Physical layer secret key generation for decentralized wireless networks

Iulia Tunaru

2015

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du signal et télécommunications

Ecole doctorale Matisse

présentée par

Iulia Tunaru

préparée à l'unité de recherche 6164 IETR et CEA-Leti
Institut d'Électronique et de Télécommunications de Rennes
Composante universitaire 917

Physical layer secret key generation for decentralized wireless networks

Thèse soutenue à Grenoble le 27.11.2015

devant le jury composé de :

Alain SIBILLE
Professeur, Télécom Paris Tech/ rapporteur

Matthieu BLOCH
Associate Professor, Georgia Tech USA/ rapporteur

Maria-Gabriella Di BENEDETTO
Professeur, Université de Rome La Sapienza/ examinateur

Claude CASTELLUCCIA
Directeur de recherche, INRIA Grenoble/ examinateur

Davide DARDARI
Professeur, Université de Bologne/ examinateur

Moe Z. WIN
Professeur, MIT/ examinateur

Benoît DENIS
Chercheur, CEA-Leti/ co-encadrant

Bernard UGUEN
Professeur, Université de Rennes 1/ directeur de thèse

UNIVERSITY OF RENNES 1

Abstract

Ecole doctorale Matisse

Doctor of Philosophy

Physical layer secret key generation for decentralized wireless networks

by Iulia Tunaru

Emerging decentralized wireless systems, such as sensor or ad-hoc networks, will demand an adequate level of security in order to protect the private and often sensitive information that they carry. The main security mechanism for confidentiality in such networks is symmetric cryptography, which requires the sharing of a symmetric key between the two legitimate parties. According to the principles of physical layer security, wireless devices within the communication range can exploit the wireless channel in order to protect their communications. Because of the theoretical reciprocity of wireless channels, the spatial decorrelation property (e.g., in rich scattering environments), as well as the fine temporal resolution of the Impulse Radio - Ultra Wideband (IR-UWB) technology, directly sampled received signals or estimated channel impulse responses (CIRs) can be used for symmetric secret key extraction under the information-theoretic source model. Firstly, we are interested in the impact of quantization and channel estimation algorithms on the reciprocity and on the random aspect of the generated keys. Secondly, we investigate alternative ways of limiting public exchanges needed for the reconciliation phase. Finally, we develop a new signal-based method that extends the point-to-point source model to cooperative contexts with several nodes intending to establish a group key.

Părinților mei

Contents

Acknowledgements

When preparing a PhD, three years is a short period: after the first year you have the impression that you finally understand your subject, the second year you are worried about the results, and the third year you realize that you were a bit too optimistic at the end of the first year. On the contrary, when writing the acknowledgement section, you realize that three years of your life have passed and this seems an awfully long time. After the first moments of panic caused by the freshly renewed acknowledgement of the unforgiving passage of time, not only are you genuinely happy about the new inspiring people that you met but also about the older still-standing bonds that reinforce your internal structure. Besides an edifying scientific experience, PhD is also a major human experience. I am therefore happy to be able to thank all the persons that have supported and encouraged me in different ways and circumstances over the last years.

My first thought goes to Benoît Denis, my supervisor from CEA-Leti, who has been actively and passionately involved in the research studies presented in this dissertation. From the first contact we had by e-mail in April 2012, I am sincerely grateful for his striving commitment to everything he said and wrote. From this point of view, I can only hope that I have managed to reply in similar terms. Scientifically, I have felt extremely fortunate to be supervised by Benoît Denis because of his skills, openness, and pedagogical nature, which allowed me to have both the freedom I was searching for and the guidance that I needed. Moreover, his optimism and especially, his patience towards my rather pessimistic nature (for which I apologize) have encouraged me and have largely contributed to keeping a serene atmosphere during the rough times that inevitably occurred. In the end, Benoît was not only my supervisor but also somebody who understood that a PhD is a human experience and offered me the subtle ingredient that everyone needs to enjoy their PhD: trust in a healthy, enriching, and productive student-supervisor relationship.

Then, I would like to express my sincere appreciation for all the efforts that Prof. Bernard Uguen has made in order to follow this work from a distance of 652 km. The regular discussions we had with him and the two work visits that I made in Rennes have been moments of great inspiration for my work and boosting remedies for my moral thanks to his enthusiastic nature. Finally, I would also like to warmly thank Prof. Bernard Uguen for both the financial support during these years and the highly appreciated help concerning the defense formalities.

Besides my two supervisors, I would also like to mention a few other persons that have contributed to my initiation to research or with whom I have had valuable discussions regarding my PhD subject. Fabrice Valois and Hervé Rivano have been my research project supervisors while I was in my final year of school. I am particularly thankful to them for having taught me how to "think" and "speak" in research terms and having introduced me to Latex. I will always value the help I received from Paul Ferrand, who has encouraged me to keep looking for interesting things to study and has patiently explained to me highly useful aspects of information theory. Similarly, I am happy to be able to mention a couple of pleasant discussions with Sharvil Patil on quantization and companding, which have later led to one of the first results of this PhD. Last but not least, I wish to thank Régis Perrier for sharing his passion for Bayesian methods most often during the coffee breaks, having insisted that I write the EM equations, and having verified them.

I would also like to sincerely thank my defense committee for accepting to review this work as well as for their remarks, especially those of Dr. Bloch, who was kind enough to send me his detailed comments.

Furthermore, I thank CEA and the line management for giving me the opportunity to do this PhD as well as to participate in diverse conferences and trainings. I am particularly thankful to Sandrine Bertola, our laboratory secretary, for her availability, understanding, and genuine care when dealing with my administrative questions or demands.

The time spent in CEA would have been at least weary and grim without people like my office colleagues, Marco and Nicolas, who taught me a set of tricks from French expressions to subtle interior decoration, "I'm-awesome-and-I-know-it" Mickael, the spiritual guide of all PhD students, and my personal social life coach Matthias. I also sincerely thank all my LESC colleagues for their warm welcome, help, and well-intended advice regarding my work.

Outside the badge-only area, I had the opportunity to make new friends and enjoy an indecent amount of laughs in their company. Despite any kind of distance, my friends from school have also offered me invaluable support during my whole time in Grenoble. For this, I can only feel lucky that I know them. Special thanks to my brother Marius and my friend Safeer for being there and for offering me homemade food supplies.

