecure channel to a secure link that can be used to transfer confidential data. The drawback of PKC is that it involves intensive calculations over large prime numbers and finite fields. The implementation of PKC has been viable for 2G, 3G, and 4G networks that consist of a small number of powerful devices such as laptops and mobile phones. However, PKC is not viable in emerging networks consisting of low-cost devices due to requiring excessive resources of power and memory [3]. In contrast to PKC, symmetric encryption has very little computational complexity and is widely accepted as a viable solution for resource-constrained networks. However, for enabling symmetric cryptography, the two communicating parties need to share a secret key. The secure link for providing two symmetric (identical) keys is often provided by PKC, e.g. by the Diffie-Hellman key exchange protocol [2].

Unable to perform PKC, current low-power/low-memory devices are subject to inadequate or absent security mechanisms. Keys are fixed and stored in hardware, or input by the user. In some systems, the least computationally complex protocols of PKC are implementable but they result in keys of a short length and low entropy [4]. Sub-optimal security methods of fixed/low entropy keys can bring security vulnerabilities not only to the device on which the attack was launched but to the whole network to which the device is connected. Taking into consideration the increasing number of interconnected devices, security enhancement for the weakest link of a network is of paramount importance.

Lastly, it is well known that PKC-based authentication such as electronic signatures do not provide protection against person-in-the-middle attacks which is the most common type of authentication attacks. In a person-in-the-middle attack, two legitimate users believe that they communicate directly, when in fact their messages have been relayed and possibly altered by an adversary. Media Access Control (MAC) and Internet Protocol (IP) addresses provide a very low level of protection in the MAC/network layer; they can be easily spoofed and altered. The most vulnerable systems to person-in-the-middle attacks in recent years are short-range communications such as keyless entry systems and contactless payments. A relay attack, for example, is a type of person-in-the-middle attack that accounts for the vast majority of car-theft crimes.

1.1.2 PLS as a potential solution

PLS is implemented in the lowest layer of a communication protocol stack (see figure 1.1) which is associated with the physical connection between two devices. Physical layer security is mainly concerned with confidentiality and authentication purposes and is subdivided into three main categories, namely keyless PLS, key-based PLS, and the most recent branch of Physical Layer Authentication (PLA).

Keyless PLS exploits the noise and channel fluctuations of the physical medium in order to “hide” information from unintended receivers. In contrast to symmetric cryptography, it has the potential to achieve confidentiality without the use of keys, and therefore, it does not require a secure link for key exchange. Furthermore, perfect secrecy is possible with keyless PLS [5] which is superior to the computational security provided by the traditional security mechanisms. As long as there is some knowledge of the adversary’s channel, keyless PLS can provide confidentiality regardless of the adversary’s computational capabilities. Notably, the method introduced in this thesis eliminates the need for knowledge of the eavesdropper’s channel, thereby enhancing the applicability and robustness of keyless PLS.

Key-based PLS extracts the unique randomness inherited in the physical link between two devices in order to provide a shared secret. This shared secret serves as two identical keys that facilitate symmetric cryptography in the upper layer of the protocol stack. Early testbed implementations prove that key-based PLS can be performed by most Radio-Frequency (RF) transceivers by measuring channel parameters on the received RF signal. It is believed to be the replacement mechanism for the computationally heavy key-exchange protocols of PKC.

It was only in 2007 when a physical layer approach for authentication (PLA) was introduced for the case of wireless networks [6], whereby a device was identified[1] simply by its hardware imperfections captured in its transmitted signals, a process commonly referred to as fingerprinting. Since then the literature and in-lab investigations have grown and promoted PLA as a promising solution against person-in-the-middle and other authentication attacks. Authentication in the physical layer is typically faster than conventional authentication as it takes place before demodulating and decoding the signal. Furthermore, it can authenticate any device that uses radio communications regardless of its computational capabilities.

1.2 Summary of Contributions

Due to being a relatively new field, there are challenges to be addressed before PLS is integrated into real-world systems. Current PLS algorithms are often based on idealistic channel models and assumptions. This thesis covers the journey of PLS, identifies challenges, and suggests novel solutions for confidentiality, and authentication attacks in short-range systems. All suggested methods for security consider one end of the legitimate communi-

[1] identification is a fundamental aspect of authentication

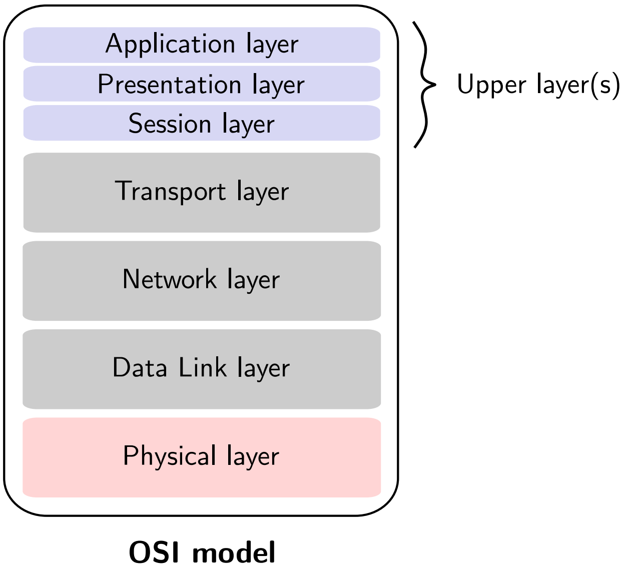


Figure 1.1: The physical layer is the first and lowest layer of the Open Systems Interconnection (OSI) networking model.

cation link being severely constrained in terms of computational capabilities. On the other hand, certain extra signal processing or computation is sometimes allowed at the second end of the link, e.g. at an access point or a base station. This thesis makes the following contributions:

1.2.1 Contributions to the state-of-the art of keyless PLS

- Introducing a novel method, namely *secret splitting*, that exploits base station cooperation for confidentiality; (chapter 5)

- Proving that secrecy splitting dramatically decreases the areas at which unintended receivers are able to convey confidential information; (chapter 5)

- Performing analytical analysis on the proposed scheme and deriving the optimal base station allocation. (chapter 5)

1.2.2 Contributions to the state-of-the-art of PLA

- Introducing a new concept for authenticating devices in short-range systems based on the spatial and temporal properties of the RF channel; (chapter 6)

- Proposing a scheme for secure authentication of co-located devices against distance fraud and replay attacks; (chapter 6)
- Proposing a scheme for validating the proximity of a non-powered device that communicates using backscattering modulation. (chapter 6)

1.2.3 Contributions in the state-of-the-art key-based PLS

- Proving that one of the most common assumptions in the field of key-based PLS brings secrecy vulnerabilities which are quantified and brought to attention; (chapter 4)
- Providing an alternative definition to *secure distance* that captures the impact of spatial channel correlation between a legitimate receiver and an eavesdropper; (chapter 4)
- Presenting a practical key agreement protocol suitable for severely resource-constrained networks. The practicality of the protocol is successfully tested in an Internet-of-Things (IoT) network composed of low-power devices. (chapter 7)

1.3 Thesis Outline

The remaining chapters of the thesis are as follows. Chapters 2 and 3 provide the underlying background and literature review of physical layer security, respectively. The literature review comprises three sections, each of which is associated with one or more of the main research chapters, i.e. chapters 4-7, as seen in figure 1.2. Chapter 4 evaluates one of the most common assumptions found in the field of key-based PLS and redefines secure distance. Chapters 5-7 consist of novel methodologies for keyless PLS for confidentiality (chapter 5), authentication in short-range systems (chapter 6), and key agreement in IoT networks (chapter 7). Lastly, chapter 8 concludes the thesis and suggests future research directions.

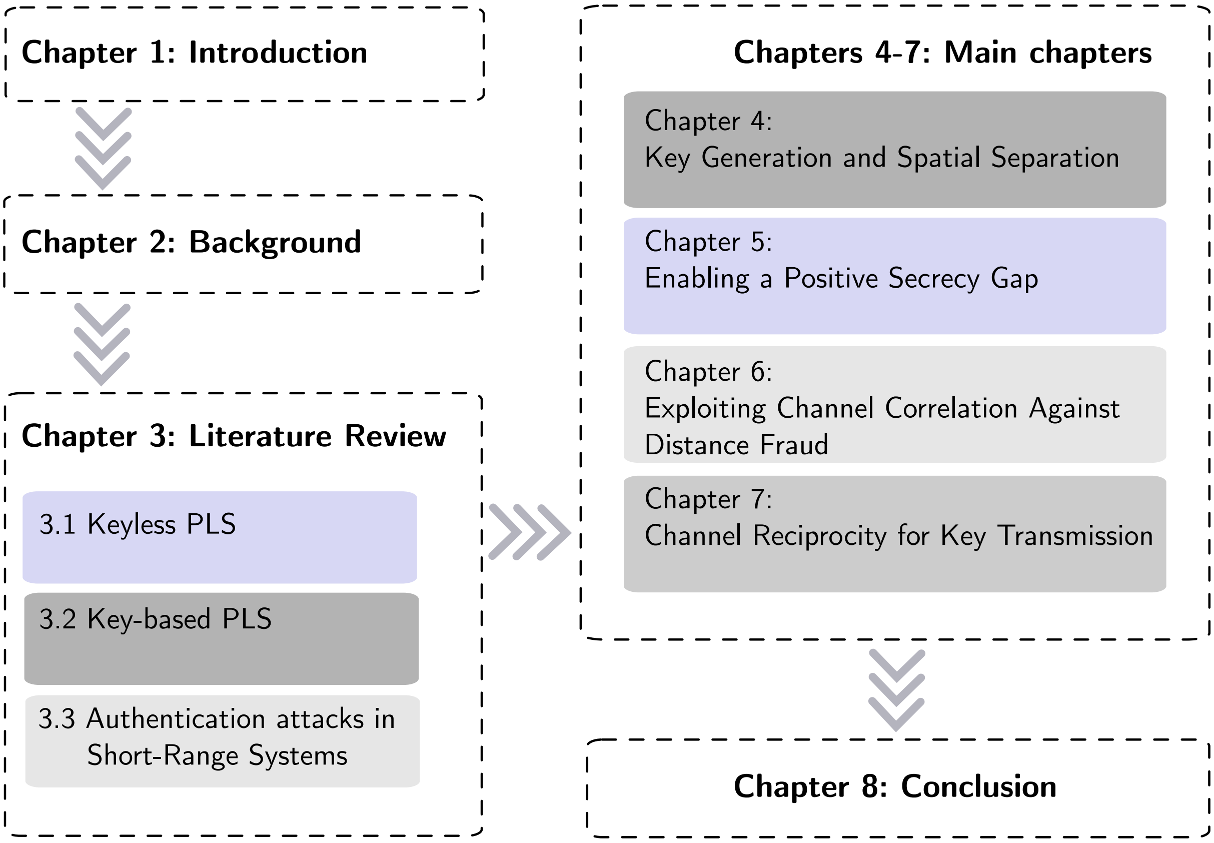

Chapter 1: Introduction
Chapter 2: Background
Chapter 3: Literature Review
3.1 Keyless PLS
3.2 Key-based PLS
3.3 Authentication attacks in Short-Range Systems
Chapters 4-7: Main chapters
Chapter 4: Key Generation and Spatial Separation
Chapter 5: Enabling a Positive Secrecy Gap
Chapter 6: Exploiting Channel Correlation Against Distance Fraud
Chapter 7: Channel Reciprocity for Key Transmission
Chapter 8: Conclusion

Figure 1.2: Thesis outline. Matching colours indicate high relevance between a section of the literature review and a main chapter.

Chapter 2

Background

Physical layer security combines the disciplines of information theory and wireless engineering. Thus, it is important to introduce the basic concepts of the two disciplines before proceeding to the main body of this thesis [1]. As this thesis promotes PLS as a facilitator for symmetric cryptography for confidentiality and a replacement for asymmetric cryptographic mechanisms for key exchange, the fundamentals of traditional cryptographic systems are also covered. This chapter is divided into three sections covering the fundamentals of information theory, cryptographic systems, and the wireless channel.

2.1 Fundamentals of Information Theory

Information Theory (IT) was invented by Shannon [7] in 1948 and shaped the digital world that we live in today. Shannon has proved that any type of information can be represented by binary data and quantified precisely. Information theory has given rise to techniques such as data compression, error correction [8], and the relatively recent study of PLS [9].

[1] A number of proofs in this chapter differ from the proofs provided by the founder of the associated theorems. These proofs have been reformed to promote consistency in style based on material content presented in previous sections.

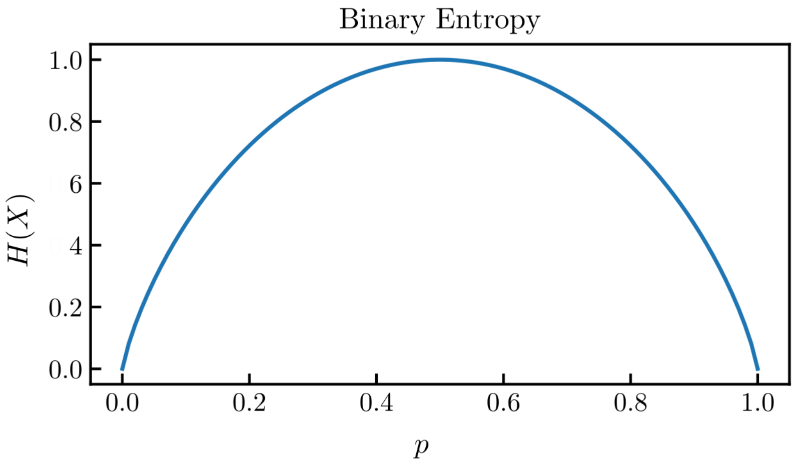


Figure 2.1: The entropy of binary random variable $X \in \{\text{Heads}, \text{Tail}\}$ against $p := P(X = \text{Heads})$.

2.1.1 Basic concepts

Information is no longer an abstract concept, but it can be quantified in bits. The definition of the information content realises a mathematical approach to communications.

Definition 2.1. Let X be a discrete random variable (r.v.), and x be a realisation of the variable. The information content gained for X by observing x is defined as

$$\mathbb{I}(X = x) = -\log_2 P(X = x). \tag{2.1}$$

One Shannon bit, or bit, is the information gained by observing the realisation of a r.v. that takes two values, e.g. 0 and 1, with equal probability.

Loosely speaking, the information content quantifies our surprise at a result. A certain event ($(P(X = x) = 1)$), has zero information content since it is associated with no surprise.

Remark 2.1. *When X is uniformly distributed, the information content is equal to $\mathbb{I}(X = x) = -\log_2 |\mathscr{X}|$, where $\mathscr{X}$ is the size of the alphabet,that random variable X is drawn from.*

For example, tossing a fair coin will convey $-\log_2(1/2) = 1$ bit of information. By rolling a fair six-sided die and observing the outcome $-\log_2(1/6) \approx 2.6$ bits of information is gained. If the die is biased such that $P(X = 1) = 0.5$, whereas $P(X = i) = 0.1$ for $i \neq 6$, then observing any side but 6 will convey more information than observing six. A natural quantity arising from the latter example is the average information gained $E(I(X))$ when an experiment is repeated.

Definition 2.2. Let r.v. X be a r.v. with probability function $P(\cdot)$. The entropy of X is defined as

$$H(X) := -\sum_{x \in \mathscr{X}} P(X = x) \log_2(P(X = x)), \tag{2.2}$$

where $\mathscr{X}$ is the set of all possible values of X.

When tossing a fair coin, the probability of "heads" is p=1/2, whereas for a biased coin $p > 0 \neq 1/2$. The entropy of the binary random variable $(X \in \{\text{Heads}, \text{Tails}\{)$ is plotted in figure 2.1 whereby it can be seen that the entropy is maximised for the uniform distribution, i.e. when $p = 1/2$. This result can be generalised for non-binary random variables.

Theorem 2.1. *The entropy of a discrete random variable, X, is maximised when X is uniformly distributed.*

The truth of the above theorem is somewhat intuitive; The uncertainty about the outcome of an experiment is maximised when all events are equiprobable. For detailed proof, the reader is referred to [8].

Corollary 2.1. *Let X be a binary string of length k. Then*

$$H(X) \leq k. \tag{2.3}$$

Proof. Let $\mathscr{X}$ be the set of all possible binary strings of length K. The maximum entropy is achieved when $P(X = x) = 1/|\mathscr{X}| = 1/2^k$, for which case $H(X) = -\sum P(X = x) \log(P(X = x)) = \sum 1/2^k \log(2^k) = k \sum P(X = x) = k$. □

2.1.2 Conditional entropy and mutual information

Communications involve two or more communicating nodes. The following definitions and theorems link two or more random variables.

Definition 2.3. Let X and Y be two r.v.s. The information gained about X given the observed value $Y = y$ is

$$\mathbb{I}(X = x | Y = y) := -\log_2 P(X = x | Y = y). \tag{2.4}$$

Definition 2.4. The average over all possible realisations of (X, Y) defines the conditional entropy of X on Y:

$$H(X|Y) := E\big(\mathbb{I}(X = x|Y = y)\big) = -\sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} P(X = x, Y = y) \log_2(P(X = x|Y = y)), \tag{2.5}$$

where $\mathscr{X}$ and $\mathscr{Y}$ are the sets comprising all possible realisations of X and Y, respectively.

From this point, the notation will be simplified by writing $P_X(x)$ instead of $P(X = x)$. Equivalently the conditional probability and joint probability are denoted by $P_{X|Y}(x|y) := P(X = x|Y = y)$ and $P_{X,Y}(x, y) := P(X = x, Y = y)$. The log-base of two is also dropped, i.e. $\log(\cdot) := \log_2(\cdot)$.

Entropy is often perceived as the amount of uncertainty or randomness of a variable X. In this context, conditional entropy of X on Y is the amount of uncertainty remaining in X given that Y is known. The amount of reduction in the uncertainty $H(X) - H(X|Y)$ is called the mutual information of X and Y and is denoted by $I(X;Y)$.

Definition 2.5. The mutual information of two r.v.s X and Y is defined as

$$I(X;Y) := H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2.6}$$

Definition 2.6. The mutual information of X and Y conditioned on Z is defined as

$$I(X;Y|Z) := H(X|Z) - H(X|(Y,Z)) \tag{2.7}$$

Theorem 2.2.

$$I(X;Y) = \sum_{y\in\mathscr{Y}} \sum_{x\in\mathscr{X}} P_{X,Y}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y))}\right) \tag{2.8}$$

$$I(X;Y|Z) = \sum_{z\in\mathscr{Z}} \sum_{y\in\mathscr{Y}} \sum_{x\in\mathscr{X}} P_Z(z) P_{X,Y,Z}(x,y,z) \log\left(\frac{P_{X,Y,Z}(x,y,z)}{P_{X,Z}(x,z)P_{Y,Z}(y,z))}\right) \tag{2.9}$$

Proof. By definition,

$$I(X;Y) := H(X) - H(X|Y) = -\sum_x P_X(x) \log P_X(x) + \sum_x \sum_y P_{X,Y}(x,y) \log P_{X|Y}(x|y). \tag{2.10}$$

In probability theory [10, Ch. 1], the Bayesian theorem states that $P_{X|Y}(x|y)P_Y(y) = P_{X,Y}(x,y)$. Applying the Bayesian theorem in equation (2.10) completes the proof. □

The definition of information content and, subsequently, the definitions of entropy and mutual information can be extended for a multivariate variable $\mathbf{X} = [X_1, \ldots, X_n]$ by simply replacing the probability function with the joint probability function $P_{X_1,\ldots,X_n}$. For example, for a bivariate r.v. $[X, Y]$, where $X \in \mathscr{X}$, and $Y \in \mathscr{Y}$, the analytical expression of the entropy is given by

$$H(X,Y) = -\sum_{x\in\mathscr{X}} \sum_{y\in\mathscr{Y}} P_{X,Y}(x,y) \log P_{X,Y}(x,y). \tag{2.11}$$

Theorem 2.3 (Chain Rule)**.**

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{2.12}$$

Proof. The proof is direct by applying the Bayesian theorem $(P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y))$ in equation (2.11). □

Corollary 2.2.

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{2.13}$$

Proof.

$$I(X;Y) := H(X) - H(X|Y) \overset{2.3}{=} H(X) + H(Y) - H(X,Y).$$

□

Corollary 2.3 (Chain rule for three variables)**.**

$$H(X,Y,Z) = H(X|Y,Z) + H(Y|Z) + H(Z). \tag{2.14}$$

Proof. The proof is derived by substituting Y with (Y,Z) in equation (2.12). □

The following lemma is the application of Jensen's inequality [8, Ch. 2] for the logarithmic function. It will be used as a tool for proving some of the theorems in this thesis such as theorem 2.4.

Lemma 2.1 (Jensen's inequality)**.**

$$E\left(\log(X)\right) \le \log\left(E\left(X\right)\right) \tag{2.15}$$

Moreover, when the equality holds, variable X is a constant number, or equivalently $E(X) = X$ with probability one.

Proof. Jensen's inequality [8, Th.2.6.2] states that when f is a convex function, then $E\left(f(x) \le f\left(E\left(X\right)\right)\right)$. It also states that if f is strictly convex, then the equality of $E\left(\log(X) = \left(E\left(X\right)\right)\right)$ implies that X is a constant number. It suffices to show that $f(x) = -\log(x)$ is a strictly convex function. Indeed, since the second derivative $\frac{\partial^2 f}{\partial x^2} = \frac{1}{x^2}$ is strictly positive, $f(x) = -\log(x)$ is strictly convex.

□

Theorem 2.4. *The following statements are equivalent:*

$$X \text{ and } Y \text{ are independent r.v.s} \tag{2.16}$$

$$H(X|Y) = H(X) \tag{2.17}$$

$$H(X,Y) = H(X) + H(Y) \tag{2.18}$$

$$I(X;Y) = 0 \tag{2.19}$$

Proof. (2.16)⇒(2.17): From probability theory, when X and Y are independent, $P_{X|Y}(x|y) = P_X(x)$ for all x and for all y. Therefore,

$$H(X|Y) = -\sum_x \sum_y P_{X,Y}(x,y) \log_2 P_{X|Y}(x|y) = -\sum_x \log_2 P_X(x) \sum_y P_{X,Y}(x,y) = -\sum_x (\log_2 P_X(x)) P_X = H(X).$$

(2.17)⇒(2.18): straightforward after applying the chain rule (2.12).
(2.18)⇒(2.19): straightforward after applying corollary 2.2.
(2.19)⇒(2.16): The proof is partitioned in two parts. First, we prove by the method of contradiction that $I(X;Y) = 0$ implies that $P_{X,Y} = cP_X P_Y$, where c is a constant number. Second, we prove that $c = 1$.

Let $I(X;Y) = 0$ and let $P_{X,Y}(x,y) \neq cP_X(x)P_Y(y)$ for all $(x,y) \in \mathscr{X} \times \mathscr{Y}$, where c is a constant number. Then,

$$I(X;Y) \overset{Th.2.2}{=} -\sum\sum P_{X,Y} \log \frac{P_X P_Y}{P_{X,Y}} \tag{2.20}$$

$$\overset{Lem.2.1}{>} -\log\left(\sum\sum P_{X,Y} \frac{P_X P_Y}{P_{X,Y}}\right) \tag{2.21}$$

$$= -\log\left(\sum\sum P_X P_Y\right) = -\log\left(\sum P_X \sum P_Y\right) \tag{2.22}$$

$$= -\log 1 = 0. \tag{2.23}$$

The strict inequality of (2.21) is derived from lemma 2.1 since $\frac{P_X P_Y}{P_{X,Y}} \neq c$.
Having assumed that $P_{X,Y}(x,y) \neq cP_X(x)P_Y(y)$ led to $I(X;Y) > 0$ which is a contradiction. Hence, given that $I(X;Y) = 0$, $P_{X,Y}(x,y) = cP_X(x)P_Y(y)$ must hold true for some $c \in \mathbb{N}$. Summing over all $x \in \mathscr{X}$ and $y \in \mathscr{Y}$ we have that $\sum P_{X,Y}(x,y) = c\sum\sum P_X(x)P_Y(y) \Rightarrow c = 1$, which completes the proof. □

2.1.3 Gaussian random variables

Until now only discrete variables have been encountered. Replacing the probability function in (2.1) with the probability density function (p.d.f.) $f_X(\cdot)$ of a continuous r.v. results in similar definitions of the information content, entropy, and mutual information.

Definition 2.7. Let X be a continuous r.v. with p.d.f. $f_X(\cdot)$. The information content of a realisation $X = x$ is defined as $I(X = x) := \log(f_X(x))$.

All the aforementioned properties of entropy and mutual information are satisfied for the continuous case, too. The proofs of theorems and corollaries are identical except for replacing the probability functions with the p.d.f.s $f_X(\cdot)$, and substituting the summations with integrals. For example, the differential entropy is given by

$$H(X) := E(I(X)) = -\int_{-\infty}^{+\infty} f_X(x)\log f_X(x)\mathrm{d}x. \tag{2.24}$$

Also known as normal r.v.s, Gaussian r.v.s will be encountered frequently throughout the thesis.

Example 2.1. *Let $x \sim \mathcal{N}(\mu,\sigma^2)$ be a Gaussian r.v. with mean μ and variance σ^2. The p.d.f. of x is given by*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{2.25}$$

Then,

$$\begin{aligned} H(X) =& \\ &-\int_{-\infty}^{+\infty} f(x)\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = \int_{-\infty}^{+\infty} f(x)\left(\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x-\mu)^2}{2\sigma^2}\right)dx \\ &= \frac{1}{2}\log(2\pi\sigma^2) + \frac{\sigma^2}{2\sigma^2} = \frac{1}{2}\log(\pi e\sigma^2) + \frac{1}{2} = \frac{1}{2}\log(2\pi e\sigma^2). \end{aligned}$$

The previous example can be generalised for the case of multivariate normal variables. The general case is provided in the form of a theorem after establishing essential terminology.

Definition 2.8. Let $\mathbf{X} = (X_1,\dots,X_n)$ be a multivariate random variable. Given that $E(X_i)$ exist, the mean of $\mathbf{X}$ is defined as

$$E(\mathbf{X}) := (E(X_1),\dots,E(X_n)). \tag{2.26}$$

The covariance matrix of $\mathbf{X}$, Σ, is the matrix the (i,j) entry of which is given by

$$\Sigma_{ij} := E\big((X_i - E(X_i))\,(X_j - E(X_j))\big). \tag{2.27}$$

Definition 2.9. A multivariate random variable $\mathbf{X} = (X_1,\dots,X_n)^T$ is said to be multivariate Gaussian with mean μ and covariance Σ -and we write $\mathbf{X} \sim N(\mu,\Sigma)$- when the probability density function of $\mathbf{X}$ is given by

$$f(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\exp\left(-\frac{1}{2}(\mathbf{X}-\mu)\Sigma^{-1}(\mathbf{X}-\mu^T)\right). \tag{2.28}$$

We give a result which generalises example 2.1

Theorem 2.5 ([8],Th. 8.4.1.)**.** *Let* $\mathbf{X} \sim N(\mu, \Sigma)$. *Then*

$$H(X_1,\ldots,X_n) = \frac{1}{2}\log\left((2\pi e)^n \det(\Sigma)\right), \tag{2.29}$$

where $\det(\Sigma)$ *is the determinant of* Σ.

The proof of theorem 2.5 can be found in [8, Ch. 8].

Theorem 2.6. *Let* $(X_1, X_2) \sim N(\mathbf{0}, \Sigma)$.

Then

$$I(X_1; X_2) = \frac{1}{2}\log\left(\frac{\Sigma_{11}\Sigma_{2,2}}{\det(\Sigma)}\right). \tag{2.30}$$

Proof.

$$I(X;Y) := H(X) - H(X|Y) \overset{(2.18)}{=} H(X) + H(Y) - H(X,Y) \overset{Ex.\ 2.1}{=} \frac{1}{2}\log(2\pi e\sigma_X^2) + \frac{1}{2}\log(2\pi e\sigma_Y^2) - H(X,Y).$$

From theorem 2.5, $H(X,Y) = \frac{1}{2}\log\left((2\pi e)^2\det(\Sigma)\right)$. Hence, $I(X;Y)$ simplifies to (2.30). □

2.1.3.1 Complex random variables

In the following chapters complex baseband signals/channels will be considered. It is therefore essential to extend the concept of entropy and mutual information to encounter complex r.v.s. Independence between the real part and the imaginary part of a complex variable is a typical assumption in wireless engineering and it will be adopted throughout this thesis. i.e. for any complex $Z = X + jY$, it is assumed that $\mathrm{Cov}(X,Y) = 0$.

Definition 2.10. The entropy of a complex r.v. $Z = X + jY$ is defined as $H(Z) = H(X,Y)$. For the multivariate case, if $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ is a complex vector, then $H(\mathbf{Z}) = H(\mathbf{X},\mathbf{Y})$.

Let $\mathbf{Z} = (Z_1,\ldots,Z_n)$. Observe that $H(Z_1,\ldots,Z_n) = H(\Re(\mathbf{Z}), \Im(\mathbf{Z}))$. Therefore, an equivalent definition to 2.10 is $H(\mathbf{Z}) := H(Z_1,\ldots,Z_n)$.

Remark 2.2. *The entropy of a random variable (complex or real) is always a real number.*

Definition 2.11. The expectation and variance of a complex vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ are defined as

$$E(\mathbf{Z}) = E(\mathbf{X}) + jE(\mathbf{Y}) \tag{2.31}$$

$$\Sigma_{ZZ} = \mathrm{Cov}(\mathbf{Z},\mathbf{Z}) = E\left((\mathbf{Z} - E(\mathbf{Z}))(\mathbf{Z} - E(\mathbf{Z}))^H\right), \tag{2.32}$$

where $(\cdot)^H$ is the hermitian transpose of a matrix.

Remark 2.3. *When $\mathbf{Z}$ is a vector with n entries covariance $\Sigma_{\mathbf{ZZ}}$ is an $n \times n$ matrix. The diagonal consists of entries $Cov(Z_i, Z_i) = E\left(Z_i Z_i^H\right) = Var(X_i) + Var(Y_i)$.*

Definition 2.12. A r.v. $Z = X + jY \in \mathbb{C}$ is said to be circularly symmetric Gaussian of zero mean and variance σ^2 when X, Y are independentally and identically distributed (i.i.d.) as $\sim \mathcal{N}(\mathbf{0}, \sigma^2/2)$.

To denote a r.v. with the above properties notation $Z \sim CN(0, \sigma^2)$ is used. For the specific case of $Z \sim CN(0,1)$, Z is described as standard complex normal.

Remark 2.4. *The covariance matrix of a circularly symmetric Gaussian $Z = X + jY \in \mathbb{C}$ is given by $\Sigma_{ZZ} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$.*

Remark 2.5. *There will be occasions when a complex number is given in polar coordinates: $X = |X|e^{i\theta}$, where $|X| = \sqrt{\Re(X)^2 + \Im(X)^2}$, and $\theta = \tan^{-1}(\Im(X)/\Re(X))$. When X is circularly symmetric complex, θ is uniformly distributed in $[0, 2\pi)$, and the amplitude $|X|$ is Rayleigh distributed [10]. Moreover, variables $|h|$ and θ are independent r.v.s. More details on circularly symmetric r.v.s are given in section 2.3.2.*

Lemma 2.2. *The entropy of a circularly symmetrical r.v. $Z \sim CN(0, \sigma^2)$ is given by*

$$H(Z) = \log(\pi e \sigma^2). \tag{2.33}$$

Let $Z = X + jY$. Then

$$H(Z) = H((X,Y)) \overset{(2.18)}{=} H(\Re(X)) + H(\Im(X)) \overset{(2.33)}{=} \log(\pi e \sigma^2).$$

Definition 2.13. Notation $\mathbf{Z} \sim CN(\mathbf{0}, \Sigma)$, where $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ is used for the case when $\mathbf{X}$ and $\mathbf{Y}$ are independent and identically distributed as $\mathbf{X}, \mathbf{Y} \sim N(\mathbf{0}, \frac{1}{2}\Sigma)$. i.e. $\mathbf{X}$ and $\mathbf{Y}$ are zero mean complex vectors with identical covariance matrices $\Sigma_{\mathbf{XX}} = \Sigma_{\mathbf{YY}} = \frac{1}{2}\Sigma$.

Theorem 2.7. *Let $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim CN(\mu, \Sigma)$*

then

$$H(\mathbf{Z}) = \log((\pi e)^n \det(\Sigma)) \tag{2.34}$$

Proof. Since $\mathbf{X}$ and $\mathbf{Y}$ i.i.d., $H(Z) := H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) = 2H(\mathbf{X}) \overset{Th.\,2.5}{=} \log((2\pi e)^n \det(\Sigma/2)) = \log((\pi e)^n \det \Sigma)$. □

Theorem 2.8. *i. Let* $\mathbf{Z} = (Z_1, Z_2) \sim CN(\mathbf{0}, \Sigma)$*, where* $\Sigma = \begin{bmatrix} \Sigma_{Z_1}^2 & \Sigma_{Z_1 Z_2} \\ \Sigma_{Z_1 Z_2} & \Sigma_{Z_2}^2 \end{bmatrix}$*. Then,*

$$I(Z_1; Z_2) = \log \frac{\sigma_{Z_1}^2 \sigma_{Z_2}^2}{\det(\Sigma)}. \tag{2.35}$$

ii. Let $\mathbf{Z} = (Z_1, Z_2, Z_3) \sim CN(\mathbf{0}, \Sigma)$*, where* $\Sigma = \begin{bmatrix} \Sigma_{Z_1}^2 & \Sigma_{Z_1 Z_2} & \Sigma_{Z_1 Z_3} \\ \Sigma_{Z_1 Z_2} & \Sigma_{Z_2}^2 & \Sigma_{Z_2 Z_3} \\ \Sigma_{Z_1 Z_3} & \Sigma_{Z_2 Z_3} & \Sigma_{Z_3}^2 \end{bmatrix}$*. Then,*

$$I(Z_1, Z_2 | Z_3) = \frac{\det(\Sigma_{Z_1 Z_3}) \det(\Sigma_{Z_2 Z_3})}{\sigma_{Z_3}^2 \det(\Sigma)}, \tag{2.36}$$

where $\Sigma_{Z_i Z_j}$ *is the matrix consisted of rows* i, j *and columns* i, j*.*

Proof. $I(Z_1; Z_2) \overset{Cor.2.2}{=} H(Z_1) + H(Z_2) - H(Z_1, Z_2) \overset{Lem.2.2}{=} \log(\pi e \Sigma_{Z_1}^2) + \log(\pi e \Sigma_{Z_2}^2) - \log((\pi e)^2 \det(\Sigma))$ which simplifies to (2.35). The proof of the second part of the theorem follows a similar line and it is omitted. □

2.1.3.2 Mutual Information and Correlation

Definition 2.14. The correlation coefficient between two r.v.s Z_1 and Z_2 is defined as:

$$\rho(Z_1, Z_2) := \frac{\Sigma_{Z_1 Z_2}}{\Sigma_{Z_1} \Sigma_{Z_2}} \tag{2.37}$$

where $\Sigma_{Z_1 Z_2} = E\left((Z_1 - E(Z_1))(Z_2 - E(Z_2))^H\right)$, and $\Sigma_{Z_i} = \sqrt{\Sigma_{Z_i Z_i}}$.

Theorem 2.9. *The correlation coefficient of two circularly symmetric random variables is real.*

The above theorem is a well-known result, the proof of which can be found in [11].

Lemma 2.3. *Let* ρ *be the correlation (coefficient) of* $Z_1 \sim CN(0, \Sigma_X^2)$ *and* $Z_2 \sim CN(0, \Sigma_Y^2)$*. Then*

$$I(Z_1; Z_2) = -\log(1 - \rho^2) \tag{2.38}$$

Proof. Let $\mathbf{Z} = (Z_1, Z_2)$. The covariance matrix of $\mathbf{Z}$ is

$$\Sigma := Cov(Z_1, Z_2) = \begin{bmatrix} \Sigma_{Z_1}^2 & \rho \Sigma_{Z_1} \Sigma_{Z_2} \\ \rho \Sigma_{Z_1} \Sigma_{Z_2} & \Sigma_{Z_2}^2 \end{bmatrix}.$$

The proof is obvious after applying theorem 2.7. □

Remark 2.6. *From probability theory [8], the definition of independence of two random variables extends to complex random variables. i.e. Z_1, Z_2 are independent $\Leftrightarrow$ $f_{Z_1,Z_2} = f_{Z_1} f_{Z_2}$. As such, theorem 2.4 also applies to complex numbers.*

Theorem 2.10. *Two Gaussian r.v.s are uncorrelated if and only if they are independent.*

Proof. Obvious from remark 2.6 and lemma 2.3 □

2.1.4 Channel Capacity

Mutual information is the mathematical tool that allowed Shannon to determine the channel capacity, i.e. the maximum transmission rate at which reliable communication over a channel is possible. A channel can be interpreted in different ways depending on the field of study. Physically, a channel is defined as the physical medium that carries information from the sender to the receiver. Through ray tracing a geometrical representation of the channel is possible by tracing the paths taken by the transmitting signal. For the design of coding schemes for reliability, computing the channel capacity necessitates a probabilistic model that inputs and outputs a single symbol.

Definition 2.15. Let $\mathscr{X}$ and $\mathscr{Y}$ be two alphabets. A discrete memoryless channel $(\mathscr{X}, \mathscr{Y}, P_{X|Y})$ inputs a symbol x from the alphabet $\mathscr{X}$ and outputs a symbol y from the alphabet $\mathscr{Y}$ with transition probability[2] function $P_{X|Y}(x|y)$.

The term "discrete" accounts for the fact the channel acts on one symbol at a time. The channel is called memoryless because the transition probability is not dependent on previous inputs or outputs. A memoryless channel is a valid assumption for narrowband communications that will be encountered in this thesis. Narrowband communications are described in section 2.3.

The communication channel is part of a communication system as viewed in figure 2.2. Variable W_k denotes the source message which is a binary sequence of length k that the transmitter wishes to send to the receiver. Before transmission, redundancy is added to the source message by the encoder. The encoder, also known as channel coding, maps W_k to a codeword $X^n = [X_1, \ldots, X_n]$ which is a sequence of n symbols. Since the channel inputs one symbol at time, the transmission of X^n requires n channel uses.

[2] For the case of continuous alphabets, a discrete memoryless channel is described by the triplet $(\mathscr{X}, \mathscr{Y}, p_{X|Y})$, where $p_{X|Y}$ is the channel's transition $p.d.f$. To avoid duplicating definitions, we merge the two notations into $(\mathscr{X}, \mathscr{Y}, P_{X|Y})$ for either the discrete-input or continuous-input case. That is, the phrase "probability function" may as well refer to "probability density function".

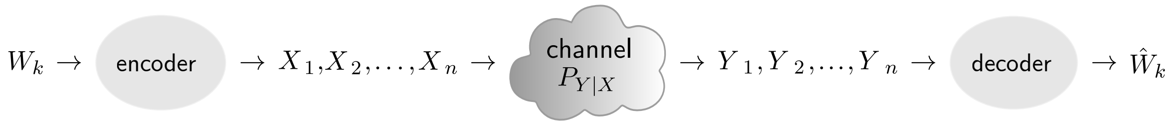

$W_k \rightarrow$
encoder
$\rightarrow X_1, X_2, \ldots, X_n \rightarrow$
channel
$P_{Y|X}$
$\rightarrow Y_1, Y_2, \ldots, Y_n \rightarrow$
decoder
$\rightarrow \hat{W}_k$

Figure 2.2: A discrete memoryless channel changes input symbol $X_i = x_i$ to symbol $Y_i = y_i$ with transition probability $P_{Y|X}(y_i|x_i)$.

The channel successively outputs n symbols and the receiver attains sequence $Y_n := [Y_1, \ldots, Y_n]$ which may be different to X^n due to the transition probability $P_{X|Y}$ that determines the channel. The receiver feeds sequence Y_n to the decoder that outputs $\hat{W}_k$. If efficient redundancy was introduced during the encoding scheme, the decoder will successfully output the source message $\hat{W}_k = W_k$. Otherwise, if $\hat{W}_k \neq W_k$ the transmission of message W_k is unsuccessful.

Definition 2.16. A communication system is said to be reliable if

$$\lim_{k \to \infty} P(W_k \neq \hat{W}_k) = 0. \tag{2.39}$$

Definition 2.17. Let W_k be mapped to $[X_1, \ldots, X_n]$ by the channel encoder. The information transmission rate for message W_k is

$$R = \frac{H(W_k)}{n}. \tag{2.40}$$

The metric for the transmission rate is bits per channel use.

In his paper [7], Shannon proved that there is a maximum transmission rate that is called the channel capacity. Transmitting with a rate higher than the channel capacity, the error probability $P(W \neq \hat{W})$ will be bounded away from zero. However, for rates below the channel capacity, reliable communication is possible.

Theorem 2.11 (Channel capacity)**.** *The capacity of a memoryless channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ is*

$$C = \sup_{P_X} I(X;Y) \tag{2.41}$$

where the supremum is taken over all input probability functions[2] for variable X.

2.1.4.1 AWGN channel

The most basic channel model in wireless engineering is the Additive White Gaussian Noise (AWGN) channel which is described in an input/output format as per the next definition. It forms the basis for the fading channel that will be examined in section 2.3.

Definition 2.18. The AWGN channel is a memoryless channel of continuous alphabet for which the output symbols are given by

$$Y_i = hX_i + N_i, \tag{2.42}$$

where $h \in C$, and $N_i \sim CN(0, \sigma^2)$ is identically and independently distributed for all i and independent of X_i.

The variance σ^2 is also known as the (average) noise power, whereas $E\left(X_i^2\right)$ is the (average) symbol power, or signal power. When a system is said to be P-power constrained, we mean that every transmitting codeword $X_n = [X_1, X_2, \ldots, X_n]$ satisfies the inequality of

$$|X_n|^2 := \sum_{i=1}^{n} |X_i|^2 \leq nP. \tag{2.43}$$

Theorem 2.12. *The channel capacity of the AWGN channel* (2.42) *with power restriction as per equation* (2.43) *and noise power* σ_N^2 *is*

$$C = \log\left(1 + \frac{P|h|^2}{\sigma_N^2}\right). \tag{2.44}$$

For a complete proof, the reader is referred to [12, Appx. B.4.2.1] whereby it is demonstrated that the channel capacity is achieved for the complex normal distribution of X with variance P, and so $Y \sim CN(0, |h|^2P + \Sigma_N^2)$. For the reader's perusal, given that $X \sim CN(0, P)$ is the optimal input distribution, then $I(X;Y) := H(Y) - H(Y|X) = H(Y) - H(N) \overset{2.33}{=} \log(\pi e(P|h|^2 + \sigma_N^2) - \log(\pi e \sigma^2) = \log((P|h|^2 + \sigma_N^2)/\sigma_N^2) = \log(1 + P/\sigma_N^2)$.

Definition 2.19. The Signal-to-Noise-Ratio (SNR) power ratio of the power limited AWGN channel is defined as

$$\text{SNR} = \frac{|h|^2 P}{\sigma_N^2} \tag{2.45}$$

With the above definition, The capacity of the AWGN channel (2.44) takes the memorable expression of $C = \log(1 + SNR)$. In practice the channel inputs come from discrete alphabets defined by the so-called constellation diagrams [12]. The capacity of the AWGN channel is used as a reference for comparing the performance of practical transmission schemes in terms of the transmission rate.

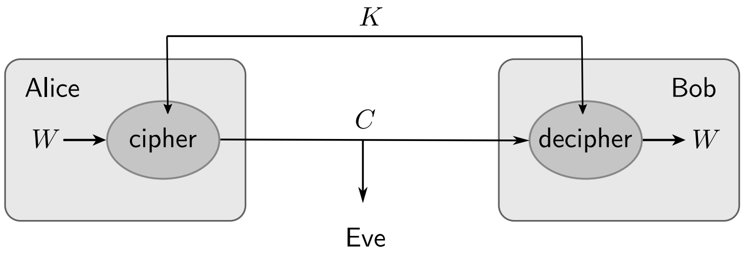


Figure 2.3: Shannon's cipher model [1].

2.2 Fundamentals of cryptographic systems

2.2.1 Shannon's cipher model

Cryptography is based on Shannon's cipher system for confidentiality that comprises three entities, namely, Alice, Bob, and Eve as seen in Figure 2.3. Alice wishes to send a confidential message, W, to Bob in the presence of the eavesdropper, Eve. To encrypt (hide) the message, Alice uses a cipher which is an injective function that inputs W and a random sequence, K. The latter is assumed to be known only by Alice and Bob and it is referred to as the key. Message W is also known as the plaintext. The output, or the ciphertext, C, is transmitted through a noiseless channel. Both Bob and Eve receive the ciphertext. To decrypt (unhide) the data, Bob inputs C and K in the decipher, i.e. the inverse encrypting function and attains W. Eve is aware of the cipher/decipher but, without the key, the plaintext remains hidden from Eve.

Definition 2.20. The sets of all possible plaintexts, ciphertexts, and keys, along with the cipher and decipher form a cryptosystem.

2.2.1.1 Perfect Secrecy

Ideally, Eve attains no useful information about the plaintext given her observation. i.e. the uncertainty about W should not be reduced given C.

Definition 2.21. A cryptosystem is said to achieve perfect secrecy when

$$H(W|C) = H(W) \Leftrightarrow I(W;C) = 0. \tag{2.46}$$

Theorem 2.13. *[1, Sec. 10] To achieve perfect secrecy, the length of the key must be at least as long as the length of the plaintext.*

Definition 2.22. Let $W, K \in \{0,1\}^k$, i.e. the confidential message and key are binary sequences of length k. The one-time pad is an additive cipher with output

$$C = W \oplus K, \tag{2.47}$$

where $\oplus$ denotes addition modulo two.

Example 2.2. *Let the binary data and key be* $W = \{11111\}$ *and* $K = \{10110\}$*, respectively. Applying one-time-pad yields the result* $C = \{01001\}$.

Shannon [1] has also proved that when the keys are (truly) random, i.e. when each key K in $\{0,1\}^k$ is chosen with probability $P_K(k) = 2^{-k}$, the one-time pad has perfect secrecy. Furthermore, he has shown that to achieve perfect secrecy, the entropy of the key must be at least as high as the entropy of the confidential message, i.e. $H(K) \geq H(W)$.

With one-time pad, the same key cannot be reused to cipher a second confidential message. To see why, let $C_1 = K \oplus W_1$ and $C_2 = K \oplus W_2$. If the eavesdropper adds her observations, the identical keys cancel out and useful information is leaked about the messages: $C_1 \oplus C_2 = (K \oplus K) + (W_1 \oplus W_2) = W_1 \oplus W_2$, so the key for a one-time pad algorithm cannot be used for two different messages.

2.2.2 Emulating the one-time pad in today's communications

2.2.2.1 Pseudo-keys

The one-time pad was been used in the second world war by undercover agents whereby the keys were agreed on ahead of time. However, providing keys as long as the plaintext is not a feasible solution for today's communications systems that exchange big volumes of data. A practical solution is to relax the requirement for perfect secrecy and replace the random keys with pseudo-random keys, also known as pseudo keys.

Pseudo-random keys appear to be random but are deterministic in nature; They are derived from functions that input a random short key. To characterise a sequence as pseudo-random and be suitable for secrecy purposes, it needs to pass a number of statistical tests. The most common tests used today are Golomb's randomness postulates [13] and the National Institute of Standards and Technology (NIST) Statistical Test Suite [14].

There are many approaches to generating long 'unpredictable' key sequences from short keys, such as using one-way functions or Linear Feedback Shift Register (LFSR) circuits

[15]. LFSRs have very low computational complexity and memory requirements; Thereby, they are found in many applications of telecommunications. With appropriate design, an LFSR can expand the length of a key from m to $2^m - 1$ in which case the output is called an $m-$sequence [13]. For example, feeding a key generator a 64-bit key, a pseudo-random sequence of 9.2×10^{18} bits (or 1.15 exa-bytes) can be attained.