And finally, I would like to thank the persons that have shown unconditional trust in my abilities since two decades ago: my mother, who went from tricking me to supporting me and finally watching me do my studies and my father, who is the main "responsible" for my coming to France.

List of Figures

List of Tables

Abbreviations

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BCS	Bayesian Compressed Sensing
BP	Basic Pursuit
CIR	Channel Impulse Response
CM	Channel Model
CRLB	Cramer-Rao Lower Bound
CS	Compressed Sensing (Compressive Sampling)
CSI	Channel State Information
CV	Cross Validation
IR-UWB	Impulse Radio - Ultra Wideband
D2D	Device-to-Device
DBPSK	Differential Binary Phase Shift Keying
DMC	Discrete Memoryless Channel
DMS	Discrete Memoryless Source
DLP	Discrete Logarithm Problem
ED	Energy Detection
EM	Expectation Maximization
FFT	Fast Fourier Transform
FRI	Finite Rate of Innovation
IoT	Internet of Things
LDPC	Low Density Parity-check Code
LOS	Line of Sight
LS	Least Squares
MANET	Mobile Ad-hoc Networks
MAP	Maximum a Posteriori
MF	Matched Filter
MIMO	Multiple Input Multiple Output
MMSE	Minimum-Mean Square Error
ML	Maximum Likelihood

MP	Matching Pursuit
MPC	Multipath Component
NB	Narrow-band
NIST	National Institute of Standards and Technology
NLOS	Non-Line of Sight
OFDM	Orthogonal Frequency-Division Multiplexing
OMP	Orthogonal Matching Pursuit
OOK	On-Off Keying
OSI	Open System Interconnection (model)
PDP	Power Delay Profile
PKC	Public Key Cryptography
PKI	Public Key Infrastructure
PN	Pseudo noise (sequence)
PPM	Pulse Position Modulation
QAM	Quadrature Amplitude Modulation
QKD	Quantum Key Distribution
RF	Radio Frequency
RMSE	Root-Mean Square Error
RSS	Received Signal Strength
RT-ToF	Round Trip-Time of Flight
RV	Random variable
Rx	Receiver
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
ToA	Time of Arrival
TR	Transmitted Reference
Tx	Transmitter
VANET	Vehicular Ad-hoc Networks
WBAN	Wireless Body Area Networks
WPAN	Wireless Personal Area Networks
WSN	Wireless Sensor Networks
WTC	Wiretap Channel

Symbols

c, n, N, λ	scalar values
$x(t)$	a continuous signal
$x[.]$, $X[.]$	arrays of scalar values
$\mathcal{I}$	a set
$\lvert\mathcal{I}\rvert$	the cardinal of a set
$std(.)$	the empirical standard deviation of a set of values
X	a random variable
$\mathbb{E}_X[.]$	the expectation over the random variable X
$\mathbb{P}$	a probability measure
$\mathbb{H}$	entropy function
$\mathbb{I}$	mutual information function
$\mathbf{x}$	a vector (in a linear algebra sense)
ψ_n	a column vector
$\mathbf{x}^T$	a transposed vector
$\lvert\lvert\mathbf{x}\rvert\rvert$	the Euclidean norm of vector $\mathbf{x}$
$\mathbf{A}$, Ψ	matrices
$\lvert\mathbf{A}\rvert$	the determinant of a matrix
$\mathrm{Tr}(\mathbf{A})$	the trace of a matrix
A, B, C	denominations of legitimate users
E	denomination of an attacker

Chapter 1

Introduction

From over-the-air to paper and wires and then back to over-the-air. Telecommunications, or the sharing of information between two spatially distant entities, includes any form of communication from smoke signals (17th century BC) to paper messages (2nd century BC), telegraphs, land-line telephones (19th century), and wireless communications, which have known a tremendous development in the last century. Although the term "wireless" strictly refers to the communication medium, wireless communications are popularly assimilated to the communications using the radio spectrum (i.e., electromagnetic waves between 9 kHz and 300 GHz), which is just one of the possible resources for transmitting information wirelessly, together with, e.g., electromagnetic optical communications or sound waves.

Between centralized and decentralized. Once the transmission technologies were in place, networks started to appear and two main architectures evolved in parallel: the centralized paradigm, implemented for example in cellular networks,[1] and the decentralized one represented by the Internet[2] and more recently by mobile ad hoc networks (MANET). Nowadays, the two trends are becoming interleaved in the context of new concepts such as the Internet of Things, ad-hoc device-to-device (D2D) communications, Cloud services linked to decentralized wireless sensor networks (WSN), vehicular ad-hoc communications (VANET) or wireless personal area networks (WPAN), including body area networks (WBAN). However, the decentralized characteristic remains representative at least under short-range connectivity in these emerging networks.

The What and The How. Telecommunications usages have also evolved across decades from military, professional or convenience-oriented to entertainment-focused (i.e., a data

[1] Cellular networks are classified as centralized with respect to the core network, which is responsible for routing, authentication etc.

[2] Although the Internet flows nowadays can be arguably considered more centralized than in its beginnings, we are herein referring to the network structure itself.

consumption model) and soon probably to a user-centric (i.e., a data production model), as suggested by the aforementioned types of networks. This implies a diversification in communication requirements such as rate, latency, reliability, and security. Whereas achieving the former three can be a layer-dependent task, ideal security solutions should holistically consider the technology, the transmission medium, the network structure and the data usage.

Emerging decentralized wireless networks

In modern device-centric applications, a considerable amount of data can be locally produced, exchanged, and collected for purposes such as energy monitoring and optimization, logistics, navigation etc. Context-aware services in Smart Cities, WSN, e-Health applications, contactless payment transactions, nomadic social networking, Intelligent Transportation Systems (ITS) are some of the examples where emerging wireless networks will have to carry sensitive information between remote users or from remote sensors to a core-network.

Most wireless networks capable of supporting such applications might require peer-to-peer interactions between end-devices or equipped users under opportunistic connectivity conditions. This type of scenarios are foreseen by WiFi Direct and D2D options in pending 5G standards, IEEE 802.11p-compliant VANET, and short range technologies (e.g., Near Field Communications (NFC), Bluetooth-Low Energy (BT-LE), IEEE 802.15.4 or Zigbee, IEEE 802.15.4a or IEEE 802.15.6 Impulse Radio - Ultra Wideband (IR-UWB)).

These emerging decentralized and/or ad-hoc networks might be subject to hardly predictable mobility patterns, erratic users' activity, and varying devices densities, hence requiring fine flexibility, reactivity and scalability. Consequently, light decentralized security and privacy solutions should accompany the establishment of this kind of networks from their early stages. Often, security is an overlay in terms of system design, sometimes considered even unnecessary unless experience proves the contrary. It is therefore essential to integrate security solutions as much as possible and by design within the communication devices. In this sense, outsourcing security to include lower layers as well (e.g., physical or MAC) could be a promising solution.

In this thesis, we study the possibility of generating security material, namely secret keys for symmetric cryptography, using the radio physical layer in the context of wireless decentralized networks. Our studies focus on the emerging Impulse Radio - Ultra Wideband (IR-UWB), a technology typically employed for low data rate transmissions or for high-precision ranging purposes in low-complexity sensor networks and potentially also in more powerful devices with localization requirements (e.g., multi-standard smartphones). The choice of this radio localization technology is justified by the unique properties of the corresponding channel impulse responses (CIRs), as illustrated in Section 1.3. Moreover, typical decentralized networks will require capabilities for providing context-aware services, which will probably encourage the proliferation of IR-UWB in such networks.

This first introductory chapter begins with a progressive state of the art on security issues. Section 1.1 describes general security schemes, including a case study on WSN, and introduces the notion of physical layer security. Section 1.2 delves into information-theoretic models for key agreement and physical layer key generation models and methods. Finally, we provide an overview of IR-UWB in Section 1.3 and present the structure as well as the main contributions of the thesis in Section 1.4.

1.1 Communications security

Historically, security systems can be classified in three categories [1]: concealment systems (e.g., steganography), privacy systems that require a physical device to recover the message, and secrecy systems that hide the information using codes or ciphers. Nowadays, information security has extended to a set of measures to counteract certain types of attacks (e.g., eavesdropping, message modification, replays, denial of service, intrusion etc.). This led to the development of a set of security goals or principles: confidentiality, integrity, availability (also known as the CIA triad), extensions of data integrity (origin authentication, non-repudiation, freshness [2]), and even accountability and assurance [3].

Security solutions follow the aforementioned general principles but they are usually defined with respect to the application constraints and to the communication protocols. In current communication systems following the layered OSI model, security is implemented separately at each layer as a supplementary cryptographic feature (e.g., IPSec or SSL protocols). In TCP/IP-alternative architectures, such as the Recursive Internetwork Architecture (RINA) [4], security is considered as part of the network design and cryptographic tools are only needed for reinforcement.

Overall, the majority of these security mechanisms involve the use of cryptographic primitives such as symmetric/asymmetric cryptography, hash functions, digital signatures. In the following, a short historical review of encryption is presented, followed by an illustration of classical security schemes for decentralized networks, and finishing with physical layer security and its two main approaches: secrecy capacity and secret key generation, the focus of this thesis.

1.1.1 A historical perspective

Since its early beginnings, the evolution of encryption has been an iterative process alternating between designing more powerful ciphers and breaking them. Starting with the ancient transposition and substitution ciphers, namely the "scytale" cipher (6th century B.C.) and Caesar's cipher (1st century B.C.), respectively, continuing in the Middle Ages with more sophisticated examples like the poly-alphabetic Vigenère cipher, and until the end of World War II, the encryption of communications remained a matter of political and military interest [5]. The only paradigm shift over this period happened at the beginning of the 20th century with the arrival of encryption and cryptanalysis machines. This also led to the invention and popularization of computers, which found their way in the private sector and created the need for publicly available encrypting methods.

If until the 1970's the encryption security relied on the secrecy of both the encryption algorithm and the key (i.e., symmetric key), the approval of the Data Encryption Standard (DES) in 1976 by the future National Institute of Standards and Technology (NIST) in the US, made the encryption algorithm public. DES is a block cipher using a 56-bit key and a mix of XOR operations, duplications, substitutions and permutations [6]. In consequence, it respects Shannon's practical security principle of "confusion" (i.e., a ciphered character depends on several parts of the key) and "diffusion" (i.e., a character change in the plaintext incurs several changes in the ciphered text). Since DES was broken for the first time in 1997, it has been replaced by Triple-DES (1999) and the Advanced Encryption Standard (AES, 2001), a substitution-permutation block cipher with a key of maximum 256 bits [6]. In parallel, faster stream ciphers based on linear/nonlinear shifting have also been developed (e.g., RC4, 1987) but they are currently considered less secure than block ciphers [6].

Since the standardization of DES, the security of symmetric cryptography relies on the secrecy of the encryption key and therefore, on its characteristics: length, randomness, distribution method. The concerns regarding symmetric key distribution catalyzed the adoption of public key cryptosystems after the invention of the key exchange protocol by Diffie, Hellman, and Merkle [7]. This led to the invention of Rivest-Shamir-Adleman (RSA) encryption algorithm (i.e., asymmetric cryptography) [6]. Contrary to the symmetric secret key approach, RSA relies on individual private and public keys with lengths between 1024 and 2048 bits and on the prime factorization of large numbers, which has not been yet proven to be efficiently solvable (i.e., they are considered computationally hard for the moment). More recent public key cryptosystems, such as Elliptic Curves [6], have replaced the factorization problem with the Discrete Logarithm Problem (DLP). Asymmetric cryptography shifts the challenge from the secret key distribution side to the certification of public keys, which is conceptually an easier task because of the inherent signature mechanism of public key cryptography (PKC). However, PKC is deemed to be more complex than symmetric cryptography regarding encryption and decryption operations, and it is thus employed mainly for symmetric key distribution and not for data encryption itself. The most prevalent example of synergy between symmetric and asymmetric cryptography is the Transport Layer Security protocol (TLS, 1999) initialized at the session layer, which allows an authenticated symmetric key exchange between a client and a server and subsequent symmetric encryption of the communications.

The symmetric key distribution problem, one of the main challenges of security implementations, has been studied from various angles and within various fields: cryptography, networking (e.g., key distribution protocols), quantum communications, and information theory. In the next section, several classical key distribution solutions will be presented and discussed through the case study of WSN.

1.1.2 Security in decentralized networks

The rapid development of wireless technologies has led to the emergence of several types of wireless ad-hoc networks such as mobile ad-hoc networks and wireless sensor networks [8]. While MANET are usually composed of easily replaceable mobile nodes, WSN are large networks consisting of resource-constrained nodes. The latter have sensing, processing and communication capabilities and are employed mainly in health, military, environmental or domestic applications [9]. The sensor nodes can be randomly deployed in an interest area in order to sense, eventually process, and route data through a multihop architecture to a gateway (the sink) connected to the Internet.