2.2.2.2 Symmetric cryptography

Shannon's cryptographic model is classified as symmetric cryptography due to the fact that both communicating parties have the same key. There are two types of symmetric cryptographic algorithms that are used today, namely, stream ciphers and block ciphers.

Trying to emulate the one-time pad, a modern stream cipher is an additive cipher that XORs the plaintext with the keystream. Stream ciphers are well suited in applications with minimal memory, power consumption, and gate count [16]. As a result, stream ciphers are popular in wireless communications such as mobile communications and local area networks. Common stream ciphers are A5 (GSM networks) [17], E0 (Bluetooth) [16], and RC4 (802.11) [18]. The size of the (truly random) key in the aforementioned examples is typically 64 or 128 bits depending on the secrecy requirements.

Whereas, the encryption in stream ciphers occurs bit by bit, block ciphers input and output blocks of data, i.e. a plaintext of a fixed length is mapped to a ciphertext of the same length. As such, the influence of one bit of plaintext is spread across many bits of ciphertext; This property is called diffusion [1] and makes a cryptosystem harder to break. Block ciphers currently used by NIST are triple-Data Encryption Standard (DES) and Advanced Encryption Standard (AES). AES is perhaps the most widely used block cipher nowadays. Blowfish is another popular cipher with fast encryption/decryption time, and minimal memory requirements [19].

Note that perfect secrecy is not achievable with pseudo-random keys. This is because key expansion algorithms do not increase the entropy of a key ($H(K) < H(W)$). Secrecy has been compromised in the interest of practicability. Assessing the resilience of such systems against security attacks is the study field of cryptanalysis. Block ciphers that input a key of 128 bits are characterised computationally secure. A cryptosystem is said to be computationally secure when there is no known method that breaks it in a reasonable amount of time, e.g. in less than hundreds of years by means of current technology.

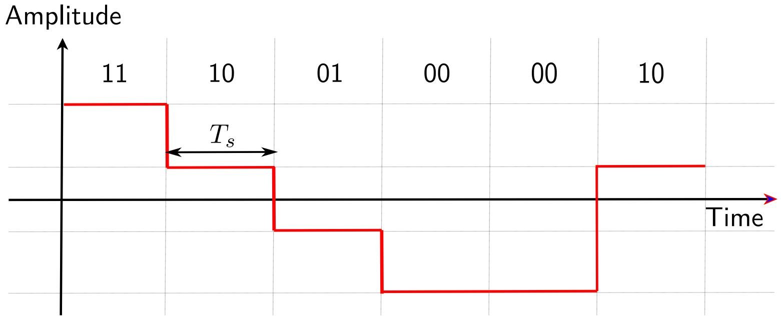


Figure 2.4: Baseband modulation converts a string of symbols into an analogue signal.

2.3 Fundamentals of the wireless channel

2.3.1 The traveling signal

2.3.1.1 At the transmitter

Recall that binary message W^k is mapped to a codeword of n complex symbols, $X_1, X_2, \ldots, X_n$, as seen in figure 2.2. To enable wireless transmission, the string of symbols needs to be converted to a real signal that is continuous in time (analogue). This happens in two stages namely, baseband modulation and carrier modulation.

Baseband modulation inputs discrete data $X_1, \ldots X_n$ and outputs a continuous complex function of time t, $x_b(t)$, referred to as the continuous baseband signal, or low pass signal. Figure 2.4 is a rectangular representation of the real (or imaginary) part of a baseband signal; The width of each rectangle represents the symbol period T_s. In practice, the continuous baseband signal does not have sharp corners but consists of ‘bell-shaped’ curves. The larger the frequency bandwidth is available, the more these bell-shaped curves resemble a rectangular shape and, subsequently, the closer they can be placed to one another in the time axis. Therefore, a larger bandwidth is equivalent to a shorter symbol period, T_s. Specifically, from Nyquist theorem [20], if W hertz is the available bandwidth, complex symbols can be spaced $1/W$ seconds apart.

In the next stage of carrier modulation, the low-pass signal is up-converted to a higher frequency, f_c, referred to as the carrier frequency. The carrier frequency and its corresponding wavelength, denoted as λ, are linked by the equation $f_c = c/\lambda$, where $c = 3 \times 10^8$m/s represents the speed of light. The real and imaginary components of the baseband signal can be simultaneously modulated using various modulation techniques, including Amplitude Shift Keying (ASK), Phase Shift Keying (PSK), and Quadrature Amplitude Modulation (QAM) [21].

ASK modulates the amplitude of the carrier signal to represent digital data. PSK, on the other hand, modulates the phase of the carrier signal to convey information, typically using different phase angles for each symbol. QAM combines both amplitude and phase modulation. It represents symbols on a complex plane, with amplitude dictating the distance from the origin and phase indicating the angle. Different QAM constellations, such as 16-QAM or 64-QAM, offer varying levels of spectral efficiency and noise resilience.

Mathematically, when performing quadrature modulation, the real part of $x_b(t)$ is multiplied by $\cos(2\pi f_c t)$, and the imaginary part is multiplied by $-\sin(2\pi f_c t)$. The transmitted passband signal, denoted as $x_p(t)$, is the sum of these two products:

$$x_p(t) = \Re(x_b(t))\cos(2\pi f_c t) - \Im(x_b(t))\sin(2\pi f_c t) \tag{2.48}$$

Note that the passband signal is real-valued because it is the summation of two real-valued components. By Euler's formula, the passband signal can be alternatively written as

$$x_p(t) = \Re\left(x_b(t)e^{j2\pi f_c t}\right). \tag{2.49}$$

2.3.1.2 The physical medium

The passband signal travels through the wireless medium in the form of electromagnetic (EM) energy and it is subject to attenuation, diffraction, reflection, and refraction [22, Ch. 2]. As a result, the signal reaches the receiver through different paths. The received signal $y_p(t)$ is the aggregation of multiple attenuated and delayed copies of the transmitted signal $x_p(t)$ and additive noise:

$$y_p(t) = \sum_{l=1}^{L} \alpha_l(t) s_p(t - \tau_l(t)) + n(t) \tag{2.50}$$

In this equation, τ_l and α_l represent the delay and attenuation associated with the l^{th} signal path, respectively. The parameter L denotes the total number of significant paths.

The root mean square (rms) delay spread, denoted as τ_{rms}, is the most common metric to quantify the delay spread of a multipath signal. It is defined as [23]:

$$\tau_{\text{rms}} = \sqrt{\frac{\sum_{l=1}^{L} (t_l - t_\alpha)^2 \, \alpha_l^2(t)}{\sum_{l=1}^{L} \alpha_l^2(t)}}, \tag{2.51}$$

where τ_α represents the time for half the power to arrive:

$$\tau_a = \sqrt{\frac{\sum_{l=1}^{L} \tau_l \alpha_l^2(t)}{\sum_{l=1}^{L} \alpha_l^2(t)}}. \tag{2.52}$$

The correspondence of the delay spread with the frequency domain is the concept of coherence bandwidth, B_c. This parameter represents the bandwidth over which the channel is considered flat, implying that the channel's influence on the transmitted signal remains relatively constant across the bandwidth of the signal. Often the coherence bandwidth is calculated as $B_c = 1/(6\tau_{\text{rms}})$, although definitions may vary depending on the geometric environment under study. In all instances, the coherence bandwidth is inversely related to the delay spread.

Of particular interest in this thesis is the concept of narrowband communications, characterised by the condition where the bandwidth of the signal is far less than the coherence bandwidth. For narrowband communication, the equation (2.50) can be simplified as follows:

$$y_p(t) = \sum_{l=1}^{L} \alpha_l(t) x_p(t) + n(t). \tag{2.53}$$

2.3.1.3 At the receiver

The receive (rx) antenna is 'tuned' in the carrier frequency f_c and receives signal $y_b(t)$. Note that even though the passband signal is real, it may contain complex information. We encounter the case of QAM. Referring to (2.49), the transmitted signal $x_p(t)$ is linked with its baseband equivalent, $x_b(t)$, as $x_p(t) = \Re\left(x_b(t)e^{j2\pi f_c t}\right)$. Substituting the latter into (2.53), the received signal takes the form:

$$y_p(t) = \Re\left(\sum_{l} \alpha_l(t) e^{-j\phi_l(t)} x_b(t) + n'(t)\right), \tag{2.54}$$

where $\phi_l(t) = 2\pi f_c \tau_l(t)$ and $n'(t) = n(t)e^{2\pi f_c t}$. To obtain the baseband (complex) signal, denoted as $y_b(t)$, the received signal is down-converted to attain [12, Ch. 2]:

$$y_b(t) = y_p(t)e^{2\pi f_c t} = \sum_{l=1}^{L} a_l(t) e^{-j\phi_l(t)} x_b(t) + n'(t) \tag{2.55}$$

The summation term is perceived as a single complex number by the receiver which is referred to as the channel coefficient:

$$h(t) := \sum_{l=1}^{L} a_l(t) e^{-j\phi_l(t)}. \tag{2.56}$$

Definition 2.23. The received baseband continuous channel of (2.53) in a narrowband communication system is

$$y_b(t) = h(t)x_b(t) + n'(t). \tag{2.57}$$

2.3.2 On the Statistics of the Channel Coefficient

The channel coefficient $h[i]$ captures the phenomena of fading. Two types of fading exist, namely large-scale and small-scale fading. Large-scale fading varies slowly as it represents the average power attenuation of the signal due to motion over extensive distances [22][3]. In contrast, small-scale fading refers to the rapid fluctuations in signal strength that occur over short distances or time intervals.

The two significant contributors to large-scale fading are shadowing and path loss. Shadowing occurs when obstacles, such as buildings, hills, or large objects, obstruct the direct path between the transmitter and receiver. This obstruction leads to a sustained decrease in signal strength, creating shadow regions with weaker signals. Path loss refers to the reduction in signal strength as it traverses free space or encounters obstacles. Contributing to overall signal attenuation, path loss results in a gradual decrease in signal strength with increasing distance.

Small-scale fading refers to the rapid fluctuations in signal strength that occur over short distances or time intervals. It is characterised by variations in both signal amplitude (power) and phase. Small-scale fading is caused by the interference of multipath components. Moving from constructive interference to destructive interference of two multipath components only requires a small change in the length of one of the paths (e.g., of the order of a wavelength). Hence, even a small movement may result in a significant change in the strength of the received signal. Small-scale fading is a facilitator of physical layer security, as will be studied in later sections. The channel coefficient will be considered a wide-sense stationary random process, meaning that the mean and auto-correlation of the random variable $h[i]$ are fixed for all i.

Definition 2.24. The fading channel in its discrete format is given by

$$y[i] = h[i]x[i] + n[i], \tag{2.58}$$

where $n[i] \sim CN(0, \sigma_N^2)$ is the noise term that changes independently from one channel use to the other. Channel coefficients $h[i] := h(i/W)$, $i = 1, 2, \ldots$ are identically distributed over some probability distribution with auto-correlation function R_{hh}.

[3] The specific range of large distances depends on the carrier frequency, spanning from a few meters to several thousand meters.

Note that (2.58) is the discrete equivalent of (2.57). Since $h[i]$ is a complex number, it changes symbol $x[i]$ in two ways; the amplitude of symbol $x[i]$ is attenuated by factor $|h[i]|$, whereas the phase of the symbol has been shifted by $\angle h[i]$.

Definition 2.25. The coherence time of the channel, denoted by T_c, is defined as the smallest value $i > 0$ for which auto correlation function $R_{hh}[i] := E\left(h^*[i]h[i+j]\right)$ differs significantly from $R_{hh}[0]$.

The word 'significantly' will not be specified further. It will be assumed that channel coefficients $h[i], h[i+j]$ are independent random variables as long as the time difference of $h[i]$ and $h[i+j]$ significantly exceed the channel's coherence time, T_c. A further simplification to ease analysis is the assumption of a block-fading channel.

Definition 2.26. Let the sequence of transmitted symbols be partitioned in blocks of size n such that the j^{th} block, $X[n]$, is an array of n symbols: $X_j = \left[x[j], x[j+1], \ldots x[j+n]\right]$. A channel is said to be block-fading when the channel coefficient changes independently from one block to the other. i.e. when the output is also partitioned in blocks such that $Y_j = h[j]X[j] + N_j$, then $h_j \in \mathbb{C}$ is a i.i.d. random variable of some distribution.

A good fit for (small-scale) fading in a rich scattering environment is the Rayleigh channel model.

Definition 2.27. When the channel coefficient in equation (2.58) follows a circular symmetric Gaussian distribution, denoted as $h \sim CN(0, \sigma^2)$, the fading is said to be Rayleigh. In this case both the real and imaginary parts of h are i.i.d., each following the normal distribution $N(0, 1/2)$.

Since the channel coefficient of a Rayleigh channel is circular symmetric, its amplitude $|h|$ is Rayleigh distributed. The p.d.f. of the Rayleigh distributed $|h|$ is given by

$$f(|h|;\sigma) = \frac{|h|}{\sigma^2} e^{-|h|^2/2\sigma^2}. \tag{2.59}$$

It is reminded from remark 2.5 that the phase of the channel coefficient h is independent of the amplitude $|h|$ and is uniformly distributed in $[0, 2\pi)$. By denoting $\phi := \angle h$, then

$$f(\phi) = 1/2\pi. \tag{2.60}$$

Statistically, a complex circular symmetric random variable h arises from a sufficiently large number of independent random variables (multipath components). For a finite number of paths, the complex gains need to be of similar amplitude in order for a Rayleigh channel to be a good fit for the resulting channel coefficient.

2.3.2.1 Channel State Information

Channel State Information (CSI) is a term commonly used in PLS in combination with the words perfect knowledge or statistical knowledge. For the case of narrowband communications, perfect knowledge of CSI, or simply CSI, means knowledge of the channel coefficient $h[i] \in C$, for all i. Therefore, for every symbol transmission, the phase shift and attenuation resulted from the wireless channel are known. Statistical knowledge of CSI refers to having some knowledge of the statistics of the random variable $h[i]$, such as its p.d.f. or its covariance matrix.

When two nodes, A and B, communicate over a wireless channel using the same carrier frequency, the channel from A to B, $h_{AB}[i]$, and the channel from B to A, h_{BA} are identical, i.e. $h_{AB} = h_{BA}$; a property that is referred to as channel reciprocity. This holds true because the signal travels through the same paths from A to B and from B to A. Given that the frequency is the same, the identical paths will result in the same attenuation and phase shift.

Pilot-based channel training is the most common method for acquiring CSI in the field of physical layer security. To acquire CSI at node A with pilot-based channel training, node B sends a pilot sequence. Also known as training sequence, a pilot sequence is a sequence of known symbols $p[1], p[2], \ldots, p[k]$. Node A receives $y[i] = h[i]p[i] + n[i]$, where $n[i] \sim CN(0, \sigma_n^2)$ is the noise term, and compares it with $p[i]$. Assuming that the channel varies much slower than the symbol period $1/W$, the channel coefficients $h[i]$ can be thought identical for $i = 1, \ldots, k$.

To eliminate the noise term $n(t)$, various channel estimation techniques exist with maximum likelihood estimation and minimum mean square error [24] being the most widely used. The former is the simplest method and is preferred for point-to-point networks.

2.3.3 Spatial correlation and multipath fading

In a multipath environment, a transmitting signal follows many paths before it reaches the receiver(s). Each multipath component is associated with a complex number $A_i \in \mathbb{C}$, that we shall refer to as *complex path gain* with phase $\angle A_i$ and amplitude $|A_i|$.

When two receivers with a uniform gain pattern are sufficiently close to one another, they observe

$$\rho(u) = \frac{A_1(u)A_2^*(u)}{|A_1(u)||A_2(u)|}, \tag{2.61}$$

where $(\cdot)^*$ is the complex conjugate, and operator $|\cdot|$ is the amplitude of the enclosed complex number. Averaging over all complex path gains, we attain the spatial channel correlation between two receivers. In practical scenarios where the exact geometry of the environment is not known, statistical models are used instead, in which case the spatial channel correlation is the statistical expectation of the path correlation

$$R := E\big(\rho(u)\big), \tag{2.62}$$

which has been studied for several statistical models [25–28].

The spatial correlation is often referred to as Angle-of-Arrival (AoA)-statistics since it depends on the distribution of the unit vector $u \in S$. Vector u is often expressed in spherical coordinates $(1, \alpha, \beta)$, where α is the polar angle and β is the azimuthal angle. The term AoA refers to the pair (α, β).

Although our scheme can be applied to any multipath channel model, for reasons of exposition we focus on Rayleigh channels, for which the spatial correlation takes a closed form. We remind the reader that a Rayleigh (fading) channel is a rich-scattering channel for which:

- the phases $\angle A(u)$ are uniformly distributed across $[0, 2\pi]$ and are independent for different $u \in S$;
- the amplitudes $|A(u)|$ are identically and independently distributed for different $u \in S$.

The summation of all complex path gains as observed at the receiver results in a Rayleigh channel coefficient (or Rayleigh channel for brevity), h, the phase of which is also uniformly distributed, whereas its amplitude is a Rayleigh distributed random variable [12].

Let d be the distance between the tag and the reader, and let λ be the wavelength of the carrier frequency. If the unit sphere, S, lives in the three-dimensional (3D) space, the spatial correlation can be expressed as a function of the distance [29]:

$$R = \operatorname{sinc}\left(\frac{2\pi d}{\lambda}\right) \qquad \text{(3D Rayleigh)}, \tag{2.63}$$

where $\operatorname{sinc}(x) = \sin(x)/x$, when $x \neq 0$, and $\operatorname{sinc}(0) = 1$. Such a model can be a good fit for indoor environments in which the ceiling and the floor act as good reflectors creating a 3D diffuse field. Eq. (2.63) is the basis of the rule-of-thumb stating that the channel decorrelates in half a wavelength.

The second most common geometry model restricts the AoA in one plane and the sphere S lives in the two-dimensional (2D) space. It is usually applied for rural environments, or when the antennas are vertically orientated and receive in the azimuthal plane. In this case, the spatial correlation can be expressed as [30]:

$$R = J_0\left(\frac{2\pi d}{\lambda}\right) \qquad \text{(2D Rayleigh)}, \tag{2.64}$$

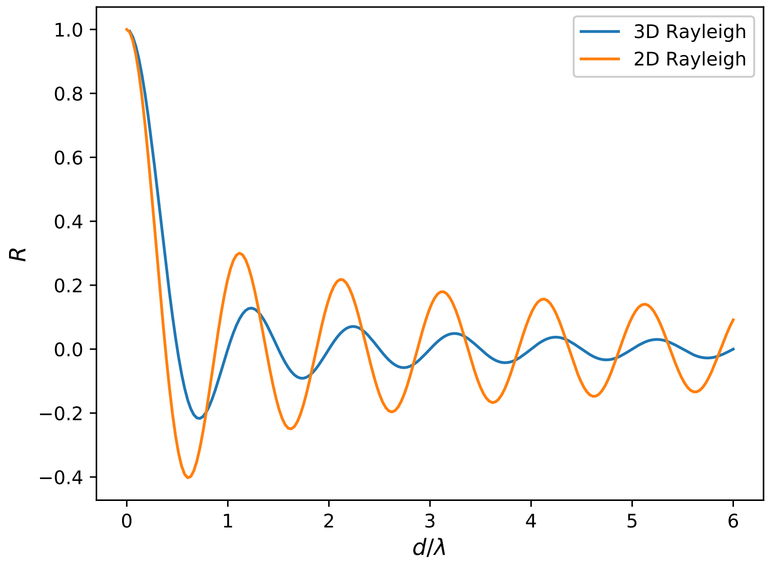


Figure 2.5: Spatial correlation against the distance normalised to the wavelength.

where J_0 is the zeroth-order Bessel function of the first kind [30]. This formula (2.64) is popular because it gives a good approximation for 3D diffuse fields as long as one of the spherical angles (e.g. the polar angle) takes values from a limited range [25]. Figure 2.5 plots the spatial correlation against the distance for two channel models. It can be seen that the first zero correlation for the 2D case occurs at 0.38 wavelengths, which translates to approximately 12 cm when the carrier frequency is 915 MHz. Observe that the spatial correlation is an oscillating function of distance.

Chapter 3

Literature Review

In the last few decades Physical Layer Security (PLS) has developed into a multi-disciplinary field that captures a wide range of techniques and security purposes. Techniques based on Wyner's work on secrecy coding for confidentiality are classified as keyless PLS, whereas Maurer's work on physical layer key generation (PLKG) has given rise to key-based PLS that aims to solve the key distribution problem. In recent years, the physical layer is also exploited for authentication purposes by exploiting not only the physical medium but also the unique Radio-Frequency (RF) characteristics associated with the hardware of a device. This chapter presents the advances, challenges and limitations of keyless PLS, key-based PLS, as well as Physical Layer Authentication (PLA) techniques specific to short-range communication systems.

Keywords: secrecy coding, secrecy capacity, physical layer key generation, secret key generation, short-range communications, distance fraud.

3.1 Keyless PLS for confidentiality

In 1975, Wyner [9, 1975] introduced an approach for securing communications without the need for cryptographic keys. He proved that, with appropriate coding, noisy channels can be exploited to 'hide' messages from unintended receivers. This realisation gave birth to PLS.

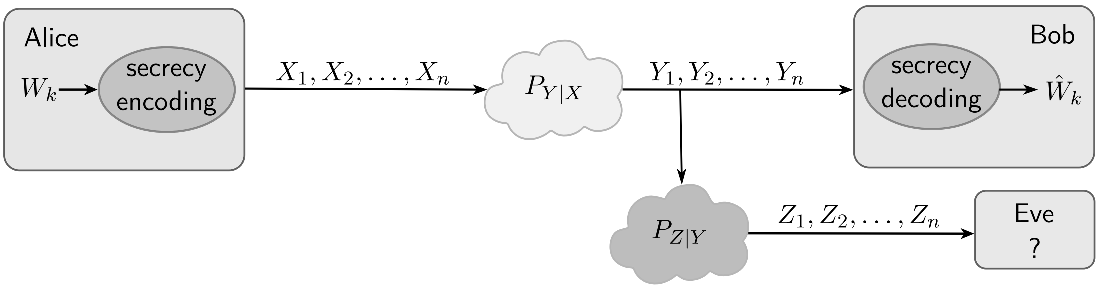


Figure 3.1: The wiretap channel as introduced by Wyner

3.1.1 The wiretap channel

This section defines the wiretap channel, its metrics, and traces its evolution. While Wyner's wiretap channel model laid the groundwork for physical layer security and perfect secrecy, it encounters limitations in wireless communications where eavesdroppers may have better signal quality than intended recipients. This chapter explores the evolution of practical channel models up to the current date of the thesis.

3.1.1.1 Wyner's wiretap channel

Definition 3.1. A three-node discrete wiretap channel $(\mathscr{X}, \mathscr{Y}, \mathscr{Z}, P_{Y|X}, P_{Z|Y})$ is the cascade of two discrete memoryless channels $(\mathscr{X}, \mathscr{Y}, P_{Y|X})$ and $(\mathscr{Y}, \mathscr{Z}, P_{Z|Y})$ (figure 3.1), which are referred to as the main channel and the wiretap channel, respectively. The main channel is the channel between the transmitter (Alice) and the intended receiver (Bob), whereas the wiretap channel is the channel between the transmitter and the unintended receiver (Eve).

Remark 3.1. *By clash of definitions, the phrase 'wiretap channel' may refer to the three-node channel model or the eavesdropper's channel in the literature. To avoid confusion, this thesis will refer to the former as the 'three-node wiretap channel', whereas 'wiretap channel' will refer to the eavesdropper's channel.*

Observe that the output, Y_i, of the main channel, is also the input of the wiretap channel. Conditioned on Y_i, variables X_i and Z_i are independent, or in statistical terminology, variables X_i, Y_i, and Z_i form a Markov chain. In this case, the wiretap channel is said to be a degraded version of the main channel. This is because codeword X 'passes through' two channels before reaching the eavesdropper.

Wyner showed [9] that when Eve has a degraded channel, perfect secrecy (see def. 2.21) can be achieved without the use of secret keys. Randomisation among multiple codewords is the key property of secrecy coding and the main difference from the error-correcting codes that solely aim for reliability [31, Ch. 1]. Randomisation is added in order to confuse the eavesdropper, thereby achieving confidentiality.

Recall that the channel capacity in Shannon's channel model is defined as the maximum rate at which information can be transmitted reliably (see sec. 2.1.4). For the transmission of confidential information, we need to achieve reliability as well as secrecy.

Definition 3.2. Let plaintext W be mapped to a codeword of n symbols after secrecy coding, i.e. a coding scheme that achieves both reliability and secrecy. The secrecy rate, R_s, is defined as

$$R_s = \frac{H(W)}{n}. \tag{3.1}$$

The maximum achievable secrecy rate, denoted by C_s, is called secrecy capacity.

Theorem 3.1. *The secrecy capacity of the discrete wiretap channel* $(\mathscr{X}, \mathscr{Y}, \mathscr{Z}, P_{Y|X}, P_{Z|Y})$ *is*

$$C_s^{DWC} = \max_{P_X}[I(X;Y) - I(X;Z)], \tag{3.2}$$

where the maximisation is over all possible probability distributions of codeword X.

As long as Eve has a noisy channel, the difference $I(X;Y) - I(X;Z)$ is positive due to the nature of cascaded channels. I.e, variable Z contains less information than Y about X, or equivalently, $X - Y - Z$ forms a Markov chain. Exempting the trivial case of a noiseless eavesdropper, the difference $I(X;Y) - I(X;Z)$, and subsequently the secrecy capacity is (strictly) positive. The difference in the mutual information is referred to as the secrecy gap.

Designing a code that achieves both reliability and perfect secrecy is a hard problem, so, Wyner relaxed the requirement of perfect secrecy to strong secrecy, and weak secrecy. Weak secrecy requires the information leaked per symbol to be asymptotically zero, whereas, with strong secrecy, the total leaked information (per codeword) is asymptotically zero.

Definition 3.3. (Weak secrecy) A system is said to achieve weak secrecy if

$$\lim_{n\to\infty} \frac{1}{n}(I(W^k; Z^n)) = 0 \tag{3.3}$$

Definition 3.4. (Strong secrecy) A system is said to achieve strong secrecy if

$$\lim_{n\to\infty} (I(W^k; Z^n)) = 0 \tag{3.4}$$

The difference between weak secrecy and strong secrecy is that the latter requires the total leaked information to be asymptotically zero.

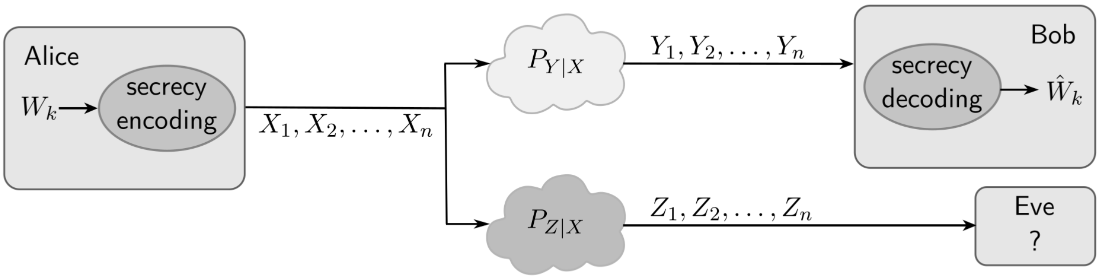


Figure 3.2: The generic wiretap channel as introduced by Csiszár and Körner.

3.1.1.2 The generic wiretap channel

Cascaded channels can be found in wired communications but not so often in wireless communications. Due to the broadcast nature of wireless communications, when Alice transmits X_i, both Bob and Eve will, most likely, observe two different noisy copies of X_i. To better match the characteristics of wireless networks, Csiszár and Körner [32] generalised Wyner's model to include non-degraded wiretap channels (figure 3.2), i.e. the case when $X_i - Y_i - Z_i$ does not necessarily form a Markov chain.

Definition 3.5. The generic wiretap model, $(\mathscr{X}, \mathscr{Y}, \mathscr{Z}, P_{Y|X}, P_{Z|X})$ comprises two discrete memoryless channels $(\mathscr{X}, \mathscr{Y}, P_{Y|X})$ (main channel) and $(\mathscr{X}, \mathscr{Z}, P_{Z|X})$ (wiretap channel).

Theorem 3.2. *The secrecy capacity of the generic discrete wiretap channel is given by*

$$C_s^{GDWC} = \max_{P_{V,X}} [I(V;Y) - I(V;Z)] \tag{3.5}$$

where the maximisation is over the probability distributions $P_{V,X}$ such that $V - X - (Y, Z)$ form a Markov chain, i.e. variables V and (Y, Z) are independent conditioned on X.

Variable V is called channel prefixing and introduces Artificial Noise (AN) in the system. The purpose of artificial noise is the degradation of the wiretap channel (the eavesdropper's channel). Only when the secrecy gap is positive ($I(V, X) - I(V, Z) > 0$), the secrecy capacity is also positive.

3.1.1.3 Continuous Wiretap channels

The Gaussian wiretap Channel S. Leung [33] extended Wyner's results for the discrete memoryless channel to the Gaussian Wiretap channel.

Definition 3.6. A three-node Additive White Gaussian Noise (AWGN) wiretap channel comprises two AWGN channels for which the output symbols are given by:

$$Y_i = h_B X_i + N_{B,i} \tag{3.6}$$

$$Z_i = h_E X_i + N_{E,i}, \tag{3.7}$$

where $h_B, h_B \in C$, and $N_{B,i} \sim CN(0, \sigma^2_{n_B})$, $N_{E,i} \sim CN(0, \sigma^2_{n_E})$ are independentally and identically distributed (i.i.d.) complex Gaussian additive noise.

Theorem 3.3. *The channel capacity of the AWGN wiretap channel with power restriction P (as per equation* (2.43)*) is given by*

$$C_S := I(X;Y) - I(X;Z) = \log\left(1 + \frac{P|h_B|^2}{\sigma^2_{n_B}}\right) - \log\left(1 + \frac{P|h_E|^2}{\sigma^2_{n_E}}\right). \tag{3.8}$$

The proof of theorem 3.3 is similar to the proof provided in [33]. The only difference is that real signals are considered in [33] which results in a 50% reduction in the secrecy capacity.

Remark 3.2. *For the case of a Gaussian channel, the requirement of a positive secrecy gap is equivalent to* $\frac{|h_B|^2}{\sigma^2_{n_B}} > \frac{|h_E|^2}{\sigma^2_{n_E}}$.

The fading wiretap channel In Leung's channel model [33], multiplicative gains h_B and h_E remained constant throughout the transmission of the (discrete) signal $\{X_i, i \in [n]\}$. If the change during the transmission of the discrete signal $\{X_i, i \in [n]\}$, the channel is characterised as a fading channel.

Definition 3.7. A three-node fading wiretap channel comprises two channels for which the output symbols are given by:

$$Y_i = h_{B,i} X_i + N_{B,i} \tag{3.9}$$

$$Z_i = h_{E,i} X_i + N_{E,i}, \tag{3.10}$$

where $h_{B,i} \in \mathbb{C}$, $h_{E,i} \in \mathbb{C}$, and the noise terms are defined as in 3.11.

If $h_{B,i}$ and $h_{E,i}$ are independent realisations of random variables $h_B \sim f_B$ and $h_E \sim f_E$ of some distribution functions f_B and f_E, the channel model is further characterised as ergodic.

It took more than two decades for extending the results for the Gaussian wiretap channel to the fading channel. With the introduction of multiple antenna systems at the beginning of the 21st century, independent research publications [34–38] gave fruitful insights into physical layer security over fading channels.

Theorem 3.4. *The secrecy capacity of the ergodic fading wiretap channel with power restriction P (as per equation* (2.43)*) is given by*

$$C_S = \max_{E_A[\gamma \leq P]} E_A \left(\log \left(1 + \frac{P|h_B|^2}{\sigma_{n_B}^2} \right) - \log \left(1 + \frac{P|h_E|^2}{\sigma_{n_E}^2} \right) \right), \tag{3.11}$$

where E_A is the expectation taken over the set $A := \{(h_B, h_E) : \frac{|h_B|^2}{\sigma_B^2} > \frac{|h_E|^2}{\sigma_E^2}\}$. The maximisation is over the power allocation, γ, such that the power constraint is satisfied.

Theorem 3.4 can be found in [37] for the case of real signals. Because of the relation between the entropy of a real and a complex random variable (r.v.) (see definition 2.10)), substituting $H(X) = 1$ with $H(X) = 2$ (similarly for $H(Y)$) gives the proof for the complex case.

To achieve a positive secrecy capacity or even a positive secrecy rate, global Channel State Information (CSI) is required at the transmitter. That is, Alice needs to know the channel quality of both the legitimate channel and the wiretap channel in all instances, i.e. all realisations of random variables h_B and h_E [37]. In practice, it is difficult to attain global CSI [31, 39]; The eavesdropper does not wish to reveal any information to Alice and remains silent. Without any transmissions from the eavesdroppers, Alice cannot track the realisations of random variable h_E.

To address the limitation imposed by the requirement of a positive secrecy gap, chapter 5 leverages concepts of keyless physical layer security without necessitating global CSI. This approach assumes that the eavesdropper can be mobile, and her signal quality can vary with her location. It introduces the method of secret splitting, which eliminates locations where the eavesdropper can gain useful information about the confidential data, even when equipped with multiple antennas.

3.1.2 Secrecy coding techniques

To give an insight into secrecy coding, we give a simplistic example of a noiseless (main) wiretap channel.

Example 3.1. *Let Alice, Bob, and Eve form a three-node wiretap channel whereby Bob experiences a noiseless channel, whereas Eve observes exactly two erasures in every codeword sent. Alice's message space is $\mathcal{W} = \{00, 01, 10, 11\}$. Her secrecy coding strategy maps each plaintext to a set of codewords as follows:*

$$w_1 = 00 \rightarrow \mathscr{X}_1 = \{0000, 0101, 1010, 1111\}$$
$$w_2 = 01 \rightarrow \mathscr{X}_2 = \{0100, 0001, 1110, 1011\}$$
$$w_3 = 10 \rightarrow \mathscr{X}_3 = \{1101, 0010, 0111, 1000\}$$
$$w_4 = 11 \rightarrow \mathscr{X}_4 = \{1100, 1001, 0110, 0011\}$$

When Alice wants to communicate message w_i to Bob, she randomly chooses a word within the set $\mathscr{X}_i$. Bob is able to decode the received message by simply looking at the codebook. Even if Eve uses the same codebook, she gains no information by observing the received word. For example, assume that Eve observes 11??. There are four possibilities of what Alice could have sent: 1100, 1101, 1110, or 1111. According to section 2.2.1.1 perfect secrecy is achieved since her uncertainty about the message sent is not reduced given her observations, i.e. $H(W|Y) = H(W)$.

The codebook in the above example has been generated using a technique [40] borrowed from error-correction theory. For practical channels, the wiretap main channel is noisy and perfect secrecy is hard to realise due to both reliability and secrecy considerations. Strong secrecy and especially weak secrecy are somewhat easier to realise.

There is a large number of codes for secrecy that are constructed for the case of a discrete wiretap channel. The most popular codes for discrete channels are low-density-parity-check codes (LDPC) codes and polar codes [41]. With their classic work [42, 43], Rathi et al construct LDPC codes that provide weak secrecy in the case of binary erasure channels. For strong secrecy, LDPC seem to require a noiseless main channel [41, 44]. Strong secrecy over noisy main channels can be provided by polar codes as described in [45–47]. However, LDPC typically have a lower encoding and/or decoding complexity when compared to polar codes [48]

Over the past two decades, most research in secrecy coding emphasises continuous wiretap channels. For the case of a Gaussian wiretap channel, lattice codes seem to be a good fit and can achieve both weak secrecy [49] and strong secrecy [49–51]. Li et al [52] have considered the case where the wiretap channel is Rayleigh distributed and the main channel is Gaussian. Achievable secrecy rates have been derived for the cases of Gaussian signalling and quadrature amplitude modulation. Belfiore et Oggier [53] have presented a criterion of design of lattice codes for secrecy when used on parallel wiretap Rayleigh channels. By parallel Rayleigh channels, it is meant that both the main channel and wiretap channel are Rayleigh. Baldi et al [54] have proposed a coding scheme over parallel Rayleigh channels that maximises the secrecy rate subject to a constraint on the maximum outage probability.

3.1.3 Secrecy Coding and Signalling processing

For many years after Wyner's paper [9], the security community doubted the practicality of (keyless) PLS due to the restricting requirement of a strictly positive secrecy gap (see section 3.1.1.1) and, as a result, the industry had little or no interest in PLS. In the last decade, PLS regained attention. Advancements in signalling processing and the employment of multiple-antenna systems can be used in a way that the secrecy gap is increased [55, Chapter 17].

With beamforming for secrecy [55, Chapter 17], the quality of the legitimate channel can be increased by exploiting spatial diversities and multiplexing gains, whereas the generation of AN can degrade the eavesdropper's channel without affecting the legitimate receiver to the same degree [56]. AN is also known as jamming noise or controlled interference in the literature. The aim is to increase the secrecy gap as much as possible in favour of the intended receiver. Beamforming for secrecy requires systems with multiple antennae or user cooperation [55, Chapter 17]. A classic work of secrecy beamforming can be found in [57].

The use of AN was introduced by Goel and Negi in 2005 [58] and many AN-based schemes have followed since. In [59], a closed-form expression was derived for the optimal power allocation over the information-bearing signal and the AN for the case of the Gaussian wiretap channel with multiple non-colluding eavesdroppers whose channels were unknown. In [60] cooperative jamming is exploited for achieving secrecy in downlink transmissions in the Internet-of-Things (IoT) networks against multiple and non-colluding eavesdroppers. A detailed survey on IoT and (keyless) PLS can be found in [61]. Security enhancement through PLS in Wireless Information and Power Transfer networks are considered in [62] and [63]. The former user cooperation for both jamming and relaying against a single eavesdropper and one destination. Many researchers consider game theory as an appropriate tool for solving problems associated with user cooperation in decentralised networks [64, 65]. Lastly, a case study for secrecy beamforming for ultra-reliable and low-latency communications in 5G and beyond is given in [66].

Many AN-based schemes are often based on the assumption of Gaussian-input signalling and they are not effective in current transmission schemes such as phase shift keying and quadrature amplitude modulation [67, 68]. Moreover, the vast majority of secrecy beamforming schemes require knowledge of the eavesdropper's CSI (perfect, imperfect, or statistical CSI) [69]. Schemes that do not require the eavesdropper's CSI exist but they are limited to the case of a single-antenna eavesdropper [66]. A detailed overview of AN-based schemes that examines theoretical and practical limitations can be found in [69].

3.2 Key-based PLS

In a key-based PLS scheme, the legitimate nodes aim to generate symmetric keys while keeping the eavesdropper(s) ignorant. By passing the keys in the upper layers of the protocol stack, key-based security can enhance and complement traditional security by reducing significantly the computational complexity of current cryptographic algorithms [44].

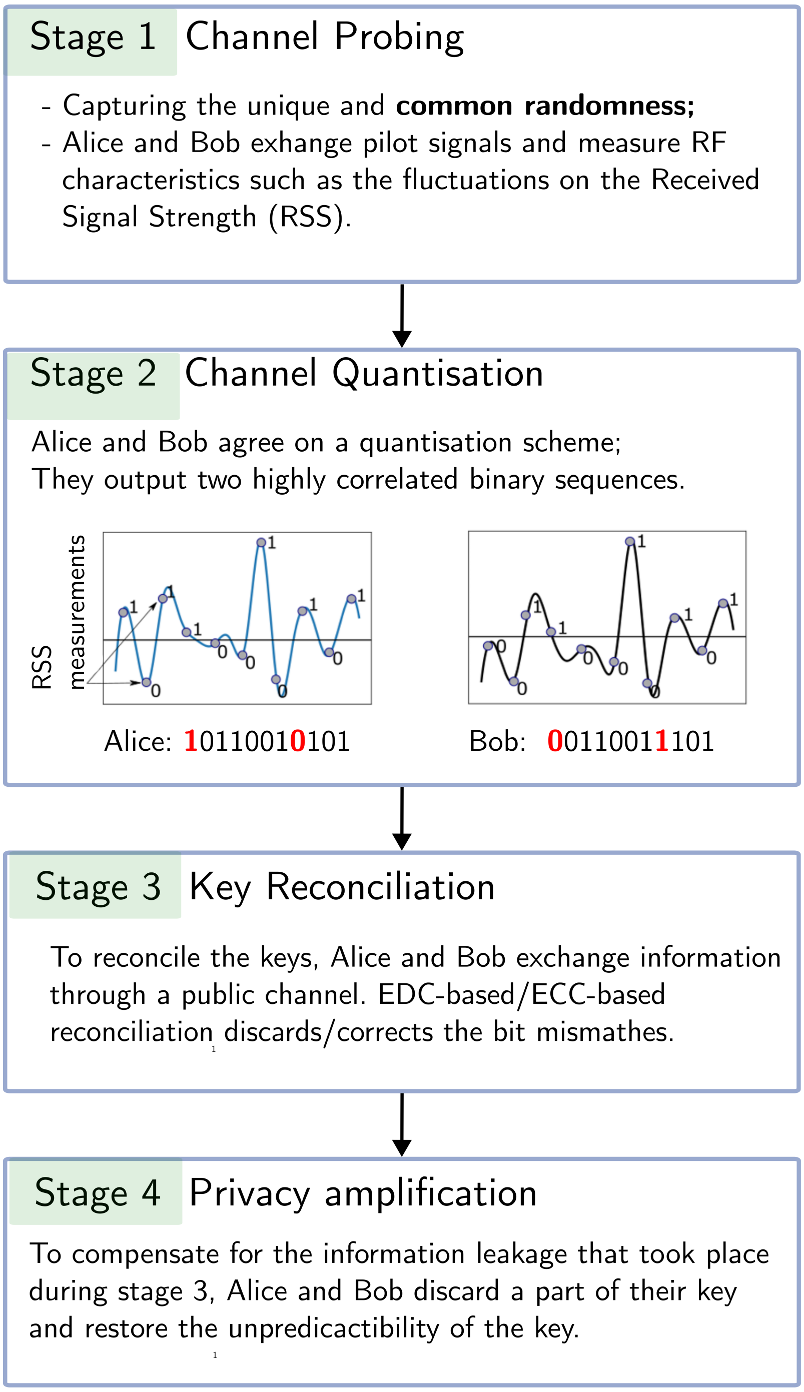
There are two models for key generation: the source model and the channel model [70, Chapter 4]. The source model uses an external source of randomness and requires feedback over a secure parallel channel. Mainly because of the requirement of a secure channel, key generation for the source model seems to have little -if any- practical value and is omitted from this thesis. For practical scenarios, the PLS community has focused on the channel model that relies on channel reciprocity.

3.2.1 Channel-reciprocity based Physical Layer Key Generation

Channel-reciprocity-based PLKG aims to "capture" the inherited randomness existing in the reciprocal channel between two nodes. Channel reciprocity is based on the fact that a signal travelling from Alice to Bob takes the same route as when it travels from Bob to Alice (see section 2.3.2.1). As long as the eavesdropper is not "too close" to a legitimate receiver, she will experience uncorrelated multipath fading. i.e. the reciprocal channel between Alice and Bob can be characterised as a unique source of common randomness.

The maximum key rate per channel realisation of a channel-reciprocity-based KG protocol is equal to $I(\hat{h}_A;\hat{h}_B)$, where $\hat{h}_A$ and $\hat{h}_B$ are the channels observed by Alice and Bob respectively [71].

The next two subsections explain the four main stages of (channel-reciprocity-based) key generation namely, channel probing, channel quantisation, key reconciliation, and privacy amplification [72], as illustrated in figure 3.3.



Stage 1 Channel Probing

- Capturing the unique and **common randomness;**
- Alice and Bob exhange pilot signals and measure RF characteristics such as the fluctuations on the Received Signal Strength (RSS).

Stage 2 Channel Quantisation

Alice and Bob agree on a quantisation scheme;
They output two highly correlated binary sequences.

Alice: **1**0110010**0**101 Bob: **0**0110011**1**101

Stage 3 Key Reconciliation

To reconcile the keys, Alice and Bob exchange information through a public channel. EDC-based/ECC-based reconciliation discards/corrects the bit mismathes.

Stage 4 Privacy amplification

To compensate for the information leakage that took place during stage 3, Alice and Bob discard a part of their key and restore the unpredicactibility of the key.

Figure 3.3: Channel-reciprocity-based key generation protocols comprise four stages.

3.2.1.1 Channel Probing

To perform channel-reciprocity-based key generation, the communicating parties, Alice and Bob, are required to operate in Time Division Duplexing (TDD) mode [1] [73]. To launch a key generation protocol Alice and Bob take turns sending a pilot signal and observing the physical interactions of each other's signal. As detailed in section 3.2.2, the most common channel characteristic under observation is the Received Signal Strength (RSS). However, when CSI measurements are available, they are preferred since they result in much higher key rates.

Ideally, the pilot exchange is realised before any change in the channel such that the observed channels are perfectly reciprocal. Most often, the channel probing phase comprises multiple rounds of a pilot exchange. The time between two pilot exchanges determines the probing rate. The time difference between the transmission of the pilots within a single round is the time lag of the pilot exchange. In reality, the inability to exchange pilots simultaneously results in correlated rather than identical observations. The level of correlation depends on how the time lag of the pilot exchange compares to the channel's coherence time. In a highly dynamic channel, the time lag cannot always be reduced below the channel's coherence time due to hardware imperfections which leads to a low correlation and, subsequently, to an increased key disagreement rate [74].

Whereas a long coherence time is beneficial for a low-key disagreement rate, it is disadvantageous for the key rate. To generate long keys of high entropy, the pilot exchanges need to span multiple channel realisations that are independent. Hence, a static environment is not desirable for key generation purposes.

3.2.1.2 Channel quantisation

After the channel probing phase, Alice and Bob proceed to the quantisation phase whereby the extracted signals are converted to binary strings. This phase involves signalling processing techniques that normalise the signal and quantise it according to a quantisation scheme. Normalising the signal means extracting the small-scale fluctuations of the signal. Various techniques exist for normalisation such as the moving average technique and "neighbouring window". In scenarios where the correlation between Alice's channel and Bob's channel is low, filtering mechanisms [75–77] may apply. Filtering mechanisms are effective in terms of reducing the number of mismatches but they can also reduce the entropy of the key [78].

[1] The operating mode TDD can be found in many communication standards such as IEEE 802.11. and LTE. Also, TDD mode is believed to be the number one candidate for 5G and beyond technologies

After the signal has been filtered and/or normalised, it is sampled and quantised. To ensure a high entropy on the quantised sequence, the sampling rate must be higher than the channel's coherence time. The level of a quantisation scheme typically varies in the range of one to four. An m-level quantisation scheme results in m bits per sample. The higher the level of the quantisation scheme is, the more the key rate increases. However, a high-level quantisation scheme reduces the entropy of the resulting keys and increases the key disagreement rate [73, 79].

3.2.1.3 Key reconciliation

After quantisation, Alice and Bob test whether their sequences are identical by sending a hash function, i.e. a message digest. If not, the binary sequences need to be reconciled before they can serve as two symmetric keys. There are two types of techniques used for reconciling the keys. Error Detection Coding (EDC)-based techniques such as Cascade and Winnow [80] borrow methods learnt from the communication theory, whereas Error Correction Coding (ECC)-based techniques such as LDPC, Bose–Chaudhuri–Hocquenghem (BCH), and turbo codes [81] use concepts from quantum cryptography. Techniques of the first type typically detect and discard the bit mismatches, whereas ECC-based techniques correct the bit mismatches.