In the following, for illustration and discussion purposes, we will sometimes take WSN as an example of decentralized networks. WSN are subject to typical design constraints that ultimately impact the entire sensor network protocol stack. First of all, sensor networks should be fault tolerant and scalable because of the densely deployed low-cost and hardware-constrained nodes. Secondly, the network topology should be maintained in the absence of any infrastructure and in the event of node failure or node redeployment. Finally, sensor lifetime maximization can be critical in some WSN. The survey in [9] offers a summary of the main physical, data link and network protocols specific to sensor networks.

Security measures in WSN

Security measures in the WSN communication stack can be implemented at several layers: physical/link layers (e.g., jamming detection), link/routing layers (e.g., securing data dissemination by link authentication, route construction authentication etc.), routing/transport/application layers (e.g., securing data aggregation or the data itself).

Asymmetric or public key cryptography has the advantages of scalability of secure communications between any two nodes of the network and digital signature support implying an intrinsic authentication capability. However, it is deemed to be a solution of high computational complexity, which also relies on the cumbersome authentication of public certificates by third-party certification authorities.

As a partial solution, lower complexity variants of PKC have been implemented on low-power sensor nodes for authentication and key distribution [10]. In order to solve the issue of authentication of public keys, ID or location based PKC relying on pairing functions has been proposed as an alternative to classical PKC. In this new PKC framework, public and private keys are either derived from identification information using a secret sharing scheme in a MANET context [11] or directly from location information in WSN [12]. Nevertheless, location-based PKC requires the secure distribution of location-based private keys in an early post-deployment phase. Secret sharing and location-based private keys can also be exploited to provide end-to-end data confidentiality, authenticity and availability in sensor networks that can benefit from a network bootstrapping phase [13].

The main solution for encrypted communications in sensor networks remains symmetric cryptography because of lower computational overhead of the encryption and decryption operations [8]. Conversely, it has no support for digital signatures and it is not very scalable because of the challenge of distributing a shared secret key between each pair of nodes.

Symmetric key management protocols

General solutions Numerous key management techniques for general networked systems have been already developed [2] and can be classified according to: i) the number of users/servers participating in the key generation phase (two-party or conference protocols); ii) the method of key establishment (key transport or key agreement protocols).

Key transport protocols imply that key generation takes place at one user or server and is then distributed among the concerned parties using symmetric or asymmetric cryptography. Despite their diversity, all the solutions based on symmetric encryption imply the existence of an initial secure channel such as a global key between all nodes or an initial key between each node and an online server (Key Distribution Center) that distributes pairwise session keys. The global or the initial keys are an important security breach with respect to an internal attacker that has compromised a node. The online server also represents a single point of failure in the network and it is not compatible with the decentralized architecture. Symmetric keys can also be transported using PKC with the advantages and disadvantages discussed before.

Key agreement protocols are key establishment protocols in which the key is a function of inputs from all users. Most of the existing schemes are based on the Diffie-Hellman-Merkle protocol, the precursor of PKC, which relies on the computational difficulty of reversing an exponential operation, also known as the Discrete Logarithm Problem. The steps of the protocol are presented in Alg. 1, where s_A and s_B are the initial private secrete keys of A and B, respectively, q is a large prime number, g is a primitive root of q,[3] and s is the shared secret key. The numbers q and g are considered public and can be therefore agreed upon by the use of a public channel. Although it is inherently adapted to decentralized networks, in order to be secure, this protocol requires the generation of sufficiently long keys and it is therefore computationally expensive to implement in networks like WSN.

```
Data: $g$, $q$, $s_A$, $s_B$
Result: $s$
A computes $p_A = g^{s_A} \bmod q$ ;
A sends $p_A$ to B;
B computes $p_B = g^{s_B} \bmod q$ ;
B sends $p_B$ to A;
B computes $s = p_A^{s_B} \bmod q$ ;
A computes $s = p_B^{s_A} \bmod q$ ;
```

Algorithm 1: Diffie-Hellman-Merkle key exchange protocol between A and B

WSN-oriented solutions A recent survey [14] makes a classification of the main methods to distributively share symmetric secret keys in WSN, recovered under the name of "key pre-distribution" or the generation of online pairwise keys using preloaded key materials. This approach has two main components: key material distribution and key agreement.

[3] A number g is the primitive root of q if for every a relatively prime with q there exists k s.t. $g^k = a \bmod q$.

Key agreement focuses on how to generate scalable shared keys based on preloaded materials. In Blom's scheme, a central authority computes a public matrix $\mathbf{P}$ and a secret matrix $\mathbf{S}$ over a finite field GF(q), where q is a large prime number. It then generates a secret matrix $\mathbf{A}$ based on $\mathbf{P}$ and $\mathbf{S}$ and loads each node indexed by i with the i-th columns from $\mathbf{P}$ and $\mathbf{A}$. When two nodes want to generate a shared key they exchange their public columns and apply a multiplication operation. A variant of Blom's scheme consists in replacing the matrix operations with bi-variate polynomial operations on the IDs of the nodes, which are now exchanged between nodes in order to generate the shared key.

Key material distribution can be classified into three categories.

- Random key material distribution is represented mainly by Random Key Predistribution (RKP), a method based on random graph theory. According to RKP, each node is preloaded with a set of keys randomly chosen from an universal pool, so that any two neighboring nodes have a certain probability of sharing the same key. If the nodes do not share a key, they have to negotiate a key through a secure route. Several extensions of RKP have been proposed in order to protect against weaknesses such as node compromise, lack of authentication because of the reuse of the same key by several nodes, or the disadvantages in terms of memory storage.
- Deterministic key material distribution replaces the random graph with a strongly connected regular graph or with a multi-dimensional grid. In the first approach, each node is preloaded with a certain number of keys or key materials (e.g., polynomials) corresponding to its adjacent links. In the second approach, each node is assigned a k-dimensional ID and is preloaded with key materials allowing it to generate shared keys with any of the nodes whose IDs differ from its own in only one dimension.
- Location-based key material distribution aims at improving the performance of aforementioned distribution schemes by taking into account the location. This can be achieved either by a cell-splitting strategy or by a pre-deployment estimation of the proximity of the nodes.

The final "key pre-distribution" solutions have different memory costs and levels of resiliency to node compromise [14], but are less scalable in time, i.e., they do not enable key renewal by design.

1.1.3 Physical layer security

Information-theoretic security

In his paper from 1949 [1], Shannon applies the mathematical concepts developed in his previous work on communication theory (i.e, entropy and conditional entropy or equivocation) to security aspects. Starting from the parallel current work in cryptography (i.e., the study of codes and ciphers), he builds a “theory of secrecy systems” that sees cryptography as the means of creating a secrecy system. Despite the information-theoretic approach, this fundamental work articulates itself around a cryptanalyst’s point of view and enriches it with a stochastic dimension: given the *a priori* knowledge of the key (K) distribution, the message (M) statistics (e.g., the redundancy of the language of the message), and the intercepted encrypted message or cryptogram ($E_K(M)$), what are the *a posteriori* probabilities of each possible message? Also, the article investigates the minimum length of the intercepted sequence for which the solution of the cryptogram becomes unique (i.e., the uncertainty of the cryptanalyst or the equivocation on the message $\mathbb{H}(M|E_K(M))$ vanishes).

Based on these considerations and the newly introduced algebraic formalism, the paper gradually studies the properties of three types of secrecy systems: perfect, ideal and practical. Perfect secrecy (i.e., information-theoretic security) supposes that the cryptanalyst gains no knowledge of the initial message from the cryptogram, or equivalently, that $\mathbb{H}(M|E_K(M)) = \mathbb{H}(M)$. This relation leads to the impractical condition that the encryption key should be at least as long as the message. An ideal system requires that the equivocation on the message and the equivocation on the key remain bounded, ideally by $\mathbb{H}(K)$, when the length of the intercepted message goes to infinity. These systems are deemed complex because they must be designed in close relation to the message space, which should have a uniform *a priori* distribution (i.e., the message should not contain any redundancies before encryption). Finally, practical secrecy is discussed by pointing out that even though the equivocation vanishes for sufficiently large intercepted lengths, the “work” needed to find a unique solution to the cryptogram can be a differentiating factor for ciphers.

The final chapter of [1] examines the fundamental trade-offs between several desirable properties of secrecy systems, most of which are still relevant today and represent fundamental metrics in various research areas: i) amount of secrecy (information-theoretic security); ii) key length (key distribution); iii) complexity and message expansion at ciphering (cryptography); iv) error propagation after deciphering (less pertinent because of the discovery of powerful error-correcting codes).

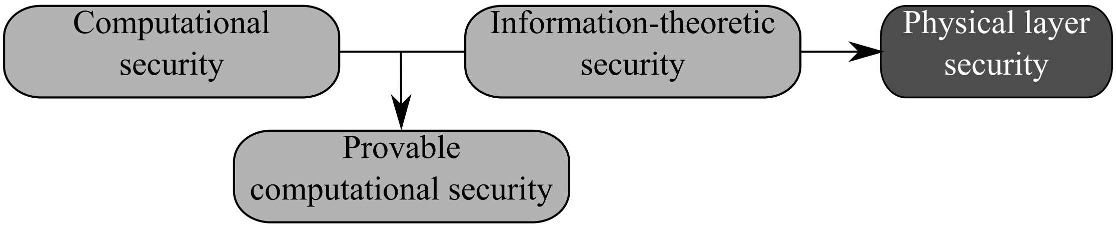

Computational security
Information-theoretic security
Physical layer security
Provable computational security

Figure 1.1: Security approaches

Currently regarded as the "father" of information-theoretic security, Shannon also initiated a general security framework outlining the links between what would later become two different approaches to security: the information-theoretic security (revisited in 1975 by Wyner [15]) and the computational security (i.e., the traditional vision of cryptographic security in today's communication networks).

Information-theoretic security refers to the property of a system that will remain secure, i.e., the leaked information to the adversary is zero (perfect secrecy) or asymptotically zero, given an adversary with unlimited computational power.[4] On the contrary, classical cryptography relies on the principle of limited computational power of an attacker: if no efficient attack algorithm is found for the considered symmetric encryption algorithm or if the involved mathematical operation is considered hard (e.g., prime factorization of large numbers or discrete algorithms over finite groups), the attacker is left with the brute force choice that will take too much time to solve.

A scheme of the various paradigms of security is given in Figure 1.1. An intermediate notion between the two approaches of security is provable computational security. One would say that RSA is provably secure if it were proved that no efficient algorithm for prime factorization exists, but the scheme would still be vulnerable in front of an attacker with unlimited computational power.

Information-theoretic security can be achieved in several contexts:

- Vernam's one-time padding encryption scheme proposed by Shannon (i.e., XORing of the message with a same-length key), which achieves perfect secrecy.
- Shamir's secret sharing scheme with perfect security (i.e., a secret is divided into unique parts that are distributed amongst users and the secret can only be reconstructed with a certain minimum number of parts).
- Private Information Retrieval schemes, where a user interrogates one or several non-cooperative servers for retrieving a piece of information while keeping it private from the servers.

[4] The equivalent term in quantum cryptography is "unconditional security".

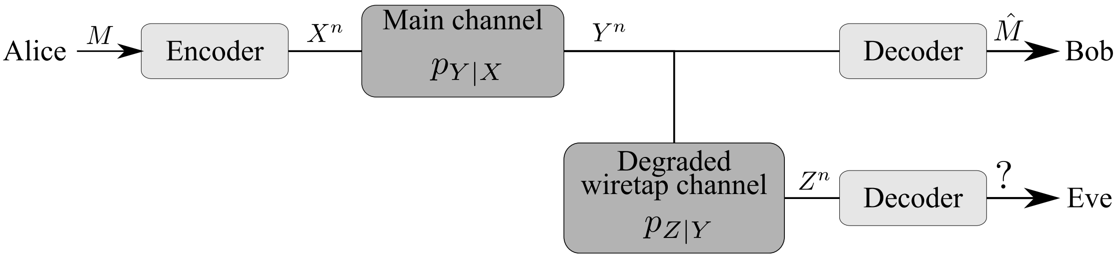


FIGURE 1.2: Degraded wiretap channel

- Secure Multi-Party Computation, in which a group of users wish to compute jointly a function over their inputs while keeping the inputs private.