When the number of keybits subject to reconciliation is high, EDC-based approaches require testing a significant amount of permutations which may not be viable in low-memory and/or low-power devices [82]. The advantage of EDC techniques is that they leak less information about the key in comparison to Error Correction Coding (ECC) [80]. However, even with an EDC approach, the information leakage increases dramatically as the number of bit mismatches increases [78]. For example, with Cascade [73] - the most popular EDC-based technique – a 1% bit mismatch between the channel sequences leaks 10% of the key, whereas a 10% bit mismatch exposes approximately 60% of the key. The table below provides a high-level characterisation of the two approaches.

Approach	Comm.Overhead	Complexity	Leakage
EDC-based [73, 80]	High	Low	Low
ECC-based [73, 81]	Low	High	High

Table 3.1: A high-level comparison between EDC-based approaches and ECC-based approaches for key reconciliation.

3.2.1.4 Privacy amplification

Quantising the leaked information during the reconciliation phase is essential for restoring the unpredictability of the key. After reconciliation, privacy amplification follows that outputs a shorter, and, ideally unpredictable key [83]. The precise quantification of the leaked information towards an eavesdropper(s) may not be practically feasible [79], and as such, deciding on the compression ratio for the key may not be an easy task. According to Calver's estimations [84] when the Cascade algorithm is used to reconcile two sequences that disagree at a rate of 1%, approximately 10% of the key bits need to be "dropped" during the privacy amplification phase. The number of exposed bits rises dramatically as the bit-mismatch rate increases. For example, with a 10% bit mismatch, 57% ~ 63% of the reconciled sequence is exposed and needs to be discarded in order to result in a secure key [84, 85].

3.2.2 Test-bed experimentation

In the classic work [86], multiple-antenna diversity is exploited for key generation between two parties in the presence of multiple eavesdroppers. Tests were conducted using off-the-shelf 802.11n equipment which resulted in high secrecy rates in both indoor and outdoor environments. In [85] and [87], RSS values were measured from multiple devices in order to generate a shared key collectively. Using multiple nodes for key extraction increases the secrecy rate even if the nodes are considered 'untrusted' [88]. A dynamic key generation scheme based on a proportional integral derivative controller was proposed in [89] according to which the probing rate was "tuned" to the channel conditions. For example, when the channel's coherence time was long, the probing rate decreased.

When the channel between transceivers varies slowly, there is not much randomness to be extracted, and as a result, the key generation suffers from slow rates which are often insufficient. With their channel being additional sources of randomness, relay nodes are employed in [83, 90, 91] to improve the secrecy rates. It was proven in [92] that the multiplexing gain of KG increases linearly as the number of relay nodes increases. KG is possible even if the relay nodes are "untrusted", but the key rate is higher when " trusted" relay nodes are employed. A jamming technique was used in [93] as a tool to change the measured values of the channel states between two users, thus increasing the secrecy key rate. A secure channel between the jammer and the transmitter was required.

A growing interest has been observed for key generation in wireless low-resource devices and especially for health-monitoring devices [94]. Authors in [85] considered reducing the reconciliation cost in body-worn devices. If the keys needed to be refreshed every hour their scheme was successful 93.5% of the time. Authors in [83] and [90] have proposed two KG schemes applicable in wireless sensor networks for the Internet-of-Things, whereas [91] specialises in wireless body area networks.

RSS-based KG is not optimal as it limits the secrecy key rate to one symbol per packet at most. However, it is widely used due to its easy access, i.e. it is available in most off-the-shelf wireless devices. In their work [95], H. Liu et al compared the RSS-based and CSI-based key generation by implementing a key generation scheme based on LDPC coding. They showed that the key rate is much higher when CSI is used. Implementation of CSI-based key generation can also be found in [96], whereby the proposed scheme is analysed in terms of entropy, key disagreement rate, and secrecy key rate. As expected, the mobile scenario was more beneficial than the static one. Experimental results in [95] also showed a remarkable difference in key rates of RSS and CSI methods. By measuring the CSI from Orthogonal Frequency-Division Multiplexing (OFDM) subcarriers, they achieved a 90bit/packet key rate in comparison to the 4bit/packet key rate achieved by previous RSS-based schemes.

3.3 Authentication attacks in Short-Range Systems

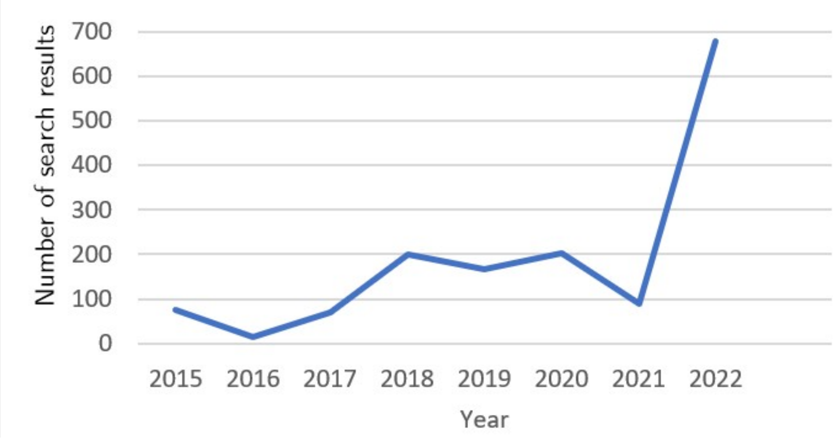


Figure 3.4: Number of search results on Google News over the years. Search string: *“replay attacks” OR “relay attacks”*. Command “before:” followed by a specific date was added to the search string to specify the year.

Short-range radio is a key technology that plays a key role in modern life. Personal area networks, body area networks, contactless payments, and keyless entry systems are used daily by many individuals around the globe. Along with the growing trend of short-range

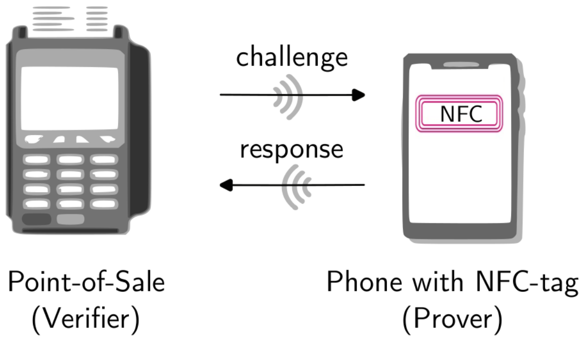


Figure 3.5: Verifier prover example

communications comes a dramatic increase in fraud. Figure 3.4 reflects the level of concern in society about two types of impersonation attacks, namely relay attacks and replay attacks. The peak value in the current year (2022) raises questions about the effectiveness of current countermeasures and necessitates attention.

This section begins by giving a brief subscription of short-range communication systems, followed by definitions of the most common attacks targetting these systems. The third part highlights the vulnerabilities of current countermeasures and reviews potential solutions.

3.3.1 Characteristics of Short-Range technology

By definition, a short-range communication system provides wireless connectivity within a local sphere of interaction [97]. There is no strict definition that limits the communication range. We, however, refer to short-range as any system that is limited to an operating distance from a few millimetres to a few tens of meters. Short-range systems are mainly designed for low-power and low-cost devices. Transceivers are typically characterised by a simple construction with built-in omnidirectional antennas [97]. The RF transmit power typically ranges from a few microwatts to 100 milliwatts. Representative examples of short-range systems are wireless personal networks, wireless body area networks, Radio Frequency Identification (RFID), and Near Field Communication (NFC) .

3.3.1.1 RFID and NFC

RFID technology can be thought of as the successor of barcodes that are used in supply-chain management. In contrast to barcodes, RFID does not require a line-of-sight link. RFID involves many different technologies based on LF, HF, or UHF spectrum [98]. Depending on the operating frequency and the power consumption, the communication range of RFID systems can be a few centimetres to approximately to tens of meters [99]. Table 1 lists some representative applications and their communication range.

An RFID system comprises a transponder (or tag), an interrogator (or reader), and a host computer or back-end database. A tag is a small, cheap, and simple device that is used for identification purposes [100]. Tags can be classified as passive or active. Passive tags transmit their data using the energy provided by the reader through a technique called backscattering modulation [101]. On the other hand, active tags use their own local power for signal transmission. Folded dipoles are a standard solution in RFID systems for minimising size whilst keeping an omnidirectional gain pattern [99].

Commonly found in contactless payments, NFC is a special case of HF RFID. Standardised by ISO 14443, NFC tags are passive with printed coiled antennas that operate in the HF region (13.56MHz). Magnetic mutual coupling -a special case of backscattering modulation- facilitates NFC data exchange within a few centimetres of distance. (<10cm) [99].

	LF	**HF**	**UHF**	**Active**
Frequency	125-134.2KHz	13.5 MHz	850-960MHz	100KHz-2.45GHz
Range	0.2-2m	< 1m (10cm for NFC)	up to 3m	up to 100m
Application	animal tracking	contactless payments	access control	tracking military assets

Table 3.2: RFID technology, operating frequency, communication range, and representative applications.

3.3.1.2 Wireless Personal Networks and Body Area Networks

A wireless Personal Area Network (PAN) is an interconnection of personal devices such as earphones, smartphones, and personal computers. The 802.15 family specifies PAN standards, the most common of which are Bluetooth, Bluetooth Low Energy (BLE), and Zigbee. Most PANs operate in the 2.4GHz band, and the transmission power is typically between a couple of milliwatts to 100mW. The typical communication range for Bluetooth applications is 10m, and longer distances can be achieved through BLE and Zigbee [97, 99]. BLE and Zigbee are specifically designed for low data rates and are a good fit for smart home automation, smart energy, and proximity sensing [102]. Bluetooth is used for applications that require continuous data transmission over long periods, such as sending the audio information of a video call to a wireless headset. Most modern mobile phones are equipped with Bluetooth and BLE, enabling entertainment and healthcare applications.

A Body Area Network (BAN) is a relatively new research area that aims to realise real-time health monitoring [103], Telemedicine [104], and M-Health [105]. In BANs, sensor devices are implanted or fixed on the surface of the body to monitor physiological changes such as changes in the heart rate or stress hormone levels. For practical reasons, these sensor devices need to be as small as possible, and they have severe memory and power constraints. Health-monitoring information needs to be highly confidential and reach the recipient (e.g. the hospital) in real-time, no matter the patient's location. As such, BANs are required to be highly reliable and secure.

3.3.2 Authentication Attacks/Distance Fraud

As the name implies, authentication attacks target the authentication phase of a communication system between two devices. The verifying device is referred to as the verifier, whereas the prover is the device that needs to prove its legitimacy. For example, in a door-access control system, the prover is a key-fob, and the verifier is a device placed by the door. In contactless payments (figure 3.5), the prover and verifier are the smartphone/ EMV card and the point-of-sale device, respectively. The verifier challenges the prover to send authentication credentials such as an encrypted identification number. When the verifier receives a valid response, it assumes that a genuine device is nearby and grants access to a service.

3.3.2.1 Mafia Fraud

Mafia Fraud [106] is a type of distance fraud whereby an adversary deceives the verifier into believing that a legitimate prover is closer than it is in reality; The adversary captures, amplifies and re-transmits the signals sent from the verifier to the prover and vice versa. The adversary does not need to know the shared secret or be able to decode the messages sent; it simply forwards the messages from the verifier to the prover and vice versa. The two legitimate entities believe that they communicate directly when, in fact, they are far away from each other. Two attack nodes often launch a Mafia Fraud attack in order to increase the distance over which the attack can be successful. One attack node relays the challenges sent by the verifier, whereas the second node relays the responses of the prover as shown in

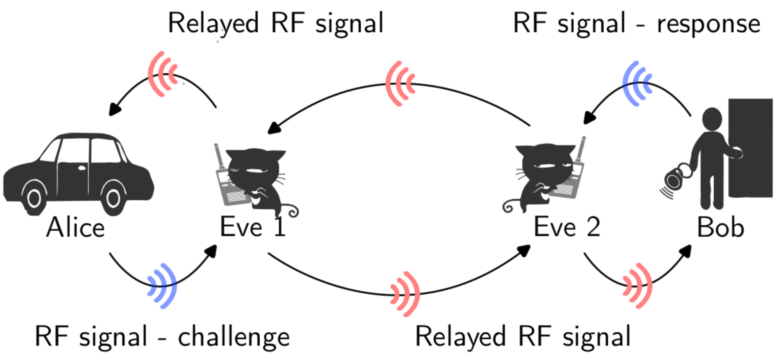


Figure 3.6: A Mafia Fraud attack with two adversary nodes.

figure 3.6. Mafia Fraud is commonly known as a relay attack. We use the term relay attack to refer to both Mafia Fraud and Terrorist Fraud.

3.3.2.2 Replay attack

A popular variant of Mafia Fraud is a replay attack. Mafia Fraud and replay attacks account for the vast majority of car-crime [107] and the two attacks are often perceived to have the same meaning in articles that address the public [108, 109]. Both attacks are impersonation attacks that involve the re-transmission difference. In academic literature, a replay attack is not a real-time attack, i.e. it is not launched during the data extraction. In a replay attack, the adversary retransmits the response that was captured during a previous authentication process between the verifier and the legitimate receiver [110]. Some types of replay attacks modify the message before re-transmission [111].

3.3.2.3 A dishonest prover

Solo Distance Fraud and Terrorist Fraud are two types of distance fraud attacks that involve a dishonest prover. In previous attacks, both the prover and the verifier were unaware of the attack. If a prover with authentication credentials attempts to 'lie' about its distance from the verifier, the prover is called dishonest.

Solo Distance Fraud In a Solo Distance Fraud, a remote dishonest prover convinces the verifier that it is in close proximity [106]. For example, a remote prover increases the transmit power beyond the standard levels so that its distance appears to be shorter; The verifier falsely believes that the prover is within the standard communication range.

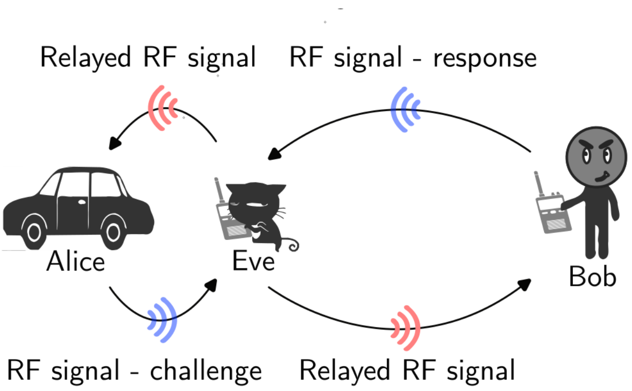


Figure 3.7: In a Terrorist Fraud attack, remote Bob cooperates with a local adversary.

Terrorist Fraud If a dishonest remote prover cooperates with an external adversary that is positioned close to the verifier, the distance fraud is known as Terrorist Fraud [106] (see figure 3.7). The cooperation does not involve sharing the long-term key with the external node; If the prover shared the long-term key, it would lose control over its own impersonation[112]. Similarly to Mafia Fraud, the Terrorist Fraud attack is a type of relay attack.

3.3.3 Current solutions and limitations

Whereas replay attacks can be dealt with in the upper layers of the protocol stack with cryptographic primitives, protection against distance fraud typically requires the exploitation of the physical medium. To date, several studies have investigated ways to verify the proximity of a transmitter, the most well-established techniques of which are based on time-of-flight measurements [106, 113, 114].

3.3.3.1 Cryptographic methods against replay attacks

In contrast to distance fraud, replay attacks can be prevented solely by cryptographic methods. Current cryptographic solutions involve one-time session keys such as rolling codes [115], timestamps, nonces [116], and tokens [117]. As the name implies, one-time session keys can only be used during an authentication process; A repeated session key should fail the authentication phase such that a replay attack is prevented. The level of security provided by one-time session keys depends on the cryptographic methods used. All one-time session keys should be encrypted or include a message digest (e.g. a Media Access Control (MAC)) to provide integrity; Otherwise, a message can be modified before being replayed, and the replay attack may be successful. As the following two descriptions demonstrate, opportunities for replay attacks exist due to flaws in existing security protocols.

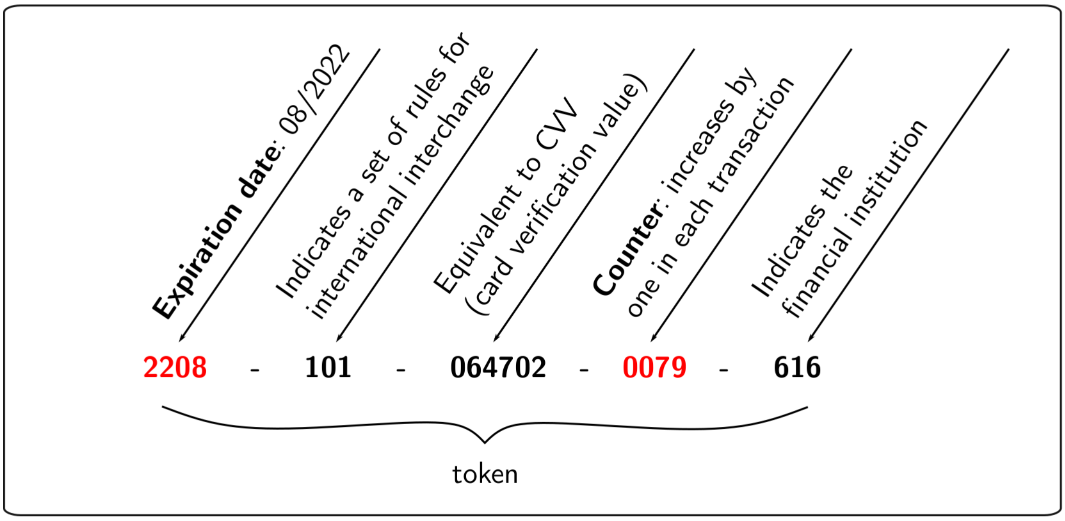

Expiration date: 08/2022
Indicates a set of rules for international interchange
Equivalent to CVV (card verification value)
Counter: increases by one in each transaction
Indicates the financial institution
2208 - 101 - 064702 - 0079 - 616
token

Figure 3.8: The last 20 digits of a token used for contactless payments with Samsung Pay.

Rolling codes Rolling codes were introduced in 2015 [115] for vehicular keyless entry systems that use RFID or Bluetooth technology. They are defined as a set of session keys that are available locally both at the verifier and the prover. Every time a key is used for authentication, it is discarded. Greenberg [115] demonstrated a replay attack that he named RollJam. In his demonstration, a rolling code is used in the authentication phase between the key-fob (prover) and the car (the verifying device in the car). An adversary node jams the signal sent by the fob, captures it and stores it at the same time. In a second trial, the fob 'rolls' to another key and authentication will be successful in future use. However, the adversary returns to the car at a later time and replays the stored signal. Authentication will be successful for the attacker since the key hasn't been discarded.

NFC payments and tokens Figure 3.8 gives an example of a token used for contactless payment with a smartphone application, as found in [117]. The token includes an expiration date and a four-digit counter. A counter is a value that increases by one for every purchase. As such, every transaction has its unique token. A purchase associated with repeated tokens alarms a replay attack and should be rejected. However, some terminals allow the same token to be used twice, which can be exploited for a replay attack as demonstrated in [118]. S. Mendoza [117] demonstrated a successful replay attack that included data modification of the token. For increased security, they [117] suggested a shorter lifespan of the token.

3.3.3.2 Physical Layer Authentication

Physical Layer Identification Physical layer identification is the process of identifying a device by its unique RF characteristics, often called RF fingerprints. Imperfection in the analogue circuitry of a device appears in the transmitted signal in such a way that it can be measured and stored at the recipient/verifier (or at a database to which the verifier has access). To authenticate an entity, the verifier compares the prover's RF fingerprints against those in their database and authenticates the prover if there is a match. Physical Layer Identification is a potential countermeasure for replay attacks, Mafia Fraud, and Terrorist Fraud attack. When the verifier receives a signal which is relayed or replayed by an adversary, the RF fingerprint captured on the received signal is that of the adversary and not of the legitimate remote user. The verifier rejects the authentication request.

There have been numerous implementations of Physical Layer Identification in applications such as RFID transponders [119], UHF sensor nodes, Wi-Fi devices [120] and Bluetooth devices [121]. For the case of mobile users, a novel scheme can be found in [122] whereby the unique hardware characteristics and time-varying carrier frequency offsets were exploited. Rahman et al. [123] observed that the time offset between the clocks of any two users is unique and used it as a tool for authentication.

RF identification is believed to be a promising solution against replay attacks and distance fraud [121, 124], but more investigation is needed to cover the gap between in-lab investigations and practical application scenarios [125]. Materials may behave differently in different temperatures and environments, and RF fingerprints of a device may not remain constant [126]. Furthermore, updating the database of fingerprints at the verifier may not be practical for some IoT applications [125].

RSS and phase-based ranging Given that there is a direct link between two communicating parties, the receiver can estimate the distance of the transmitter by simply measuring the path attenuation of the signal. Given that the transmit power is regulated by standards or a set of rules, the receiver measures the RSS or Received Signal Strength Indicator (RSSI) and compares it with the transmit power. Due to RSS/RSSI measurements being widely available, RSS(I)-based distance estimation is a popular technique. For example, the NHS app uses the RSS indicator to estimate the proximity of another mobile user who is registered as a COVID-19 carrier. Some car manufacturers have also chosen this technique to prevent Mafia Fraud attacks [127].

For low-power networks, the phase-based ranging method is often preferred due to its extremely low power requirements [128]. Similarly to radar-ranging systems, phase-based techniques measure the difference between the phase of the transmit sinusoidal signal - which is fixed and known at the receiver- and the phase of the received signal. When the wavelength of the operating frequency is less than the distance, the multi-carrier phase ranging technique eliminates the need to keep track of the elapsed cycles of the sinusoidal wave. Multi-carrier phase distance estimation is a cost-optimised technique [128], and it is deployed by many UHF RFID, Zigbee, and BLE ranging systems [128–130].

Papers [127, 128, 131] suggest that techniques based on RSS(I) or phase measurements are not secure against distance fraud and that they should be avoided. In RSS-based ranging, a transmitter can 'lie'about their distance by simply ignoring the regulations and transmitting with a higher power, in which case the system does not protect against Solo Distance fraud. In a Mafia Fraud or Terrorist Fraud attack, the attack node may not even need to amplify the signal above the regulations if it is placed closer to the verifier. Attacking phase-based ranging systems is also straightforward. This time, instead of changing the transmit power, the adversary changes the phase of the transmit signal, which causes an erroneous estimation of the phase shift at the verifier. Hence, the verifier cannot calculate the real distance and a relay attack is possible [128].

Time-of-flight distance bounding There exist many distance-bounding protocols [113, 132–136] but the key idea of all is the same: The verifier transmits a set of challenges at which the prover responds as soon as possible in a ping-pong manner. Relying on the fact that information cannot travel faster than the speed of light, measuring the round-trip-time allows Bob to give an upper bound of his distance to the prover. Multiple bit exchanges are needed to reduce the estimation error. If the round-trip-time is larger than a threshold, the prover is believed to be remote and his messages are relayed. In this case, the verifier rejects the prover.

Despite the conceptual simplicity, distance bounding protocols are hard to realise because they require specialised hardware for time accuracy. To understand why accuracy in timing is important, think that a time offset of just 1μs translates to a 30-meter estimation error. The effectiveness of distance bounding protocol relies heavily on the estimation accuracy. Whereas time-off-flight measurements typically give a much less accurate estimation of the distance than RSS or phase measurements, they are believed to provide a much more resilient scheme against distance fraud [128].

The first practical implementation of a distance bounding protocol took place in 2007 by S. Drimmer and S. Murdock for the case of contactless payments [113]. Existing protocols for EMV payments based on this method are resilient to Mafia Fraud when the attacker introduces a relay delay greater than 5ms [132]. However, some Mafia Fraud attacks that target vehicles can achieve relay speeds of a few microseconds [113]. Solo Distance Fraud attacks are much faster than relay attacks simply because the signal is transmitted directly to the verifier. Increasing the resilience of distance bounding techniques against 'fast ' distance fraud is equivalent to increasing the time accuracy of ToF measurements. Achieving a nanosecond precision may be feasible with current technology but costly due to requiring specialised hardware as well as an ultra-wideband channel [132].

Ambient conditions Authentication based on ambient conditions checks the proximity between the verifier and the prover in order to tackle distance fraud. When the prover is close to the verifier, the environment around them will be similar. For example, the two nodes shall 'listen' to the same sounds [137], 'sense' the same temperature [138], and observe the same entities around them such as WiFi and Bluetooth devices [139]. These observations require the deployment of sensing technology at the two entities. The prover quantifies a set of observations and sends it to the verifier along with a cryptographic signature. If the prover's response demonstrates a similar environment to that observed by the verifier, the proximity verification is successful.

The prover's responses expire when the environment changes, limiting the time frame over which a replay attack can be successful. For real-time attacks, a Mafia Fraud attack and a Solo Distance Fraud attack will also fail as long as the remote prover observes a different environment to that of the verifier. However, the method cannot protect against Terrorist Fraud. An adversary node that is local to the verifier will make the same observations as the verifier, and upon cooperation with the remote dishonest prover, the observations will be signed before being sent to the verifier.

For a resilient scheme that offers protection against replay attacks, Solo Distance fraud and Mafia Fraud, scholars seem to agree that multiple sensors are required. Shrestha et al. [140] use an off-the-shelf ambient sensing platform to capture the humidity, altitude, precision gas, and temperature of the physical environment. The fact that mobile phones are equipped with multiple sensors (sometimes more than ten) is exploited in [141]. The authors conclude that using the location sensor, accelerometer sensor, microphone, magnetometer sensor, and attitude sensor in iPhone 6S has some potential countermeasures to secure Apple Pay from relay attacks.

The vulnerabilities of ambient-based protocols are best described by Y.Tu et al. in their recent work [142]. They point out that most of the ambient conditions could be manipulated by an attacker leading to impersonation attacks. They also add that conducting a sufficient number of measurements in a short time is often challenging. Protocols based on the ambient environment are still in their early stage of research [142]. Whether they can result in resilient and practical authentication is yet to be investigated.

3.4 Conclusion

The physical medium can be exploited in many ways for providing secrecy. Keyless PLS uses secrecy coding and signal processing techniques in order to “hide” a confidential message, whereas the aim of key-based PLS is to extract symmetric keys and enable upper-layer cryptography. Physical layer authentication is a newly emerging field and a promising solution to distance fraud.

Until recently, the study of keyless PLS was limited to theoretical models and had no practical value due to the requirement of a positive secrecy gap, i.e. the requirement that the main channel is “better” than the wiretap channel. Recent advancements in wireless technologies can be used in a way that increases the secrecy gap. The study of secrecy coding has since regained attention and is no longer limited to discrete channel models. A small but growing number of publications on secrecy coding for fading channels have emerged in recent years. Nevertheless, it seems like there are many challenges to be addressed before keyless PLS is viable in real-life networks. For example, most artificial-noise-based schemes assume Gaussian signalling and are insufficient in current signalling schemes.

The practicality of key-based PLS in systems that operate in TDD mode has been proved by numerous test beds. The vast majority of implementations have used RSS measurements due to being widely available by off-the-shelf devices. However, RSS-based key generation results in a much lower key rate when compared to CSI-based key generation. In slow varying fading channels, both RSS-based and CSI-based key generation suffer from low key rates. Although a highly dynamic RF channel improves the key rate, it can introduce bit-mismatches in the key sequences. Reconciling a high number of mismatches is costly in terms of computation overhead, computational complexity, and information leakage.

Lastly, motivated by real-life fraud in emerging short-range communications, a literature review of physical layer authentication against distance fraud and replay attacks has been given. For ultra-wideband transceivers, time-of-flight distance bounding is a tested and efficient approach against distance fraud. Physical layer identification is a promising solution but more research is needed before it is adopted by the industry.

CHAPTER 4

KEY GENERATION AND SPATIAL SEPARATION

Channel Reciprocity-based Key Generation (CRKG) is an attractive technique for generating symmetric cryptographic keys due to low memory and processing power requirements. As a relatively new field, physical layer security is often limited to idealistic scenarios. Most CRKG protocols assume that the extracted key is secure from an eavesdropper as long as she is distanced by half a wavelength or more from the legitimate users. This chapter convinces us that the typical definition of coherence distance is inappropriate for secrecy purposes. The half-wavelength decorrelation assumption brings secrecy vulnerabilities which are quantified and brought to attention. Secure distance is redefined, and numerical examples are given for both directional and non-directional RF channels. The main results and conclusions of this chapter have been published in [143].

Keywords: Physical layer Key generation, physical layer security, Rayleigh channels, secret key capacity, spatial channel correlation

Prerequisite: Fundamentals of Information Theory 2.1 (with a focus on Mutual Information and Correlation 2.1.3.2), Spatial correlation and multipath fading 2.3.3, Key-based PLS 3.2.

4.1 Introduction

The spatial channel correlation between receive antennas is an essential topic for antenna design and spatial diversity techniques [144, Ch. 6], [145], and has been extensively studied by wireless engineers. Typically [146, Ch. 4], the coherence distance is defined as the minimum distance required so that the spatial channel correlation between two receive antennas falls below 0.5. Depending on the application and the quality-of-service requirements, the threshold may vary from lower to higher values. For example, J. Salz [25] has shown that a correlation as high as 0.5 has little impact on the performance of an adaptive array that combats fading and suppresses interference. When maximal ratio combining is employed at a multi-antenna receiver, a correlation as high as 0.7 is acceptable [147].

Based on the typical definition of the coherence distance, the channels at two receive antennas half wavelength (0.5λ) apart are considered “essentially” decorrelated as long as the antennas are surrounded by multiple scatterers. Channel decorrelation within half a wavelength is a rule adopted by the community of physical layer security. Specifically, the vast majority of key generation protocols promise security as long as the eavesdropper is placed at a distance greater than 0.5λ from either legitimate user. This chapter convinces us that channel correlation has a significant impact on the secret key capacity and that a distance of 0.5λ is not perfectly secure even within idealistic environments of isotropic scattering. In the scenario of directive channels, we show that the secrecy performance is decreased dramatically along with the beamwidth of the received signals.

4.1.1 Contributions

The major contributions of this chapter are as follows.

- A thorough study of the impact of spatial channel correlation on the secret key generation capacity;

- Proof that even under idealistic scenarios of rich scattering, the half-wavelength distance is not perfectly secure: an eavesdropper may attain crucial information even when positioned at a distance much greater than 0.5λ.

- A novel definition of secure distance that couples distance with spatial correlation and facilitates quantification of spatial correlation and secrecy performance.

4.1.2 Relevant Work

Authors in [148] set up an experimental platform consisting of a base station and a mobile user that exchanged packets following the 802.15.4 standard. Their results exhibit strong cross-correlation on the RSSI between the base station and a distanced eavesdropper (up to 7.5 wavelengths). However, RSSI measurements are expected to result in a high correlation. This is because RSSI is the mean value of the instantaneous received signal within a packet; the sample mean of a random variable is less uncertain than the random variable. Authors in [149] demonstrated through simulations that multiple collaborative eavesdroppers were able to reconstruct the key through correlated observations. The key was obtained from envelope measurements under Rappaport's channel model, whereby the energy arrived from certain directions.

Whereas the mentioned works show that correlation attacks are possible in certain environments, our work aims to give explicit definitions and formulae that quantify the secrecy degradation and secure distance. Moreover, our work does not focus solely on poor scattering environments. We show that a decrease in the key rate exists even in rich scattering environments and when the complex information of the signal is considered.

This work is based on results found in [150]. The authors in [150] provide an information-theoretic analysis of the secret key generation rate under correlated observation at the eavesdropper. We extend this work by quantifying secrecy performance degradation and defining secure distance when the assumption of spatial decorrelation is lifted.

4.2 Preliminaries and channel model

4.2.1 The legitimate channel

When launching a Channel Reciprocity Key Generation (CRKG) protocol [73, 79] between two users, say, Alice and Bob, pilot sequences are sent to either end of the link. When Alice sends a pilot sequence, Bob estimates the channel between him and Alice, h_{AB}. Similarly, when Bob sends a pilot sequence, Alice estimates her channel h_{BA}. We assume that the two pilot sequences are sent before the channel changes. Specifically, we assume perfect reciprocity, and we denote the common channel between Alice and Bob by $h_L := h_{AB} = h_{BA}$. Channel h_L is assumed to be a Rayleigh channel with zero mean and unit power, i.e. $h_L \sim CN(0,1)$. Alice's and Bob's estimations are noisy copies of the true channel:

$$\hat{h}_A = h_L + n_A \text{ and } \hat{h}_B = h_L + n_B, \tag{4.1}$$

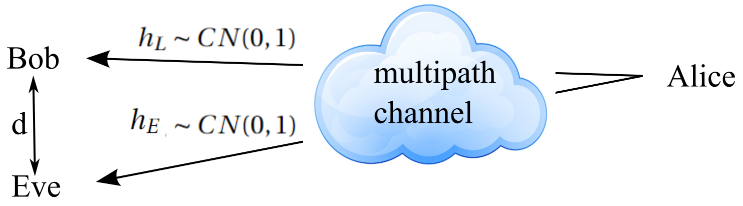


Figure 4.1: To get an estimate of the key generated by Alice and Bob, Eve approaches Bob at distance d. Both channels h_L and h_E are Rayleigh channels with zero mean and unit power.

where $n_A \sim CN(0,\sigma_A^2)$ and $n_B \sim CN(0,\sigma_B^2)$. As such, the estimations are Gaussian channels of zero mean and variance of $\sigma_u^2 + 1$, $u \in \{A, B\}$.

After the channel estimation, a pre-agreed quantisation scheme takes place. The process is repeated when the channel changes until a sufficient key length is attained. Due to noise, the keys at Bob and Alice may not match perfectly, and a reconciliation phase takes place to ensure symmetrical keys [79]. During the reconciliation phase, Alice and Bob exchange information about the key through a public channel. A final phase, called privacy amplification, takes place to compensate for the information leakage. The key rate is therefore limited. The maximum key rate per channel realisation of CSI-based KG protocols is equal to $I(\hat{h}_A;\hat{h}_B)$. Observe that $\hat{h}_v \sim CN(0, 1+\sigma_v^2)$, $v \in \{A,B\}$. Also, due to channel reciprocity, $E\left(\hat{h}_A\hat{h}_B^*\right) = E\left(h_A h_B^*\right) = 1$. The covariance matrix of $\hat{h}_A$ and $\hat{h}_B$ is therefore $\Sigma_{AB} = \begin{bmatrix} 1+\sigma_A^2 & 1 \\ 1 & 1+\sigma_B^2 \end{bmatrix}$.

From theorem 2.8 i., we have that

$$I(\hat{h}_A;\hat{h}_B) = \log\left(1 + \frac{1}{\det(\Sigma_{AB})}\right). \tag{4.2}$$

4.2.2 The eavesdropper's correlated channel

Dropping the assumption of an uncorrelated channel at the eavesdropper, the maximum achievable secret key rate may be lower than $I(\hat{h}_A;\hat{h}_B)$ as seen in section 3.2.1. Having a correlated channel to Bob's channel, Eve gains information about the key. For a secret key, vulnerable keybits must be discarded, hence the secret key rate is lower than $I(\hat{h}_A;\hat{h}_B)$. Bloch et al. have shown [70] that the secret key capacity is bounded as:

$$C_{sk} \geq I(\hat{h}_A;\hat{h}_B) - \min I(\hat{h}_A;\hat{h}_E), I(\hat{h}_B;\hat{h}_E) \tag{4.3}$$

$$C_{sk} \leq \min(I(\hat{h}_A;\hat{h}_B), I(\hat{h}_A;\hat{h}_B|\hat{h}_E)) \tag{4.4}$$

Without loss of generality, let the eavesdropper, Eve, be closer to Bob than to Alice as seen in figure 4.1. When Alice transmits a pilot sequence, Eve estimates channel $h_E \sim CN(0,1)$. Similar to Bob's channel, the channel between Alice and Eve is also Rayleigh with unit power. Eve's estimation for the channel is denoted by $\hat{h}_E = h_E + n_E$, where $n_E \sim CN(0, \sigma_E^2)$. Let $\rho := E\left(h_L h_E^*\right)$ denote the spatial channel correlation between the legitimate channel and the eavesdropper's channel. Observe that this is the correlation coefficient between two circularly symmetric random variables. Hence, from theorem 2.9, ρ is a real number: $\rho \in \mathbb{R}$. Applying theorem 2.8, we have that

$$I(\hat{h}_B; \hat{h}_E) = \log\left(1 + \frac{\rho^2}{\det(\Sigma_{BE})}\right) \tag{4.5}$$

and

$$I(\hat{h}_A; \hat{h}_B | \hat{h}_E) = \log\left(\frac{\det(\Sigma_{AE})\det(\Sigma_{BE})}{(1+\sigma_E^2)\det(\Sigma_{ABE})}\right), \tag{4.6}$$

where $\Sigma_{uE} = \begin{bmatrix} 1+\sigma_u^2 & \rho \\ \rho & 1+\sigma_E^2 \end{bmatrix}$, $u \in \{A, B\}$, and

$$\Sigma_{ABE} = \begin{bmatrix} 1+\sigma_A^2 & 1 & \rho \\ 1 & 1+\sigma_B^2 & \rho \\ \rho & \rho & 1+\sigma_E^2 \end{bmatrix}.$$

Note that for the case of Rayleigh channels, the terms of independence and zero-correlation are interchangeable. As such, when $\rho = 0$, the eavesdropper observes independent channel realisations to Bob's. In other words, when $\rho = 0$, since $I(\hat{h}_A; \hat{h}_E) = I(\hat{h}_B; \hat{h}_E) = 0$ and $I(\hat{h}_A; \hat{h}_B)|\hat{h}_E) = I(\hat{h}_A; \hat{h}_B)$, the bounds of (4.3) and (4.4) become equal. The secret key capacity is then given by

$$C_{sk} = I(\hat{h}_A, \hat{h}_B) \quad \text{when } \rho = 0 \tag{4.7}$$

On the other hand, when $|\rho|$ takes its maximum value of $|\rho| = 1$, the upper bound (4.4) becomes zero since $I(\hat{h}_A; \hat{h}_B)|\hat{h}_E) = 0$. As such,

$$C_{sk} = 0 \quad \text{when } |\rho| = 1 \tag{4.8}$$

4.2.3 SNR at the receivers

By Signal-to-Noise-Ratio (SNR), we mean the power ratio between useful signal and noise. Since the channel powers are normalised to one, the SNR at Alice, Bob, and Eve, are given by $SNR_A = 1/\sigma_A^2$, $SNR_B = 1/\sigma_B^2$, and $SNR_E = 1/\sigma_E^2$, respectively. Although the formulae given are not restricted to specific values, numerical results will be given for the following two cases:

1. $\sigma_A^2 = \sigma_B^2 > 0, \sigma_E^2 = 0$

2. $\sigma_A^2 = \sigma_B^2 = \sigma_E^2$

The first case corresponds to the worst-case scenario where the eavesdropper has a noiseless channel. This could apply to scenarios where Eve has multiple antennas and/or the thermal noise at her device is negligibly small. The second case captures the valid assumption that if Bob and Eve are in close proximity and equipped with similar devices, they shall observe similar SNR. For both cases, the SNR at the legitimate receivers are set to be equal: $\text{SNR}_A = \text{SNR}_B = \text{SNR}_L$. This could be the case when they both transmit their pilots with the same power.

Remark 4.1. *A MIMO antenna could potentially benefit an eavesdropper by providing a better SNR. This case is covered by a single-antenna eavesdropper with an infinitely large SNR, i.e., the case when* $\sigma_E^2 = 0$.

4.3 Spatial channel correlation and Secrecy Degradation

For given σ_u^2, $u \in \{A, B, E\}$, the secret key capacity can be thought of as a function of the spatial channel correlation ρ. The case when $\rho = 0$ is the idealistic case of independent/uncorrelated observation at the eavesdropper. In this case, the secret-key capacity is equal to $I(\hat{h}_A; \hat{h}_B)$ (observe that the two bounds in (4.3) and (4.4) become equal when $|\rho| = 0$). When $|\rho| > 0$ the secret key capacity degrades. A metric to facilitate comparison against the idealistic case of $|\rho| = 0$ is the normalised difference of $C_{sk}(0) - C_{sk}(\rho)$.

Definition 4.1. The secrecy degradation for given σ_A^2, σ_B^2, and σ_E^2 is defined as a function of ρ:

$$\text{DEG}_{sk}(\rho) = \frac{C_{sk}(0) - C_{sk}(\rho)}{C_{sk}(0)} \tag{4.9}$$

Although it can be expressed as a function of secret key capacity, secrecy degradation does not take a closed-form expression in terms of the spatial channel correlation. The following theorem derives the upper bound and lower bound of secrecy degradation.

Theorem 4.1. *The lower and upper bounds for the secrecy degradation are given by:*

$$DEG_{sk}(\rho) \leq \log\left(1+\frac{\rho^2}{\det(\Sigma_{BE})}\right)/\log\left(1+\frac{1}{\det(\Sigma_{AB})}\right) \tag{4.10}$$

$$DEG_{sk}(\rho) \geq \max\left(0, 1-\log\left(\frac{\det(\Sigma_{AE})\det(\Sigma_{BE})}{(1+\sigma_E^2)\det(\Sigma_{ABE})}\right)/\log\left(1+\frac{1}{\det(\Sigma_{AB})}\right)\right) \tag{4.11}$$

Proof. From (4.5) and (4.6) for $\rho = 0$, we have that $I(\hat{h}_B;\hat{h}_E) = 0$ and $I(\hat{h}_A;\hat{h}_B)|\hat{h}_E) = I(\hat{h}_A;\hat{h}_B)$. As such, the bounds in (4.3) and (4.4) are equal to $I(\hat{h}_A;\hat{h}_B)$. Thus, $C_{sk}(0) = I(\hat{h}_A;\hat{h}_B)$. From (4.3) we have that $-C_{sk} \leq -I(\hat{h}_A;\hat{h}_B) + I(\hat{h}_B;\hat{h}_E)$. By adding constant $C_{sk}(\rho) = I(\hat{h}_A;\hat{h}_B)$ and dividing by the same constant, we attain that $\frac{C_{sk}(0)-C_{sk}(\rho)}{C_{sk}(0)} \leq \frac{I(\hat{h}_B;\hat{h}_E)}{I(\hat{h}_A;\hat{h}_B)}$. By the definition (4.1), the left-hand side is the system degradation. By substituting (4.2) and (4.5) on the right-hand side, the proof of 4.10 is complete. Similar steps are followed for proving (4.11). We start with the upper bound of secret key capacity:

$$\begin{aligned} C_{sk}(\rho) &\leq \min(I(\hat{h}_A;\hat{h}_B), I(\hat{h}_A;\hat{h}_B)|\hat{h}_E)) \leftrightarrow \\ \frac{C_{sk}(0)-C_{sk}(\rho)}{C_{sk}(0)} &\geq \frac{C_{sk}(0)-\min(I(\hat{h}_A;\hat{h}_B), I(\hat{h}_A;\hat{h}_B)|\hat{h}_E))}{C_{sk}(0)} \leftrightarrow \\ \mathrm{DEG}_{sk}(\rho) &\geq \frac{I(\hat{h}_A;\hat{h}_B)-\min(I(\hat{h}_A;\hat{h}_B), I(\hat{h}_A;\hat{h}_B)|\hat{h}_E))}{I(\hat{h}_A;\hat{h}_B)} \leftrightarrow \\ \mathrm{DEG}_{sk}(\rho) &\geq \frac{\max(0, I(\hat{h}_A;\hat{h}_B)-I(\hat{h}_A;\hat{h}_B)|\hat{h}_E))}{I(\hat{h}_A;\hat{h}_B)}. \end{aligned}$$

Substituting (4.2) and (4.6) on the right-hand side completes the proof. □

4.3.1 Secrecy Degradation and asymptotic behaviour

Figures 4.2 and 4.3 plot the secrecy degradation as a function of the SNR_L and $|\rho|$ for the two cases of $\sigma_E^2 = 0$ and $\sigma_E^2 = \sigma_B^2$, respectively. Observe that the bounds become tight as SNR_L increases. Such behaviour is expected; Authors in [151] have proved that the bounds of secret key capacity given in (4.4) and (4.3) merge towards the value of

$$C_{\mathrm{sk}}(\rho) = \log((1-\rho^2)/(\sigma_A^2+\sigma_B^2)) \tag{4.12}$$

as $\sigma_u^2 \to 0$, $\forall u \in \{A,B,E\}$. By substituting (4.12) in (4.9), we have that

$$\mathrm{DEG}_{sk} \to 1-\log\left(\frac{1-\rho^2}{\sigma_A^2+\sigma_B^2}\right)/\log\left(\frac{1}{\sigma_A^2+\sigma_B^2}\right) \quad \text{as } \sigma_u \to 0,\ \forall u \in \{A,B,E\}. \tag{4.13}$$

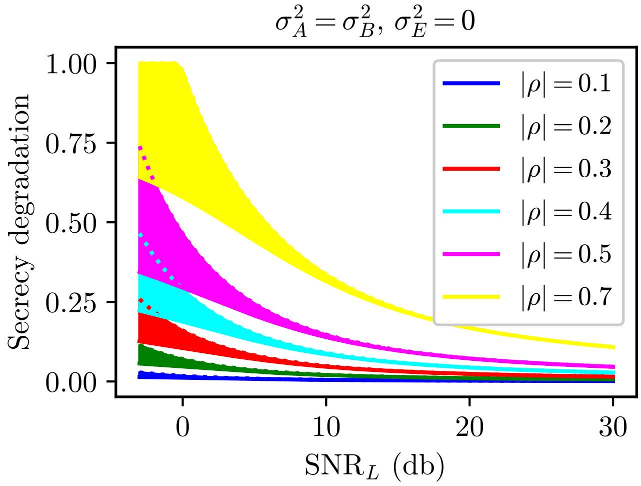


Figure 4.2: Secrecy degradation as a function of ρ and SNR at the legitimate users ($SNR_L = SNR_A = SNR_B$) when the eavesdropper has a noiseless channel.

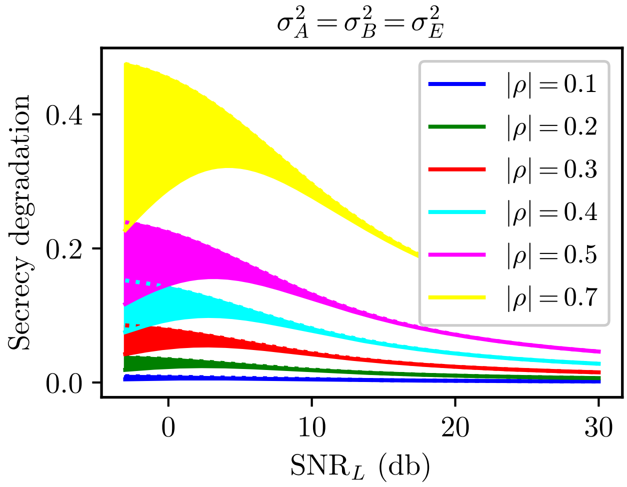


Figure 4.3: Secrecy degradation as a function of $|\rho|$ and SNR at the legitimate users ($SNR_L = SNR_A = SNR_B$)

.

4.3.2 On the value of spatial correlation

Figures 4.3 and 4.2 suggest that the typical definition of coherence distance, i.e. the distance at which the spatial channel correlation is $\rho = 0.5$ may be inappropriate for secrecy purposes. When the absolute value of spatial channel correlation is equal to $|\rho| = 0.5$ and the SNR at Eve is equal to Bob's SNR, figure 4.3 suggests that the secrecy degradation is higher than 20% when $\text{SNR}_L(=\text{SNR}_E) \leq 20dB$. In the case of a noiseless eavesdropper, the secrecy degradation is higher than 10%,when $|\rho| = 0.5$ and $\text{SNR}_L \leq 20$ (figure 4.2). The degradation increases as the SNR at the legitimate receivers decreases; Specifically, for the case of $\text{SNR}_L = 0dB$, the degradation can reach up to 47% when the eavesdropper is noiseless (figure 4.2), and up to 23% when $\sigma_E^2 = \sigma_B^2$ (figure 4.3).

Assuming a positive value of SNR_L, figures 4.2 and 4.3 suggest that spatial channel correlation needs to have an absolute value below 0.2 in order to guarantee a system degradation below 10%. Note that a secrecy degradation of 10% means that 10% of the key is vulnerable and needs to be dropped at the amplification phase.

To further decrease the secrecy degradation, we suggest a spatial channel correlation with an absolute value of 0.1 or less. When $|\rho| = 0.1$ the secrecy degradation is essentially zero as it falls below 3% even for the case of a noiseless eavesdropper.

4.4 Secure Distance

Having studied the impact of spatial channel correlation on the secret key capacity, we define the secure distance as a function of the spatial channel correlation. Such a definition allows numerical evaluations and secrecy guarantees. Before the explicit definition, we give a relaxed interpretation of secure distance.

Imagine that Bob is surrounded by a fence of a circular arrangement of radius d_s. Being unable to approach Bob at a distance less than d_s, Eve cannot achieve an arbitrarily high spatial channel correlation. If the absolute value of spatial channel correlation, $|\rho(h_E, h_b)|$, is guaranteed to be less than ϵ due to the space limitation, distance d_s is said to be ϵ-secure. The formula definition of e-secure distance follows.

Definition 4.2. In a Cartesian coordinate system, let Eve be positioned at $E(x, y, z)$ and let $\rho(E(x, y, z))$ be the spatial channel correlation between Eve's channel and Bob's channel when Eve is positioned at $E(x, y, z)$. A distance d_s is said to be ϵ-secure if

$$|\rho(E(x, y, z))| < \epsilon \text{ for all } E(x, y, z) \notin S(B, d_s), \tag{4.14}$$

where $S(B, d_s)$ is the sphere with centre Bob's location and radius d_s.

Observe that if distance d_s is ϵ-secure, so does any distance greater to d_s. However, a distance of less than d_s may not be ϵ- secure. The minimum ϵ- secure distance is denoted by $d_{s_o}(\epsilon)$:

$$d_{s_o}(\epsilon) := \min\{d_s | d_s \text{ is } \epsilon\text{-secure}\}. \tag{4.15}$$

Corollary 4.1. *The minimum ϵ-secure distance is a decreasing function of ϵ.*

Proof. Let $\epsilon_1 < \epsilon_2$, and $d_{s_o}(\epsilon_1) = d_1$ and $d_{s_o}(\epsilon_2) = d_2$. We will prove that $d_1 < d_2$. Since d_1 is ϵ_1-secure, and $\epsilon_1 < \epsilon_2$, then $\rho(d) < \epsilon_1 < \epsilon_2 \ \forall d > d_1$, or equivalently, d_1 is also ϵ_2-secure. But d_2 is the minimum ϵ_2-secure distance, hence $d_2 \geq d_1$ which completes the proof. □

The problem of finding the minimum ϵ-secure distance is analogous to the problem of finding the smallest possible radius of the circular fence around Bob. When d_{s_o} is the minimum ϵ-secure distance, the highest value of spatial correlation that Eve can achieve is ϵ.

Remark 4.2. *When distance d_{s_o} is the minimum ϵ-secure distance, then ϵ is the highest spatial channel correlation that the eavesdropper can achieve when positioned at a distance $d > d_{s_o}$ from Bob.*

Being a function of the spatial channel correlation, the secure distance depends on the geometry of the environment. We perform analysis and give numerical examples for three geometrical models:

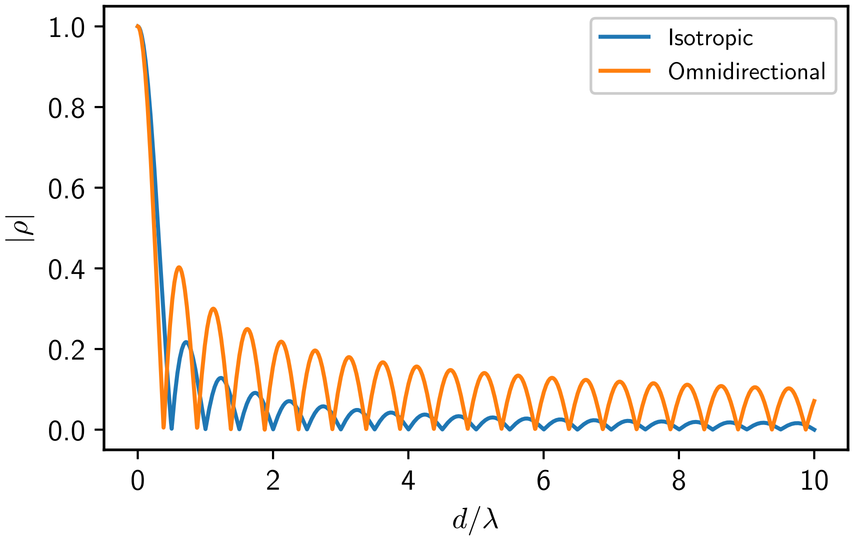
1. Isotropic model;
2. Omnidirectional model;
3. Restricted-uniform AoA model.

For all three geometrical models, we assume that the multipath components of the received signal have a similar amplitude. All three geometrical models result in Rayleigh channels (2.3.3, [152]), thus, we keep consistent with the statistics of h_A, h_B, and h_E described in 4.2.

4.4.1 Isotropic model

In this model, the Angle-of-Arrival (AoA) at the two receivers, Bob and Eve, is uniformly distributed across the unit sphere. This is the case where Bob and Eve are equipped with isotropic antennas and are placed in a rich scattering environment. Recall that the spatial channel correlation of the isotropic model is a sinc function (2.3):

$$\rho = \operatorname{sinc}(2\pi d/\lambda), \tag{4.16}$$



co

Figure 4.4: The spatial channel correlation against the normalised distance between two receivers for the cases of isotropic diffuse field and omnidirectional diffuse field.

where λ is the wavelength of the operating frequency, and d is the distance between Bob and Eve. Due to the uniform radiation patterns, the spatial channel correlation depends only on the distance between the two receivers. Hence, definition 4.2 can be simplified to:

$$d_s \text{ is } \epsilon\text{-secure} \iff |\mathrm{sinc}(2\pi d/\lambda)| < \epsilon \text{ for all } d \geq d_s \tag{4.17}$$

In figure 4.4, the graph labeled ‘isotropic’ uses equation (4.16) and plots the spatial channel correlation against the distance. The first zero spatial decorrelation appears at $d = 0.5\lambda$. However, distance $d = 0.5$ is not 0-secure; if the eavesdropper positions herself at distance $d = 0.7\lambda$ from Bob, her channel will correlate by $\rho = \mathrm{sinc}(2\pi \times 0.7) = 0.22$.

By definition, if d_o is ϵ-secure then every distance larger to d_o is also ϵ-secure. Determining the minimum ϵ-secure distance is an interesting problem for providing secrecy guarantees.

Lemma 4.1. *The minimum ϵ-secure distance, denoted by d_{s_o} is upper bounded by*

$$d_{s_o} \leq \frac{\lambda}{2\pi\epsilon} \tag{4.18}$$

Proof. It suffices to show that $\frac{\lambda}{2\pi\epsilon}$ is a ϵ-secure distance, or equivalently (by definition of secure distance) that: $|\rho| < \epsilon$ for all $d > \frac{\lambda}{2\pi\epsilon}$. Indeed, let $d > \frac{\lambda}{2\pi\epsilon}$, then:

$$|\rho| = \left|\text{sinc}\left(\frac{2\pi}{\lambda}d\right)\right| := \frac{\left|\sin\left(\frac{2\pi}{\lambda}d\right)\right|}{\frac{2\pi}{\lambda}} \leq \frac{1}{\frac{2\pi}{\lambda}d} < \frac{1}{\frac{2\pi}{\lambda}\frac{\lambda}{2\pi\epsilon}} = \epsilon. \tag{4.19}$$

□

Given lemma 4.1 and the fact that the |sinc| function is continuous in the range of $(0,1]$, the minimum secure distance is the maximum common solution of the system $y = \epsilon$ and $y = |\text{sinc}(2\pi d/\lambda)|$. Or, equivalently:

$$d_{s_o} = \underset{d>0}{\text{argmax}}(|\text{sinc}(2\pi d/\lambda)| = \epsilon) \tag{4.20}$$

A closed-form expression of the minimum secure distance as a function of ϵ appears to be a hard problem due to the nature of the sinc function; the local maxima of any Bessel function are given by infinite summations. Fortunately, since the minimum secure distance is limited in $(0, \frac{\lambda}{2\pi\epsilon})$, an algorithm that returns the minimum secure distance in a finite number of steps exists. Algorithm 1 first sets the minimum secure distance equal to the upper bound. Then it decreases the distance by a segment of c and updates the new minimum secure distance as long as the spatial correlation does not exceed the threshold ϵ. The algorithm terminates when $|\rho| > \epsilon$.

Algorithm 1 Calculate the minimum Secure Distance for the isotropic model

```
Inputs: $c$: step; $\lambda$: wavelength; $\epsilon$ : upper bound for $|\rho|$
Outputs: minimum secure distance at $c$ accuracy
$\rho(d) := \text{sinc}(\frac{2\pi}{\lambda}d)$                ▷ define function
$d \leftarrow \lambda/(2\pi\epsilon)$                            ▷ initiate distance
while $|\rho(d)| < \epsilon$ do
    $d_{s_o} \leftarrow d$                                  ▷ update minimum secure distance
    $d \leftarrow d - c$                                    ▷ decrease distance by $c$
end while
Print $d_{s_o}$
```

4.4.2 Omnidirectional model

An isotropic antenna is an abstract antenna model that does not exist in reality. Practically, the least-directive gain patterns are those of dipole antennas and are 'donut' shaped. Assuming that Bob is equipped with a vertically-oriented dipole, Eve's best strategy is to also employ a vertically-oriented dipole, so that she attains similar observations. Note that only

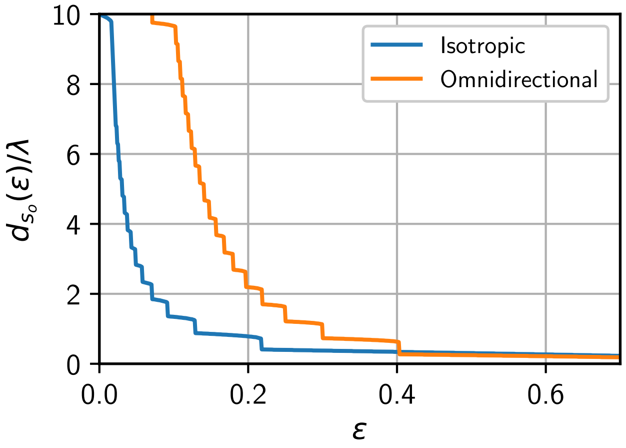


Figure 4.5: The minimum secure distance for different thresholds ϵ.

the most beneficial locations for Eve determine the secure distance. The spatial correlation is maximised (for a given distance) when Eve is inline with the communication link AB, where AB is the imaginary line connecting Alice and Bob [25]. The spatial channel correlation is [30, 153]:

$$\rho = J_o(2\pi d_{EB}/\lambda) \tag{4.21}$$

Finally, definition 4.2 can be simplified to:

$$d_s \text{ is } \epsilon\text{-secure} \iff |J_o(2\pi d/\lambda)| < \epsilon \text{ for all } d \geq d_s. \tag{4.22}$$

Although the first zero correlation occurs at 0.38λ, which is smaller in comparison to the isotropic case (see Fig. 4.4), the next numerical example shows that the minimum secure distance increases significantly.

Example 4.1. *Figure 4.4 suggests that, when Bob and Eve are equipped with dipole antennas, it takes approximately 10λ for the channel correlation to drop below* 0.1 *(plot labelled as "omnidirectional "). On the other hand, it takes less than* 1.5λ *when isotropic antennas are employed. Equivalently, the 0.1-secure distances for the two cases are approximately 10λ and 1.5λ.*

Functions $J_0(\cdot)$ and sinc$(\cdot)$ both belong to the family of Bessel functions, and as such, they share common properties. They are both oscillating continuous functions with decreasing local maxima, and no closed-form expressions are known for the extrema. The minimum secure distance is determined by following a similar approach to the isotropic case. The minimum secure distance, d_{s_o}, for the case of omnidirectional antennas can be simplified to:

$$d_{s_o} = \underset{d>0}{\operatorname{argmax}}(|J_o(2\pi d/\lambda)| = \epsilon) \tag{4.23}$$

Lemma 4.2. *The minimum ϵ-secure distance is upper bounded by:*

$$d_{s_o} \leq \left(\frac{2\pi}{\lambda}\right)^{-1}\left(\frac{c}{\epsilon}\right)^3, \tag{4.24}$$

where $c = 0.785746704...$

Proof. It suffices to show that $\left(\frac{2\pi}{\lambda}\right)^{-1}\left(\frac{c}{\epsilon}\right)^3$ is a ϵ-secure distance, or equivalently (by definition of secure distance) that: $|\rho| < \epsilon$ for all $d > \left(\frac{2\pi}{\lambda}\right)^{-1}\left(\frac{c}{\epsilon}\right)^3$. It is known [154] that $J_o(x) \leq c|x|^{-1/3}$, where $c = 0.7857\ldots$ Thus,

$$|\rho| = |J_o\left(\frac{2\pi}{\lambda}d\right)| < c\left(\frac{2\pi}{\lambda}d\right)^{-1/3} \tag{4.25}$$

It can be seen that when $d_{s_o} \leq \left(\frac{2\pi}{\lambda}\right)^{-1}\left(\frac{c}{\epsilon}\right)^3$, $|\rho| \leq \epsilon$. □

The algorithm that calculates the minimum secure distance for the omnidirectional gain patterns is almost identical to algorithm 1 and it is omitted. Only lines #1 and #2 differ; the function for the spatial channel correlation changes to $\rho := J_o(2\pi d/\lambda)$, and the distance initiation becomes $d \leftarrow \frac{2\pi}{\lambda}^{-1} c/\epsilon^3$.

Utilising the algorithms, we plot the minimum secure distance for the two cases of isotropic and omnidirectional antennas as shown in figure 4.5. At a first glance, the minimum secure distance decreases at a slower rate when omnidirectional antennas are employed. The next two examples combine the results of figures 4.5, 4.2 and 4.3.

Example 4.2. *The half-wavelength distance is 0.2-secure when isotropic antennas are employed. With dipoles, a half-wavelength distance is 0.4-secure which may be unacceptable; the secrecy degradation reaches above the value of* 10% *for many instances and even when the SNR is high.*

Example 4.3. *It has already been noted that the 0.1-secure distance is* 1.5λ *for the isotropic antennas and* 10λ *for the omnidirectional antennas. These results agree with the plot of* 4.5. *Assuming that the SNR at the legitimate users is at least 10dB, these distances achieve an essentially zero secrecy degradation (*$< 0.5\%$*).*

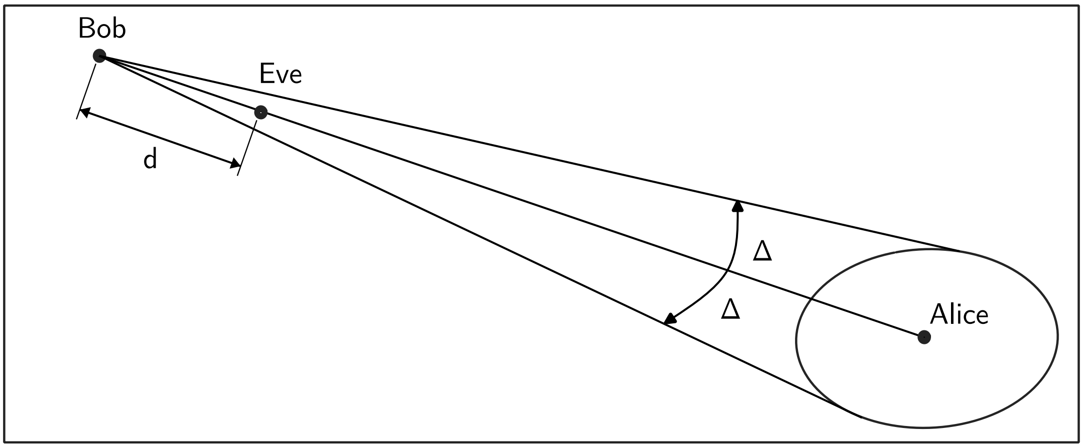


Figure 4.6: The restricted uniform AoA model. The spatial channel correlation maximises when Eve is inline with AB.

4.5 Non-omnidirectional environments

The lack of scatterers around a receiver restricts the AoA of the incoming signal to a subset of the set $[0, 2\pi)$. Among the different statistical models of directional channels, we choose the restricted uniform distribution model [152] which is a good fit in macrocell environments [25, 152]. Figure 4.6 shows a graphical representation of the geometrical model; Bob and Eve are placed in an open space whereas the transmitter, Alice, is surrounded by local scatters. The magnitude and AoA of each multipath observed by the receivers is uniformly distributed in the sets $[0, 1]$ and $[-\Delta + \phi, +\Delta + \phi]$, respectively.

To determine the secure distance of this model, it suffices to consider the most beneficial places for Eve in terms of maximising the spatial channel correlation for a given distance. Eve's best strategy is to be placed inline the communication link as seen in figure 4.6. For the inline placement, the spatial channel correlation (see 2.3.3) is simplified to

$$\rho(d) = R_{xx}(d) + jR_{xy}(d), \tag{4.26}$$

where

$$R_{xx}(d) = \frac{1}{2\Delta}\int_{\Delta}^{-\Delta} \cos\left(\frac{2\pi d}{\lambda}\sin(\phi_i)\right) \mathrm{d}\phi_i, \text{ and} \tag{4.27}$$

$$R_{xy}(d) = \frac{1}{2\Delta}\int_{\Delta}^{-\Delta} \sin\left(\frac{2\pi d}{\lambda}\sin(\phi_i)\right) \mathrm{d}\phi_i \tag{4.28}$$

Although 4.26 simplifies to a Bessel function of the second order when the two receivers are placed broadline [147], there is no known closed-form expression for the case of inline placement. This fact prevents the derivation of an upper bound for the minimum secure distance. The algorithm used for deriving the secure distance relies on the user's input that initiates the distance d. If the distance given is not ϵ- secure, i.e. $|\rho(d)| \geq \epsilon$ then the algorithm terminates and the user needs to repeat the process by entering a larger value. Once the input d is an ϵ-secure distance, the algorithm returns the minimum ϵ-secure distance.

Algorithm 2 Calculate the minimum Secure Distance for the restricted uniform AoA model

```
Inputs: $c$: step; $\lambda$: wavelength; $\epsilon$ : upper bound for $|\rho|$, $d$, $\Delta$
Outputs: minimum secure distance at $c$ accuracy
$\rho(d) := R_{xx}(d) + jR_{yy}(d)$
ensure $|\rho(d)| \geq \epsilon$
while $|\rho(d)| < \epsilon$ do
    $d_{s_o} \leftarrow d$                ▷ update minimum secure distance
    $d \leftarrow d - c$                  ▷ decrease distance by $c$
end while
Print $d_{s_o}$
```

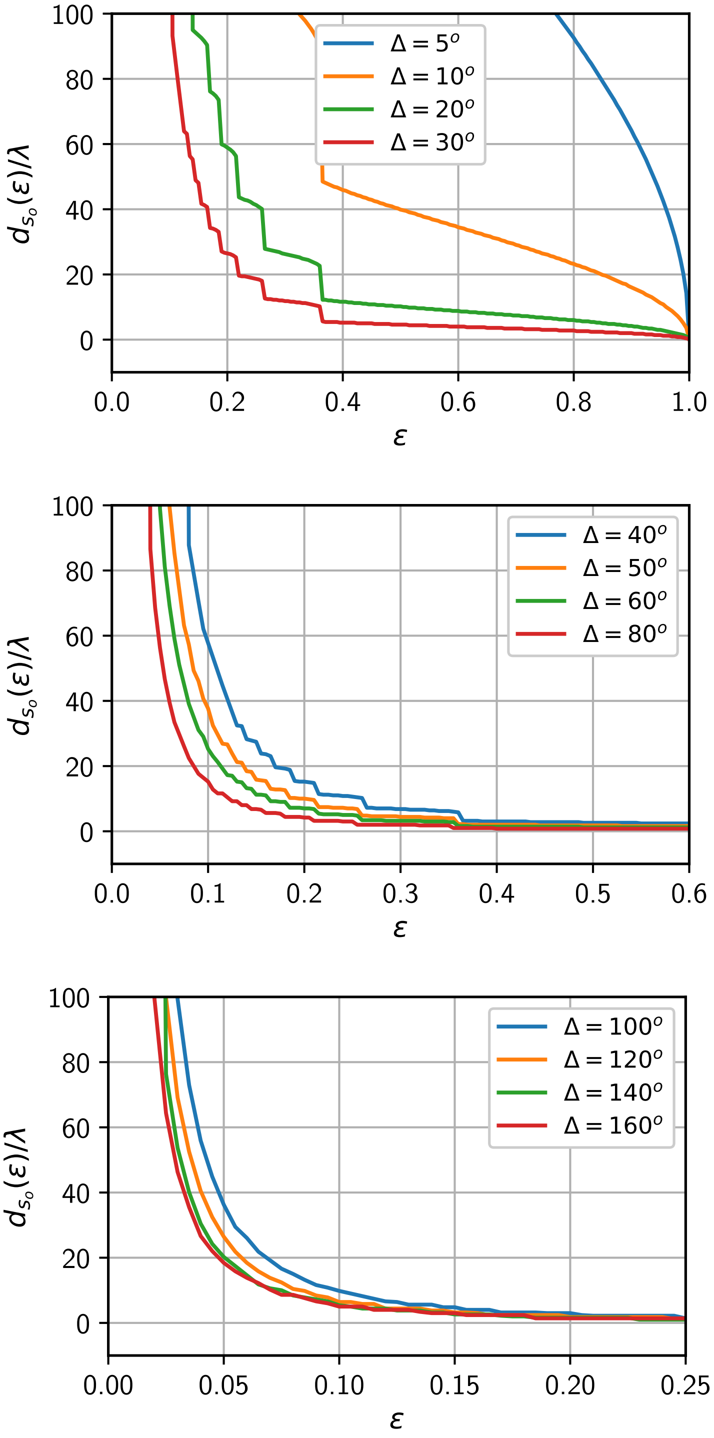


Figure 4.7: Minimum secure distance such that the spatial correlation falls below ϵ.

Figure 4.7 plots the minimum ϵ-secure distance normalised to the wavelength for different angles Δ. The unit for the angles has switched to degrees to avoid congestion in the legends. Recall that the beamwidth is double the size of angle Δ. The minimum secure distance is dramatically affected for small values of Δ. For example, when $\Delta = 20^o$, the minimum 0.5-secure distance is approximately eleven wavelengths, whereas doubling the angle results in a quadruple minimum 0.5-secure distance ($d_{s_o}(0.5) = 40\lambda$). When $\Delta = 5^o$, it takes 100 wavelengths for the maximal spatial correlation (ϵ) to drop below 0.5.

Recall that a spatial channel correlation as high as 0.5 results in a high secrecy degradation as explained in section 4.3. Next, we focus on wider beamwidths 4.7) and give numerical examples of minimum 0.1-secure distance.

Example 4.4. *Let the operating frequency be 2.4GHz, and let the AoA-beamwidth be* 80^0 *(*$\Delta = 40^o$*). To achieve a maximal spatial correlation of 0.1, the eavesdropper needs to be distanced by at least 7m (*57λ*) from Bob. If the beamwidth doubles in size (*160^o*), the 0.1-secure distance decreases to 2m (19λ).*

Example 4.5. *To achieve a minimum 0.1-secure distance of less than 1m* $d_{s_o}(0.1) < 1m$*), the beamwidth needs to be at least* 200^o *when the operating frequency is* f_c*=2.4GHz. When* f_c*=2.4GHz, a beamwidth of size* 160^o *suffices.*

4.6 Conclusion

We showed that a spatial channel correlation as high as 0.5 can significantly degrade the secrecy performance. Thus, we conclude that the typical definition of coherence distance is not appropriate for secrecy purposes. In order to effectively design a secure secrecy system, we suggest that the impact of spatial channel correlation must be considered so that the vulnerable key bits are dropped and the unpredictability of the key is restored.

For the case of a rich scattering environment whereby multiple scatterers surround the legitimate user close to the eavesdropper, a distance of a couple of wavelengths can be considered secure only when isotropic antennas are employed. For the practical case of dipole antennas, an eavesdropper needs to be placed at least ten wavelengths away from the legitimate users so that the secret key capacity is negligibly affected. For the case of directive channels, the secure distance increases dramatically as the angle-of-arrival deviates from being uniformly distributed across $(0, 2\pi]$.

CHAPTER 5

ENABLING A POSITIVE SECRECY GAP

Although key-based Physical Layer Security (PLS) is the dominant subfield of PLS in terms of employability in current systems (see section 3.2), it requires a dynamic channel. As seen in the literature review, a slow-fading channel results in a low-key rate, and the key rate is zero for static channels. Chapter 4 also demonstrated that typical key generation protocols are vulnerable when the spatial channel correlation remains high over long distances. This chapter suggests a method of confidential data transmission which does not rely on a dynamic channel or the spatial decorrelation property. The channel model of this chapter is fixed and non-varying. The presented method addresses the main challenge of keyless PLS which is the requirement of a positive secrecy gap (section 3.1.1.1). Based on secrecy coding (see section 3.1.1.1) and base station cooperation the method facilitates the transmission of small confidential messages such as keys. The main results and conclusions of this chapter have been published in [155].

Keywords: physical layer security, secrecy coding, base station cooperation, reverse training, maximal-ratio transmit beamforming.

Prerequisite: Wyner's wiretap channel 3.1.1.1, Fundamentals of Information Theory 2.1, Perfect Secrecy 2.2.1.1

5.1 Introduction

Secrecy coding is a branch of physical layer security that is not based on channel correlation but requires a positive secrecy gap, i.e. the eavesdropper's channel needs to be worse than the legitimate channel. In the presence of a passive eavesdropper whose Channel State Information (CSI) is unknown, secrecy coding fails to provide a design with information-theoretic guarantees. When the eavesdropper's CSI is unknown, typical methods of physical layer security use artificial noise in order to degrade the eavesdropper's channel.

This paper suggests an alternative way for achieving confidentiality which is based on Base-Station (BS) cooperation on the downlink as supported in 3GPP LTE-advanced [156]. BS cooperation is expected to be the core attribute of many 5G/6G technologies such as coordinated multipoint 5G communications [157–160], unmanned aerial vehicles, [161, 162], and reconfigurable intelligent surfaces [163, 164].

With the proposed method, each BS sends a sequence to the legitimate receiver who is able to reconstruct the information message by XoR-ing the received sequences. As long as the eavesdropper(s) is not at the same location as the legitimate receiver, there is a likelihood that one of the links will not be of high quality, thus she will not be able to acquire all sequences required for decoding the message. The proposed scheme has low complexity at the receiver and can be used in systems with finite-alphabet input, whereby most artificial noise-based schemes are ineffective.

Existed schemes of key-based PLS perform well when the channel between the legitimate pair fluctuates fast and practical implementations have already been recorded. However, in slow-fading channels, key-based PLS suffers from low secrecy-key rates; The secrecy key rate for a static channel is zero in static channels. Our analysis revolves around a single realisation of the fading channel coefficient, and as such, it takes no advantage of the channel's rate of fluctuations. With our scheme, it is possible to achieve a positive secrecy key rate even when the channel is static. Our channel model can be viewed as a special case of an Additive White Gaussian Noise channel, the capacity of which is given by $C = \log(1 + SNR)$ measured in bps (we set the channel bandwidth equal to one).

The drawback of our scheme is that the secrecy capacity is decreased as the number of BSs increases. For applications that demand a high data rate, the proposed scheme can be used in order to exchange symmetric keys. If the transmitter securely transmits a key to the legitimate receiver in the physical layer, the so-called Symmetric Key Cryptography (SKC) can be applied in the upper layers of the protocol stack. Note that a cross-layer approach that generates keys in the physical layer and enables SKC in the upper layer is believed by many to be a promising method to secure complex future wireless networks such as IoT, HetNets, and reduced cell-size 5G networks.

The scheme aims to provide a positive secrecy gap by degrading the eavesdropper's channel without the use of AN. Therefore, it can be applicable in communication systems where AN schemes are vulnerable, such as those with discrete-alphabet inputs and scenarios involving a multiple-antenna eavesdropper[1] [165]. We demonstrate that the deployment of multiple base stations (BSs) along with an encoding scheme called 'secret splitting' can significantly increase the probability of a positive secrecy gap and enable secure transmissions.

5.1.1 Related work

The main idea of the scheme, *secret splitting* (also known as *secret sharing*), has its origins in *network coding* whereby the confidential message is split into M 'splits' and are sent to the legitimate receiver through different paths. In Capar's work [166, 167], a large network of trusted relay nodes is considered and the splits (or shares) travel through parallel paths after appropriate relaying in a multi-hop network. Loosely speaking, parallel paths mean that the transmission links do not cross at any other location but only at the legitimate receiver. As such, the eavesdropper(s) will not acquire all 'splits' and will fail to decode the message.

Motivated by recent advancements in distributed massive-MIMO and BS cooperation, the work examines secret splitting under links created solely in the physical layer. In contrast to secret splitting in network coding, we do not examine the choice of paths/routes for which secrecy is guaranteed. Communications happen in a one-hop manner, the number of BSs is fixed as well as their location. As such, the paths may not be parallel in the sense that the secret splits may travel via beams that overlap. Lastly, our analysis revolves around a single realisation of the fading channel coefficients, and as such, it takes no advantage of the fading properties of the channel [57].

5.1.2 Organisation

Sections 5.2 defines and explains secret splitting and secrecy gap under secret splitting, respectively. In section 5.3, the probability of a positive secrecy gap is derived and analysed under a specific channel setting and transmission scheme. A comparison between conven-

[1] A multiple-antenna eavesdropper can use spatial diversity to separate the noise from the desired signal

tional wiretap coding and secret splitting follows in section 5.4 and numerical results are presented. The chapter concludes with a discussion in section 5.5.

5.2 Secret Splitting

Let $\mathbf{w}$ denote the confidential binary message of length k that Alice wishes to send to Bob in the presence of an eavesdropper, Eve. Alice is able to control M base stations, namely $A_1, A_2, \ldots, A_M$. Alternatively, A_i can also be considered to be a relay node with which Alice can communicate through a secure network.

The transmitter generates $M-1$ uniform independent binary sequences of length k, namely, $\mathbf{w_1}, \ldots, \mathbf{w_{M-1}}$. An M^{th} sequence is generated as

$$\mathbf{w}_M = \bigoplus_{i=1}^{M-1} \mathbf{w_i} \oplus \mathbf{w}. \tag{5.1}$$

We call $\{\mathbf{w}_i,\ i \in [M]\}$ the *secret splits* of $\mathbf{w}$. Secret split $\mathbf{w}_i$ is sent to Bob through base station A_i. After collecting all M secret splits, Bob XoRs the sequences and attains the confidential message. Indeed, it is evident from (5.1) that $\bigoplus_{i=1}^{M} \mathbf{w_i} = \mathbf{w}$.

Note that the confidential message $\mathbf{w}$ may not have a uniform distribution, e.g. it may correspond to an English word or to a user's predictable password. However, when random sequence $\bigoplus_{i=1}^{M-1} \mathbf{w_i}$ is XoRed to $\mathbf{w}$, the resulting split, $\mathbf{w}_M$, is also random and independent of $\mathbf{w}$. *Secret splitting* can also be thought of as a *one-time pad* encryption [1] with $\bigoplus_{i=1}^{M-1} \mathbf{w}_i$ being the secret key and $\mathbf{w}_M$ being the codeword.

Lemma 5.1. *Let $\mathscr{W}_s$ denote a proper subset of $\{\mathbf{w}_i,\ i \in [M]\}$. Then*

$$H(\mathbf{w}|\mathscr{W}_s) = H(\mathbf{w}). \tag{5.2}$$

Proof. If $\mathbf{w}_M \notin \mathscr{W}_s$, then

$$H(\mathbf{w}, \mathscr{W}_s) = H(\mathbf{w}) + H(\mathscr{W}_s),$$

since the choices of $\mathbf{w}_1, \ldots \mathbf{w}_{M-1}$ are independent of the choice of message $\mathbf{w}$.

We now examine the case when $\mathbf{w}_M \in \mathscr{W}_s$. Set $\mathscr{W}_s$ can be expressed as $\{\mathscr{W}_T, \mathbf{w}_M\}$, where $T \subset \{1, \ldots, M-1\}$. The complementary of T is $T^c = \{1, \ldots M-1\} \setminus T \neq \emptyset$. Then,

$$
\begin{aligned}
H(\mathbf{w},\mathscr{W}_s) &= H(\mathbf{w},\mathscr{W}_T,\mathbf{w}_M) \\
&= H(\mathbf{w},\mathscr{W}_T,\mathbf{w}_M \oplus \mathbf{w} \oplus \bigoplus_{i\in T}\mathbf{w}_i) \\
&= H(\mathbf{w},\mathscr{W}_T,\bigoplus_{i\in T^c}\mathbf{w}_i) \\
&= H(\mathbf{w}) + H(\mathscr{W}_T) + H(\bigoplus_{i\in T^c}\mathbf{w}_i) \qquad (5.3)\\
&= H(\mathbf{w}) + H(\mathscr{W}_T) + k. \qquad (5.4)
\end{aligned}
$$

We also have that,

$$
\begin{aligned}
H(\mathscr{W}_s) &= H(\mathscr{W}_T,\mathbf{w}_M) \\
&= H(\mathscr{W}_T,\mathbf{w}_M \oplus \bigoplus_{i\in T}\mathbf{w}_i) \\
&= H(\mathscr{W}_T,\mathbf{w}\bigoplus_{i\in T^c}\mathbf{w}_i) \\
&= H(\mathscr{W}_T) + H(\mathbf{w}\bigoplus_{i\in T^c}\mathbf{w}_i) \qquad (5.5)\\
&= H(\mathbf{w}) + H(\mathscr{W}_T) + k. \qquad (5.6)
\end{aligned}
$$

From (5.4) and (5.6), we have that $H(\mathbf{w},\mathscr{W}_s) = H(\mathbf{w}) + H(\mathscr{W}_s)$ which completes the proof.

□

Theorem 5.1. *As long as the eavesdropper attains less than M secret splits, she gains no information about the confidential message* $\mathbf{w}$*:*

$$I(\boldsymbol{w};\mathscr{W}_s) = 0 \quad \textit{for all}\, \mathscr{W}_s \subset \{\mathbf{w}_i, i \in [M]\}. \qquad (5.7)$$

Proof.

$$I(\mathbf{w};\mathscr{W}_s) := H(\mathbf{w}) - H(\mathbf{w}|\mathscr{W}_s) = 0. \qquad (5.8)$$

□

In information theoretical terms, when equation (5.7) is satisfied, *strong secrecy* is achieved which guarantees zero information leakage regardless of the length, k, of the message. That is, it only takes one weak link between the eavesdropper and a base station in order to achieve confidentiality.

5.2.1 Secrecy Gap

With appropriate wiretap coding, secret split $\mathbf{w}_i$ can be securely transmitted as long as $C_{B_i} - C_{E_i} > 0$. As theorem 5.1 implies, the secure transmission of one secret split is sufficient for securing message $\mathbf{w}$. Thus, for secrecy purposes, it is required that $C_{B_i} - C_{E_i} > 0$ for some $i \in [M]$. The latter is equivalent to requiring $\max_{i\in[M]}(C_{B_i} - C_{E_i}) > 0$ which motivates the following definition.

Definition 5.1. The secrecy gap under *secret splitting* is defined as

$$\mathrm{SG}_{split} := \max_{i\in[M]}(C_{B_i} - C_{E_i}) \tag{5.9}$$

The secret splits are transmitted by spatially separated transmitting nodes. Even if the eavesdropper is positioned very close to a transmitting node, her ability to decode is limited by her weakest channel. For instance, when there are two transmitting nodes, A_1 and A_2 (resulting in two secret splits), and the eavesdropper is in close proximity to A_1 such that $C_{E_1} > C_{E_2}$, the secrecy gap under secret splitting is defined as $\mathrm{SG}split = (C_{B_2} - C_{E_2})$. Therefore, secret splitting can reduce the eavesdropper's decoding capability without the necessity of generating artificial noise.

5.3 Secrecy Gap in Rayleigh Channels

When the secrecy gap SG_{split} is positive with a probability equal to one or zero, secure communication is possible, or not possible, respectively. When the channels are not deterministic but random processes, quantity $P\left(\mathrm{SG}_{split} > 0\right)$ can take any value in the interval $[0,1]$. This section studies the probability of a positive secrecy gap under a Rayleigh channel and transmit beamforming.

5.3.1 Channel Model

In our channel model, the legitimate receiver is a single-antenna device, whereas the adversary and transmitter may have multiple antennas. We denote by N_E and N_A the number of antennas at Eve and A_i, respectively. The base stations have the same number of antennas (N_A) for simplicity.

Vector $\mathbf{h}_i = (h_1^{(i)}, \dots h_{N_A}^{(i)}) \in \mathscr{C}^{1\times N_A}$, $i \in [M]$ comprises the channel coefficients $h_j^{(i)}$ of the channel between the j^{th} antenna of A_i and Bob. The matrix $\mathbf{G}_i = (\mathbf{g}_1^{(i)}, \dots, \mathbf{g}_{N_A}^{(i)}) \in \mathscr{C}^{N_E\times N_A}$

indicates the channel between Eve and base station A_i. Column $\mathbf{g}_j^{(i)}$ is the channel vector between base station A_i and the j^{th} antenna at Eve. All channels are assumed to be reciprocal, i.e. communication takes place in a time-division-duplex manner.

Bob's and Eve's channels are independent and drawn from a Rayleigh distribution:

$$\mathbf{h}_i \sim CN(\mathbf{0}_{N_A}, \sigma_{B_i}^2 \mathbf{I}_{N_A}) \text{ and } \mathbf{g}_j^{(i)} \sim CN(\mathbf{0}_{N_E}, \sigma_{E_i}^2 \mathbf{I}_{N_E}), \tag{5.10}$$

for all $i \in [M]$ and $j \in [N_E]$.

When base station A_i transmits $\mathbf{x} \in \mathbb{C}^{N_A \times 1}$, the received signal at Bob and Eve are given by

$$y = \mathbf{h}_i \mathbf{x} + n_B^{(i)} \quad \text{and} \quad \mathbf{z} = \mathbf{G}_i \mathbf{x} + \mathbf{n}_E^{(i)}, \tag{5.11}$$

respectively. Variables $n_B^{(i)}$ and $\mathbf{n}_E^{(i)}$ denote additive white Gaussian noise of zero mean and unit variance/covariance- matrix that vary independently for different $i \in [M]$ and from the transmission of one symbol to the other:

$$n_B^{(i)} \sim \mathscr{CN}(0,1) \quad \text{and} \quad \mathbf{n}_E^{(i)} \sim \mathscr{CN}(\mathbf{0}_{N_E}, \mathbf{I}_{N_E}). \tag{5.12}$$

5.3.2 Transmission scheme

5.3.2.1 Wiretap Coding and modulation

With Bob being a single-antenna node, the base stations transmit the secret splits successively. Before transmission, reliability and equivocation bits may be added to each one of the secret splits resulting in longer binary words. Modulation such as QAM or PSK modulation maps the binary words to a sequence of signals ready for transmission through the medium.

For example, after wiretap coding, secret split $\mathbf{w}_1$ is transmitted as $\mathbf{s}_1 = (s_1, \ldots, s_n) \in \mathbb{C}^{1 \times n}$ for some $n \in \mathbb{N}$. Note that the length n may differ at other BSs depending on the wiretap coding and modulation scheme used. Without loss for generality, the signal power is normalised to one: $\mathbb{E}(|s_j|^2) = 1$.

5.3.2.2 Transmit beamforming

Transmit beamforming is preferred for secrecy purposes since it avoids CSI leakage at the eavesdropper [168–170]. Being unaware of her own channel, the eavesdropper is unable to increase her decoding capabilities, e.g. by performing receive-beamforming. No CSI of the wiretap channel is available at Alice, either. For example, this is the case when the eavesdropper is passive and remains silent. Under this scenario, the best transmit beamforming strategy for secrecy purposes is Maximal-Ratio-Transmit (MRT) beamforming [171, Corr. 2]; With MRT the signal is sent towards the channel direction of the legitimate receiver and, as such, his SNR is maximised.

With the channel remaining static throughout the transmission of a secret split, the MRT beamforming vector $\mathbf{h}_1^H/||\mathbf{h}_1||$ is applied to every symbol of $\mathbf{s}_i = (s_1, \dots, s_n)$. To avoid a complicated notation, we drop the subscript at the symbols. When A_i transmits

$$\mathbf{x} = \frac{\mathbf{h}_i^H}{||\mathbf{h}_i||}s, \tag{5.13}$$

substitution in (5.11) shows that the received signals at Bob and Eve are

$$y_i = ||\mathbf{h}_i||s + n_B^{(i)} \quad \text{and} \tag{5.14}$$

$$\mathbf{z}_i = \frac{\mathbf{G}_i\mathbf{h}_i^H}{||\mathbf{h}_i||}s + \mathbf{n}_E^{(i)}, \tag{5.15}$$

respectively.

5.3.3 Probability of positive secrecy gap

Given the unit variance receiver-noise and the normalised to unit power signal, the average SNRs for sequence $\mathbf{s}_i$ at the two receivers are given by

$$\gamma_{B_i} = ||\mathbf{h}_i||^2 \quad \text{and } \gamma_{E_i} = ||\mathbf{G}_i\mathbf{h}_i^H||^2/||\mathbf{h}_i||^2 \tag{5.16}$$

Theorem 5.2. *Distribution of SNR at two receivers*

1. *Variable γ_{B_i} follows the Gamma distribution with shape parameter N_A and scale parameter $\sigma_{B_i}^2$:*

$$\gamma_{B_i} \sim \Gamma(N_A, \sigma_{B_i}^2). \tag{5.17}$$

2. *Variable γ_{E_i} is independent of γ_{Bi} and follows the gamma distribution with shape parameter N_{E_i} and scale parameter $\sigma_{E_i}^2$:*

$$\gamma_{E_i} \sim \Gamma(N_E, \sigma_{E_i}^2). \tag{5.18}$$

3. *The expected values of γ_{Bi} and γ_{E_i} are*

$$\bar{\gamma_{Bi}} := N_A\sigma_{B_i}^2 \textit{ and } \bar{\gamma_{E_i}} = N_E\sigma_{Ei}^2. \tag{5.19}$$

Proof. The components of $\mathbf{h}_{B_i}$ are complex numbers, the parts of which are independent and Gaussian distributed with zero mean and variance $\sigma^2_{B_i}/2$. As such, they are exponentially distributed with parameter $\sigma^2_{B_i}$.

The channel gain $||\mathbf{h}_{B_i}||^2$ is the summation of N_A independent exponential variables, and as such, it follows the Gamma distribution with shape parameter N_A and scale parameter $\sigma^2_{B_i}$: $||\mathbf{h}||^2 \sim \Gamma(N_A, \sigma^2_{B_i})$. □

The probability density function (p.d.f.) and cumulative density function (c.d.f.) of γ_{B_i} are given by:

$$f_{B_i}(\gamma_{Bi}) = \frac{\gamma_{Bi}^{N_A-1}\exp\{-\gamma_{Bi}/\sigma_B^2\}}{\sigma_B^{2N_A}(N_A-1)!} \qquad (p.d.f) \tag{5.20}$$

$$F_{B_i}(\gamma_{Bi}) = \frac{1}{(N_A-1)!}\gamma_{\text{inc}}(N_A, \gamma_{Bi}/(\sigma^2_{Bi})) \qquad (c.d.f) \tag{5.21}$$

and the variance of γ_{Bi} is

$$\mathbb{V}(\gamma_{Bi}) = N_A\sigma_B^4. \tag{5.22}$$

Observe that both functions increase linearly with the number of antennas employed by Alice.

Proof. From (5.15), the SNR at Eve, denoted by γ_{E_i}, is

$$\gamma_{E_i} = \frac{\mathbf{G}_i\mathbf{h}_i^H}{||\mathbf{h}_i||} = \sum_{j=1}^{N_E}\frac{||\mathbf{g}_j^{(i)}\mathbf{h}_i^H||^2}{||\mathbf{h}_i||^2}, \tag{5.23}$$

where $\mathbf{g}_j^{(i)}$ denotes the j^{th} row of $\mathbf{G}_i$.

Observe that $||\mathbf{g}_j^{(i)}\mathbf{h}_i^H||^2 = \mathbf{g}_j^{(i)}\cdot\mathbf{h}_i$, where '·' is the inner product of two vectors. Since Bob's and Eve's channels are independent Rayleigh channels, their channel vectors can be expressed as

$$\mathbf{g}_j^{(i)} = ||\mathbf{g}_j^{(i)}||\mathbf{u}_i \quad \text{and} \quad \mathbf{h}_i = ||\mathbf{h}_i||\mathbf{v}_i,$$

where $\mathbf{u}_i, \mathbf{v}_i \in \mathscr{C}^{N_A}$ are two unit random vectors, uniformly distributed in the N_A-1 dimensional complex sphere $\mathscr{S}^{N_A-1}$. Indeed, the four quantities $(||\mathbf{g}_j^{(i)}||, \mathbf{u}_i, ||\mathbf{h}_i||, \mathbf{v}_i)$ are independent.

Eve's SNR can be simplified to

$$\gamma_{E_i} = \sum_{j=1}^{N_E}||\mathbf{g}_j^{(i)}||^2(\mathbf{u}_i\cdot\mathbf{v}_i)^2, \tag{5.24}$$

which is therefore independent of $||\mathbf{h}_i||$, and hence independent of Bob's SNR.

Observe that for a fixed $\mathbf{v}_i$, the geometrical interpretation of $\mathbf{u}_i \cdot \mathbf{v}$ is the projection of $\mathbf{u}_i$ to the direction of $\mathbf{v}_i$. The distribution of the projection is independent of the choice of $\mathbf{v}_i$. Hence, $f(\gamma_{E_i}|\gamma_{Bi}) = f(\gamma_{E_i}|\mathbf{v}_i) = f(\gamma_{E_i})$ which proves the independence.

Similar to $||\mathbf{h}_i||^2$, $||\mathbf{g}_j^{(i)}||^2$ follows the gamma distribution:

$$||\mathbf{g}_j^{(i)}||^2 \sim \Gamma(N_A, \sigma^2_{E_i}).$$

The distribution of $(\mathbf{u} \cdot \mathbf{v})^2$ is a Beta distribution with shape parameters 1 and $N_A - 1$. I.e.,

$$(\mathbf{u} \cdot \mathbf{v})^2 \sim \mathrm{B}(1, N_A - 1).$$

According to [172, theo 1], the product of two random variables that follow a $\Gamma(N_A, \sigma^2_{Ei})$ distribution and a $\mathrm{B}(1, N_A - 1)$ is a Γ distribution with shape 1 and rate σ^2_{Ei}. Lastly, the summation of N independent and identical distributed Gamma functions is also a Gamma function with parameters N_E and σ^2_{Ei}.

$$\gamma_{E_i} \sim \Gamma(N_E, \sigma^2_{Ei}) \tag{5.25}$$

□

As such, the p.d.f. and c.d.f. of γ_{E_i} are given by:

$$f_{E_i}(\gamma_{E_i}) = \frac{\gamma_{E_i}^{N_E - 1} \exp\{-\gamma_{E_i} / \sigma^2_{Ei}\}}{\sigma_E^{2N_E}(N_E - 1)!} \qquad (p.d.f) \tag{5.26}$$

$$\tag{5.27}$$

and the variance of γ_{E_i} is

$$\mathbb{V}(\gamma_{E_i}) = N_E^2 \sigma^4_{E_i}. \tag{5.28}$$

Observe that Bob's average SNR is a linear function of N_A whilst Eve's average SNR is a linear function of N_E. Only Bob benefits from an increase in the number of antennas at Alice.