- quantum key distribution (QKD) [16] based on quantum properties: polarization properties of individual photons or transmitted light pulses cannot be reliably read by a passive eavesdropper without knowing certain parameters of the transmission or without revealing his presence. More details about QKD are provided in Section 1.2.2.

- physical layer security, in which the assumption of an imperfect communication channel offers the opportunity of information-theoretically secure communications with weaker assumptions than perfect secrecy. Physical layer security has two main study components: secure communication and secret key agreement, the focus of the present thesis.

Secure communication over imperfect channels

For the definition of "perfect secrecy", Shannon [1] considers that the attacker can have access to a perfect copy of the encrypted message, meaning that the transmission is done over a noiseless public channel. In physical communications, especially in the wireless case, the transmissions are affected by noise and other random phenomena, which must be compensated by channel encoding in order to achieve *reliability* [17]. Based on this observation, Wyner [15] imagines a new type of channel model called the "wiretap channel model" (WTC, 1975) represented in Figure 1.2.

The perfect secrecy condition would be expressed as: $\mathbb{I}(M; Z^n) = 0$. Instead, Wyner [15] proposes the less stringent and more tractable definition of "weak" (Eq. (1.1)) secrecy based on the behavior of the leaked information $\mathbb{I}(M; Z^n)$. This definition is later extended to "strong" secrecy (Eq. (1.2)).

$$\lim_{n\to+\infty} \frac{1}{n}\mathbb{I}(M; Z^n) = 0 \quad (1.1)$$

$$\lim_{n\to+\infty} \mathbb{I}(M; Z^n) = 0 \quad (1.2)$$

The secrecy capacity is defined as the maximum communication rate at which Alice and Bob can communicate while guaranteeing both *reliability* (Eq. (1.3)) and strong or weak *secrecy* of their communication with respect to Eve, also called a *wiretapper* [18]. In this model, it is assumed that Eve knows the input message space as well as the encoding and the decoding functions that define the WTC code.

$$\lim_{n \to +\infty} \mathbb{P}(M \neq \hat{M}) = 0 \tag{1.3}$$

Wiretap channel models. For the mentioned WTC model, also called the degraded WTC, the secrecy capacity can be computed as shown in Eq. (1.4). In the case of binary symmetric channels, it is strictly equal to the difference between the capacity of the main channel and that of the wiretapper channel. Intuitively, we can see the fundamental trade-off between reliability, which demands more redundancy to improve error-correction performance on the legitimate side ($\mathbb{I}(X;Y)$), and secrecy, which calls for less redundancy in order to limit the eavesdropper's decoding capabilities ($\mathbb{I}(X;Z)$) [18].

$$C_s^{DWTC} = \max_{p_X}(\mathbb{I}(X;Y) - \mathbb{I}(X;Z)) \tag{1.4}$$

In the following years, several other channel models for the study of the secrecy capacity have been proposed. These include the following models:[5]

- the broadcast channel with confidential messages [21], in which the wiretap channel is not a degraded version of the main channel (Figure 1.3). In this case, the question of the quality of the main and wiretap channels arises. For example, when the wiretap channel is noisier than the main channel, the secrecy capacity is the same as for the degraded WTC.
- the Gaussian wiretap channel [22], which is a degraded WTC in which the main and wiretap channels alter the input signal by adding white Gaussian noise. The secrecy capacity is a function of the SNR of the main channel (γ_m) and of the wiretap channel (γ_w) as shown in Eq. (1.5). So, if $\gamma_m \leq \gamma_w$ the secrecy capacity is 0.
$$C_s^{GWTC} = \frac{1}{2}\log(1+\gamma_m) - \frac{1}{2}\log(1+\gamma_w) \tag{1.5}$$
- the fading wiretap channel [23], for which the authors show that the secrecy rate can be positive even when eavesdropper has a higher average SNR than the legitimate receiver, given that the transmitter knows only the instantaneous CSI of the main channel. This is possible because of the different instantaneous realizations

[5] A more detailed review of these channel models and others can be found in [19] and [20].

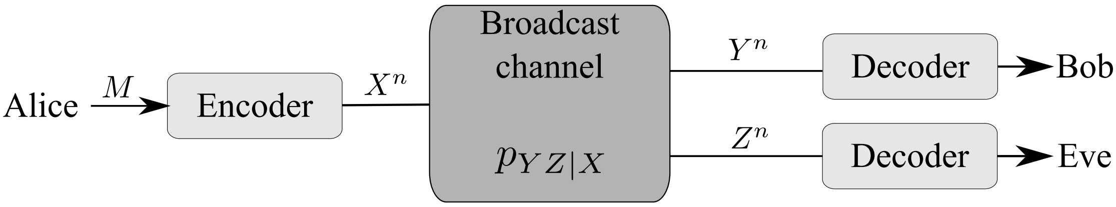


FIGURE 1.3: Broadcast channel with confidential messages

of the fading coefficients on the main and wiretap channels, guaranteeing a non-zero probability to have a higher instantaneous SNR on the main channel. This secrecy capacity can also be used to share a symmetric key that can be later used for encryption [24].

Secrecy-achieving strategies. The secrecy capacity expressions are generally proved using the random-coding argument, a method that does not provide explicit code construction because it relies on averaging over all messages in all possible codebooks. Wiretap code design is also challenging because their performance cannot be measured by an objective metric like the BER in the case of error correction codes [18]. For example, LDPC, polar, and lattice codes can be used for building wiretap codes for binary erasure channels [18], binary symmetric channels, and Gaussian channels, respectively.

It can be observed that secrecy can be achieved when the legitimate users have a so-called “advantage”, which can be designed as a coding advantage, a signal processing advantage or both. Although wiretap codes are hard to construct, a generic code structure that guarantees both secrecy and reliability can be inferred: each message should be mapped to a pool of codewords, from which the transmitter randomly draws the transmitted codeword in order to confuse the eavesdropper (i.e., a binning structure [18]). A signal processing advantage can be designed by employing multiple antennas, beamforming or jamming [20]. The signal processing only approach cannot be translated in an information-theoretic weak/strong secrecy conditions but simplifies the system design [20].