With Eve's SNR and Bob's SNR being independent, their joint p.d.f., f_{BE_i}, is equal to:

$$f_{BE_i}(\gamma_{Bi}, \gamma_{E_i}) = f_{B_i}(\gamma_{Bi}) f_{E_i}(\gamma_{E_i}). \tag{5.29}$$

Figure 5.2 plots the joint p.d.f. for the case whereby Bob and Alice have both single antenna devices. The theoretical p.d.f. is validated by empirical p.d.f.

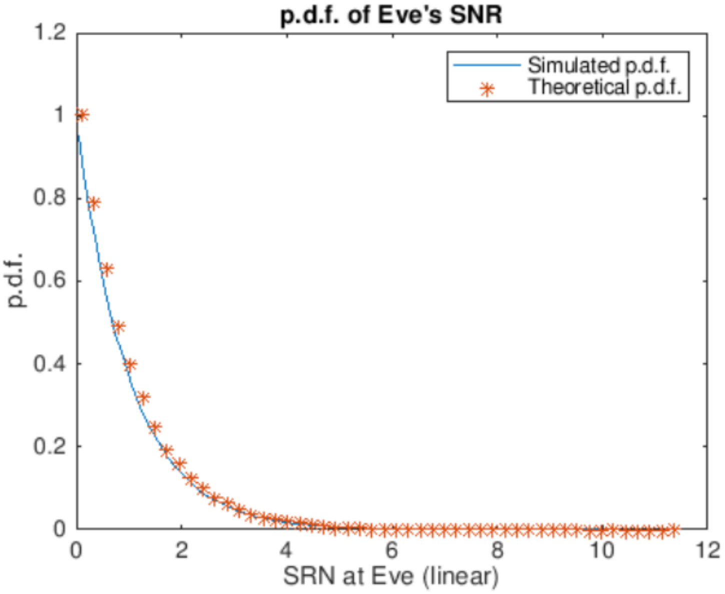


Figure 5.1: P.d.f. of Eve's SNR for = 1, $\sigma^2_{Ei} = 1$, and $N_E = 1$. The graph is invariant to N_A;

Theorem 5.3. *The probability of positive secrecy gap under secret splitting is*

$$P\left(SG_{split} > 0\right) = 1 - \prod_{i=1}^{M} P\left(\gamma_{E_i} \geq \gamma_{B_i}\right). \tag{5.30}$$

Proof.

$$P\left(\text{SG}_{split} > 0\right) = P\left(\max_{i \in [M]}(C_{B_i} - C_{E_i}) \leq 0\right)$$
$$P\left(\cap_{i=1}^{M}\{C_{B_i} - C_{E_i} \leq 0\}\right)$$

By invoking the independence of γ_{E_i} and γ_{B_i}, we have that $P\left(\text{SG}_{split} \leq 0\right) = \prod P\left(\gamma_{E_i} \geq \gamma_{B_i}\right)$. □

Theorem 5.4. *The probability of positive secrecy gap under secret splitting, MRT, and independent Rayleigh channels is equal to:*

$$P\left(SG_{split} > 0\right) = 1 - \prod_{i=1}^{M} \int_0^\infty \frac{{\gamma_{B_i}}^{N_A-1} \exp\left(\frac{-\gamma_{B_i}}{\sigma^2_{B_i}}\right) \sum_{k=1}^{N_E-1} \frac{1}{k!}\left(\frac{\gamma_{B_i}}{\sigma^2_{E_i}}\right)^k}{{\sigma_{B_i}}^{2N_A}(N_A-1)!} d\gamma_{B_i} \tag{5.31}$$

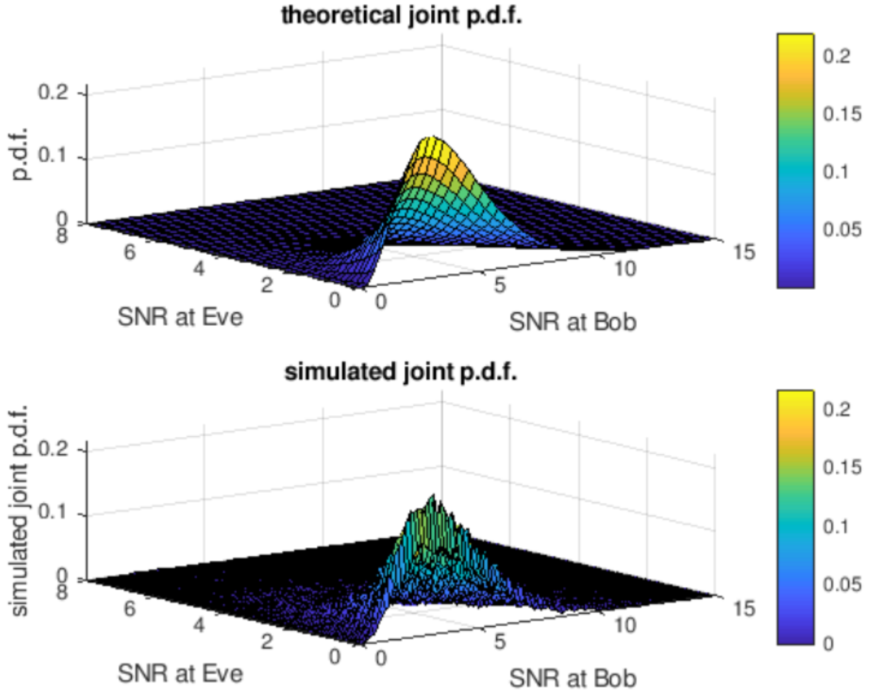


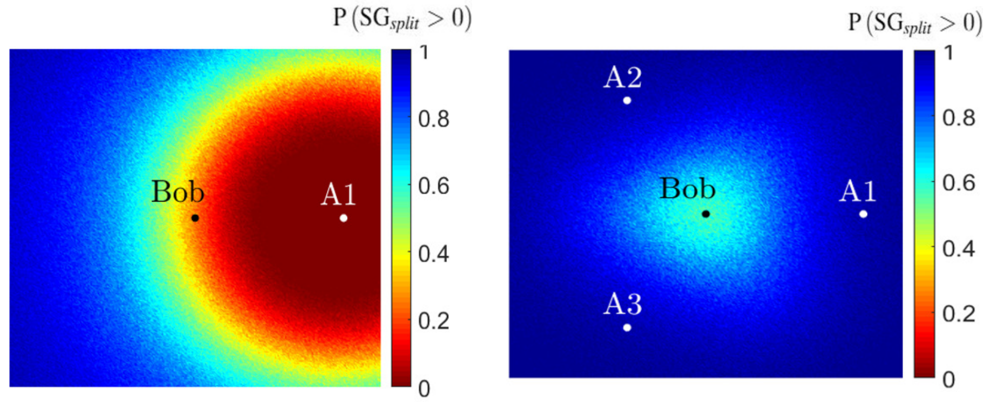
Figure 5.2: Joint p.d.f. over Bob's and Eve's SNR plane. Parameters are set as: $\sigma^2_{Bi} = \sigma^2_{Ei} = 1,\ N_A = 4,\ N_E = 1$.

Proof.

$$
\begin{aligned}
P\left(\gamma_{Bi} > \gamma_{E_i}\right) &= \int_0^\infty \int_{\gamma_{Bi}}^\infty f_{BE}(\gamma_{Bi}, \gamma_{E_i}) d\gamma_{E_i} d\gamma_{Bi} \\
&\overset{(5.29)}{=} \frac{\int_0^\infty {\gamma_{Bi}}^{N_A-1} \exp\left(\frac{-\gamma_{Bi}}{\sigma^2_{Bi}}\right) \int_{\gamma_{Bi}}^\infty {\gamma_E}^{N_E-1} \exp\left(\frac{-\gamma_{E_i}}{\sigma^2_{Ei}}\right) d\gamma_{E_i} d\gamma_{Bi}}{{\sigma^2_{Bi}}^{2N_A} {\sigma^2_{Bi}}^{2N_E} (N_A-1)!(N_E-1)!} \\
&= \frac{\int_0^\infty {\gamma_{Bi}}^{N_A-1} \exp\left(\frac{-\gamma_{Bi}}{\sigma^2_{Bi}}\right) \Gamma_{\text{inc}}\left(N_E, \frac{\gamma_{Bi}}{\sigma^2_{Bi}}\right) d\gamma_{Bi}}{{\sigma^2_{Bi}}^{2N_A} (N_A-1)!(N_E-1)!}
\end{aligned}
$$

Expressing the incomplete upper Gamma function Γ_{inc} as a summation according to [173, equation 2] and applying lemma 5.3 completes the proof. □

Note that the integration in Eq. (5.4) is with respect to γ_{Bi}. As such, the probability of $P\left(\text{SG}_{split} > 0\right)$ is a function of the channel statistics, σ^2_{Bi} and σ^2_{Ei}, and the number of antennas, N_A and N_E.



(a) $M = 1, N_A = 1, N_E = 2$

(b) $M = 3, N_A = 1, N_E = 2$

Figure 5.3: The red colour indicates areas at which a 2-antenna adversary node has a better signal than Bob with high probability. The blue colour indicates the opposite.

Corollary 5.1. *When Eve is a single antenna node ($N_E = 1$), Eq.* (5.31) *can be expressed as*

$$P\left(SG_{split} > 0\right) = 1 - \prod_{i=1}^{M}\left(1 + \frac{\sigma_{B_i}^2}{\sigma_{E_i}^2}\right)^{-N_A}. \tag{5.32}$$

Proof. The cor follows after substituting $N_E = 1$ in Eq. (5.31) and using the standard result of $\int_0^\infty x^n e^{-\alpha x} dx = n!/\alpha^{n=1}$. □

From a user's location point of view, by invoking the relationship between average signal power and distance [174], the channel statistics can be expressed as

$$\sigma_{Bi}^2 = k/d(A_i, B)^\alpha \text{ and } \sigma_{Ei}^2 = k/d(A_i, E)^\alpha, \tag{5.33}$$

for some $k \in \mathbb{R}$, where $d(A_i, B)/d(A_i, E)$ is the distance between A_i and Bob/Eve and α is the path-loss exponent. For example, Eq. (5.32) is equivalent to

$$P\left(\text{SG}_{split} > 0\right) = 1 - \prod_{i=1}^{M}\left(1 + \left(\frac{d(A_i, E)}{d(A_i, B)}\right)^\alpha\right)^{-N_A}. \tag{5.34}$$

Although the probability of a positive secrecy gap is a function of the path-loss exponent α, the differences in the graphs for different values of $\alpha \in [3, 5]$ were hardly noticeable. All numerical results of this paper consider the case when $\alpha = 4$ only.

In Figure 5.3.3 the red area indicates the locations at which the eavesdropper has an advantage over Bob. i.e. locations at which the probability of a positive gap is low. When three single-antenna BSs (or relay nodes) are employed, the likelihood that Eve attains a better signal than Bob is decreased dramatically (Fig. 5.3.3).

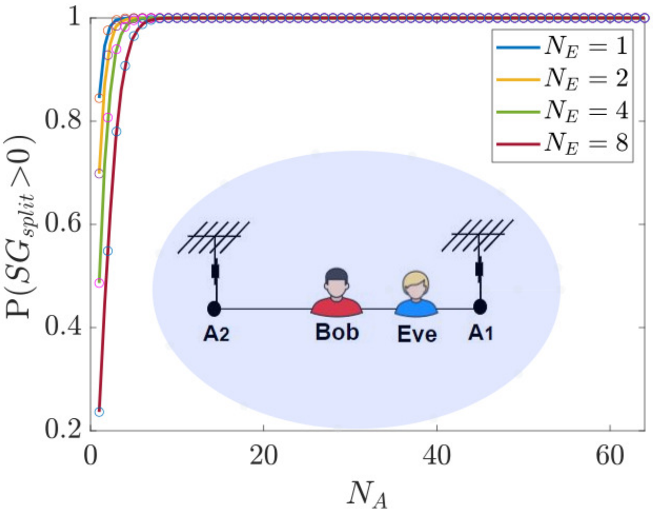


Figure 5.4: Probability of positive/negative SG$_{split}$ against N_A when two BSs are employed ($M = 2$) and Eve is at the middle between Bob and A_1. Solid lines/scattered plots are derived theoretically/empirically.

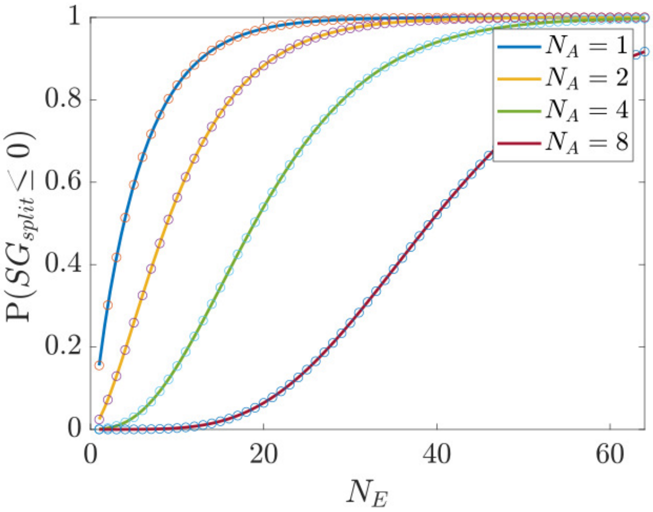


Figure 5.5: Probability of positive/negative SG$_{split}$ against N_E when two BSs are employed ($M = 2$) and Eve is at the middle between Bob and A_1. Solid lines/scattered plots are derived theoretically/empirically.

5.3.4 Asymptotic behaviour

It is evident from Th. 5.3 that the probability of positive secrecy gap is an increasing function of the number of base-stations, M; If Eve is equipped with a finite number of antennas then it is a strictly increasing function. In the latter case, when M becomes asymptotically large, the secrecy gap under secret splitting is positive with probability one:

$$\lim_{M\to\infty} P\left(\mathrm{SG}_{split} > 0\right) = 1. \tag{5.35}$$

On the other hand, for a fixed number BSs, M, secure communication is not possible when $N_E \to \infty$. Indeed, with an asymptotically large number of antennas available at Eve only, she always experiences a better SNR than Bob. Since $P\left(\gamma_{E_i} \geq \gamma_B\right) = 1$ for all $i \in [M]$, it follows that

$$\lim_{N_E\to\infty} P\left(\mathrm{SG}_{split} \leq 0\right) = 1. \tag{5.36}$$

Consider the metrics $P\left(\mathrm{SG}_{split} > 0\right)$ and $P\left(\mathrm{SG}_{split} \leq 0\right)$, i.e. the probability of Bob being successful and Eve successful in terms of achieving a better signal, respectively. For the setting as illustrated in Fig. 5.3.3 whereby two base stations are deployed, the probability of Bob being successful converges much faster than the probability of Eve being successful. Indeed, even when the adversary is equipped with $N_E = 8$ antennas, ten antennas at each BS is sufficient to provide a positive secrecy gap with probability approximate to one (0.9999). On the other hand, when $N_A = 8$, the eavesdropper needs at least forty antennas for a 50% chance to get a better signal than the legitimate receiver (Fig. 5.3.3).

Lastly, for the case when $N_A \to \infty$, it is evident from Eq. (5.17) and (5.18) that the employment of an infinite number of antennas N_A increases Bob's SNR asymptotically. As such, for a fixed number of antennas at Eve, we have that $P\left(\gamma_{E_i} \geq \gamma_{Bi}\right) = 0$ for all $i \in [M]$ which results in a certain positive secrecy gap:

$$\lim_{N_A\to\infty} P\left(SG > 0\right) = 1. \tag{5.37}$$

The above equation implies that the employment of a single base station and conventional wiretap coding are sufficient to secure the communication from Alice to Bob when N_A is asymptotically large.

5.3.5 Secrecy Outage Probability

For secure transmissions, the condition $\mathrm{SG}_{split} > 0$ is necessary but not sufficient. Depending on the choice of the equivocation rates, information leakage may occur even when

$\text{SG}_{split} > 0$. Recall that the eavesdropper gains no information as long as one secret split is secure (Th. 5.1). As such, the secrecy requirement of our scheme is the following:

$$\textit{There exists } i \in [M] \textit{ such that } C_{E_i} < R_{E_i} \quad \text{(sec. req.)} \tag{5.38}$$

When the above equation is not satisfied, we say that an *outage event* occurs.

Definition 5.2. We define the *secrecy performance metric* as the probability of a secrecy outage event:

$$\mathscr{P}_{\text{out}}^{(s)} := P\left(\cup_{i=1}^{M} C_{E_i} \geq R_{E_i}\right) \tag{5.39}$$

As the secrecy performance metric is defined as the probability of an event, it does not itself guarantee secrecy. Nevertheless, it serves as a valuable metric for comparison between the conventional method and secrecy splitting.

With γ_{E_i}'s, being independent, so do the channel capacities C_{E_i}'s. For our channel model, Def. 5.2 is equivalent to

$$\mathscr{P}_{\text{out}}^{(s)} := \prod_{i=1}^{M} P\left(C_{E_i} \geq R_{E_i}\right) \tag{5.40}$$

A direct result of the definition is that the secrecy outage probability is a decreasing function of M. For an asymptotically large number of base stations and as long as Eve employees a finite number of antennas, the probability of a secrecy outage is zero.

$$\lim_{M\to\infty} \mathscr{P}_{\text{out}}^{(s)} = \begin{cases} 0 & \text{if } N_E \not\to \infty \\ 1 & \text{if } N_E \to \infty \end{cases} \tag{5.41}$$

Theorem 5.5. *Under secret splitting, MRT, and independent Rayleigh channels, the secrecy performance metric is given by*

$$\mathscr{P}_{out}^{(s)} = \prod_{i}^{M} \left(1 - \frac{\gamma_{inc}(N_E, (2^{R_E} - 1)/(p_i \sigma_{E_i}^2))}{(N_E - 1)!}\right) \tag{5.42}$$

Proof. Direct after expressing Eq. (5.2) as

$$\mathscr{P}_{\text{out}}^{(s)} = \prod_{i}^{M} \left(1 - F_{E_i}(R_{E_i})\right). \tag{5.43}$$

□

5.3.6 Secrecy Capacity

Without knowledge of the wiretap channel, the secrecy capacity cannot be derived. However, when $\min C_{E_i}$ is known and when $\min C_{E_i} < \min C_{B_i}$, we show that an explicit formula can be given as well as the strategy that achieves the secrecy capacity. Having assumed perfect main CSI, Alice utilises the base stations A_i, $i \in [M]$ for which $C_{B_i} > \min C_{E_i}$ only. Thus, the assumption of $C_{B_i} > \min C_{E_i}$ is coherent.

Recall that the secrecy capacity is the maximum achievable secrecy rate. As such, it is essential to derive the secrecy rate for our scheme first.

The secrecy rate at which secret split $\mathbf{w}_i$ was wiretap-coded is $R_{S_i} = R_{B_i} - R_{E_i}$. Individual *secret splits* do not convey any information. It takes M transmissions to transmit confidential message $\mathbf{w}$. Since the length of secret split is k, k/R_{S_i} seconds are required for Bob to attain $\mathbf{w_i}$. Due to successive transmission, $\sum_{i=1}^{M} k/R_{S_i}$ seconds are required in total for attaining message $\mathbf{w} = \bigoplus_{i=1}^{M} \mathbf{w}_i$. With k also being the length of confidential message $\mathbf{w}$, the secrecy rate for our scheme is equal to

$$R_S = \left(\sum_{i=1}^{M} \frac{1}{R_{S_i}}\right)^{-1} = \left(\sum_{i=1}^{M} \frac{1}{R_{B_i} - R_{E_i}}\right)^{-1}. \tag{5.44}$$

Recall that secrecy capacity is the maximum rate at which secret $\mathbf{w}$ is transmitted both reliably and securely. To derive an explicit formula for the secrecy capacity, it is essential to state the reliability and secrecy requirements of our scheme first.

Lemma 5.2. *If* $\min_i C_{E_i} \leq R_E$*, strong secrecy is achieved.*

Proof. Without loss of generality let $\min_i C_{E_i} = C_{E_1}$. Since $C_{E_1} \leq R_E$, Eve is unable to attain secret split w_1. With Eve obtaining only a subset of secret splits, Eq. (5.7) is satisfied and $\mathbf{w}$ is strongly secure. □

Hence, for secrecy constraints, we require

$$\min_{i \in [M]} C_{E_i} \leq R_E \qquad \text{(secrecy requirement)} \tag{5.45}$$

In order to maximise the secrecy rate, R_S, we need to maximise the differences $R_{B_i} - R_{E_i}$, $i \in [M]$ under reliability and secrecy constraints. As per Shannon [1], the maximum transmission rate for reliable transmissions is equal to the channel's capacity: We set $R_{B_i} = C_{B_i}$, for all $i \in [M]$.

As for the choice of the equivocation rates, observe that the fixed value of $R_{E_i} = \min C_{E_i}$ for all $i \in [M]$ satisfies the secrecy requirement (Eq. (5.38)). Choosing zero equivocation rates for the $M-1$ least noisy wiretap channel and $\min C_{E_i}$ for the channel A_i, $i = \arg\min C_{E_i}$ would also satisfy the secrecy requirement and increase the secrecy rate, but it would not guarantee secrecy: Alice does not know which channel is the worst for Eve. As such, she fixes the equivocation rate to $\min C_{E_i}$ for all channels so that the secrecy requirement is met with probability one. We have showed that the maximum secrecy rate under $M-$secret splitting is achieved when $R_{B_i} = C_{B_i}$ and $R_{E_i} = \min_i C_{E_i}$, for all $i \in [M]$.

Theorem 5.6. *When* $\min C_{E_i} < \min C_{B_i}$*, the secrecy capacity under $M-$secret splitting is equal to*

$$C_S = \left(\sum_{i=1}^{M} (C_{B_i} - \min_{i \in [M]} C_{E_i})^{-1} \right)^{-1}. \tag{5.46}$$

When $M = 1$, Eq. (5.46) simplifies to the classical definition of secrecy capacity for the case $C_{B_1} > C_{E_1}$ [9]. For cases when $C_S = 0$ and $M = 1$, adding a second base station will increase the secrecy capacity with high probability. However, for $M > 2$, the secrecy capacity is (always) a decreasing function of M. When choosing $M > 2$, the trade-off between secrecy capacity and probability of secrecy outage events must be taken into consideration.

Another important consideration when determining M for practical applications is the overhead costs associated with BS cooperation. BS cooperation necessitates additional signalling between cooperating base stations and the core network. This signalling involves exchanging information about user equipment, resource allocation, and coordination. Moreover, coordinating data transmission and reception between multiple base stations will introduce data overhead. This includes the need to exchange and process data, leading to increased data transmission. It is worthwhile to note that the numerical results in the following section do not account for overhead costs, simplifying the analysis.

5.4 Base Station Allocation and Numerical Results

5.4.1 Secret Splitting Vs Conventional Wiretap Coding

With the probability of a positive secrecy gap being an increasing function of both N_A and M, the question arising is whether giving Alice more antennas is more beneficial than employing more BSs for secrecy purposes or vice versa. Besides, when taking into account the transmission rate, a small M is preferred given that Bob is a single antenna device and receives the secret splits successively. Two strategies are considered:

Strg 1: Alice employs $M > 1$ base stations each equipped with K antennas. ($M > 1$ & $N_A = K$).

Strg 2: Alice employs one base station with MK antennas.

($M = 1$ & $N_A = MK$).

The first strategy is referred to as the M-secret splitting strategy, whereas the trivial case ($M = 1$) is the case of conventional wiretap coding. The total number of antennas is MK for both cases to facilitate comparison. Whether the first strategy outperforms the second in terms of providing a positive secrecy gap depends on the channel statistics of the two receivers. It can be shown that conventional wiretap coding outperforms secret splitting when the eavesdropper channel or location is known at Alice.

When the wiretap CSI is known at Alice, M-secret splitting is unnecessary: Alice can simply transmit with the BS that maximises the ratio $\gamma_{B_i}/\gamma_{E_i} > 1$. Even when only the location of the eavesdropper is known, the trivial case whereby Alice transmits with the BS minimises the ratio of the distances $d(A_i, B)/d(A_i, E)$ maximises the probability of a positive secrecy gap. However, in a practical scenario, the location of a passive eavesdropper is unknown. It will be shown that, in the case of a passive eavesdropper, M-secret splitting is a better strategy in terms of secrecy.

With no information on the eavesdropper's location, E, the comparison between the two strategies will be made by evaluating the average performance, $\mathscr{P}_0$, over a set of possible locations for Eve, $\mathscr{E}$:

$$\mathscr{P}_0 := \mathbb{E}[P\left(\mathrm{SG}_{split} > 0 | E \in \mathscr{E}\right)]. \tag{5.47}$$

The set of possible locations, $\mathscr{E}$, is taken to be either the interior of a square or the interior of a circle:

- $\mathscr{E} = C(B, \rho_E)$: the interior of the circle of radius ρ_E and centre B, i.e. Bob's location, or
- $\mathscr{E} = S(B, \rho_E)$: the interior of some square of base $2\rho_E$ and centre B.

Due to the infinite cardinality of the sets and the complexity of the formulae, the evaluation of the performance $\mathscr{P}_0$ will be derived empirically by sampling the eavesdropper's location in $\mathscr{E}$ uniformly. Note that in this paper Eve and Bob lie on the same plane. The simulation methods have been validated a priori by considering discrete sets of small cardinality for which the theoretical results matched the empirical ones.

With Bob being at the origin, $B(0,0)$, of a polar coordinate system, let A_i be placed at $A_i(\rho_{A_i}, \theta_{A_i})$. As Fig. 5.4.1 demonstrates, the average performance under 2-secret splitting

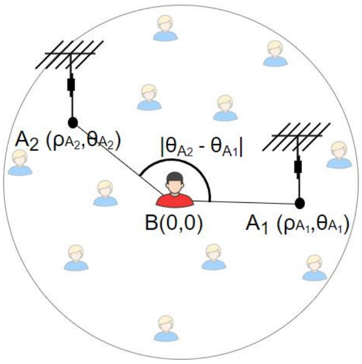


Figure 5.6: The set of possible locations for Eve is $\mathscr{E} = C(B, 1.5\rho_{A_1})$ (5.4.1).

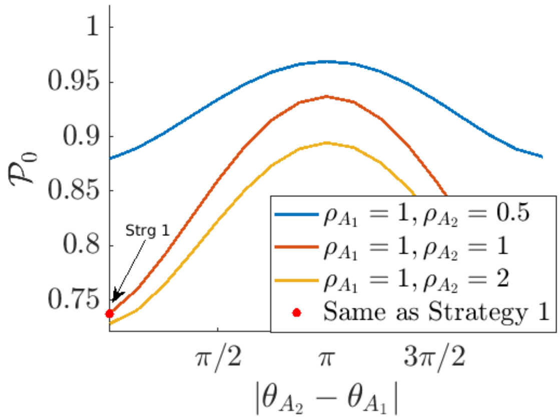


Figure 5.7: The average performance:
$\mathscr{P}_0 = \mathbb{E}[P\left(\text{SG}_{split} > 0 | E \in \mathscr{E}\right)]$ against the angle difference of the two BSs $|\theta_{A_1} - \theta_{A_2}|$ (5.4.1).

increases as the difference of the angles of the two base stations approaches π. The angle difference of $|\theta_{A_2} - \theta_{A_1}| = \pi$ will be referred to as the optimal angle-difference. Observe that a near-optimal angle difference (e.g., $\pi \pm \pi/4$) achieves performances near the maximum. This is an encouraging result for real-life communication systems when considering that the angle difference will most likely differ from the optimal.

The performance of conventional wiretap coding can also be extracted from the graph in Fig. 5.4.1. When $\mathscr{E} = C(0, 1.5)$ and $N_E = 1$ Strategy 1 achieves a positive secrecy gap with probability $\mathscr{P}_0 = 0.73$. As for the second strategy, even when A_2 is placed at double the distance from Bob than A_1 ($\rho_{A_2} = 2$), the probability, $\mathscr{P}_0$, increases remarkably (up to 27%).

Table 5.1 lists five examples for a set of different parameters. The average performance

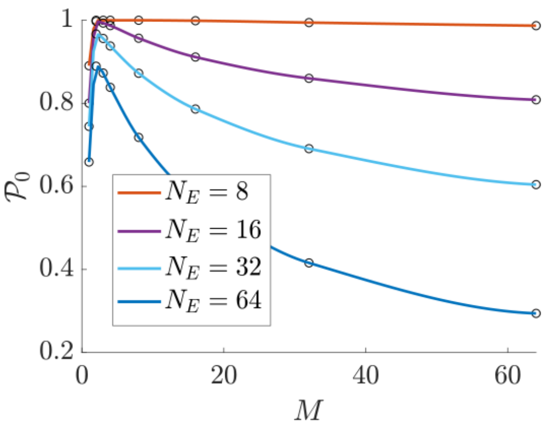


Figure 5.8: Average performance:
$\mathscr{P}_0 = \mathbb{E}[P\left(\text{SG}_{split} > 0|E \in \mathscr{E}\right)]$ against the number of BSs (M) when the total number of antennas is fixed to $\sum N_A = 64$. The set of the eavesdropper's possible locations is the square $S(B,1.5)$ and for every M, the BSs are placed optimally at distance one from Bob.

has been evaluated over the circle $C(B,1.5\rho_{A_1})$. For the case when $M = 2$, the second base station is placed at distance $\rho_{A_2} = \rho_{A_1} = 1$ from Bob as illustrated in Fig. 5.4.1. Column 'optimal' lists the average performance, $\mathscr{P}_0$, when the BSs are placed diametrically opposed to Bob ($|\theta_{A_2} - \theta_{A_1}| = \pi$). The average performance is also recorded for the case when the angle-difference differs far from the optimal: $|\theta_{A_1} - \theta_{A_2}| = 2\pi/3$. For all cases, Strategy 1 is the best strategy in terms of providing a positive secrecy gap.

Table 5.1:
Average Performance $\mathscr{P}_0 = P\left(\text{SG}_{split} > 0|E \in \mathscr{E}\right)$ under Strategy 1 and Strategy 2. Bob is at the origin $B(0,0)$ and the set of possible locations for Eve is the circle $C(B,1.5)$. Two cases are considered in Strategy 1: (a) the two BSs are at $A_1(1,0)$ and $A_2(1,\pi)$ forming an 'optimal' angle with Bob (b) the two BSs are at $A_1(1,0)$ and $A_2(1,3\pi/4)$ forming a 'non-optimal' angle.

$\mathscr{P}_0$	**Strategy 1:** **2 BSs** with **K antennas** each		**Strategy 2:** **1 BS** with **2K antennas**
	optimal	**non-optimal**	
$\mathbf{K=2,\ N_E=1}$	0.981	0.992	0.815
$\mathbf{K=3,\ N_E=1}$	0.999	0.995	0.846
$\mathbf{K=32,\ N_E=1}$	1.00	1.00	0.950
$\mathbf{K=2,\ N_E=64}$	0.098	0.097	0.084
$\mathbf{K=32, N_E=64}$	0.950	0.894	0.595

Observe that even when Eve is a single-antenna node, beamforming with $2K = 64$ antennas at one BS does not guarantee a positive secrecy gap ($\mathscr{P}_0 = .95$). On the other hand,

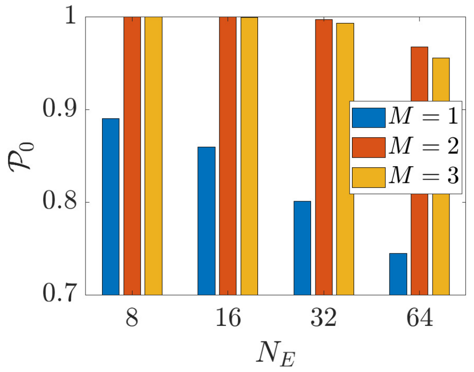


Figure 5.9: A closer look at the data of Fig. 5.8 for the cases where $M \leq 3$ BSs are employed. The performance, $\mathscr{P}_0$, is plotted against the number of antennas at the eavesdropper.

distributing the antennas in two BSs (case $K = 32$, $N_E = 1$) results in a positive secrecy gap with probability one. Simulations suggest that when the eavesdropper is a single-antenna node, two BSs with just three antennas each can almost certainly provide the legitimate pair with a positive secrecy gap (case $K = 3, N_E = 1$). Lastly, both strategies perform poorly when the adversary has a much bigger number of antennas than Alice (case $K = 2, N_E = 64$).

5.4.2 M=2 Vs M>2

It has been shown that security can significantly be enhanced by distributing the antennas at two base stations when there is no knowledge of the wiretap channel. This section examines the case of multiple BSs ($M \geq 2$) and compares non-trivial secret-splitting strategies when the total number of antennas is fixed for the two cases. i.e. having established that under an appropriate base station allocation secret splitting outperforms conventional wiretap coding for secrecy, we now examine what is the optimal number of BSs. For example, Alice is concerned about whether three BSs with two antennas each perform better than two BSs with three antennas each. The multiple BSs are placed in a way such that they form a regular polygon with Bob being at the centre:

$$A_i \text{ is placed at } (1, 2\pi(i-1)/M). \tag{5.48}$$

For example, when $M = 3$, the BSs form an equilateral triangle. Assuming that the BSs can have a minimum distance of one from Bob, the above BS allocation is optimal in terms of increasing the probability $\mathscr{P}_0$. Indeed, by separating the BSs as far as possible from each other whilst the distance between each of them and Bob is kept to the minimum, there is always one BS to which Bob is closer than Eve. As such, the probability of Eve 'missing' a secret split is maximised.

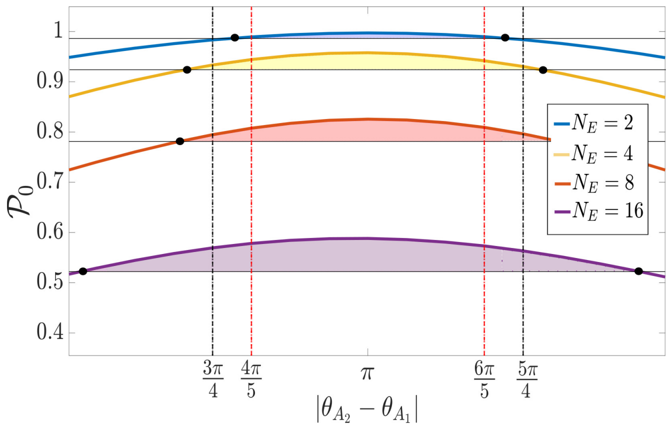


Figure 5.10: The curved lines indicate the performance $\mathscr{P}_0$ against the optimal and sub-optimal angle-difference between two BSs. The vertical lines correspond to the case when $M = 3$ are placed optimally. The highlighted segments indicate the angle-differences for which 2 BSs outperform the employment of 3 BSs.

As seen in Figure 5.8, when the total number of antennas is fixed, the performance is maximised for $M = 2$ and degrades gradually with $M > 2$. In particular, the more antennas employed at Eve, the faster the performance of $\mathscr{P}_0$ degrades with $M > 2$. Therefore, if there exist two BSs that are placed diametrically opposed to Bob, transmitting two splits with two 3-antenna BSs is a better strategy than transmitting three splits with three 2-antenna BSs. This is an encouraging finding for practical situations where the cost overhead for base station cooperation increases with the number of base stations.

Extracting the data from Fig. 5.8, for $M \leq 3$, Fig. 5.9 is plotted. Since the difference in the performance of the cases $M = 2$ and $M = 3$ is very small, transmitting with three splits may be more beneficial if the two BSs are not placed optimally. Simulations suggest that $M = 2$ is the optimal number of BSs as long as the angle difference doesn't differ more than $\pi/5$ from the optimal angle difference (π). The simulations were run for a different set of parameters: $\sum N_A = 6, 12, 60, 120$ and $N_E \in [10 \sum N_A]$. Figure 5.10 is an example of the performance for the two cases $M = 2$ Vs $M = 3$ when $\sum N_A = 6$. The curved lines indicate the performance of the case $M = 2$ against the angle difference whilst the vertical lines indicate the maximum performance for the case $M = 3$. In most cases, the case $M = 2$ performs better even when $3\pi/4 < |\theta_{A_2} - \theta_{A_1}| < 5\pi/4$.

5.5 Conclusion

This chapter presents the secret splitting scheme, aiming to reduce the eavesdropping capabilities of unintended receivers. While the definitions and scheme provided are broad and applicable to any channel model, the chapter specifically explores secret splitting for the case of narrowband Rayleigh channels and transmit beamforming.

The derived formulae facilitate theoretical analysis and the presentation of numerical results, focusing on the scheme's impact on secrecy outage probability. In comparison to conventional wiretap coding methods, secrecy splitting significantly reduces secrecy outage probability. It is important to note that the evaluation of secrecy rates did not account for additional overhead costs arising from base station cooperation. Nonetheless, secrecy splitting creates a positive secrecy gap in areas where the secrecy rate with conventional ways of secrecy coding would have been zero.

The examination of the scheme also covers the optimal number of base stations and their allocation. Subject to the constraint $\sum N_A \leq K$, distributing K antennas among a small number of base stations, with two being optimal as long as the legitimate receiver is positioned between them, proves to be more advantageous. For example, the proposed scheme could find a good fit when the legitimate receiver moves along streets or railways. Finally, as the secrecy capacity decreases linearly with the number of employed BSs, we suggest secret splitting be used for the confidential transmission of short messages such as symmetric keys.

CHAPTER 6

EXPLOITING CHANNEL CORRELATION AGAINST DISTANCE FRAUD

Whereas channel correlation has a negative impact on physical layer key generation, as seen in chapter 4, this chapter demonstrates how it can be exploited to our benefit for protecting short-range systems against distance fraud. Two novel methods are presented that can be employed by narrowband low-cost transceivers. The method introduced in section 6.2 addresses relay attacks and replay attacks, while the method in section 6.3.3 exploits backscattering modulation to protect against solo distance fraud. The main results and conclusions of this chapter have been published in [175, 176]. The first method has also been patented in [177].

Keywords: Authentication, multipath fading, physical layer security, relay attack, replay attack, short-range communications, spatial channel correlation.

Prerequisite: Fundamentals of the wireless channel 2.3, Authentication attacks in Short-Range Systems 3.3, Spatial correlation and multipath fading 2.3.3

6.1 Introduction

Short-range communications systems such as Bluetooth and Radio Frequency Identification (RFID) systems have become an essential part of everyday life. Contactless payments, for example, have replaced cash payments for many individuals, and the same is expected to happen with key-fobs or access cards replacing physical keys [175]. A common assumption of such systems is that the physical constraints of the communication channel implicitly prove the proximity of a device. However, as seen in section 3.3, this assumption is far from true and reports of crime have dramatically increased in the last year.

This chapter demonstrates how the fundamental correlation properties of the multipath Radio-Frequency (RF) channel can be exploited to protect against distance fraud and replay attacks. Two methods are presented, namely, CHannel Reflection Yields Secure Proximity (CHRYSP) and (against) Solo Distance Fraud for RFID (SDF-RFID). The first method, CHRYSP, protects against relay attacks (Mafia Fraud) and replay attacks, whereas the second method, SDF-RFID, protects RFID systems against solo-Distance Fraud and replay attacks.

Both methods are facilitated by the cooperation of an external node whose transmitting signal is the reference signal for capturing the small-scale fading at two receivers: Alice and Bob. Bob plays the role of the prover whereas Alice is the verifier. If Alice and Bob are in close proximity, they shall experience correlated fading. On the other hand, if they are placed at a distance much greater than the wavelength, they shall experience uncorrelated fading. The spatial correlation properties of the RF channel are exploited to protect against distance fraud, and the temporal correlation properties add protection against replay attacks.

6.1.1 Related Work

	Replay attack	Solo Dist. Fraud	Mafia Fraud	Terrorist Fraud
RF fingerprints	✓	X	✓	✓
Ambient Conditions	✓	✓	✓	X
Time-of-flight dist. bound.	X	✓	✓	depends on protocol
CHRYSP	✓	✓	✓	X
SDF-RFID	✓	✓	X	X

Table 6.1: Possibility for protection against authentication attacks in short-range communication systems

To the author's best knowledge, it is the first time that channel correlation is exploited for proximity verification. Different methodologies against authentication attacks in short-range authentication systems have been reviewed in section 3.3. Table 6.1 lists some of the existing methods along with the type of fraud that they address. As discussed in section 3.3 the most resilient method against distance fraud is the Distance Bounding (DB) technique that measures the time of flight of an incoming signal. However, due to their requirement for time accuracy, DB protocols require specialised hardware and an Ultra-WideBand (UWB) channel [132, 133] in order to be effective. Moreover, DB techniques do not protect against replay attacks.

The proposed method CHRYSP naturally protects against relay and replay attacks and can be employed in both narrowband and UWB communications (the focus of this chapter is narrowband communications). The advantage of the second method, SDF-RFID, against other methods that protect against Solo Distance Fraud (SDF) and replay attacks is that it requires no further action from the prover. Equipped with an RFID transponder, the prover is only required to send typical RFID data to the verifier. As such SDF-RFID requires no extra power or memory resources from the RFID transponder. The latter statement is of high importance when considering the extremely limited resources of RFID transponders (see section 3.3.1.1).

6.1.2 Organisation

The remainder of this chapter is segmented into two main sections: section 6.2 that presents the method CHRYSP and section 6.3 that presents method SDF-RFID. Each main section begins with the channel model and the estimation of the channel correlation. Section 6.3 also includes background information on backscattering modulation (section 6.3.1) which is crucial for our understanding of how the verifier is able to estimate the channel correlation in UHF-RFID systems. The proceeding sections discuss the possibility of protection against distance fraud. The numerical examples in section 6.2.6 apply to both methods CHRYSP and SDF-RFID. Lastly, the concluding remarks are found in section 6.4.

6.2 CHRYSP

6.2.1 Channel model and channel correlation

In a short-range communication system, Alice allows access to a service only when a user with valid cryptographic primitives is in close proximity. Bob is a user with valid cryptographic primitives, whereas Eve is an attacker who tries to impersonate Bob. As explained in section 3.3, there are four common types of impersonation attacks or distance fraud in short-range communication systems, namely replay attacks, Solo-Distance Fraud, Mafia

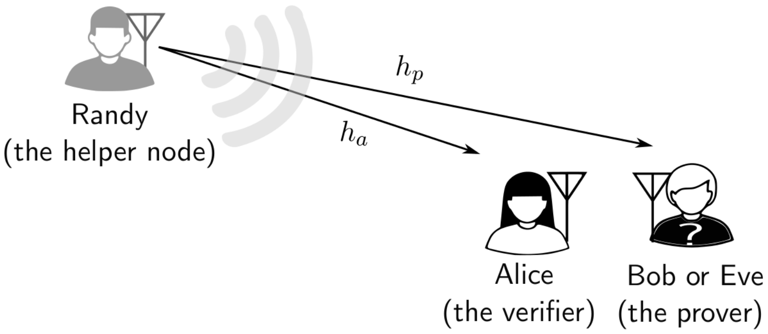


Figure 6.1: Randy, Alice and Bob/Eve play the role of the helper node, the verifier, and the prover, respectively.

fraud (or relay attack), and Terrorist Fraud. The proposed method, CHRYSP, offers protection against the first three types of attacks, a brief description of which is as follows:

- Replay attack: Eve replays Bob's signal at a later time.
- Solo-Distance Fraud: Bob is remote and dishonest; He tries to deceive Alice into believing he is in close proximity. Eve is not present in this type of attack.
- Mafia-Fraud (relay attack): Eve relays Bob's signal in real-time (if Mafia Fraud is launched by two attackers, Eve is the attacker closer to Alice).

A helper node, Randy is employed to facilitate the authentication scheme. Let $h_a, h_b, h_e \in \mathbb{C}$ denote the complex fading channels between Randy and Alice, Bob, and Eve, respectively. It's important to highlight that single-antenna devices are chosen due to their common use in short-range systems, driven by considerations related to space and power constraints.

A dynamic flat fading multipath channel is assumed between Randy and Alice. When the identity of the prover is not known, notation h_p is used to refer to the prover's channel, i.e. $p \in \{b, e\}$. Figure 6.1 is a representation of the channel model. Channels h_a and h_p are assumed to be wide stationary, meaning that the first and second-order statistics do not change. The mean value and variance of $h_u, u \in \{a, b, p\}$ are denoted by μ_u and σ_u^2, respectively. As explained in section 2.3, when the prover is in close proximity to Alice, channels h_a and h_p are correlated. The channel correlation is given by

$$R(h_a, h_p) = \frac{E[(h_a - \mu_a)(h_p - \mu_p)^*]}{\sigma_a \sigma_p}. \tag{6.1}$$

When a distance fraud occurs (Mafia Fraud or Solo-Distance Fraud) it is assumed that Bob is positioned at a significantly greater distance than the wavelength from Alice, causing the correlation between their channels to be nearly zero. To aid in our analysis, we assume that the channel correlation is exactly zero, denoted as $R(h_a, h_b) = 0$. This assumption will allow us to better comprehend the effects of an inaccurate estimation of the channel correlation on Alice's end, upon which the calculation of the decision threshold will take place. Further, it is assumed that when there is no attack, Bob approaches Alice in close proximity such that $|R(h_a, h_b)| \geq R_o > 0$.

While the initial aim of this chapter was to mitigate distance fraud, it has become evident that the proposed scheme can also safeguard against replay attacks, provided that the replay attack occurs considerably later. By "considerably later", we mean that the time interval between Bob's transmission and the replayed signal transmission exceeds the channel's coherence time, such that their channels decorrelate. Again, for analysis purposes, we assume $R(h_a, h_b) = 0$, where h_a represents Alice's channel during the replay attack and h_b denotes Bob's channel when Eve captured and stored her signal. It's important to recognise that there might be replay attacks that happen "immediately after" Bob's transmission, rendering the above assumption invalid. In this case, the suggested scheme should only be used as a countermeasure against distance fraud.

Summarising our assumptions we have the following conditions:

$$|R(h_a, h_b)| \geq R_o > 0 \qquad \text{(no attack: Bob is in close proximity)} \tag{6.2}$$

$$|R(h_a, h_b)| = 0 \qquad \text{(attack: Solo-Distance Fraud, Mafia Fraud, or replay attack)} \tag{6.3}$$

6.2.2 Received signal

As will be seen in Section 6.2.4, CHRYSP begins with channel measurements made upon Randy's signal transmission. We assume that both Alice and the prover can perfectly track their channel over a period of time in a synchronised manner[1]. The assumption of perfect channel estimation allows us to solely focus on the impact of estimation errors of the channel coefficient $R(h_a, h_b)$.

Assuming block fading, we partition Randy's transmitting signal into a sequence of blocks: $\mathbf{s} = \mathbf{s}_1, \ldots, \mathbf{s}_n$ such that the fading channel remains static during the transmission of each block and changes randomly from one block to the other. Let $h_u[i]$ be the channel realisation during the transmitting block $\mathbf{s}_i$ at receiver $u \in \{a, b, e\}$. The received block of

[1] Perfect synchronisation is not required as long as the time-offset between the measurements is less than the coherence time of the channel

symbols at Alice/prover is

$$\mathbf{y}_u[i] = h_u[i]\mathbf{s}_i + \mathbf{n}_u, \quad u \in \{a, b, e\}, \tag{6.4}$$

where $\mathbf{n}_u$ is additive noise. To evaluate $h_u[i]$, the simplest technique is maximum likelihood estimator whereby the receivers multiply $\mathbf{y}_u[i]$ with $\mathbf{s}_i^H$, where $(\cdot)^H$ is the transpose conjugate operator.

The longer the block sequence, the more accurate the estimation of h_i is. To communicate the key components of CHRYSP, the channel estimation error is not taken into account and perfect channel estimation is assumed at each receiver. By repeating the process over different blocks each receiver attains a sequence of N independentally and identically distributed (i.i.d.) samples. Alice's channel sequence is

$$\{h_a[i]\} = (h_a[1], \dots h_a[N]), \text{ where } h_a[i] \text{ are i.i.d.} \tag{6.5}$$

The channel sequence at the prover is

$$\{h_p[i]\} = (h_p[1], \dots h_p[N]), \text{ where } h_p[i] \text{ are i.i.d.} \tag{6.6}$$

Note that independence refers to the samples within the receiver's channel sequence. Sequence $\{h_a[i]\}$ can be correlated to $\{h_p[i]\}$. The sample correlation based on the observed channel sequences $\{h_a[i]\}$ and $\{h_p[i]\}$ is given by:

$$\hat{R}(h_a, h_p) = \frac{\sum_{i=1}^{N} \left(h_a[i] - \hat{\mu}_a\right) \overline{\left(h_p[i] - \hat{\mu}_p\right)}}{N \hat{\sigma}_a \hat{\sigma}_p}, \tag{6.7}$$

where $\hat{\mu_u}$ and $\hat{\sigma_u^2}$ are the sampled mean and sampled variance:

$$\hat{\mu_u} = \frac{1}{N} \sum_{i=1}^{N} h_u[i] \tag{6.8}$$

$$\hat{\sigma_u^2} = \frac{1}{N} \sum_{i=1}^{N} (h_u[i] - \mu_u)(h_u[i] - \mu_u)^*. \tag{6.9}$$

6.2.3 Valid tags

Alice and Bob share a secret key, k, i.e. a random sequence that is known to them only. To pass the legitimacy test, a prover needs to demonstrate knowledge of the key by computing a valid keyed cryptographic hash function such as a message authentication code, or a keyed SHA-2 [178].

Definition 6.1. Let $tag_k(\mathbf{s})$ be a (keyed) cryptographic hash function with inputs the secret key, k, and sequence $\mathbf{s}$. The pair $(\mathbf{t}, \mathbf{s})$ is said to be valid if and only if $\mathbf{t} = tag_k(\mathbf{s})$. For a valid pair $(\mathbf{t}, \mathbf{s})$, we say that tag $\mathbf{t}$ is valid for sequence $\mathbf{s}$.

Remark 6.1. *Commonly used keyed hash functions require binary inputs. If the sequence* $\mathbf{s}$ *is not binary, we need to take its binary representation, say* $f(\mathbf{s})$*, to a certain level of precision and calculate the hash function of* $tag_k(f(\mathbf{s}))$*. We simplify the notation by merging the notation of the composite function* $tag_k \cdot f$ *to* tag_k*.*

We assume that only the holder of the pre-shared key, k, can produce valid pairs. However, Eve may eavesdrop the communications between Alice and Bob and attain valid tags for some strings.

6.2.4 Methodology

ChRYSP blends together two tests, namely the proximity test and the legitimacy test. The proximity is based on the channel correlation, whereas the legitimacy test requires the computation of a valid tag for the prover's channel sequence. Overall, the method comprises four stages.

Stage 1:*Channel measurements.* Randy transmits a sequence of known symbols. Alice and the prover record N independent realisations of h_a and h_p. Alice's channel sequence is $\{h_a[i]\}$, and the prover's channel sequence is $\{h_p[i]\}$ as per Eq. (6.5) and (6.6).

Stage 2: *Signing the channel sequence.* The prover computes tag, $\mathbf{t}$, for their channel sequence $\{h_p[i]\}$. The prover sends $\mathbf{t}$ followed by $\{h_p[i]\}$ to Alice.

Stage 3: *Legitimacy test.* Upon reception of $(\mathbf{t}, \{h_p[i]\})$, Alice computes $tag_k(\{h_p[i]\})$ and checks whether it is equal to $\mathbf{t}$. If it is, the received tag is valid, the prover passes the legitimacy test, and Alice proceeds to the last stage. If the tag is not valid, Alice rejects the prover and the authentication process terminates.

Stage 4: *Proximity test.* Alice estimates the channel correlation $R(h_a, h_p)$ as per Eq. (6.7). For a given threshold τ, if $|\hat{R}| \geq \tau$, the prover passes the proximity test and authentication has been successful. If $|\hat{R}| \geq \tau$, the prover fails the test and Alice rejects the prover.

$$\text{If } |\hat{R}| \geq \tau, \text{ accept the prover;}$$
$$\text{Otherwise, reject the prover.}$$

6.2.5 Protection against authentication attacks and distance fraud

The purpose of the legitimacy test in stage 3 of the methodology is the prevention of attacks whereby Bob is not involved in the authentication process. Without the legitimacy test in stage 3, any node in close proximity to Alice would be successfully authenticated. Next, we examine the possibility of protection against replay attacks, Solo-Distance Fraud, Mafia Fraud, and Terrorist Fraud.

6.2.5.1 CHRYSP and Terrorist Fraud

CHRYSP does not protect against Terrorist Fraud. Recall that in such an attack, dishonest remote Bob cooperates with Eve who is close to Alice. For more details, the reader is referred to 3.3.

Terrorist Fraud can be launched as follows:

- In stage 1 of the methodology 6.2.4, local (to Alice) Eve collects channel sequence $\{h_e[i]\}$. Eve is in such proximity that $R(h_a, h_e) \geq R_o$.
- In stage 2, Eve relays $\{h_e[i]\}$ to Bob, Bob computes a valid tag $\mathbf{t}$ for $\{h_e[i]\}$ and sends $(\mathbf{t}, \{h_e[i]\})$ to Eve who then relays the message to Alice.
- In stage 3, $(\mathbf{t}, \{h_e[i]\})$ passes the legitimacy test since it has been provided by Bob.
- In stage 4, $\{h_e[i]\}$ demonstrates a high correlation with Alice's channel $\{h_e[i]\}$ and Alice falsely grants access to Eve.

To protect against Terrorist Fraud, CHRYSP could be combined with RF fingerprints based techniques. Such a combination could protect against any type of distance fraud.

6.2.5.2 CHRYSP against distance fraud and replay attacks

In the case of a Solo-Distance Fraud, distanced Bob will send valid tags and be successful in passing the legitimacy test. In Mafia Fraud/replay attack, Eve relays/replays Bob's valid tag and passes the legitimacy test. Observe, that all successful attacks up to this stage involve Bob's channel sequence $\{h_b\}$ and not $\{h_e\}$. If Eve modified Bob's signal $(\mathbf{t}, h_b)$, and replaced h_b with another sequence, tag $\mathbf{t}$ would no longer be valid. Proceeding to the last stage of the proximity test, Alice possesses Bob's channel sequence $\{h_b\}$.

All three types of attacks can successfully pass the legitimacy test. Since SD Fraud, Mafia Fraud, and replay attacks can pass stage 3, their detection depends on the last stage. The proximity test in stage 4 examines whether channel sequence $\{h_b\}$ is spatially/temporally correlated with Alice's channel sequence $\{h_a\}$. As such, it can detect whether Bob's channel sequence is relayed (spatial separation due to Mafia fraud), replayed (temporal separation due to replay attack), or sent by a remote prover (spatial separation due to Solo-Distance Fraud). Alice's binary decision between accepting the prover and rejecting the prover relies on choosing between " $R \geq R_o$" (no attack) versus "$R = 0$" (relay attack or replay attack). The probability of taking the right decision depends on the accuracy of the estimation of the channel correlation.

6.2.5.3 Performance Metrics

By the term False Negative Rate (FNR), we refer to the probability of missed detection of a relay attack, a replay attack, or Solo-Distance Fraud. It can be thought of as the probability of falsely accepting the prover.

$$\text{FNR} := P(|\hat{R}| \geq \tau | R = 0). \tag{6.10}$$

The True Positive Rate (TPR) is the complement of FNR. It is the probability of detecting an attack when it occurs.

$$\text{TPR} := 1 - \text{FNR} = P(|\hat{R}| < \tau | R = 0). \tag{6.11}$$

By the term False Positive Rate (FPR), we refer to the probability of falsely assuming an attack (replay, Mafia Fraud, or SD Fraud), i.e. falsely rejecting Bob.

$$\text{FPR} := P(|\hat{R}| < \tau || R| \geq R_o). \tag{6.12}$$

The complement of FPR is referred to as the True Negative Rate and can be thought of as the probability of correctly accepting Bob.

$$\text{TNR} := P(|\hat{R}| \geq \tau | R \geq R_o). \tag{6.13}$$

Remark 6.2. *As channel correlation increases, e.g. due to Bob approaching Alice closer, the performance of the scheme in terms of FPR increases. FPR is upper bounded by*

$$FPR \leq P(|\hat{R}| < \tau || R| = R_o). \tag{6.14}$$

Choosing a pessimistic approach, the graphs of Section 6.2.6 only consider the maximum value of FPR which is denoted by FPR_{max}.

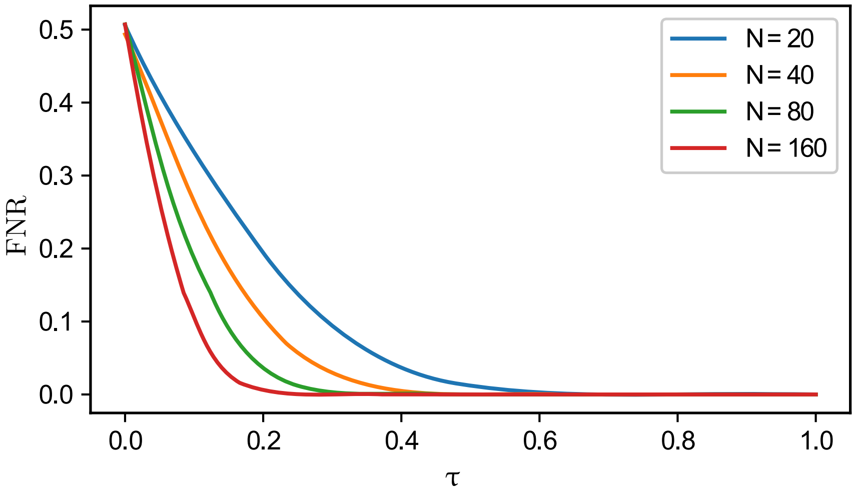


Figure 6.2: False-negative rate against the decision threshold

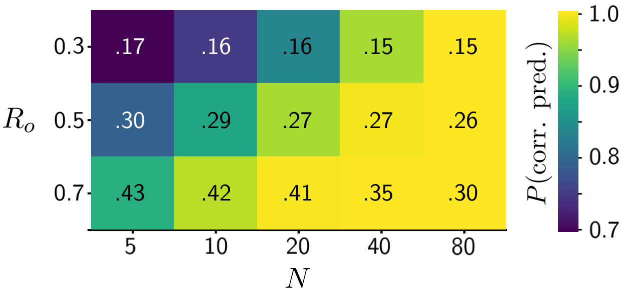


Figure 6.3: Optimal thresholds for increasing the probability of correct predictions

6.2.6 Numerical Examples

The choice of the decision threshold, τ, and the sample size, N, is a problem that needs to be studied case by case depending on the system's requirements and restrictions. Let us consider the restriction where the time of channel probing is at most T_p. Then N can be at most $\lfloor T_p/T_c \rfloor$, where T_c is the channel's coherence time. Asymptotically, as $N \to \infty$, the sample correlation converges to the true correlation, and Alice's predictions will be 100% correct for any decision threshold $0 < \tau \leq R_o$. For a fixed N, the choice of the decision threshold will determine the values of FNR and FPR.

To give numerical examples, we run simulations whereby the events of "attack" and " no attack" occur in the same frequency. I.e. $P(\text{"attack"}) = P(\text{"no attack"}) = 0.5$. Standard complex channels are considered: $h_a,\ h_b \sim \mathbf{CN}(0,1)$. In the case of an attack, Bob's re(p)layed channel is decorrelated from Alice's channel such that $R(h_a, h_b) = 0$, otherwise, $R(h_a, h_b) = R_o$.

Figure 6.2 suggests that for $N = 160$, any threshold $\tau \geq 0.17$ meets the requirement, whereas, for $N = 20$, the decision threshold needs to be at least 0.5. Observe from (6.10) and (6.12) that there is a trade-off between the values of FPR_{max} and FNR. As such, for minimising FPR whilst meeting the requirement of the form FNR $< \epsilon$, the smallest possible decision threshold needs to be considered.

To find the threshold that maximises the probability of correct predictions, $P(\text{corr. pred.})$, observe that this is equal to the summation of:

$$\begin{aligned}
P(\text{corr. pred.}) &= P(|\hat{R}| \geq \tau, R \geq R_o) + P(|\hat{R}| < \tau, R = 0) \\
&= \frac{1}{2}P(|\hat{R}| \geq \tau | R \geq R_o) + \frac{1}{2}P(|\hat{R}| < \tau | R = 0) \\
&= \frac{1}{2}(\text{TPR} + \text{TNR}). \qquad (6.15)
\end{aligned}$$

Figure 6.3 demonstrates that the threshold that maximises $P(\text{corr. pred.})$, denoted by τ_o, is roughly half the size of the correlation coefficient R_0 when $N = 40$. The optimal threshold slowly decreases with N; Halving the size of channels samples (N) results in less than 1% change in τ_o. A sample size of $N = 80$ guarantees 99.95% of correct predictions even when the channel correlation is as low as $R_o = 0.3$.

To examine the overall performance of a binary classification problem (“ fraud” or “ no fraud?”) without fixing the decision threshold, Receiver Operator Characteristic (ROC) graphs are commonly used. A ROC graph is plotted by considering TPR and FPR as functions of τ and plotting TPR(τ) against FPR(τ) for all $0 \leq \tau \leq 1$. Typically, a binary classifier is considered to be “accurate” when the Area Under the Curve (AUC) is 0.9 or higher. In the ideal case where AUC = 1, the system, i.e. Alice, makes 100% correct predictions. Plots in figure 6.4 are the ROC curves for the cases of $R_o = 0.3, 0.5$ and 0.7 respectively. The level of accuracy for the values of AUC is three decimal places. For $N \geq 40$, the binary classifier is accurate as long as $R_0 \geq 0.3$. When $R_0 < 0.3$, a fast varying dynamic channel is needed to ensure a sufficiently large sample size, N.

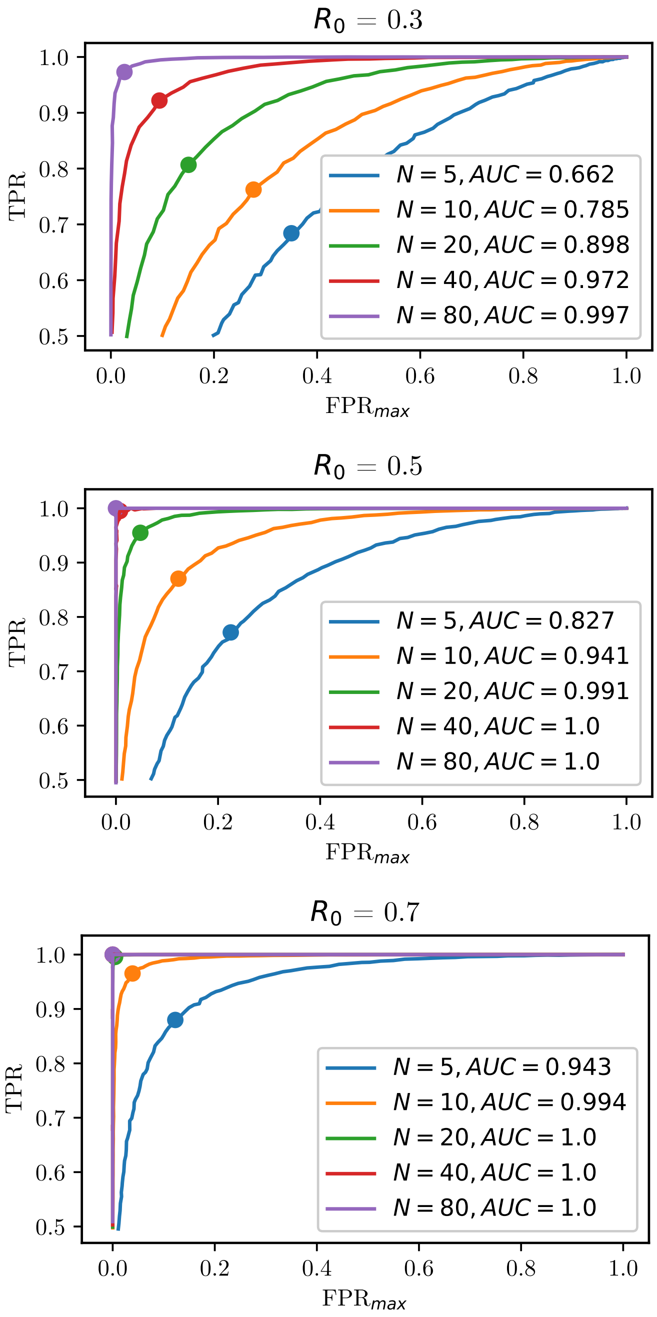


Figure 6.4: ROC graphs for $R_0 = 0.3, 0.5$, and 0.7. The marked coordinates correspond to the optimal performance of the binary classifier in terms of increasing the probability of correct predictions.

6.3 Verifying proximity through backscattering modulation in RFID systems

This section presents a method against Solo Distance fraud that can be applied to UHF RFID systems. We refer to this method as SDF-RFID. Similarly to CHRYSP, SDF-RFID exploits the spatial/temporal correlation properties of the RF channel and is facilitated by the employment of the helper node, Randy. The main difference is that the prover (Bob) does not take any further action rather than communicating typical RFID data such as ID and cryptographic primitives. Through backscattering modulation, the channel between Bob and Randy, h_b, is "reflected" back to the verifier (Alice) when Bob sends data. As such, the verifier is able to estimate channel h_b through backscattering modulation, thereby enabling her to perform the task of channel measurements for both channels h_a and h_b.

6.3.1 Backscattering modulation in UHF RFID systems

SDF-RFID relies on radiative coupling as currently employed by RFID systems that operate in the Ultra-high frequency (UHF) spectrum [101]. UHF RFID systems are often found in secure access control, file tracking, supply chain management, and smart labelling. Although our scheme can be applied to any systems that use radiative backscattering modulation, we make UHF RFID systems our case study.

Recall from 3.3.1.1 that RFID transponders, also known as tags, have memory constraints due to their low cost and they may (semi-passive case) or may not (passive case) have local power. Once excited by the interrogator, the transponder responds by sending data through backscattering modulation. Upon reception, the interrogator forwards the data from the transponder to the host computer for further processing. The computing power can, therefore, be thought to be concentrated on the interrogator.

The transponder's antenna is not a typical radio transmitter, in the sense that it does not transmit its own electromagnetic (EM) wave. To send the data requested by the interrogator, a transponder performs EM backscattering modulation. In this type of modulation, the EM wave that carries the transponder's baseband message is provided by the interrogator who transmits a continuous sinusoidal wave. The transponder reflects back the carrier wave after modulating the signal, usually by means of Amplitude Shift Keying (ASK) or Phase Shift Keying (PSK) modulation. Our scheme uses 100% ASK which means that the digital data is represented as the presence or absence of the backscattered carrier wave. Such modulation is commonly found in protocols compliant with the *Electronic Product*

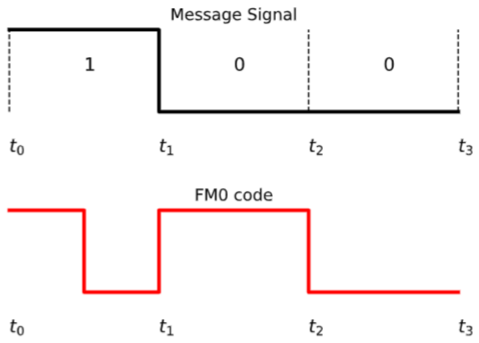


Figure 6.5: Example of FM0 encoding at the transponder.

Code (EPC) Gen2 UHF specification: the de facto specification for UHF RFID transponders [179].

Most EPC Gen2 UHF RFID protocols use the FM0 code or Miller code as encoding schemes. With FM0, a binary 0 is represented by a high or low voltage occupying the entire bit window whereas a binary 1 is represented by a transition in the middle of the bit window. Miller coding guarantees a transition in every other bit [98] which results in larger pulse width, hence less bandwidth to be transmitted.

Example 6.1. *To send binary string 100, the baseband message with FM0 coding is*

$$m(t) = \begin{cases} 1 & \textit{for } t \in [t_0, t_0/2) \\ 0 & \textit{for } t \in [t_0/2, t_1) \\ 1 & \textit{for } t \in [t_1, t_2) \\ 0 & \textit{for } t \in [t_2, t_3) \end{cases} \tag{6.16}$$

where $(t_{i+1} - t_i)$ *is the bit duration, primarily defined by the reader. The bit duration usually takes a value in the range between* $6\mu s$ *and* $25\mu s$. *Figure 6.5 gives a visual representation of the encoded message.*

To modulate a binary one/zero of the encoded message with 100% ASK, the transponder closes/opens its circuit for the duration of the binary bit. As such, we can associate the values of $m(t) = 1$ and $m(t) = 0$, as reflection and non-reflection, or simply as ‘on’ and ‘off’, respectively.

UHF transponders operate in the region around 915MHz or 433MHz with corresponding wavelengths of 33cm and 69cm. The transponder’s antenna in these frequencies comes in many shapes such as dipoles, folded dipoles, printed dipoles or patch antennas [12, 98].

Table 6.2: Notation

h_a:	channel Randy → Alice
h_b:	channel Randy→ Bob
g:	channel Bob → Alice
γ:	reflection coefficient
$c(t)$:	carrier wave
$m(t)$:	baseband signal
$y(t)$:	receive signal at Alice
$b(t)$	backscattered signal

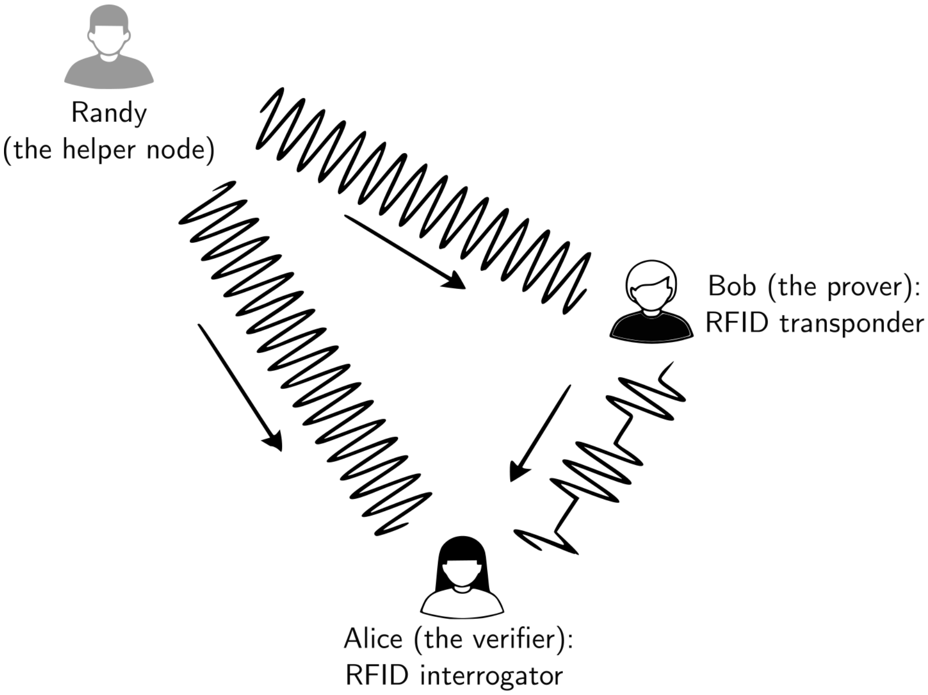


Figure 6.6: The interrogator receives the superposition of two transmitting signals. At times when the transponder does not reflect the interrogator evaluates the channel h_a, whereas h_b is measured during reflection ($m(t) = 1$).

Our scheme requires undirected gain patterns which can be provided by dipoles or folded dipoles.

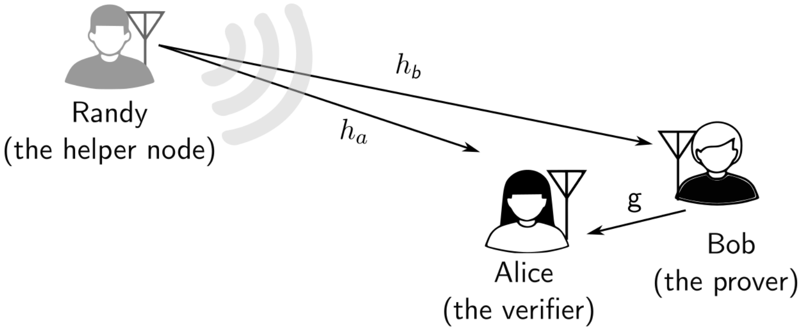


Figure 6.7: Channel model: For authentication purposes, Bob needs to be displayed close to Alice, in which case both h_a and h_b are Rayleigh channels, whereas $g \in \mathbb{C}$ is a fixed LoS channel.

6.3.2 The communication channel and channel correlation

6.3.2.1 Channel model

With SD-RFID, the verifier, Alice, and the prover, Bob, are the interrogator and transponder of a UHF-RFID system, respectively. Different to typical RFID systems, when the interrogator (Alice) requests data transfer from the transponder (Bob), the EM energy is provided by the helper (Randy) instead of the interrogator himself.

The same assumptions as with CHRYSP are made for the channels between the helper node and receivers Alice and Bob. Helper, Randy, is distanced from Alice and Bob. The channel between Randy and Alice, and the channel between Randy and Bob, denoted by $h_a \in \mathbb{C}$ and $h_b \in \mathbb{C}$, are assumed to be Rayleigh and dynamic in nature. When Bob is distanced from Alice (at a distance much bigger than the wavelength), h_b is independent of h_a. With SD-RFID, the channel between Bob and Alice, $g(t)$, is also required. When Bob is close to Alice, the channel between Bob and Alice is deterministic due to a strong LoS component between the two entities. That is:

$$g(t) = g \in \mathbb{C}, \tag{6.17}$$

for all t during the period of transmission. figure 6.7 illustrates the channel model.

The helper transmits a continuous wave sinusoid of constant phase and amplitude. For simplicity, the amplitude is normalised to one and the phase is set to zero. That is, at time t the helper node transmits:

$$c(t) = e^{j2\pi f_c t}, \tag{6.18}$$

where f_c is the carrier frequency. The corresponding wavelength is denoted by λ. The signal $c(t)$ reaches the interrogator and the transponder $r_1(t)$ and $r_2(t)$, respectively:

$$r_1(t) = h_a(t)c(t) + n_1(t) \tag{6.19}$$

$$r_2(t) = h_b(t)c(t) + n_2(t) \tag{6.20}$$

Components $n_1(t)$ and $n_2(t)$ are Additive White Gaussian Noise (AWGN) of zero mean that vary independently from one another and for different samples.

6.3.2.2 Backscattered signal

The transponder (Bob) modulates the baseband signal, $m(t)$, on the received carrier frequency resulting in the passband signal of $m(t)[h_b(t)c(t)) + n_2(t)]$. Let γ be the antenna reflection of the transponder. This is a complex number fixed at the time of manufacture.

The backscattered signal reaches the interrogator as

$$b(t) = \gamma g \underbrace{m(t)[h_b(t)c(t) + n_2(t)]}_{passband} + n_3(t), \tag{6.21}$$

for some AWGN noise $n_3(t) \sim \mathscr{CN}(0, \sigma_3)$, independent to $n_2(t)$. The transponder employs 100% ASK modulation and FM0 (or Miller) coding scheme. As such, the baseband signal fluctuates between two values.

With the helper node transmitting continuously during the data transfer, Alice observes the superposition of two signals $r_1(t)$ (from the helper) and the backscattered signal $b(t)$ (from Bob), as seen in figure 6.6. Alice receives:

$$y(t) = \underbrace{h_a(t)c(t) + n_1(t)}_{r_1(t)} + \underbrace{m(t)[h_b(t)c(t) + n_2(t)] + n_3(t)}_{b(t)} \tag{6.22}$$

$$= \begin{cases} h_a(t)c(t) + n(t), & \text{when } m(t) = 0 \\ [h_a(t) + \gamma g h_b(t)]c(t) + n(t), & \text{when } m(t) = 1 \end{cases}, \tag{6.23}$$

where $n(t)$ captures all noise components. Observe the amplitude of the received signal is higher when $m(t) = 1$ than when $m(t) = 0$. As such, the interrogator is able to demodulate by 'observing' the envelope of the received signal.

The observations of the received signal can also be utilised to estimate channels h_a and h_b at Alice. Even though the transponder, Bob, is unaware of his channel, the following section demonstrates that perfect channel estimation at Alice is feasible; Given a sufficient number of sample measurements, Alice can extract both h_a and h_b from the backscattered signal. Then, similar to the scheme in section 6.2, estimation errors are introduced in the second-order statistics, i.e., when estimating the channel correlation.

6.3.2.3 Estimating the channel correlation

Alice tracks the channel between herself and the helper at times when transponder Bob does not reflect, i.e. when $m(t) = 0$, whereas, during reflection, Alice is able to track the channel between the helper and Bob since this information is apparent on the backscattered signal. We show how by focusing first on one coherence block: a time interval, T_c, at which the channel h_a remains static. If Bob is co-located with Alice, channels h_a and h_b remain static for the same period of time: $h_a(t) = h_a(t_0)$ and $h_b(t) = h_b(t_0)$, for all $t \in T_c$ and some $t_0 \in T_c$.

Given the nature of FMO (or Miller) coding, there are no long runs of zeros or ones. Furthermore, the channel(s) typically change(s) much slower than the bit rate. We can therefore say that within the coherence time, there exist two sub-intervals, $T_1 \subset T_c$ and $T_2 \subset T_c$ for which $m(t)$ takes the value zero when $t \in T_1$, whereas, when $t \in T_2$, $m(t) = 1$. The channel $h_a(t_0)$ and $h_b(t_0)$ are estimated at times $t \in T_1$ and $t \in T_2$, respectively.

Estimating h_a Referring (6.19), when $m(t) = 0$ the received signal at the interrogator is $y(t) = h_a(t)c(t) + n_1(t)$. A sample for the channel coefficient h_a is taken at time t_i by multiplying the received signal with the conjugate of the carrier:

$$\widehat{h_a}(t_i) = y(t_i)c^*(t_i) = h_a(t_0) + n(t_i)c^*(t_i) \tag{6.24}$$

Having collected a number of samples within the time interval T_1, the estimation of $h_a(\tau_0)$ is attained by taking the sample mean of $\hat{h}_a(t_i)$, i.e. $h_a(\tau_0) = \widehat{\mathbb{E}}(\hat{h}_a(t_i))$.

Lemma 6.1. *As the number of samples of $h_a(t_0)$ increases, the sample mean $\widehat{\mathbb{E}}(\hat{h}_a)$ converges to the true channel coefficient $h_a(t_0)$.*

Proof. Variable $n(t_i)$ denotes the aggregation of three noise components as seen at (6.23): $n(t) = n_1(t) + n_2(t) + n_3(t)$. Since all three noise components have a mean of zero, so does their summation. Applying linearity of expectation (once again) on (6.24), $E\left(\widehat{h_a}(t_i)\right) = E(h_a(t_0)) + E(n(t_i)c^*(t_i))$. The variable of noise is independent of the helper's signal, $c(t)$, hence, the last term is equal to zero: $(E(n(t_i)c^*(t_i)) = E(n(t_i))E(c^*(t_i)) = 0)$ which completes the proof. □

Given the low data rate of RFID systems (e.g. $m(t)$ remains zero for at least 5μs), we assume that a sufficient number of samples are taken resulting in an accurate evaluation of $h_a(\tau_0)$.

6.3.2.4 Estimating h_b

Within the coherence block of duration T_c, there will be times where $m(t) = 1$, $t \in T_2 \subset T_c$. From (6.20) the reader receives $y(t) = [h_a(t) + \gamma g h_b(t)]c(t) + n(t)$. Channel coefficient $h_a(t) = h_a(\tau_0)$ has been already evaluated. It is assumed that coefficients γ and g are assumed to be known by the interrogator; A priori direct communication with the transponder can achieve such knowledge. The reader applies simple operations on the received signal to attain a sample of $h_b(\tau_0)$:

$$\widehat{h_b}(t_i) = \frac{(\gamma g)^*}{|\gamma g|^2}(y(t)c^*(t) - h_1) = h_b + \frac{\gamma^* g^*}{|\gamma g|^2} n'(t). \tag{6.25}$$

Similarly to lemma 6.1, it can be shown that for a sufficiently large sample size taken within the time interval T_2, an accurate estimation of $h_b(\tau_0)$ can be attained by taking the sample mean $\widehat{\mathbb{E}}(\widehat{h_b}(t_i))$. Similarly to h_a, we assume that a sufficient number of samples are collected and an accurate evaluation of $h_b(\tau_0)$ is attained. The assumption of perfect channel estimation simplifies the theoretical analysis.

Repeating the process M times The process presented for attaining $h_a(t_0)$ and $h_b(t_0)$ is repeated for different coherence blocks resulting in two sequences of size M:

$$\mathscr{H}_1 := [h_a(t_0), \ldots, h_a(\tau_{M-1})] \tag{6.26}$$

$$\mathscr{H}_2 := [h_b(t_0), \ldots, h_b(\tau_{M-1})] \tag{6.27}$$

The elements within each $\mathscr{H}_i$ are independent of one another, but the two sequences will be correlated if the transponder is in close proximity to the reader. Based on $\mathscr{H}_1$ and $\mathscr{H}_2$, Alice finds the sample mean of $h_a {h_b}^*$, and $|h_i|^2$, $i \in a, b$, and applies equation (6.7) to attain an estimation of the spatial correlation between herself and Bob. M can be thought of as the number of independent channel realisations and the sample size for estimating the spatial correlation. After estimating the spatial correlation, Alice will:

- validate the tag's proximity if $\hat{R} \geq \tau$;
- reject the transponder if $\hat{R} < \tau$,

for some decision threshold $\tau \in \mathbb{R}$.

6.3.3 Distance-Fraud and replay attacks

SDF-RFID is a scheme that can detect solo-distance fraud as well as replay attacks. When Bob launches a solo-distance fraud, his channel is decorrelated with Alice's channel: $R(h_a, h_b) =$

0. Similarly, the channel correlation is also zero ($R(h_a, h_b) = 0$) when h_a and h_b are sufficiently temporally separated. As such, a replay attack will most likely fail given that it is launched at a time greater than the channel's coherence time.

SDF-RFID, however, does not protect against Mafia Fraud or Terrorist Fraud, whereby Eve is close to Alice. Different to CHRYSP, a legitimacy test through cryptographic primitives cannot be combined with the proximity test in an inseparable manner. Cryptographic primitives can only be added separately to the scheme which does not necessarily prevent Mafia Fraud or Terrorist Fraud attacks. Indeed, if Alice requests cryptographic primitives from the prover, local Eve can relay the right response from Bob to Alice and be mistaken for Bob (in close proximity).

Different to CHRYSP, the prover in SDF-RFID method does not create a pair of dependencies between cryptographic methods and channel measurements. Specifically, as the channel measurements are made solely by Alice, Bob is not required to know his own channel and therefore he does not “sign" it. For applications that require security against most types of distance fraud, SD-RFID can be combined with RF-fingerprint-based techniques which can offer protection against Mafia Fraud and Terrorist Fraud.

When the definitions of false negative rate and false positive rate are modified, figures 6.2, 6.3, and 6.4 of section 6.2 can be reused for the demonstration of the performance of current scheme SDF-RFID.

By the term False Negative Rate (FNR), we refer to the probability of missed detection of a Solo-Distance Fraud (instead of a relay attack, a replay attack, or a Solo-Distance Fraud).

$$\text{FNR} := P(|\hat{R}| \geq \tau | R = 0). \tag{6.28}$$

The True Positive Rate (TPR) is the complement of FNR. It is the probability of detecting a Solo-Distance Fraud when it occurs.

By the term False Positive Rate (FPR), we refer to the probability of falsely assuming a Solo Distance fraud.

$$\text{FPR} := P(|\hat{R}| < \tau || R| \geq R_o). \tag{6.29}$$

The complement of FPR, i.e. the probability of correctly assuming an SD fraud is the True Negative Rate.

Note that only the semantic meaning of FPR, FNR, TPR, and TNR has changed when compared to section 6.2. The mathematical expressions have remained the same for the two schemes of CHRYSP and SDF-RFID. Therefore, graphs, analytical results, and numerical examples of section 6.2 are applicable as long as the focus is on the attacks of Solo-Distance Fraud and replay attacks.

6.4 Conclusion

This chapter presents two novel methods that utilise the spatial channel correlation to protect against distance fraud. The first method, CHRYSP, protects short-range communication systems against Solo-Distance Fraud and Mafia Fraud. The second method, SDF-RFID, protects against Solo-Distance Fraud and is applicable in UHF RFID systems. Both methods naturally protect against replay attacks due to the temporal decorrelation property of the RF channel.

The case of narrowband communications has been studied, which is typical in short-range communication systems. Under perfect channel estimation, numerical results suggest that CHRYSP and SDF-RFID have great potential in non-static environments. A channel rich in entropy allows the required minimum channel correlation to be a small value, thereby enabling authentication over longer distances. Because of the small computational requirements, CHRYSP and SD-RFID are believed to be a good fit for resource-constrained networks.

CHAPTER 7

CHANNEL RECIPROCITY FOR KEY TRANSMISSION

Physical Layer Key Generation (PLKG) may not be a practical method of key agreement in resource-constrained devices due to a high reconciliation cost. This chapter modifies a typical PLKG protocol by using an encoding function to achieve an arbitrarily low key disagreement rate. As such, the reconciliation stage of common PLKG protocols can be eliminated without increasing the computational complexity. Analytical results allow design optimisation and derive the entropy requirement for perfect secrecy. The practicality of the modified protocol, CRicKET, is successfully demonstrated on a series of Internet-of-Things (IoT) boards connected in a wireless network. The peer-reviewed paper [180] has been accepted for presentation at the IEEE 2023 International Conference on Communications.

Keywords: Channel reciprocity, IoT networks, key agreement, physical layer key generation, reconciliation cost.

Prerequisite: Channel-reciprocity based Physical Layer Key Generation 3.2.1, Fundamentals of Information Theory 2.1, Perfect Secrecy 2.2.1.1

7.1 Introduction

Over the last few years, the research community has given attention to physical layer key generation (PLKG) as a potential solution to the key distribution problem [72]. Although PLKG has much lower computational complexity in comparison to conventional key agreement protocols, i.e. those based on public key cryptography, it is not always feasible on low-power and low-memory devices. In particular, when the early stages of PLKG output keys disagree in many bits, reconciling the keys may require an non-viable amount of resources.

As described in section 3.2, PLKG exploits the channel reciprocity between two communicating entities, say, Alice and Bob, for extracting correlated randomness. PLKG comprises four main stages: channel probing, quantisation, reconciliation, and privacy amplification. In the first stage, Alice and Bob exchange pilot signals to measure their common RF channel in terms of received signal strength, phase information, or other characteristics of the RF channel. Small-scale fading attributes are gleaned from the channel measurements and then converted to binary strings in the quantisation stage. In the next stage, reconciliation, mismatched bits are discarded or corrected. After reconciliation, the keys become identical with high probability in order to facilitate upper-layer encryption.

In the ideal case (when the first two stages output identical keys), the reconciliation cost is zero. However, as seen in section 3.2.1.3, if there are many mismatches between the two keys, reconciliation may be too costly in terms of communication overhead, information leakage, memory, or computational complexity. A high reconciliation cost results in the impracticability of PLKG in resource-constrained networks.

Recall from section 3.2.1.3 that there are two approaches for reconciling the keys: (a) Error Detection Coding (EDC) based and (B) Error Correction Coding (ECC)-based. Key reconciliation based on ECC is typically not a viable solution for resource-constrained devices [181] due to the relatively high computational complexity. EDC may find a better fit in low-cost devices, but not always; When the number of keybits subject to reconciliation is high, EDC-based approaches require testing a significant amount of permutations which may not be viable in low-memory devices [82].

7.1.1 Related Work

A low reconciliation cost is equivalent to reducing the number of mismatches between Alice's and Bob's keys. Mismatches between the two keys mainly occur due to the inability to exchange pilot signals simultaneously during the channel probing phase (channel reciprocity occurs in time-division duplexing communication systems). As such, a common method for reducing the number of mismatches is a reduction in the time lag between pilot exchanges [73, 74]. Although this is an effective method for reducing the reconciliation cost, two devices can exchange pilot signals only as fast as the available hardware or the medium access control protocol permits. Besides, even if the exchange of pilots occurs before any change in the RF channel, a relatively high bit-mismatch ratio due to low signal-to-noise ratio may be inevitable in power-constrained networks.

Other approaches include filtering mechanisms [75–77] that take place between the channel probing phase and quantisation. Filtering mechanisms are effective in terms of reducing the number of mismatches but they can also reduce the entropy of the key, thus resulting in a vulnerable scheme [78]. Lastly, in the quantisation phase, a low-level quantisation scheme is preferred [73] for reducing the reconciliation cost (as well as for a high entropy of the key) but results in low key rates.

7.1.2 Overview and contributions

Similar to previous work, our proposed method, Channel Reciprocity for KEy Transmission (CRicKET), aims to reduce the number of mismatches subject to reconciliation. Instead of focusing on the stages of channel probing and quantisation, CRicKET focuses on the stages after the generation of the keys. The keys derived from the reciprocal channel do not serve as the final keys but facilitate an encoding/decoding mechanism for transmitting a third key generated by Alice, as seen in figure 7.1. That is, CRicKET exploits channel reciprocity in order to “hide” a key and not to extract a key.

The main difference to existing work that aims to reduce the reconciliation cost is that CRicKET’s efficiency -in terms of decreasing the Key Disagreement Rate (KDR)- does not depend on the outputs of the quantisation phase. The encoding/decoding parameters can be adjusted in order to compensate for many mismatches resulting from the quantisation phase. As such, its performance is guaranteed. The significant contributions of this study are as follows:

- A scheme is described to reduce the key disagreement rate to an arbitrary low level;
- An analytical study provides optimal parameters, in terms of maximising encoding rate, for a variety of scenarios;
- Proof that perfect secrecy is achievable even when there is redundancy in the channel sequences;

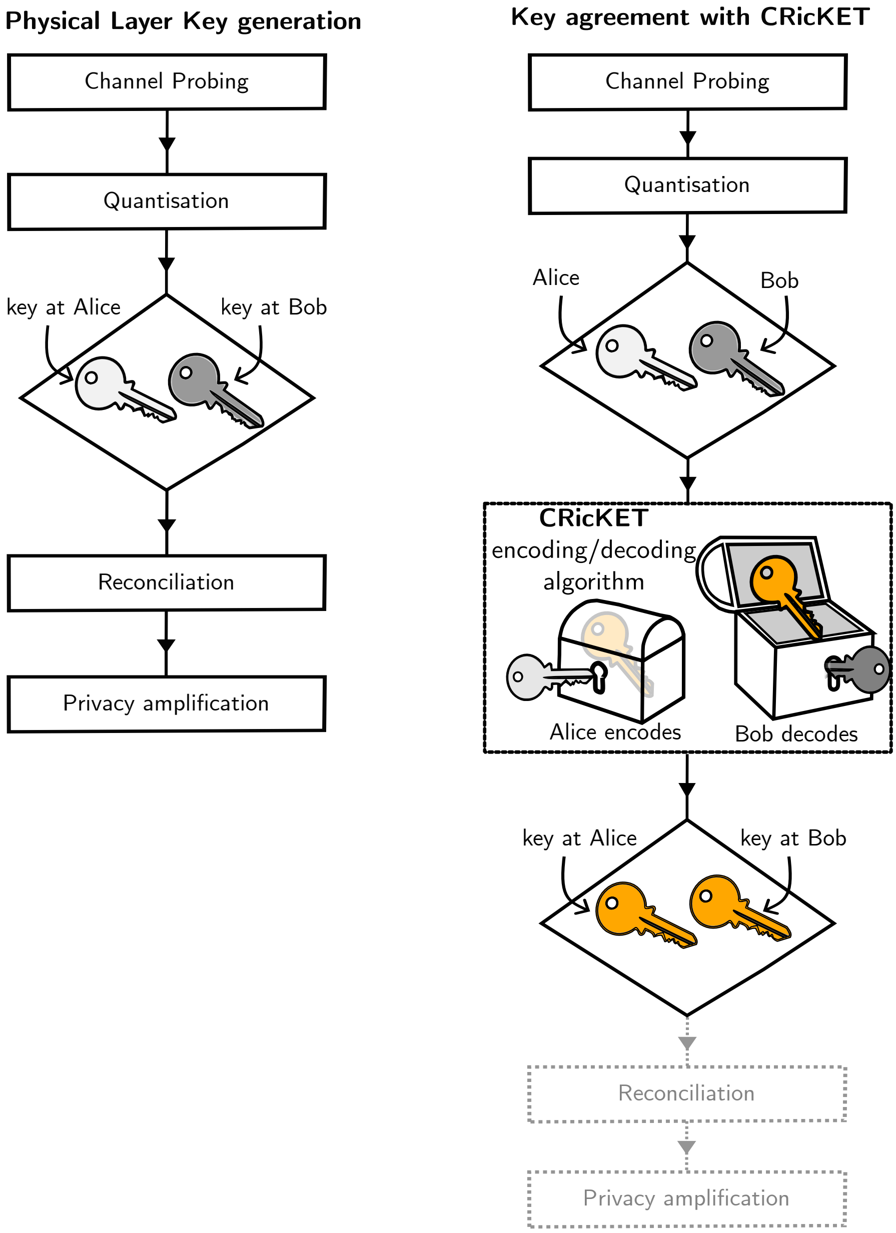


Figure 7.1: CRICKET vs PLKG

- A demonstration of the practicality of the proposed scheme by using single-chip IoT devices.

7.2 System model

Let Alice and Bob be the two entities that wish to agree on a key by communicating over a public, hence, insecure channel. Similarly to PLKG, CRicKET starts with the stages of channel probing and channel quantisation. To provide a flexible key agreement protocol, the exact methods used in any of those first two stages are left unspecified. It is assumed that Alice and Bob have quantified their reciprocal channel and attained two binary sequences that will be referred to as *channel sequences*- we reserve the term “keys” for different sequences as will be seen in sections 7.2.2 and 7.3.3.

Definition 7.1. A binary sequence that is derived from measuring and quantifying features of the reciprocal RF channel between two nodes as described in section 3.2 is referred to as *channel sequence.*

7.2.1 Properties of channel sequences

7.2.1.1 Channel Mismatch

It is a common practice to shuffle the channel sequences (using the same permutation at each end) so that there are no runs of mismatches. Let $\mathbf{s}_a = \{s_{a_1}, \dots s_{a_l}\}$ and $\mathbf{s}_b = \{s_{b_1}, \dots s_{b_l}\}$ be the (shuffled) channel sequences at Alice and Bob, respectively. Since the mismatches are uniformly distributed, the probability of a bit mismatch is independent of its position in the sequence:

$$P\left(s_{a_i} = s_{b_i} | s_{a_j} = s_{b_j}\right) = P\left(s_a(i) = s_b(i)\right), \text{ for all } i \neq j. \tag{7.1}$$

Probability $p_{\text{ch}} := P\left(s_{a_i} = s_{b_i}\right)$ is referred to as the *channel mismatch* and is assumed to be less than 0.5 (if $p_{ch} > 0.5$, the channel mismatch is set to 1-p_{ch}). It is assumed that Alice and Bob are aware of the value of p_{ch}.