In order to illustrate how secrecy could be possible over a noisy channel without any encryption mechanism, consider a degraded AWGN wiretap channel.[6] According to the definition, the degraded wiretap channel [Alice-Eve] has lower SNR than the main channel [Alice-Bob]. Consequently, for a given binary constellation employed by Alice and Bob, the BER for the main channel is lower than the BER for the wiretap channel. Consider that the difference is large enough so that after decoding of the repeated symbols sent by Alice, Bob is able to identify an unique symbol, while Eve can only

[6] The following example is just for intuition purposes. It does not achieve strong secrecy.

see a cloud of points all over the constellation, which would make her unable to decode the sent symbol. This is one of the simplest examples of exploiting a physical layer advantage to securely transmit information between two parties. Despite the inherent existence of the advantage in this scheme, the choice of the constellation by Alice and Bob depends on the knowledge they have about the state of the wiretap channel, which, along with code design, is one of the main challenges in the field of physical layer secrecy. Usually, physical layer secrecy studies consider either complete or partial knowledge of CSI or no knowledge at all [20].

Next, we will briefly describe the second approach in physical layer security, i.e., secret key agreement, which stems from exploiting or creating the same kind of advantage while using an authenticated noiseless public channel.

Secret key agreement

The second approach for providing information-theoretic security using the physical layer is to generate secret keys from communication channels and use them as one-time pads. The legitimate users employ the noisy communication channel to generate correlated (i.e., reciprocal) observations and a noiseless public feedback channel to correct the mismatches and obtain a secure key. Consequently, the secrecy and reliability requirements can be treated separately, making secret key agreement strategies simpler to implement than wiretap code design [18]. The public channel is commonly assumed to be bidirectional, authenticated and rate-unlimited. In wireless communications, correlated observations can be obtained in two main ways.[7]

- Two-way channel probing: owing to the reciprocal nature of electromagnetic wave propagation, physical layer measurements are a common source of information for two legitimate users and can be therefore processed in order to obtain common bits. Ideally, the two users share an inherent advantage over any possible eavesdropper situated at more than a few wavelengths of any of them because the eavesdropper's radio channel is uncorrelated to the main direct and reverse channels. This is a possible instantiation of the "source model" for secret key agreement and it represents the approach taken in this thesis.
- One-way transmissions over the wireless channel: the observed outputs at the receiver and eavesdropper correspond to the outputs of a classical wiretap channel. Therefore, the so-called "channel model" for key agreement can be seen as a wiretap channel reinforced by a public channel, but with a different purpose (i.e.,

[7] A more detailed view of secret key agreement models is given in Section 1.2.1.

generate a secret key for encryption purposes at higher levels instead of obtaining secrecy directly at the physical layer) [18].

Conclusion

In this section, we have briefly described how the physical communication channel can be used either for keyless secure communication or for secret key agreement. These notions are intrinsically linked to information-theoretic security, which does not place any computational restrictions on the eavesdropper. In order to get close to or achieve information-theoretic security, systems should use suitably long wiretap codes or utilize the generated secret keys as one-time pads. Meanwhile, even though the key rate might not be large enough for one-time padding, generating a shared secret key on the fly between two devices is a promising alternative for symmetric key distribution, especially in the context of decentralized networks. Moreover, physical layer security does not require additional equipment than that already employed for communication [18].

Nevertheless, physical layer security is also facing several challenges [25]. First of all, compared to cryptographic approaches, most physical layer studies consider only passive attackers (at both physical and applicative layers) with few observations and the same type of communication equipment as the legitimate users. Then, a set of assumptions on the wireless channel are necessary for some physical security schemes: i) total or partial knowledge of the eavesdropper's channel state for secure keyless communications; ii) reciprocity of the bidirectional communications and in the same time temporal decorrelation of the channels for continuous key generation; iii) spatial decorrelation or difficulty of prediction of the wireless channel by the attacker.[8] Finally, information-theoretic channel models or suggested signaling (e.g., Gaussian signaling instead of practical QAM) might not always be realistic.

Several of these issues have already been or start to be addressed. For example, in the case of key agreement, secret key generation under active attacks has recently been considered in practical key generation schemes [26]. Several experimental studies in various conditions and with various radio technologies confirm the approximate reciprocity of communication channels ([27], [28], [29]) or propose methods to post-process measurements for improving it [30]. Also, attackers with powerful ray-tracing software cannot gain sufficient information about the legitimate signals in rich scattering environments, as shown in [31].

[8] This particular assumption could be questioned if powerful ray-tracing tools and the necessary knowledge about the environmental conditions are available to the attacker.

Physical layer security can be extended to include authentication schemes based on the uniqueness of the channel between the legitimate users [25] or on location- and device-dependent features.[9] While the first two physical layer security components dealing with confidentiality are situated at the crossroads of information theory and signal processing, the latter consists in signal processing techniques for linking the identity of a device with its wide-sense physical context (e.g., radio channel, absolute positions, relative connectivity).

To conclude with, as already suggested by the need of prior authentication of the public channel, physical layer security solutions are not meant to be a replacement for cryptographic solutions, but rather a complementary solution in order to enforce security, especially in emerging networks with decentralized architecture or limited computational capabilities.

1.2 Physical layer secret key agreement

The present section focuses on the main contributions in the state of the art of physical layer secret key agreement. Firstly, the seminal papers from 1993 introducing the information-theoretic secret key agreement from common randomness and the two standard key generation models are presented, followed by the description of the sequential key distillation procedure for the source model. The link between the information-theoretic models and the wireless fading channel is also discussed, which leads to a classification of the identified key generation approaches. Finally, we outline various channel characteristics or configurations usually employed for key generation with wireless channels.

1.2.1 Source and channel models

In the field of information theory, two main papers [32] [33] provide the transition from secrecy capacity studies, initiated by the introduction of the wiretap channel model, towards the notion of secret key capacity. The contributions of these studies are summarized below.

Firstly introduced as the secrecy capacity of broadcast channels with public discussion [32] (or equivalently, the *channel model* capacity [33]), the secret key capacity is then defined in the context of a broader model (or equivalently, the *source model* [33]). In

[9] Studies regarding the generation of location- and device-dependent pseudonyms have been accomplished in parallel of the present thesis and are reported in Appendix E.