7.2.1.2 Confidentiality

We assume that the eavesdropper has no information about the channel sequences $\mathbf{s}_A$ and $\mathbf{s}_B$ to facilitate a clear description of our scheme. Such an assumption is found in the majority of PLKG protocols [74]. According to the findings of chapter 4, this assumption can be considered valid as long as Alice and Bob are situated in a rich-scattering environment and distanced from potential eavesdroppers by several wavelengths. However, our scheme may be extended to include scenarios where the eavesdropper has some information about the channel sequences, as described in section 4.2.2.

7.2.2 The initial key

Differently from PLKG, the channel sequences are not used to derive a key but to securely transmit a key generated by Alice. Equipped with a source of randomness, Alice generates a random key with m binary inputs:

$$\mathbf{k}_a := \{k_{a_1}, \ldots, k_{a_m}\} \in \{0,1\}^m. \tag{7.2}$$

Sequence $\mathbf{k}_a$ will be referred to as the *initial key.*

7.3 CRicKET

CRicKET is an encoding/decoding algorithm comprising three stages at each end:

Alice	**Bob**
1. Setting up parameters	1. Setting up parameters
2. Encoding	2. Decoding
3. Deriving the final key	3. Deriving the final key

To ease the description of our method, the following two definitions are introduced:

Definition 7.2. A binary sequence, $\mathbf{x}$, is said to be flipped when every bit of $\mathbf{x}$ is XoR-ed with 1. The flipped sequence of $\mathbf{x}$ is denoted by $\mathbf{x}_{\text{flip}}$:

$$\mathbf{x}_{\text{flip}} := \mathbf{x} \oplus \{1, \ldots, 1\}, \tag{7.3}$$

where $\oplus$ denotes pairwise exclusive OR operation.

Definition 7.3. Let $\mathbf{x}, \mathbf{y}$ be two binary sequences of the same length. The hamming distance of $\mathbf{x}$ and $\mathbf{y}$, denoted by $\text{dist}(\mathbf{x}, \mathbf{y})$, defines the number of places where $\mathbf{x}$ and $\mathbf{y}$ disagree. The hamming distance can be computed as:

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \text{sum}(\mathbf{x} \oplus \mathbf{y}), \tag{7.4}$$

where, $\text{sum}(\cdot)$ is the function that sums the unitary digits within the binary sequence.

Table 7.1: Basic Notation

$\mathbf{k}_a = \{k_{a_i} \| i \in [m]\}$	the initial key at Alice
$\mathbf{k}_b = \{k_{b_i} \| i \in [m]\}$	Bob's estimate of $\mathbf{k}_a$
$\mathbf{k}'_a \subseteq \mathbf{k}_a$	the final key at Alice
$\mathbf{k}'_b \subseteq \mathbf{k}_b$	the final key at Bob
m	size of initial key
$\mathbf{s}_a = \{s_{a_i} \| i \in [nm]\}$	channel sequence at Alice
$\mathbf{s}_b = \{s_{b_i} \| i \in [nm]\}$	channel sequence at Bob
p_{ch}	channel mismatch
$\mathbf{a}_i = \{s_{a_j} \| j \in [i+n]\}$	i^{th} channel block at Alice
$\mathbf{b}_i = \{s_{b_j} \| j \in [i+n]\}$	i^{th} channel block at Bob
$\mathbf{c}_i = \mathbf{a}_i \oplus \{k_i, \ldots, k_i\}$	the i^{th} cipher block
n	size of block $\mathbf{a}_i$, ($\mathbf{b}_i$, or $\mathbf{c}_i$)
τ	decision threshold
KDR:	key disagreement rate
R:	encoding rate of (final) key

7.3.1 Setting-up parameters

Based on the value of channel mismatch, p_{ch}, and the system's requirements (see Sec. 7.5), Alice and Bob decide on two parameters, namely blocksize: $n \in \mathbb{N}^+$, and decision threshold: $\tau \in \mathbb{N} < n/2$. The functionality of the decision threshold, τ, will be regarded in section 7.3.2.3, i.e. at the decoding phase.

7.3.2 Encoding/Decoding

7.3.2.1 Preparation (grouping into blocks)

Integer n is used for grouping the channel sequences into blocks (of size n). Only the first mn bits of sequences $\mathbf{s}_A$ and $\mathbf{s}_B$ will be used for encryption/decryption. Any extra bits are discarded or kept for future use. Alice and Bob now have exactly m blocks. The first block at Alice is $\mathbf{a}_1 = \{s_a(1), \ldots, s_a(n)\}$, the second block comprises the next n bits of the sequences, etc. Using, an equivalent notation for Bob, the i^{th} block at Bob is $\mathbf{b}_i = s_b(in), \ldots s_b((i+1)n)$. These are the *channel blocks* at Alice and Bob.

7.3.2.2 Encoding

Alice uses channel sequence $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ and key sequence $\{k_1, \ldots, k_m\}$ to generate a new sequence of blocks which are denoted by $\{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$ and referred to as the cipher blocks.

The cipher blocks are generated by flipping or not the channel blocks: if $k_i^a = 1$, then $\mathbf{c}_i = \mathbf{a}_i'$, otherwise $\mathbf{c}_i = \mathbf{a}_i$. Alice sends the m cipher blocks to Bob through the public channel.

Algorithm 3 Encoding

```
Inputs: $\{k_1^a \ldots, k_m^a\}$: key sequence; $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$: channel sequence at Alice
Output: a sequence of $m$ blocks $\{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$
for i=1,2,...,m do
    if $k_i^a = 1$ then
        $\mathbf{c}_i = \text{flip}(\mathbf{a}_i)$
    else
        if $k_i^a = 0$ then
            $\mathbf{c}_i = \mathbf{a}_i$
        end if
    end if
end for
```

Example 7.1. *Let* $\mathbf{k}_a = \{1,0,1\}$ *and* $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \{\{0,0,1\},\{0,1,0\},\{1,1,0\}\}$. *The ciphertext consists of three cipher blocks:*

$$\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\} = \{\{1,1,0\},\{0,1,0\},\{0,0,1\}\}$$

7.3.2.3 Decoding

Upon reception of the cipher blocks, Bob compares each cipher block $\mathbf{c}_i$ with his channel block $\mathbf{b}_i$ in order to guess which blocks $\mathbf{b}_i$ and $\mathbf{c}_i$ $\mathbf{a}_i$ have been flipped or not, or equivalently which keybits are units and which are not. Observe that:

$$\text{dist}(\mathbf{c}_i, \mathbf{b}_i) = \begin{cases} \text{dist}(\mathbf{a}_i, \mathbf{b}_i) & \text{when } k_{a_i} = 0 \\ n - \text{dist}(\mathbf{a}_i, \mathbf{b}_i) & \text{when } k_{a_i} = 1 \end{cases}$$

For a decoder that can correct up to τ mismatches between the channel blocks $\mathbf{a}_i$ and $\mathbf{b}_i$, Bob estimates k_{a_i} as follows:

$$k_{b_i} = \begin{cases} 0 & \text{when } \text{dist}(\mathbf{b}_i, \mathbf{c}_i) \leq \tau \\ 1 & \text{when } \text{dist}(\mathbf{b}_i, \mathbf{c}_i) \geq n - \tau \\ ? & \text{otherwise} \end{cases},$$

where k_{b_i} is Bob's estimation of the keybit k_{a_i}. The value for τ is referred to as the decision threshold and its value needs to be less than $n/2$. If $\tau \geq n/2$ cases $dist(\mathbf{a}_i, \mathbf{b}_i) \leq \tau$ and

Algorithm 4 Decoding

Inputs: $\{\mathbf{b}_1,\ldots,\mathbf{b}_m\}$: channel sequence at Bob $\{\mathbf{c}_1,\ldots,\mathbf{c}_m\}$: channel sequence sent by Alice; τ: decision threshold

Output: $\mathbf{k}_b = \{k_1^b,\ldots,k_m^b\}$: an estimate of the key

```
$n = |\mathbf{b}_1|$                                        ▷ $n$ is the blocksize
for i=1,2,...,m do
    if dist($\mathbf{b}_i,\mathbf{c}_i$) $\leq \tau$ then
        $k_i^b = 0$
    else if dist($\mathbf{b}_i,\mathbf{c}_i$) $\geq n-\tau$ then
        $k_i^b = 1$
    else
        $k_i^b =$ “?″
    end if
end for
```

$dist(\mathbf{a}_i,\mathbf{b}_i) \geq n-\tau$ may be satisfied simultaneously and would cause confusion at the decoder. By repeating the process for $i \in [m]$, Bob attains an estimate of the key, denoted by $\mathbf{k}_b$.

7.3.3 Deriving the final keys

The final key at Bob, $\mathbf{k}'_b$, is attained by dropping the uncertain bits of his estimate, $\mathbf{k}_b \in \{0,1,?\}^m$. Bob sends Alice the positions of “?″ and Alice drops the key bits of those positions to attain her final key $\mathbf{k}'_a$. The two keys are of the same length and may be shorter than the initial key, i.e. $|\mathbf{k}'_a| = |\mathbf{k}'_b| \leq m$.

Example 7.2. *Let*

$$\{\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3\} = \{\{0,1,1,0,0\},\{1,1,0,1,1\},\{1,0,0,1,0\}\}$$
$$\{\mathbf{c}_1,\mathbf{c}_2,\mathbf{c}_3\} = \{\{0,1,1,0,1\},\{1,0,1,0,0\},\{1,0,0,1,1\}\},$$

be Bob's channel blocks and the cipher blocks received by Alice, respectively. The blocksize is equal to $n = 4$. *Let the decision threshold at Bob be equal to* $\tau = 1$. *The Hamming distance for the first pair of blocks is dist*$(\mathbf{b}_1,\mathbf{c}_1) = 1$. *Since it is less than or equal to* τ, *Bob guesses that Alice must have not flipped her first channel block and assigns the first keybit equal to* $\mathbf{k}_1^b = 0$. *For the second pair dist*$(\mathbf{b}_2,\mathbf{c}_2) = 4 \geq n-\tau = 4$ *and Bob's estimate is* $\mathbf{k}_1^b = 1$. *For the last channel block, Bob is unable to make a confident guess since dist*$(\mathbf{b}_2,\mathbf{c}_2) = 2$ *and sets* $\mathbf{k}_1^b =?$. *Bob's estimate is* $\mathbf{k}_b = \{0,1,?\}$, *and the final keys are* $\mathbf{k}_a = \{0,1\}$ *(at Alice) and* $\mathbf{k}_b = \{0,1\}$ *(at Bob) which are matching.*

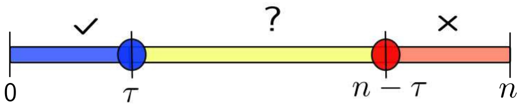


Figure 7.2: When dist($\mathbf{a}_i, \mathbf{b}_i$) lies on the blue, yellow, or red area, the i^{th} keybit is decoded correctly is uncertain, or is decoded to an erroneous keybit, respectively.

7.4 Performance Analysis

To allow the analytical study of the final keys, this section starts with the performance on the key-bit level on Bob's estimation of the key. Then, we derive the probability of two identical final keys, expected length of final keys, Key Disagreement Rate (KDR), and Encoding Rate (R).

7.4.1 Performance on the bit level

The final keys at Bob and Alice, $\mathbf{k}'_a$ and $\mathbf{k}'_b$, are extracted from $\mathbf{k}_a$ and $\mathbf{k}_b$. Before analysing the performance of the final keys, it is essential to focus on $\mathbf{k}_b \in \{0,1,?\}$, i.e. on Bob's estimation of $\mathbf{k}_a$ (before dropping the uncertain keybits). When Alice sends keybit k_{a_i}, Bob decodes it as $0, 1$ or ?. Whether Bob decodes k_{a_i} correctly, wrongly, or is indecisive depends on how well the channel blocks $\mathbf{a}_i$ and $\mathbf{b}_i$ match. In particular:

1. Bob attains a correct keybit exactly when blocks $\mathbf{a}_i$ and $\mathbf{b}_i$ disagree by τ or less:

$$k_i^a = k_i^b \iff \text{dist}(\mathbf{a}_i, \mathbf{b}_i) \leq \tau \tag{7.5}$$

2. Bob attains an erroneous keybit $k_b(i)$ exactly when blocks $\mathbf{a}_i$ and $\mathbf{b}_i$ disagree by $n-\tau$ or more:

$$k_i^a \neq k_i^b \iff \text{dist}(\mathbf{a}_i, \mathbf{b}_i) \geq n-\tau \tag{7.6}$$

3. Bob assign a keybit equal to "?" if and only if $\mathbf{a}_i$ and $\mathbf{b}_i$ disagree by more than τ but less than $n-\tau$:

$$k_i^a = \text{“?"} \iff \tau < \text{dist}(\mathbf{a}_i, \mathbf{b}_i) < n-\tau \tag{7.7}$$

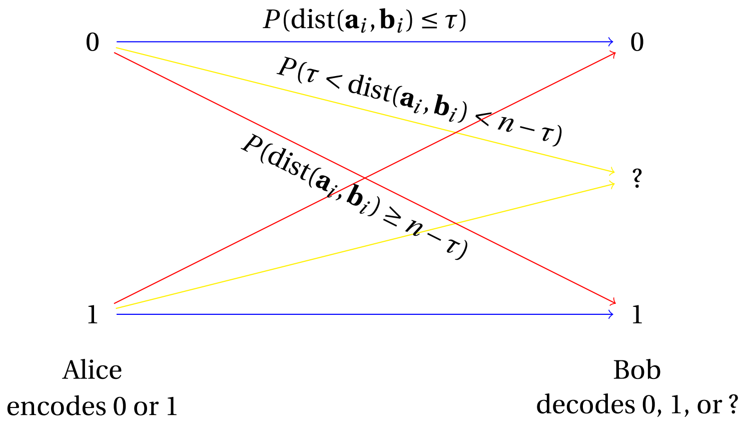

$P(\text{dist}(\mathbf{a}_i, \mathbf{b}_i) \le \tau)$
0
0
$P(\tau < \text{dist}(\mathbf{a}_i, \mathbf{b}_i) < n - \tau)$
$P(\text{dist}(\mathbf{a}_i, \mathbf{b}_i) \ge n - \tau)$
?
1
1
Alice
encodes 0 or 1
Bob
decodes 0, 1, or ?

Figure 7.3: Whether Bob decodes correctly or is indecisive about a keybit, depends on the Hamming Distance between the channel blocks, the blocksize n and the decision threshold τ.

Lemma 7.1. *The probabilities of decoding a keybit erroneously, correctly, or being indecisive are given by:*

$$P(k_{b_i} \neq k_{a_i}) = \sum_{j=n-\tau}^{n} \binom{n}{j} p_{ch}^{j} (1 - p_{ch})^{n-j} \tag{7.8}$$

$$P(k_{b_i} = k_{a_i}) = \sum_{j=0}^{\tau} \binom{n}{j} p_{ch}^{j} (1 - p_{ch})^{n-j} \tag{7.9}$$

$$P(k_{b_i} = \text{“?"}) = \sum_{j=\tau+1}^{n-\tau-1} \binom{n}{j} p_{ch}^{j} (1 - p_{ch})^{n-j} \tag{7.10}$$

Proof. Let $X := \text{dist}(\mathbf{a}'_i, \mathbf{b}'_i)$ the number of matches between $\mathbf{a}_i$ and $\mathbf{b}_i$). Variable X can be thought of as the number of successes of m Bernoulli trials with success probability p_{ch}. Thus, $P(X \le \tau)$ is the cumulative binomial probability. Hence, $P(k_{a_i} = k_{b_i}) = P(\text{dist}(\mathbf{a}_i, \mathbf{b}_i) \le \tau) = \sum_{j=n-\tau}^{n} \binom{n}{j} p_{\text{ch}}^{j} (1 - p_{\text{ch}})^{n-j}$. Equations (7.9) and (7.10) follow a similar proof. □

7.4.2 Performance Analysis on the final keys

7.4.3 Key Disagreement Rate

Two identical final keys occur when Bob's estimation $\mathbf{k}_b$ has no erroneous keybits, or equivalently, when the event of $k_{b_i} \neq k_{a_i}$ does not happen for any $i \in [m]$. Hence, the probability of two identical final keys is equal to

$$P(\mathbf{k}'_a = \mathbf{k}'_b) = (1 - P(k_{b_i} \neq k_{a_i}))^m \tag{7.11}$$

The probability of two matching keys are performance metrics that depend on the length of the keys. For example, the probability of two 64-bit matching keys is much higher than the probability of two matching 128-bit keys. A metric that is independent to the length and is the Key Disagreement Rate.

Definition 7.4. The key disagreement rate, KDR, is defined as the ratio of the keybit mismatches divided by total length of the (final) key:

$$\text{KDR} := \frac{\text{number of mismatches}}{\text{length of final key}} = \frac{\text{dist}(\mathbf{k}'_a, \mathbf{k}'_b)}{|\mathbf{k}'_b|} \tag{7.12}$$

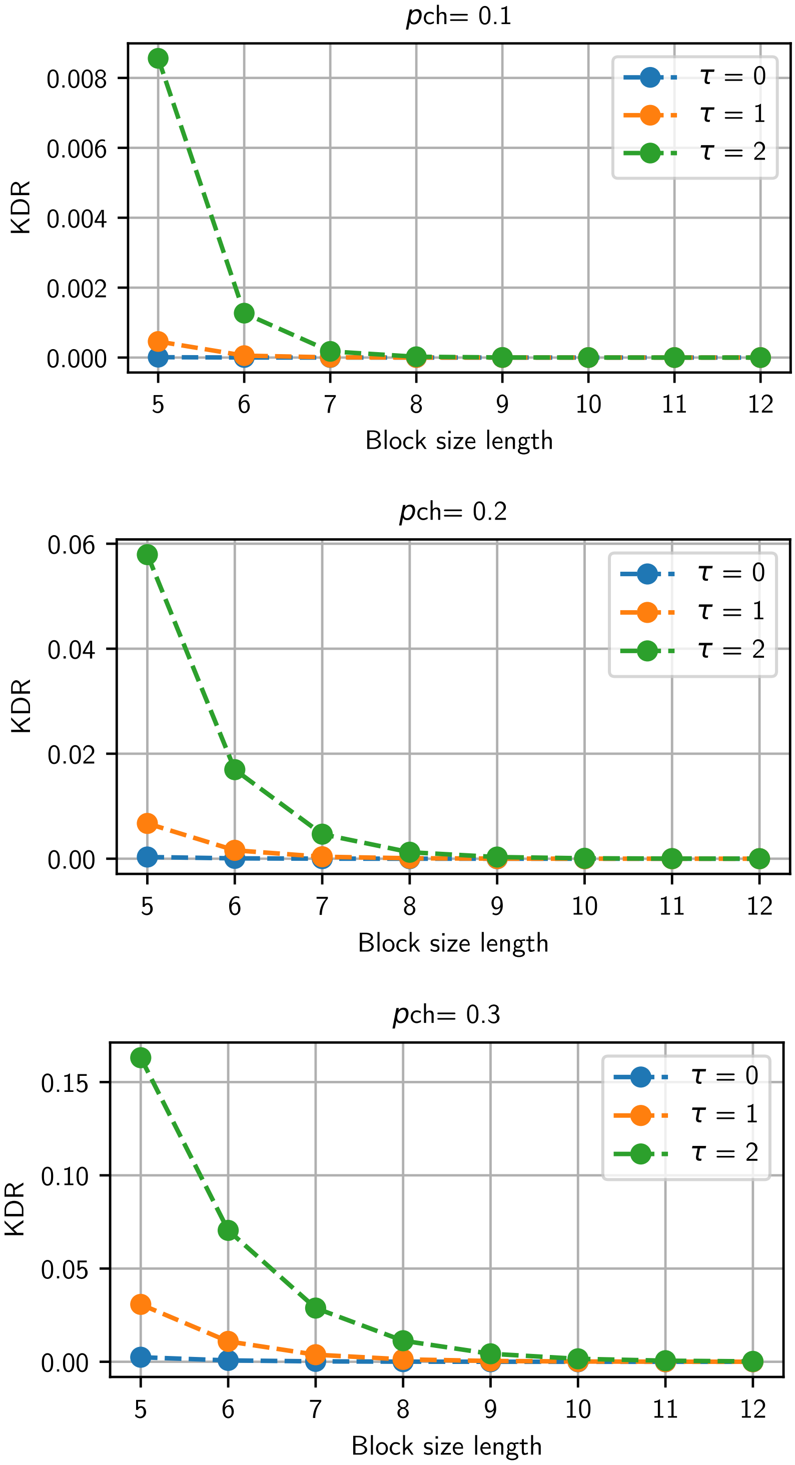


Figure 7.4: Graphs for the Key Disagreement Rate as functions of the blocksize (n) and decision threshold (τ) for three different values of channel mismatch (p_{ch}).

Theorem 7.1.

$$E(KDR) = \frac{P(k_{a_i} \neq k_{b_i})}{P(k_{a_i} \neq k_{b_i}) + P(k_{a_i} = k_{b_i})}, \tag{7.13}$$

where $P(k_{a_i} \neq k_{b_i})$ and $P(k_{a_i} = k_{b_i})$ are given by (7.8) *and* (7.9).

Proof. Observe that $\text{dist}(\mathbf{k}'_a, \mathbf{k}'_b) = X + Y$, where X and Y are the number of mismatches and the number of matches between keys $\mathbf{k}_a$ and $\mathbf{k}_b$, respectively. Variables X and Y can be thought as two random variables that follow the trinomial distribution function of m trials and corresponding probabilities p_1 and p_2. The joint probability mass function for (X, Y) is known[] to be:

$$P(X = x, Y = y) = \frac{m!}{x!y!(m-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{m-u}.$$

The length of the final key $\mathbf{k}'_b$ (or $\mathbf{k}'_a$) is the total number of matches and mismatches between $\mathbf{k}_a$ and $\mathbf{k}_b$. It is sufficient to show that $E\left(\frac{X}{X+Y}\right) = \frac{p_1}{p_1+p_2}$.

For $0 \leq x \leq u \leq m$,

$$\begin{aligned} P(X = x | X + Y = u) &= \frac{P(X = x, Y = u - x)}{P(X + Y = u)} \\ &= \frac{\frac{m!}{x!(u-x)!(n-u)!} p_1^x p_2^{u-x} (1-p_1-p_2)^{m-u}}{\binom{m}{u}(p_1+p_2)^u (1-p_1-p_2)^{m-u}}. \end{aligned}$$

That is, the conditional distribution of X given $X + Y = u$ is $\text{Binom}(u, p_1/(p_1+p_2))$

We next apply the tower rule [cite] on the expectation:

$$E\left(\frac{X}{X+Y}\right) = E\left(E\left(\frac{X}{X+Y}|X+Y\right)\right) \tag{7.14}$$

$$= E\left(\frac{1}{X+Y} E(X|X+Y)\right) \tag{7.15}$$

$$= \frac{p_1}{p_1+p_2} E\left(\frac{1}{X+Y} \cdot (X+Y)\right) \tag{7.16}$$

$$= \frac{p_1}{p_1+p_2} \tag{7.17}$$

□

The key disagreement rate is a decreasing function of n and τ. This behaviour is revealed after substituting (7.9) and (7.8) to Eq. (7.13). Figure 7.4 is a visual representation of the decreasing behaviour of the key disagreement rate for the cases of $p_{\text{ch}} = 0.1, 0.2$ and 0.3.

7.4.4 Encoding Rate

Recall that after decoding, Bob drops the uncertain bits and shares those positions with Alice. The final keys are a subset of the initial keys, i.e. $\mathbf{k}'_a \subset \mathbf{k}_a$ and $\mathbf{k}'_b \subset \mathbf{k}_b$. The expected length of the final keys is equal to

$$E(|\mathbf{a}_i|) = E(|\mathbf{b}_i|) = m(1 - P(k_{b_i} = ?)) \tag{7.18}$$

Dividing the expected length by the total number of channel bits used for encoding defines the encoding rate. A formal definition follows.

Definition 7.5. The encoding rate, denoted by R, is the ratio of the length of the final key $\mathbf{k}'_b$ to the total number of channel bits used for encoding. I.e.:

$$R := \frac{|\mathbf{k}'_b|}{nm} \tag{7.19}$$

Theorem 7.2. *The expected value of the encoding rate is given by*

$$E(R) = \frac{1}{n}(1 - P(k_i^b = \text{“?''})), \tag{7.20}$$

where $P(k_i^b = \text{“?"})$ *is given by* (7.10)

Proof. At decoding, if a keybit k_{b_i} is decoded as an uncertain bit (“?"), it is discarded from the final key-sequence. The expected number of uncertain bits is $mP(k_i^b = \text{“?"})$, and so, the expected length of the key is equal to $E(|\mathbf{k}'_a|) = m - mP(k_i^b = \text{“?"})$. Hence, $E(R) = \frac{m - mP(k_i^b = \text{“?"})}{nm} = \frac{1 - P(k_i^b = \text{“?"})}{n}$. □

Same as the key disagreement rate (see figure 7.4), the encoding rate, R, is also a decreasing function of n and τ. This behaviour is revealed after substituting (7.9) and (7.8) to Eq. (7.20). Figure 7.5 is a visual representation of the decreasing behaviour of the encoding rate for the cases of $p_{\text{ch}} = 0.1, 0.2$ and 0.3.

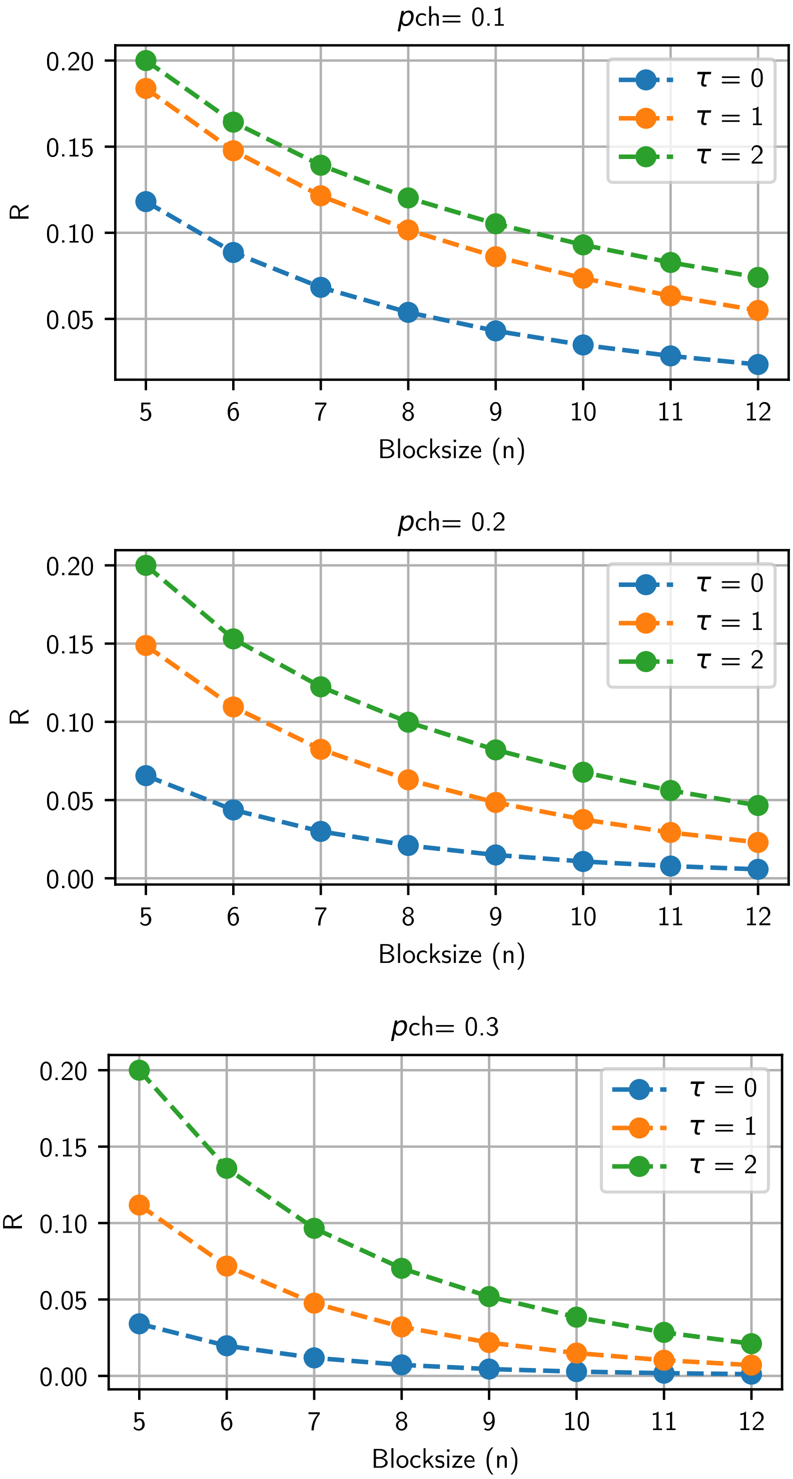


Figure 7.5: Graphs for the Encoding Rate as functions of the blocksize (n) and decision threshold (τ) for three different values of channel mismatch (p_{ch}).

7.5 Choosing parameters

The choice of τ and n is not a trivial problem. A high value of τ may decrease the key disagreement rate but also decrease the encoding rate. The same applies to the choice of n. If a communication system prioritises KDR over the encoding rate, the optimal choice of n and τ is given by:

$$(n,\tau) = \arg[\max_{n,\tau}(R\,|\text{such that } KDR \le \epsilon_1)], \tag{7.21}$$

where ϵ_1 is the maximum acceptable value for KDR. Figure 7.6 plots the achievable rates under optimal choice of parameters for the cases when KDR $< 0.01, 10^{-4}, 10^{-6}$ and 10^{-8}. Note that the sharp corners of each plot occur due to the integer nature of the tuple (n, τ). As the channel mismatch $p_c h$ increases, the optimal parameters that result in the corresponding R do not always increment simultaneously, but, when they do, they change the gradient of the plot.

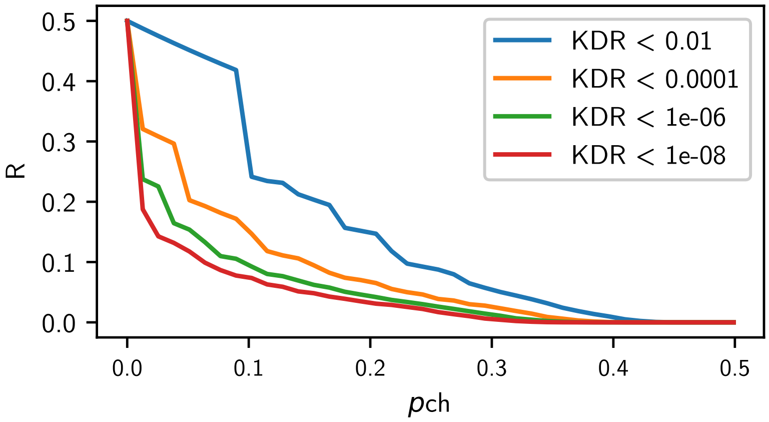


Figure 7.6: Achievable encoding rates against channel mismatch for different requirements for the key disagreement rate.

The analytical forms of KDR and R may look cumbersome, but they can be computed fast given that the summations run for a relatively small number ($n << \infty$). Although the optimisation problem is not a hard one, providing lookup tables is a convenient way to keep the amount of computations to the very minimum. The table below is an example for the case when $KDR < 10^{-4}$. Given a look-up table, Alice and Bob run through the first row to find the value that is closest to their channel mismatch. Then, they fix the encoding/decoding parameters to the corresponding values.

p_{ch}	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
n	3	5	4	10	15	19	28	35	54
τ	0	0	0	3	5	6	9	9	9

(a)

p_{ch}	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
n	4	6	7	12	21	28	32	40	64
τ	0	1	1	3	7	9	9	9	9

(b)

Table 7.2: Optimal parameters for different values of p_{ch} when the system's requirement is (a) KDR < 0.001 (b) KDR < 0.0001.

7.5.1 Information Theoretic Guarantees

CRicKET can securely transmit an unpredictable key when two requirements are met:

- Alice generates a key with high entropy (requirement for unpredictable key);
- The normalised entropy of the channel sequence is $1/n$ or more (requirement for transmitting with perfect secrecy).

The first requirement can be met when Alice is equipped with a source of true randomness. If such a source is unavailable, pseudo random key generators with good properties are widely available. For example, there exist key generators [182, 183] that pass all sixteen tests of the NIST test suit [14] and can execute on severely resource-constrained devices.

The second requirement is much more relaxed than the typical key generation requirement of independent bits within the channel sequence, i.e. $H(s_{a_i}) = 1$. The following theorem proves that CRicKET achieves perfect secrecy even when the normalised entropy of the channel is as low as $H(s_{a_i}) = 1/n$.

Theorem 7.3. *To securely transmit a key, CRicKET does not require independent channel measurements. Specifically, when $H(s_{a_i}) \geq 1/n$ CRicKET achieves perfect secrecy.*

Proof. Since $\mathbf{a}_i$ comprises of n channel bits and $H(s_{a_i}) \geq 1/n$, then $H(\mathbf{a}_i) \geq 1$. We will show that keybit k_{a_i} can be transmitted with perfect secrecy. As per the definition of perfect secrecy [], we need to show that $I(k_{a_i}; \mathbf{c}_i) = 0$, or equivalently that $H(\mathbf{c}_i | k_{a_i}) - H(\mathbf{c}_i) = 0$. We simplify the notation $P(X = x)$ to $P(x)$ to avoid congestion. Recall that $k_{a_i} \in \{0, 1\}$ is random, and so $P(k_{a_i}) = 0.5$. Also, cipher block $\mathbf{c}_i$ is equal to channel block $\mathbf{a}_i$ if $k_{a_i} = 0$, otherwise, if $k_{a_i} = 1$, $\mathbf{c}_i$ and $\mathbf{a}_i$ disagree in every single place. In other words, $a_i = \mathbf{c}_i \oplus (k_{a_i}, \ldots, k_{a_i})$. Thus,

$$P(\mathbf{c}_i, k_{a_i}) = P(\mathbf{c}_i | k_{a_i}) P(k_{a_i}) = 0.5 P(a_i = \mathbf{c}_i \oplus (k_{a_i}, \ldots, k_{a_i})) \tag{7.22}$$

By the definition of conditional entropy:

$$\begin{aligned} H(\mathbf{c}_i | k_{a_i}) &= - \sum_{k_{a_i} \in \{0,1\}} \sum_{\mathbf{c}_i \in \{0,1\}^m} P(\mathbf{c}_i, k_{a_i}) \log(P(\mathbf{c}_i | k_{a_i})) \overset{(7.22)}{=} \\ &- \sum \sum 0.5 P(\alpha_i = \mathbf{c}_i \oplus (k_{a_i}, \ldots, k_{a_i})) \log(\alpha_i = \mathbf{c}_i \oplus (k_{a_i}, \ldots, k_{a_i}) \\ &= -0.5 \sum_{\mathbf{c}_i \in \{0,1\}^m} P(\alpha_i = \mathbf{c}_i \oplus (0, \ldots, 0)) \log(\alpha_i = \mathbf{c}_i \oplus (0, \ldots, 0) + \\ &\quad - 0.5 \sum_{\mathbf{c}_i \in \{0,1\}^m} P(\alpha_i = \mathbf{c}_i \oplus (1, \ldots, 1)) \log(\alpha_i = \mathbf{c}_i \oplus (1, \ldots, 1) \\ &= \sum_{a_i \in \{0,1\}^m} P(\alpha_i) = H(k_{a_i}). \end{aligned}$$

□

7.6 Discussion

7.6.1 On the requirement of long sequences

The proposed scheme may require longer channel sequences for encoding/decoding purposes than a typical key generation protocol. For example, assuming a one-level quantisation scheme, a typical physical key generation protocol would require one channel measurement for extracting one key-bit, whereas, CRicKET would need n channel measurements for securely transmitting one key-bit. Having proved that CRicKET allows some redundancy in the channel sequence, the probing rate can be increased without compromising security. E.g., assuming a block fading channel model, up to n multiple channel measurements can be taken in each coherence block, and so CRicKET does not necessitate an increase in the duration of channel probing. Hence, even if the proposed scheme requires longer channel sequences for encoding/decoding purposes than a typical key generation protocol, it is not less competitive than typical key generation protocols when encountering slow varying channel models.

7.6.2 CRicKET as a corrective mechanism

Theorem 7.3 motivates a second usage of CRicKET, that of a corrective mechanism for secrecy enhancement. As aforementioned, multi-level quantisation or filtering mechanisms may decrease the entropy of a channel sequence. Keys extracted from low-entropy channel sequences are more vulnerable to brute-force attacks to truly random keys. On the other hand, it has been proved that CRicKET can provide information theoretic guarantees even when the channel sequences have a low entropy. As such, if CRicKET is used in combination with a high level quantisation scheme for increasing the key rate, and/or a "smoothening" filtering mechanism for decreasing p_{ch}, perfect secrecy is still achievable.

7.7 Implementation

In collaboration with Synergia[1], CRicKET has been implemented in a typical IoT network comprising devices that exchange environmental temperature and humidity measurements. The IoT devices selected by the collaborators were nRF8240-DK boards [184] which are typically defined as resource constrained, due to the limited flash/RAM memory, reduced processing capabilities, and their battery-sourced nature (see figure 7.7). The communication between the devices over the wireless network relied on the 802.15.4 standard (Zigbee).

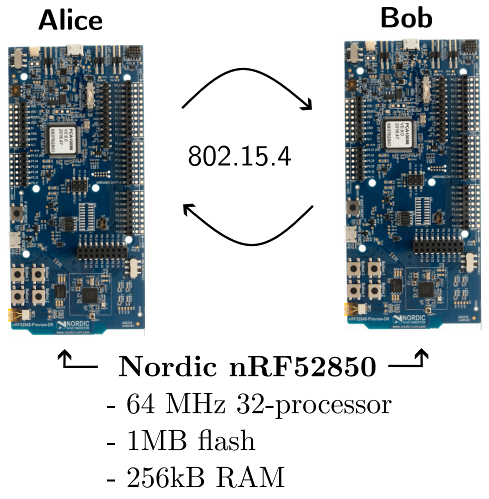


Figure 7.7: The hardware utilised to implement CRicKET.

Collaborators from Synergia were interested in implementing CRicKET seamlessly on the ongoing traffic of the IoT network. Not having a dedicated channel for the key agreement protocol meant that CRicKET had to be implemented in multiple rounds until a 128-bit key was derived at each end. The target KDR was set to be at most 0.001. The feature of the Radio-Frequency (RF) channel measured was Received Signal Strength Indicator (RSSI)

[1] Synergia is a project of University of Bristol, funded by the UK's national innovation agency and led by Toshiba Europe Limited. Synergia's aim is to increase the resilience and security of industrial IoT items.

values. RSSI values were measured every few seconds when application messages were sent by the devices subject to the key agreement protocol.

The author was asked to extend the lookup table to include the parameter of the length of the initial key at Alice, m, so that the resulted final key of each round had an average length of eight bits ($E(|\mathbf{k}'_a|) = E(|\mathbf{k}_b|') = 8$). Thus, the system's requirements were:

- KDR < 0.001
- $E(|\mathbf{k}'_a|) = E(|\mathbf{k}'_b|) = 8$

Choosing encoding parameters τ and n such that KDR < 0.001 has already been detailed in section 7.5. As for m, equation (7.18) suggests that

$$m = \frac{8}{1 - P(k_{b_i} = ?)}.$$

Table 7.2(a) of the previous section forms a subset of the lookup table used for the experiment. The complete lookup table can be found in appendix A.

7.7.1 Flow diagram

As seen in section 2.2, a 128-bit key is considered to be computationally secure at the time of writing this thesis. Aiming for 128 keybits, multiple rounds ($\approx$ 16) were repeated until the desired length was attained. The "final keys", $\mathbf{k}'_a$ and $\mathbf{k}'_b$, of each round were appended to those of the previous round until a 128-bit key was formed. Figure 7.8 illustrates the flow diagram used. Note that the term *final keys* is used for consistency with the previous section.

The algorithm CRicKET, as explained in section 7.3, constitutes a subset of the flow diagram 7.8. The remaining procedures are essential for creating the system model that facilitates CRicKET (see 7.2). The first two stages of the flow diagram supply the channel sequences. Shuffling is essential for distributing the bit-mismatches uniformly across the length of the channel sequences so that equation (7.1) is satisfied. The channel sequences resulted from RSSI measurements and a 2-level quantisation scheme. The realisation of the first two stages has been a collaborative work [185]. For details and access in the software used, the reader is referred to [185].

The requirement of Alice and Bob being aware of their channel mismatch ratio is realised in the third stage of the flow diagram "estimate p_{ch}". In this stage, Alice and Bob exchange a subset of their channel sequences so that each end can estimate p_{ch}. Simulations with different sample sizes (see figure 7.9) suggested that sixty four channel bits

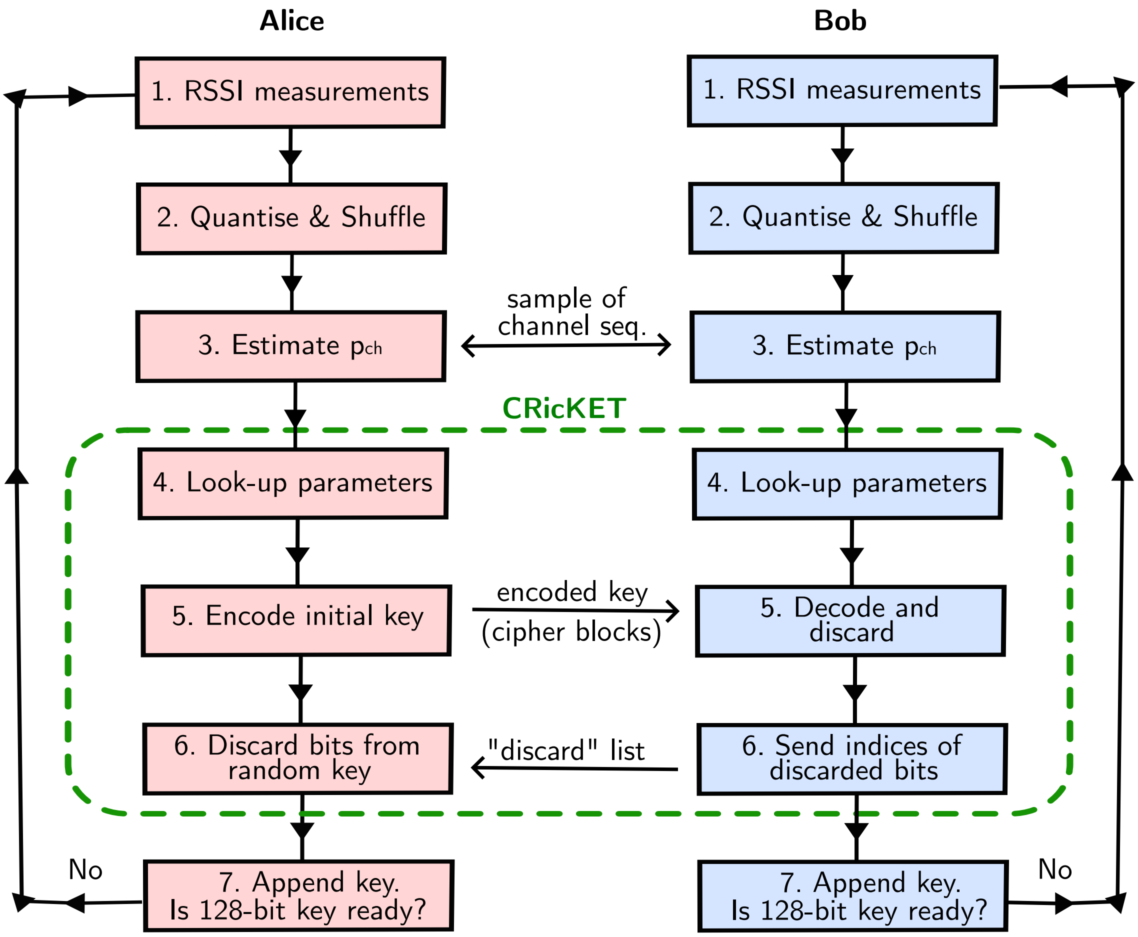


Figure 7.8: The flow diagram represents the algorithm used to program devices Alice and Bob to establish a 128-bit key.

were sufficient for an accurate estimation of the channel mismatch ratio. As such, the two boards, Alice and Bob, exchange a subset of sixty four bits from their channel sequences for estimating p_{ch}. The exposed subsets are then discarded from the channel sequences.

Lastly, the requirement of a true source of randomness at Alice for deriving the initial key (7.2) is satisfied by the features of the hardware used. That is, true random generators are available at the nRF8240-DK board [184]. Having derived the channel sequence and channel mismatch p_{ch}, Alice and Bob perform the CRICKET algorithm (in stages 4, 5, and 6). The last stage of the flow diagram appends the key to previous and checks the length. If the length is shorter than 128-bits, the algorithm is repeated.

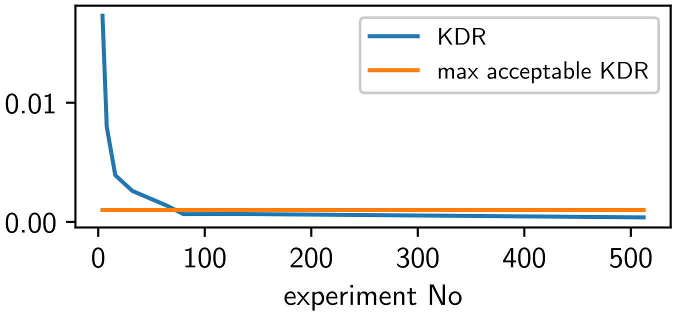


Figure 7.9: Simulations suggest that a sample size of at least sixty four bits is needed for eliminating the performance degradation due to estimation errors

7.7.2 Experimental results

Out of the twenty one pairs of collected keys, only one bit-mismatch was found in one pair. Specifically, the key disagreement rate of the experiment with a confidence interval of 95% was .0003 ± .0007 which meets the system's requirement of $KDR < 0.001$.

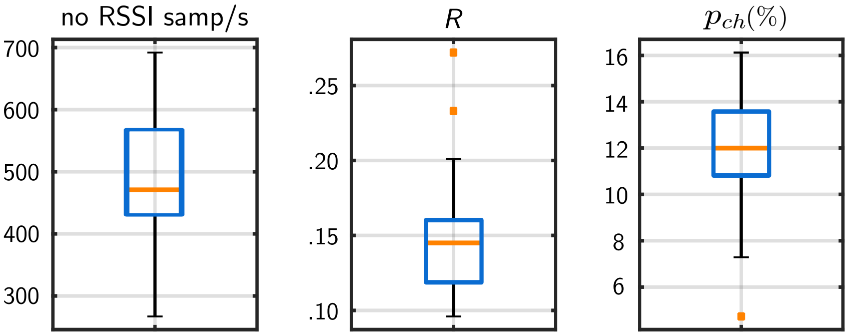


Figure 7.10: Box plots for the number of RSSI samples, encoding rate (R), and channel mismatch (p_{ch}).

The average channel mismatch ratio varied from 7.8% to 16.2% with an average value of $\overline{p_{ch}} = 11.82\%$. Subsequently, the encoding rate varied significantly from 0.1 to 0.27 with an average value of $R = 0.15$. Hence, a 128-bit key required 853 channel bits on average or equivalently, given the 4-level quantisation scheme, 427 RSSI samples. For more statistical metrics, figure 7.10 provides the box plots for the quantities of p_{ch}, R, and number of RSSI samples required for each 128-bit key.

To examine whether the keys have been transferred with perfect secrecy, the channel bit entropy, $H(s_{a_i})$, has been approximated using the "approximate entropy" function provided by NIST [14]. From theorem 7.3, the requirement of perfect secrecy requires $H(s_{a_i}) \geq$

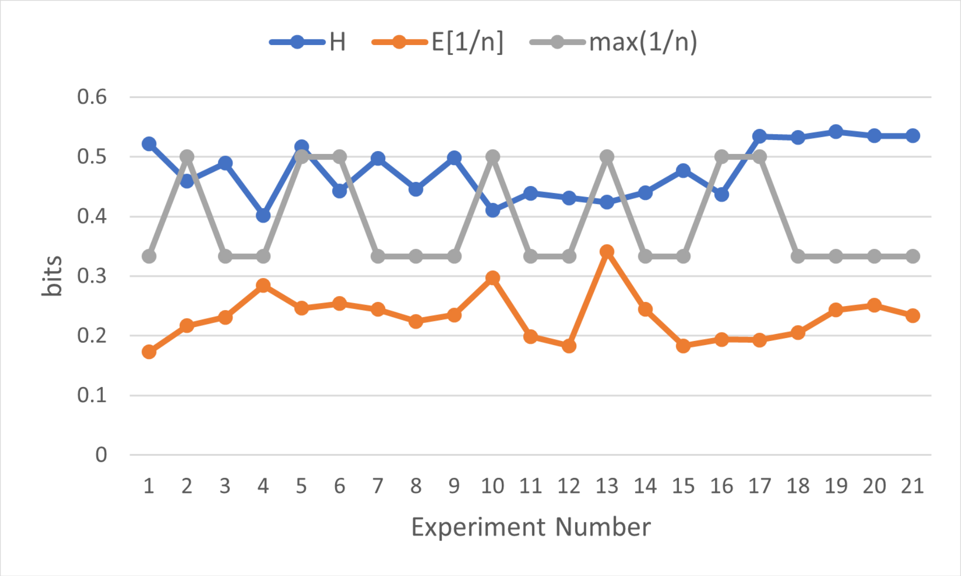


Figure 7.11: Testing the requirement for perfect secrecy.

$1/n$. Since a 128-key bit may have required multiple rounds (as seen in figure7.8), the blocksize n may have taken different values to compensate for a change in the channel mismatch.

To give an insight into how $1/n$ compares to the channel's entropy, figure 7.11 compares $H(s_{a_i})$ with the average and maximum value of $1/n$ for each experiment. Note that when $H(s_{a_i}) > \max(1/n)$, the requirement of perfect secrecy ($H(s_{a_i}) > 1/n$) is satisfied. Sixteen out of twenty-one 128-bit keys have been transmitted with perfect secrecy, whereas for the remaining five keys, approximately 12% of their keybits were transmitted with perfect secrecy. Note that these five keys were agreed upon during non-operating times when the office was quiet and the entropy $H(s_{a_i})$ dropped below 0.5.

7.8 Conclusion

CRicKET addresses the key disagreement issue that arises during the early stages of PLKG, where a substantial number of bits may disagree, necessitating resource-intensive reconciliation. The main distinction from existing work that aims to reduce reconciliation costs is CRicKET's efficiency in decreasing the KDR, which is independent of the outputs of the quantisation phase. The encoding/decoding parameters can be adjusted to compensate for many mismatches resulting from the quantisation phase, guaranteeing its performance.

Assuming that an eavesdropper observes decorrelated fading compared to a legitimate node, CRicKET is proven to be a secure protocol. This assumption holds validity in location-sensitive channels, commonly found in indoor environments. Specifically, it has been proven that perfect secrecy is attainable even when there is redundancy in the sequences obtained during the quantisation phase.

It is crucial to note that CRicKET may not be viable in static environments or scenarios where the key needs to be refreshed frequently within an hour. The time required for key generation increases linearly with the desired key length, and as a consequence, a slow varying channel may not facilitate a fast key generation scheme. The advantageous aspect is that a longer key does not trade off computational complexity. Given sufficient time, CRicKET can offer security even in resource-constrained devices.

CRicKET, as effectively demonstrated through real-world implementation on single-chip IoT devices, underscores its practicality and applicability. This successful deployment solidifies CRicKET's potential as a practical and viable protocol for secure key generation in delay-tolerant, resource-constrained environments, particularly within the realm of IoT and personal area networks.

Chapter 8

Conclusions and Further Work

This thesis examines the physical layer for confidentiality and authentication purposes, specific to short-range systems. The literature review in chapter 3 covers the evolution of Physical Layer Security (PLS) and its expansion to different disciplines. The main objective of keyless PLS and key-based PLS is confidentiality, whereas, the field of Physical Layer Authentication (PLA) is concerned with authentication. The literature review also reveals that PLS has been mostly considered by the community of information-theoretic (IT) security. The consideration of PLS by wireless engineers is a recent phenomenon followed by advances in wireless technology. Building practical and robust secrecy systems requires the coupling of wireless engineering and IT security. On that account, chapter 2 and part of chapter 3 provide a comprehensive background of IT security and wireless communications.

Whereas some believe that PLS can replace the overall security mechanisms for confidentiality, the literature review in chapter 3 indicates that there are many challenges to be addressed before such a belief can be proven true. For this reason, this thesis promotes a cross-over approach to confidentiality, exploiting the physical layer for deriving or transferring keys which can be used to facilitate upper-layer encryption. Focusing on key-based PLS, chapter 4 enhances our understanding of how spatial channel correlation impacts typical Physical-Layer (PL) key generation protocols in terms of secret key capacity and invites researchers to be judicious in their assumptions. It establishes that the typical definition of coherence distance is not appropriate for secrecy purposes, and re-defines *secure distance*.

Two novel methods for secure key transfer are proposed in chapters 5 and 7. The first method, namely secrecy coding, borrows concepts from keyless PLS, whereas the method of chapter 7, Channel Reciprocity for KEy Transmission (CRicKET), builds upon key-based PLS. Secret splitting (chapter 5) addresses one of the main challenges of keyless PLS which is the requirement of a positive secrecy gap. A positive secrecy gap is provided through base station cooperation and transmit beamforming. As long as they are placed diametrically opposite to each other with respect to the receiver, two base stations suffice for dramatically degrading the decoding capabilities of unintended receivers. CRicKET is an encoding/decoding mechanism that can achieve an arbitrarily low key disagreement rate without increasing the computational complexity or leaking information. CRicKET (chapter 7) solves the problem associated with a high reconciliation cost in resource-constrained networks. Optimal encoding/decoding parameters are provided via analytical forms and lookup tables. The practicality of the method has been successfully tested in Internet-of-Things (IoT) devices. Under the assumption that an eavesdropper observes decorrelated fading to that of a legitimate node, CRicKET is proven to achieve perfect secrecy.

Contributing to the literature of PLA, chapter 6 introduces the novel concept of exploiting spatial channel correlation for authenticating co-located devices. Two methods originate from this concept, namely CHannel Reflection Yields Secure Proximity (CHRYSP) and (against) Solo Distance Fraud for RFID (SDF-RFID). Under the assumption of perfect channel state information, the two methods detect distance fraud and replay attacks with high accuracy. Specifically, with SDF-RFID, an Radio Frequency Identification (RFID) reader can verify the proximity of an RFID tag by applying signal processing techniques on the backscattered signal. The RFID tag is only required to send typical RFID data, thus, the application of the method is seamless from the tag's point of view. CHRYSP can be employed by a wider range of short-range systems for authenticating co-located devices. CHRYSP integrates signal properties with cryptographic mechanisms for providing protection against most types of distance fraud.

Further Work

To advance the methodologies presented in this thesis, the following paragraphs recommend exploring additional research directions.

Secure distance in MIMO Key Generation Systems. Chapter 4 defines secure distance as the required physical separation between the eavesdropper and a legitimate node in physical layer key generation systems. The findings of chapter 4 could also provide insights into Multiple Input Multiple Output (MIMO) key generation systems regarding the required antenna separation within a legitimate node. Specifically, the concept of secure distance could be expanded to explore vulnerabilities arising from assuming independent key streams generated by a MIMO system, as well as how the key rate is influenced as a function of the distance between the antennas within a node. Considering vulnerabilities stemming from spatial channel correlation, attack scenarios could be designed from the eavesdropper's perspective, assessing the ease or difficulty of exploiting this vulnerability to gain insights into the secret key.

CRicKET for FDD systems. Chapter 7 shows that the key agreement protocol, CRicKET, is a viable solution for resource-constrained networks that operate in Time Division Duplexing (TDD) mode. An interesting direction is the implementation of CRicKET over Frequency Division Duplexing (FDD) communication systems. The frequency separation of the pilot exchange during the first stage of CRicKET (see chapter 7.3) needs to be such that Alice and Bob observe correlated multipath angles and delays on the incoming pilot signals. Then, the angle and/or delay information can serve as the correlated observations of Alice and Bob. Note that even if the correlation is very low, CRicKET does not result in a high reconciliation cost, i.e. as seen in chapter 7 CRicKET does not trade off between bit-mismatch-rate and reconciliation cost. To date, a practical physical key agreement protocol for FDD systems is still missing. CRicKET is believed to be a viable solution for low power/low memory and delay-tolerant applications that communicate using FDD.

CRicKET and directive channels. Under the assumption that the eavesdropper experiences decorrelated fading, CRicKET is proven to achieve perfect secrecy. According to the findings of chapter 4, this assumption can be considered valid in a rich scattering environment as long as the eavesdroppers are positioned several wavelengths away from either legitimate receiver. Future research could take into consideration the channel correlation between the eavesdropper's channel and the legitimate channel. To compensate for the information leakage resulting from the correlation, it is suggested that a privacy amplification phase follows after CRiCKET, the compression ratio of which is dependent on the level of channel correlation.

CHRYSP/SDF-RFID and practical considerations. Numerical results in Chapter 6 suggest that CHRYSP and SDF-RFID show promise as solutions against distance fraud and replay attacks. The analysis assumes that there is zero correlation between the intruder's channel and the legitimate channel. However, in reality, the channels may correlate by a small value. In the case of a “fast” replay attack, the channels may even correlate by a considerable amount. To ensure the continued effectiveness of CHRYSP and SDF-RFID schemes, the sample size of channel measurements will likely need to be increased for an even more accurate estimation of the channel correlation at the verifier. Depending on the quality of service requirements (e.g. given an acceptable false negative rate), the number of additional sample sizes could be determined as a function of the channel correlation.

The scenario of imperfect channel estimation could also be considered. Imperfect channel estimation will lead to larger errors in estimating the channel correlation. To compensate for imperfect channel estimation, a greater number of independent channel measurements may be needed at the verifier. To facilitate a greater number of independent channel measurements within a short period, the suggested schemes could be combined with Reconfigurable Intelligent Surfaces (RIS) technology or MIMO systems. For instance, a set of M sufficiently spaced antennas at the verifier could capture M times higher entropy than a single antenna within the same time interval. If RIS technology is employed, it could facilitate a dynamic multipath environment, which will also increase the channel entropy.

Secret Splitting and Air-to-Ground Communications. The method of secret splitting for confidentiality in chapter 5 has considered a channel model in a two-dimensional (2D) environment. That is, the legitimate users and the eavesdropper move on a horizontal plane. Motivated by the growing field of air-to-ground communications, an interesting research direction is the analysis of secrecy splitting in a three-dimensional (3D) setting, whereby a legitimate transceiver is equipped with airborne base stations. A line=of-sight component between the base station(s) and the legitimate receiver is expected to be beneficial for increasing the secrecy gap between the legitimate receiver and the eavesdropper(s). On the other hand, if the eavesdropper is equipped with airborne equipment, the areas over which she has an advantage in signal quality may increase.

Through the results presented in this thesis, the state-of-the-art in PLS has been furthered, thus providing a viable basis for security enhancement to wireless networks of small devices.

Appendix A

Lookup Table for CRicKET

The lookup table used for the implementation of CRicKET (see section 7.7) is shown in the first four columns. Variables p_{ch}, n, τ are chosen such that the encoding rate, R, is maximised whilst satisfying the requirement that the key disagreement rate (KDR) is less than or equal to 0.001, i.e KDR $\leq$ 0.001. The achievable rate is given in the rightmost column. The length of the initial key is such that the final keys at the two communicating parties have an average length of 8 bits.

channel mismatch (p_{ch})	blocksize (n)	decision threshold (τ)	length of initial key (m)		encoding rate (R)
0.01	2	0	9		0.4901
0.02	2	0	9		0.4804
0.03	2	0	9		0.4709
0.04	3	0	10		0.2949
0.05	3	0	10		0.2858
0.06	3	0	10		0.2769
0.07	3	0	10		0.2682
0.08	3	0	11		0.2597
0.09	3	0	11		0.2514
0.1	5	1	9		0.1838
0.11	5	1	9		0.1808
0.12	4	0	14		0.15
0.13	4	0	14		0.1433
0.14	4	0	15		0.1368
0.15	4	0	16		0.1306
0.16	6	1	11		0.1256
0.17	8	2	10		0.1074
0.18	8	2	10		0.105
0.19	10	3	9		0.0897
0.2	10	3	10		0.088
0.21	9	2	12		0.0791
0.22	9	2	12		0.0761
0.23	11	3	11		0.0698
0.24	13	4	10		0.063
0.25	15	5	10		0.0568
0.26	12	3	13		0.0515
0.27	14	4	12		0.0487
0.28	18	6	11		0.0435
0.29	22	8	10		0.0383
0.3	19	6	12		0.0351

Bibliography

[1] C. E. Shannon, “Communication theory of secrecy systems,” *The Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[2] S. D. Galbraith, *Mathematics of public key cryptography*. Cambridge University Press, 2012.

[3] A. Sanenga, G. A. Mapunda, T. M. L. Jacob, L. Marata, B. Basutli, and J. M. Chuma, “An overview of key technologies in physical layer security,” *Entropy*, vol. 22, no. 11, p. 1261, 2020.

[4] C. Lv, H. Li, J. Ma, and Y. Zhang, “Vulnerability analysis of elliptic curve cryptography-based RFID authentication protocols,” *Transactions on Emerging Telecommunications Technologies*, vol. 23, no. 7, pp. 618–624, 2012.

[5] L. Sun and Q. Du, “Physical layer security with its applications in 5G networks: A review,” *China Communications*, vol. 14, no. 12, pp. 1–14, 2017.

[6] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, “Fingerprints in the ether: Using the physical layer for wireless authentication,” in *2007 IEEEInternational Conference on Communications*, pp. 4646–4651, IEEE, 2007.

[7] C. E. Shannon, “A mathematical theory of communication,” *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[8] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[9] A. D. Wyner, “The wire-tap channel,” *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[10] Y. V. Prohorov and Y. A. Rozanov, *Probability theory: basic concepts, limit theorems, random processes*, vol. 18. Springer, 1969.

[11] C. B. Do, “The multivariate gaussian distribution,” *Section Notes, Lecture on Machine Learning, CS*, vol. 229, 2008.

[12] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[13] G. Gong, T. Helleseth, and P. V. Kumar, “Solomon w. golomb—mathematician, engineer, and pioneer,” *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2844–2857, 2018.

[14] J. Soto, *Statistical testing of random number generators*, vol. 10. NIST Gaithersburg, MD, 1999.

[15] K. Zeng, C.-H. Yang, D.-Y. Wei, and T. Rao, “Pseudorandom bit generators in stream-cipher cryptography,” *Computer*, vol. 24, no. 2, pp. 8–17, 1991.

[16] M. D. Galanis, P. Kitsos, G. Kostopoulos, N. Sklavos, and C. E. Goutis, “Comparison of the hardware implementation of stream ciphers.,” *Int. Arab J. Inf. Technol.*, vol. 2, no. 4, pp. 267–274, 2005.

[17] O. D. Jensen and K. A. Andersen, “A5 encryption in gsm,” 2017.

[18] A. H. Lashkari, M. M. S. Danesh, and B. Samadi, “A survey on wireless security protocols (wep, wpa and wpa2/802.11 i),” in *2009 2nd IEEE international conference on computer science and information technology*, pp. 48–52, IEEE, 2009.

[19] P. Patil, P. Narayankar, D. Narayan, and S. M. Meena, “A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and blowfish,” *Procedia Computer Science*, vol. 78, pp. 617–624, 2016.

[20] H. Landau, “Sampling, data transmission, and the nyquist rate,” *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.

[21] L. Hanzo, W. Webb, and T. Keller, *Single-and multi-carrier quadrature amplitude modulation.* John Wiley & Sons West Sussex, England, 2000.

[22] J. H. Schiller, *Mobile communications.* Pearson education, 2003.

[23] J. G. Proakis, *Digital communications.* McGraw-Hill, Higher Education, 2008.

[24] H. Arslan and G. E. Bottomley, "Channel estimation in narrowband wireless communication systems," *Wireless Communications and Mobile Computing*, vol. 1, no. 2, pp. 201–219, 2001.

[25] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE transactions on Vehicular Technology*, vol. 43, no. 4, pp. 1049–1057, 1994.

[26] L. Tian, X. Yin, X. Zhou, and Q. Zuo, "Spatial cross-correlation modeling for propagation channels in indoor distributed antenna systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–11, 2013.

[27] U. Schilcher, J. F. Schmidt, M. K. Atiq, and C. Bettstetter, "Autocorrelation and coherence time of interference in poisson networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 7, pp. 1506–1518, 2019.

[28] P. D. Teal, T. D. Abhayapala, and R. A. Kennedy, "Spatial correlation for general distributions of scatterers," *IEEE signal processing letters*, vol. 9, no. 10, pp. 305–308, 2002.

[29] M. K. Ozdemir, E. Arvas, and H. Arslan, "Dynamics of spatial correlation and implications on MIMO systems," *IEEE Communications Magazine*, vol. 42, no. 6, pp. S14–S19, 2004.

[30] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell system technical journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[31] X. Zhou, L. Song, and Y. Zhang, *Physical layer security in wireless communications*. Crc Press, 2013.

[32] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE transactions on information theory*, vol. 24, no. 3, pp. 339–348, 1978.

[33] S. Leung-Yan-Cheong and M. Hellman, "The gaussian wire-tap channel," *IEEE transactions on information theory*, vol. 24, no. 4, pp. 451–456, 1978.

[34] P. Parada and R. Blahut, "Secrecy capacity of simo and slow fading channels," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pp. 2152–2155, IEEE, 2005.

[35] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4687–4698, 2008.

[36] J. Barros and M. R. Rodrigues, “Secrecy capacity of wireless channels,” in *2006 IEEE International Symposium on Information Theory*, pp. 356–360, IEEE, 2006.

[37] Y. Liang, H. V. Poor, and S. Shamai, “Secure communication over fading channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2470–2492, 2008.

[38] Z. Li, R. Yates, and W. Trappe, “Secrecy capacity of independent parallel channels,” in *Securing Wireless Communications at the Physical Layer*, pp. 1–18, Springer, 2009.

[39] J. D. V. Sánchez, L. Urquiza-Aguiar, M. C. P. Paredes, and D. P. M. Osorio, “Survey on physical layer security for 5g wireless networks,” *Annals of Telecommunications*, vol. 76, no. 3, pp. 155–174, 2021.

[40] M. Bloch, O. Günlü, A. Yener, F. Oggier, H. V. Poor, L. Sankar, and R. F. Schaefer, “An overview of information-theoretic security and privacy: Metrics, limits and applications,” *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 5–22, 2021.

[41] W. K. Harrison, J. Almeida, M. R. Bloch, S. W. McLaughlin, and J. Barros, “Coding for secrecy: An overview of error-control coding techniques for physical-layer security,” *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 41–50, 2013.

[42] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund, “Performance analysis and design of two edge-type LDPC codes for the BEC wiretap channel,” *IEEE transactions on information theory*, vol. 59, no. 2, pp. 1048–1064, 2012.

[43] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund, “Rate-equivocation optimal spatially coupled LDPC codes for the BEC wiretap channel,” in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2393–2397, IEEE, 2011.

[44] Y. Wu, A. Khisti, C. Xiao, G. Caire, K.-K. Wong, and X. Gao, “A survey of physical layer security techniques for 5G wireless networks and challenges ahead,” *IEEEJournal on Selected Areas in Communications*, vol. 36, no. 4, pp. 679–695, 2018.

[45] E. Şaşoğlu and A. Vardy, “A new polar coding scheme for strong security on wiretap channels,” in *2013 IEEE International Symposium on Information Theory*, pp. 1117–1121, IEEE, 2013.

[46] T. C. Gulcu and A. Barg, “Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component,” *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1311–1324, 2016.

[47] R. A. Chou and M. R. Bloch, “Polar coding for the broadcast channel with confidential messages: A random binning analogy,” *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2410–2429, 2016.

[48] C. Yang, M. Zhan, Y. Deng, M. Wang, X. H. Luo, and J. Zeng, “Error-correcting performance comparison for polar codes, LDPC codes and convolutional codes in high-performance wireless,” in *2019 6th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, pp. 258–262, IEEE, 2019.

[49] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, “Semantically secure lattice codes for the gaussian wiretap channel,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6399–6416, 2014.

[50] L.-C. Choo and C. Ling, “Superposition lattice coding for gaussian broadcast channel with confidential message,” in *2014 IEEE Information Theory Workshop (ITW 2014)*, pp. 311–315, IEEE, 2014.

[51] X. He and A. Yener, “The gaussian many-to-one interference channel with confidential messages,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2730–2745, 2011.

[52] Z. Li, R. Yates, and W. Trappe, “Achieving secret communication for fast rayleigh fading channels,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2792–2799, 2010.

[53] J.-C. Belfiore and F. Oggier, “Lattice code design for the rayleigh fading wiretap channel,” in *2011 IEEE International Conference on Communications Workshops (ICC)*, pp. 1–5, IEEE, 2011.

[54] M. Baldi, F. Chiaraluce, N. Laurenti, S. Tomasin, and F. Renna, “Secrecy transmission on parallel channels: Theoretical limits and performance of practical codes,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 1765–1779, 2014.

[55] T. Q. Duong, X. Zhou, and H. V. Poor, *Trusted communications with physical layer security for 5G and beyond*, vol. 76. IET, 2017.

[56] H. V. Poor and R. F. Schaefer, “Wireless physical layer security,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 1, pp. 19–26, 2017.

[57] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE transactions on wireless communications*, vol. 7, no. 6, pp. 2180–2189, 2008.

[58] R. Negi and S. Goel, "Secret communication using artificial noise," in *IEEE Vehicular Technology Conference*, vol. 62, p. 1906, Citeseer, 2005.

[59] J. Xiong, K.-K. Wong, D. Ma, and J. Wei, "A closed-form power allocation for minimizing secrecy outage probability for MISO wiretap channels via masked beamforming," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1496–1499, 2012.

[60] L. Hu, H. Wen, B. Wu, F. Pan, R.-F. Liao, H. Song, J. Tang, and X. Wang, "Cooperative jamming for physical layer security enhancement in internet of things," *IEEEInternet of Things Journal*, vol. 5, no. 1, pp. 219–228, 2018.

[61] A. Mukherjee, "Physical-layer security in the internet of things: Sensing and communication confidentiality under resource constraints," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1747–1761, 2015.

[62] H. Xing, K.-K. Wong, A. Nallanathan, and R. Zhang, "Wireless powered cooperative jamming for secrecy multi-af relaying networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7971–7984, 2016.

[63] M. R. Khandaker, K.-K. Wong, Y. Zhang, and Z. Zheng, "Probabilistically robust swipt for secrecy MISOME systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 211–226, 2017.

[64] J. Li, S. Chang, X. Fu, L. Zhang, Y. Su, and Z. Jin, "A coalitional formation game for physical layer security of cooperative compressive sensing multi-relay networks," *Sensors*, vol. 18, no. 9, p. 2942, 2018.

[65] K. Wang, L. Yuan, T. Miyazaki, D. Zeng, S. Guo, and Y. Sun, "Strategic antieavesdropping game for physical layer security in wireless cooperative networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9448–9457, 2017.

[66] X. Li, H.-N. Dai, M. K. Shukla, D. Li, H. Xu, and M. Imran, "Friendly-jamming schemes to secure ultra-reliable and low-latency communications in 5g and beyond communications," *Computer Standards & Interfaces*, vol. 78, p. 103540, 2021.

[67] Y. Huo, Y. Tian, L. Ma, X. Cheng, and T. Jing, "Jamming strategies for physical layer security," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 148–153, 2018.

[68] K. Cumanan, H. Xing, P. Xu, G. Zheng, X. Dai, A. Nallanathan, Z. Ding, and G. K. Karagiannidis, “Physical layer security jamming: Theoretical limits and practical designs in wireless networks,” *IEEEAccess*, vol. 5, pp. 3603–3611, 2017.

[69] K. Cumanan, H. Xing, P. Xu, G. Zheng, X. Dai, A. Nallanathan, Z. Ding, and G. K. Karagiannidis, “Physical layer security jamming: Theoretical limits and practical designs in wireless networks,” *IEEE Access*, vol. 5, pp. 3603–3611, 2016.

[70] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering.* Cambridge University Press, 2011.

[71] U. M. Maurer, “Secret key agreement by public discussion from common information,” *IEEE transactions on information theory*, vol. 39, no. 3, pp. 733–742, 1993.

[72] G. Li, C. Sun, J. Zhang, E. Jorswieck, B. Xiao, and A. Hu, “Physical layer key generation in 5G and beyond wireless communications: Challenges and opportunities,” *Entropy*, vol. 21, no. 5, p. 497, 2019.

[73] K. Zeng, “Physical layer key generation in wireless networks: challenges and opportunities,” *IEEE Communications Magazine*, vol. 53, no. 6, pp. 33–39, 2015.

[74] M. S. Kumar, R. Ramanathan, M. Jayakumar, and D. K. Yadav, “Secret key generation schemes for physical layer security.,” *Defence Science Journal*, vol. 71, no. 4, 2021.

[75] S. N. Premnath, S. Jana, J. Croft, P. L. Gowda, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, “Secret key extraction from wireless signal strength in real environments,” *IEEE Transactions on mobile Computing*, vol. 12, no. 5, pp. 917–930, 2012.

[76] G. Margelis, X. Fafoutis, G. Oikonomou, R. Piechocki, T. Tryfonas, and P. Thomas, “Physical layer secret-key generation with discreet cosine transform for the internet of things,” in *2017 IEEE international conference on communications (ICC)*, pp. 1–6, IEEE, 2017.

[77] W. Xu, S. Jha, and W. Hu, “Exploring the feasibility of physical layer key generation for LoRaWAN,” in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 231–236, IEEE, 2018.

[78] G. Li, A. Hu, J. Zhang, L. Peng, C. Sun, and D. Cao, “High-agreement uncorrelated secret key generation based on principal component analysis preprocessing,” *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 3022–3034, 2018.

[79] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, “Key generation from wireless channels: A review,” *IEEE access*, vol. 4, pp. 614–626, 2016.

[80] G. Li, Z. Zhang, Y. Yu, and A. Hu, “A hybrid information reconciliation method for physical layer key generation,” *Entropy*, vol. 21, no. 7, p. 688, 2019.

[81] G. Brassard and L. Salvail, “Secret-key reconciliation by public discussion,” in *Workshop on the Theory and Application of of Cryptographic Techniques*, pp. 410–423, Springer, 1993.

[82] S. T. Ali, V. Sivaraman, and D. Ostry, “Eliminating reconciliation cost in secret key generation for body-worn health monitoring devices,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2763–2776, 2013.

[83] Z. Sheng, C. Mahapatra, C. Zhu, and V. C. Leung, “Recent advances in industrial wireless sensor networks toward efficient management in iot,” *IEEE access*, vol. 3, pp. 622–637, 2015.

[84] T. I. Calver, “An empirical analysis of the cascade secret key reconciliation protocol for quantum key distribution,” 2011.

[85] S. T. Ali, V. Sivaraman, and D. Ostry, “Eliminating reconciliation cost in secret key generation for body-worn health monitoring devices,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2763–2776, 2014.

[86] K. Zeng, D. Wu, A. Chan, and P. Mohapatra, “Exploiting multiple-antenna diversity for shared secret key generation in wireless networks,” in *2010 Proceedings IEEEINFOCOM*, pp. 1–9, IEEE, 2010.

[87] H. Liu, J. Yang, Y. Wang, Y. J. Chen, and C. E. Koksal, “Group secret key generation via received signal strength: Protocols, achievable rates, and implementation,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2820–2835, 2014.

[88] C. D. T. Thai, J. Lee, and T. Q. Quek, “Physical-layer secret key generation with colluding untrusted relays,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1517–1530, 2016.

[89] Y. Wei, K. Zeng, and P. Mohapatra, “Adaptive wireless channel probing for shared key generation based on PID controller,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1842–1852, 2013.

[90] C. T. Zenger, M.-J. Chur, J.-F. Posielek, C. Paar, and G. Wunder, "A novel key generating architecture for wireless low-resource devices," in *2014 International Workshop on Secure Internet of Things*, pp. 26–34, IEEE, 2014.

[91] R. Cavallari, F. Martelli, R. Rosini, C. Buratti, and R. Verdone, "A survey on wireless body area networks: Technologies and design challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1635–1657, 2014.

[92] L. Lai, Y. Liang, and W. Du, "Cooperative key generation in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, pp. 1578–1588, 2012.

[93] D. Chen, Z. Qin, X. Mao, P. Yang, Z. Qin, and R. Wang, "Smokegrenade: An efficient key generation protocol with artificial interference," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1731–1745, 2013.

[94] M. Talal, A. Zaidan, B. Zaidan, A. Albahri, A. Alamoodi, O. Albahri, M. Alsalem, C. Lim, K. Tan, W. Shir, *et al.*, "Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: Multi-driven systematic review," *Journal of medical systems*, vol. 43, no. 3, p. 42, 2019.

[95] H. Liu, Y. Wang, J. Yang, and Y. Chen, "Fast and practical secret key extraction by exploiting channel response," in *2013 Proceedings IEEEINFOCOM*, pp. 3048–3056, IEEE, 2013.

[96] W. Xi, X.-Y. Li, C. Qian, J. Han, S. Tang, J. Zhao, and K. Zhao, "Keep: Fast secret key extraction protocol for d2d communication," in *2014 IEEE 22nd International Symposium of Quality of Service (IWQoS)*, pp. 350–359, IEEE, 2014.

[97] A. Bensky, *Short-range wireless communication.* Newnes, 2019.

[98] N. C. Karmakar, *Handbook of smart antennas for IEEE systems.* John Wiley & Sons, 2011.

[99] K. Finkenzeller, *RFID handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication.* John wiley & sons, 2010.

[100] H. Knospe and H. Pohl, "RFID security," *Information security technical report*, vol. 9, no. 4, pp. 39–50, 2004.

[101] D. M. Dobkin, *The RF in IEEE: UHF IEEE in practice.* Newnes, 2012.

[102] E. Callaway, P. Gorday, L. Hester, J. A. Gutierrez, M. Naeve, B. Heile, and V. Bahl, "Home networking with IEEE 802.15. 4: a developing standard for low-rate wireless personal area networks," *IEEE Communications magazine*, vol. 40, no. 8, pp. 70–77, 2002.

[103] Y. Hao and R. Foster, "Wireless body sensor networks for health-monitoring applications," *Physiological measurement*, vol. 29, no. 11, p. R27, 2008.

[104] A. S. Albahri, A. Zaidan, O. S. Albahri, B. Zaidan, and M. Alsalem, "Real-time fault-tolerant mhealth system: Comprehensive review of healthcare services, opens issues, challenges and methodological aspects," *Journal of medical systems*, vol. 42, no. 8, pp. 1–56, 2018.

[105] F. Gonçalves, J. Macedo, M. J. Nicolau, and A. Santos, "Security architecture for mobile e-health applications in medication control," in *2013 21st international conference on software, telecommunications and computer networks-(SoftCOM 2013)*, pp. 1–8, IEEE, 2013.

[106] G. Avoine, M. A. Bingöl, I. Boureanu, S. Čapkun, G. Hancke, S. Kardaş, C. H. Kim, C. Lauradoux, B. Martin, J. Munilla, *et al.*, "Security of distance-bounding: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–33, 2018.

[107] Y. E. Post, "How to prevent your car being stolen without keys, as thefts continue to rise." https://www.yorkshireeveningpost.co.uk/read-this/how-to-prevent-your-car-being-stolen-without-keys-as-thefts-continue-to-rise-3639548.

[108] T. car expert, "To key or not to key?." https://www.thecarexpert.co.uk/to-key-or-not-to-key/.

[109] I. news, "Insurance crime bureau warns of rise in car thefts through relay attacks." https://www.iol.co.za/business-report/economy/insurance-crime-bureau-warns-of-rise-in-car-thefts-through-relay-attacks-b48accbf-ba27-4668-9d80-9577302324af.

[110] S. Akter, S. Chellappan, T. Chakraborty, T. A. Khan, A. Rahman, and A. A. Al Islam, "Man-in-the-middle attack on contactless payment over NFC communications: design, implementation, experiments and detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 3012–3023, 2020.

[111] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: The road ahead," *Computer networks*, vol. 76, pp. 146–164, 2015.

[112] C. Cremers, K. B. Rasmussen, B. Schmidt, and S. Capkun, "Distance hijacking attacks on distance bounding protocols," in *2012 IEEE Symposium on Security and Privacy*, pp. 113–127, IEEE, 2012.

[113] S. Drimer, S. J. Murdoch, *et al.*, "Keep your enemies close: Distance bounding against smartcard relay attacks.," in *USENIX security symposium*, vol. 312, 2007.

[114] T. Chothia, J. De Ruiter, and B. Smyth, "Modelling and analysis of a hierarchy of distance bounding attacks," in *27th USENIX Security Symposium (USENIX Security 18)*, pp. 1563–1580, 2018.

[115] A. Greenberg, "This hacker's tiny device unlocks cars and opens garages," *Wired*, 2015.

[116] B. C. Neuman and S. G. Stubblebine, "A note on the use of timestamps as nonces," *ACM SIGOPS Operating Systems Review*, vol. 27, no. 2, pp. 10–14, 1993.

[117] S. Mendoza, "Samsung pay: Tokenized numbers, flaws and issues," *Proc. Black Hat USA*, pp. 1–11, 2016.

[118] D. Giese, K. Liu, M. Sun, T. Syed, and L. Zhang, "Security analysis of near-field communication (nfc) payments," *arXiv preprint arXiv:1904.10623*, 2019.

[119] J. Han, C. Qian, P. Yang, D. Ma, Z. Jiang, W. Xi, and J. Zhao, "Geneprint: Generic and accurate physical-layer identification for uhf RFID tags," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 846–858, 2015.

[120] G. Li, J. Yu, Y. Xing, and A. Hu, "Location-invariant physical layer identification approach for wifi devices," *IEEE Access*, vol. 7, pp. 106974–106986, 2019.

[121] E. Uzundurukan, Y. Dalveren, and A. Kara, "A database for the radio frequency fingerprinting of bluetooth devices," *Data*, vol. 5, no. 2, p. 55, 2020.

[122] W. Hou, X. Wang, J.-Y. Chouinard, and A. Refaey, "Physical layer authentication for mobile systems with time-varying carrier frequency offsets," *IEEE Transactions on Communications*, vol. 62, no. 5, pp. 1658–1667, 2014.

[123] M. M. U. Rahman, A. Yasmeen, and J. Gross, "Physical layer authentication via drifting oscillators," in *2014 IEEEGlobal Communications Conference*, pp. 716–721, IEEE, 2014.

[124] M. Nair, T. Cappello, S. Dang, V. Kalokidou, and M. A. Beach, "RF fingerprinting of lora transmitters using machine learning with self-organizing maps for cyber intrusion detection," in *2022 IEEE/MTT-S International Microwave Symposium-IMS 2022*, pp. 491–494, IEEE, 2022.

[125] L. Cui, Z. Zhang, N. Gao, Z. Meng, and Z. Li, "Radio frequency identification and sensing techniques and their applications—a review of the state-of-the-art," *Sensors*, vol. 19, no. 18, p. 4012, 2019.

[126] B. Danev, H. Luecken, S. Capkun, and K. El Defrawy, "Attacks on physical-layer identification," in *Proceedings of the third ACM conference on Wireless network security*, pp. 89–98, 2010.

[127] A. Ranganathan and S. Capkun, "Are we really close? verifying proximity in wireless systems," *IEEE Security & Privacy*, vol. 15, no. 3, pp. 52–58, 2017.

[128] H. Ólafsdóttir, A. Ranganathan, and S. Capkun, "On the security of carrier phase-based ranging," in *International Conference on Cryptographic Hardware and Embedded Systems*, pp. 490–509, Springer, 2017.

[129] R. Miesen, F. Kirsch, P. Groeschel, and M. Vossiek, "Phase based multi carrier ranging for UHF RFID," in *2012 IEEE International Conference on Wireless Information Technology and Systems (ICWITS)*, pp. 1–4, IEEE, 2012.

[130] X. Lu, Y. Yin, N. Zhao, and H. Wei, "Indoor positioning experiment based on phase ranging with bluetooth low energy (BLE)," in *Journal of Physics: Conference Series*, vol. 1971, p. 012044, IOP Publishing, 2021.

[131] P. De Cauwer, T. Van Overtveldt, J. Doggen, F. Van der Schueren, M. Weyn, and J. Bracke, "Study of RSS-based localisation methods in wireless sensor networks," in *European Conference on the Use of Modern Information and Communication Technologies (ECUMIT), Ghent*, 2010.

[132] I. Boureanu, T. Chothia, A. Debant, and S. Delaune, "Security analysis and implementation of relay-resistant contactless payments," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pp. 879–898, 2020.

[133] D. H. Yum, J. S. Kim, S. J. Hong, and P. J. Lee, “Distance bounding protocol for mutual authentication,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 592–601, 2010.

[134] G. P. Hancke and M. G. Kuhn, “An RFID distance bounding protocol,” in *First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM’05)*, pp. 67–73, IEEE, 2005.

[135] Y.-J. Tu and S. Piramuthu, “RFID distance bounding protocols,” in *First International EURASIP Workshop on RFID Technology*, pp. 67–68, Citeseer, 2007.

[136] C. H. Kim, G. Avoine, F. Koeune, F.-X. Standaert, and O. Pereira, “The swiss-knife RFID distance bounding protocol,” in *International Conference on Information Security and Cryptology*, pp. 98–115, Springer, 2008.

[137] W. Choi, M. Seo, and D. H. Lee, “Sound-proximity: 2-factor authentication against relay attack on passive keyless entry and start system,” *Journal of Advanced Transportation*, vol. 2018, 2018.

[138] P. Urien and S. Piramuthu, “Elliptic curve-based RFID/NFC authentication with temperature sensor input for relay attacks,” *Decision Support Systems*, vol. 59, pp. 28–36, 2014.

[139] H. T. T. Truong, X. Gao, B. Shrestha, N. Saxena, N. Asokan, and P. Nurmi, “Comparing and fusing different sensor modalities for relay attack resistance in zero-interaction authentication,” in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 163–171, IEEE, 2014.

[140] B. Shrestha, N. Saxena, H. T. T. Truong, and N. Asokan, “Drone to the rescue: Relay-resilient authentication using ambient multi-sensing,” in *International Conference on Financial Cryptography and Data Security*, pp. 349–364, Springer, 2014.

[141] G. Haken, K. Markantonakis, I. Gurulian, C. Shepherd, and R. N. Akram, “Evaluation of apple idevice sensors as a potential relay attack countermeasure for apple pay,” in *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*, pp. 21–32, 2017.

[142] Y.-J. Tu and S. Piramuthu, “On addressing RFID/NFC-based relay attacks: An overview,” *Decision Support Systems*, vol. 129, p. 113194, 2020.

[143] C. Paschou, O. Johnson, and A. Doufexi, “Re-defining secure distance for csi-based key generation protocols,” in *2022 IEEE Globecom*, pp. 1–6, IEEE, 2022.

[144] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*, vol. 2. prentice hall PTR New Jersey, 1996.

[145] W. Jakes, “New techniques for mobile radio,” *Bell Laboratory Rec.*, pp. 326–330, 1970.

[146] R. Steele and L. Hanzo, *Mobile radio communications: Second and third generation cellular and WATM systems: 2nd.* IEEE Press-John Wiley, 1999.

[147] J. Salz and J. H. Winters, “Effect of fading correlation on adaptive arrays in digital wireless communications,” in *Proceedings of ICC’93-IEEE International Conference on Communications*, vol. 3, pp. 1768–1774, IEEE, 1993.

[148] M. Edman, A. Kiayias, and B. Yener, “On passive inference attacks against physical-layer key extraction?,” in *Proceedings of the Fourth European Workshop on System Security*, pp. 1–6, 2011.

[149] X. He, H. Dai, W. Shen, and P. Ning, “Is link signature dependable for wireless security?,” in *2013 Proceedings IEEE INFOCOM*, pp. 200–204, IEEE, 2013.

[150] J. Wallace, “Secure physical layer key generation schemes: Performance and information theoretic limits,” in *2009 IEEE International Conference on Communications*, pp. 1–5, IEEE, 2009.

[151] F. Rottenberg, T.-H. Nguyen, J.-M. Dricot, F. Horlin, and J. Louveaux, “Csi-based versus RSS -based secret-key generation under correlated eavesdropping,” *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1868–1881, 2020.

[152] Z. Mingyue and L. Yuanan, “An overview of spatial channel models used in smart antenna system analysis,” in *2001 International Conferences on Info-Tech and Info-Net. Proceedings (Cat. No. 01EX479)*, vol. 2, pp. 542–548, IEEE, 2001.

[153] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, “Overview of spatial channel models for antenna array communication systems,” *IEEE personal communications*, vol. 5, no. 1, pp. 10–22, 1998.

[154] L. Landau, “Bessel functions: monotonicity and bounds,” *Journal of the London Mathematical Society*, vol. 61, no. 1, pp. 197–215, 2000.

[155] C. Paschou, O. Johnson, A. Doufexi, Z. Zhu, and W. H. Chin, "Increasing the secrecy gap in quasi-static rayleigh channels with secret splitting," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–7, IEEE, 2020.

[156] Y.-H. Nam, L. Liu, Y. Wang, C. Zhang, J. Cho, and J.-K. Han, "Cooperative communication technologies for LTE-advanced," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5610–5613, IEEE, 2010.

[157] H. Pang, J. Liu, X. Fan, and L. Sun, "Toward smart and cooperative edge caching for 5g networks: A deep learning based approach," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, pp. 1–6, IEEE, 2018.

[158] S. Rezvani, N. M. Yamchi, M. R. Javan, and E. A. Jorswieck, "Resource allocation in virtualized comp-noma hetnets: Multi-connectivity for joint transmission," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4172–4185, 2021.

[159] K.-C. Chang, K.-C. Chu, H.-C. Wang, Y.-C. Lin, and J.-S. Pan, "Energy saving technology of 5G base station based on internet of things collaborative control," *IEEE Access*, vol. 8, pp. 32935–32946, 2020.

[160] P. Georgakopoulos, T. Akhtar, C. Mavrokefalidis, I. Politis, K. Berberidis, and S. Koulouridis, "Coalition formation games for improved cell-edge user service in downlink noma and MU-MIMO small cell systems," *IEEE Access*, vol. 9, pp. 118484–118501, 2021.

[161] H. Zhang, L. Song, Z. Han, and H. V. Poor, "Cooperation techniques for a cellular internet of unmanned aerial vehicles," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 167–173, 2019.

[162] A. Fotouhi, M. Ding, and M. Hassan, "Flying drone base stations for macro hotspots," *IEEE Access*, vol. 6, pp. 19530–19539, 2018.

[163] R. Liu, M. Li, Y. Liu, Q. Wu, and Q. Liu, "Joint transmit waveform and passive beamforming design for ris-aided dfrc systems," *IEEE Journal of Selected Topics in Signal Processing*, 2022.

[164] A. S. Abdalla, T. F. Rahman, and V. Marojevic, "Uavs with reconfigurable intelligent surfaces: Applications, challenges, and opportunities," *arXiv preprint arXiv:2012.04775*, 2020.

[165] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, “Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT,” *IEEEJournal on Selected Areas in Communications*, vol. 36, no. 4, pp. 918–931, 2018.

[166] Ç. Çapar, D. Goeckel, B. Liu, and D. Towsley, “Secret communication in large wireless networks without eavesdropper location information,” in *2012 Proceedings IEEE INFOCOM*, pp. 1152–1160, IEEE, 2012.

[167] Ç. Çapar and D. Goeckel, “Network coding for facilitating secrecy in large wireless networks,” in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, IEEE, 2012.

[168] T.-Y. Liu, P.-H. Lin, S.-C. Lin, Y.-W. P. Hong, and E. A. Jorswieck, “To avoid or not to avoid CSI leakage in physical layer secret communication systems,” *IEEE Communications Magazine*, vol. 53, no. 12, pp. 19–25, 2015.

[169] P.-C. Lan, Y.-W. P. Hong, and C.-C. J. Kuo, “Enhancing secrecy in fading wiretap channels with only transmitter-side channel state information,” in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1314–1319, IEEE, 2014.

[170] T.-Y. Liu, S.-C. Lin, and Y.-W. P. Hong, “On the role of artificial noise in training and data transmission for secret communications,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 516–531, 2016.

[171] S. Gerbracht, C. Scheunert, and E. A. Jorswieck, “Secrecy outage in MISO systems with partial channel information,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 704–716, 2011.

[172] G. Yeo and R. Milne, “On characterizations of beta and gamma distributions,” *Statistics & probability letters*, vol. 11, no. 3, pp. 239–242, 1991.

[173] E. W. Weisstein, “Incomplete gamma function,” *MathWorld, A Wolfram Web Resource.,[Online]. Available: http://mathworld. wolfram. com/Incomplete Gamma-Function. html*, 2006.

[174] T. S. Rappaport, “Wireless communications–principles and practice,” *Microwave Journal*, vol. 45, no. 12, pp. 128–129, 2002.

[175] C. Paschou, O. Johnson, Z. Zhu, and A. Doufexi, “A lightweight protocol for validating proximity in UHF RFID systems,” in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–7, IEEE, 2021.

[176] C. Paschou, O. Johnson, A. Doufexi, Z. Zhu, and W. H. Chin, “Physical layer protection against relay/replay attacks in short-range systems,” in *2022 IEEE Wireless Communications and Networking Conference (WCNC*, pp. 1–7, IEEE, 2022.

[177] C. Paschou, “Preventing replay/relay attacks in keyless entry systems,” May 5 2022. Patent 543229US.

[178] N. Sklavos and O. Koufopavlou, “On the hardware implementations of the sha-2 (256, 384, 512) hash functions,” in *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS’03.*, vol. 5, pp. V–V, IEEE, 2003.

[179] M. Burmester and B. De Medeiros, “The security of EPC Gen2 compliant IEEE protocols,” in *International Conference on Applied Cryptography and Network Security*, pp. 490–506, Springer, 2008.

[180] C. Paschou, R. Francesco, G. Michael, D. McEwan, P. James, and O. George, “Cricket: A practical physical layer key agreement protocol for iot networks,” in *2023 IEEE international conference on communications (ICC)*, pp. 1–6, IEEE, 2023.

[181] S. T. Ali, V. Sivaraman, and D. Ostry, “Secret key generation rate vs. reconciliation cost using wireless channel characteristics in body area networks,” in *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, pp. 644–650, IEEE, 2010.

[182] K.-F. Krentz, C. Meinel, and H. Graupner, “Secure self-seeding with power-up sram states,” in *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1251–1256, IEEE, 2017.

[183] D. MCEwan, *Behaviour Analysis in Binary SoC Data.* PhD thesis, University of Bristol, 2022.

[184] Nordic Semiconductor, *nRF52840*, 11 2021. Rev. 7.

[185] C. Paschou, R. Francesco, and G. Michael, “CRicKET.” Available at: https://github.com/fraimondo18/cricket, 11 2022.

INDEX