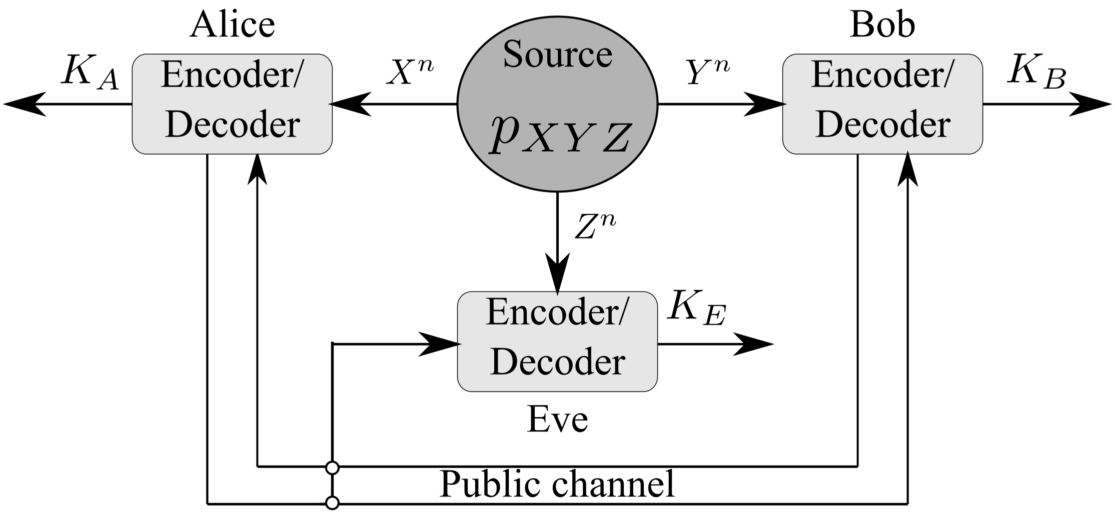


FIGURE 1.4: Secret key agreement source model

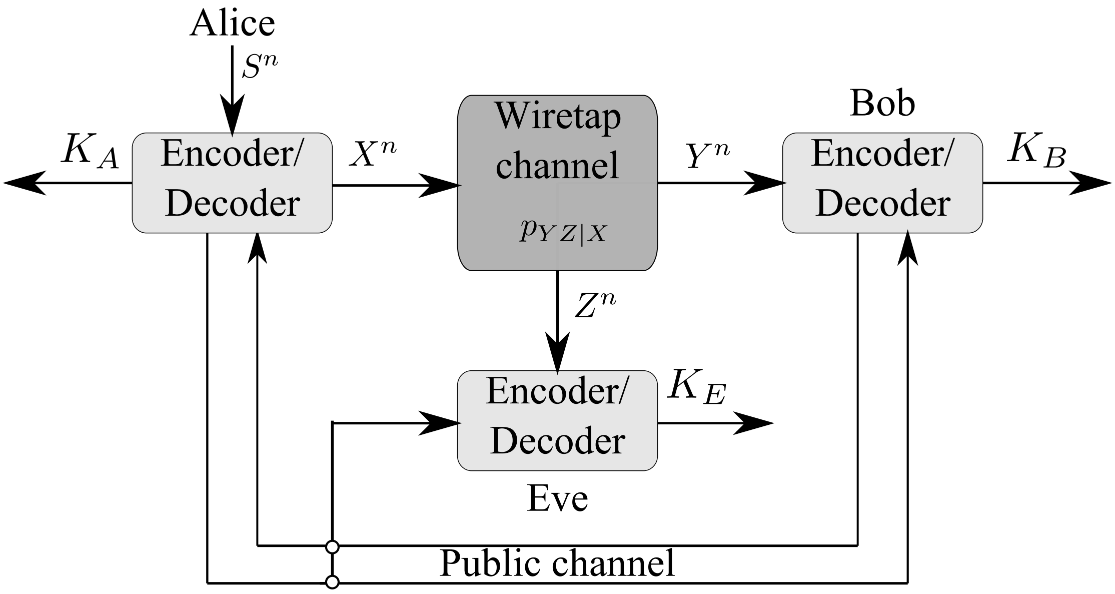


FIGURE 1.5: Secret key agreement channel model

parallel, Ahlswede and Csiszàr [33] shows similar results on the secret key capacity of the so-called *source model* and *channel model* with or without wiretapper/eavesdropper.

To summarize, the *source model* (Figure 1.4) assumes the existence of a discrete memoryless source (DMS) defined by p_{XYZ} with components (X^n, Y^n, Z^n) observed by Alice, Bob, and Eve respectively. In the *channel model* (Figure 1.5), Alice sends a random sequence X^n over a discrete memoryless channel (DMC) defined by $p_{YZ|X}$ and Bob and Eve observe the outputs Y^n and Z^n. Indeed, the broadcast or wiretap channel (characterized by $p_{YZ|X}$) is a particular case of common information (characterized by $p_{XYZ} = p_{YZ|X}p_X$) when Alice chooses p_X and generates n i.i.d. realizations according to p_X. Both models include a two-way authenticated public channel with no rate limitation, unless stated otherwise.

The secret key capacity ($\mathbb{S}(X;Y||Z)$) represents the maximum achievable rate at which Alice and Bob can generate a key out of their observations X^n and Y^n, while keeping Eve's key information rate from her observation Z^n and public messages (ϕ_A, ϕ_B) arbitrarily small. Given a space of keys $\mathcal{K}$, a secret key rate R is achievable if for every $e > 0$ and sufficiently large n, there exists a secret sharing strategy such that [33]:

$$\begin{aligned} &\mathbb{H}(K_A) > R - e && \text{(key rate)} \\ &\mathbb{P}(K_A \neq K_B) < e && \text{(reliability)} \\ &\tfrac{1}{n}\mathbb{I}(K_A; Z^n, \phi_A, \phi_B) < e && \text{(secrecy)} \\ &\tfrac{1}{n}\mathbb{H}(K_A) > \tfrac{1}{n}log|\mathcal{K}| - e && \text{(key uniformity)} \end{aligned} \tag{1.6}$$

Maurer [32] derives upper and lower bounds for the secret key capacity of the source model in the case of independent repeated realizations of each random variable X, Y, and Z (Eq. (1.7)-(1.8)). Finally, it is shown that for a particular case of joint probability distribution p_{XYZ} (i.e., when the three parties receive a binary symmetric source over independent binary symmetric channels), secret key agreement is possible even when the eavesdropper's observations are on average less noisy than the legitimate observations. This is rather counter-intuitive given the results on secrecy capacity but nonetheless possible because of the public channel, which the legitimate parties can use to create an advantage. Later, Maurer and Wolf [34] finds a tighter upper bound for the secret key rate of the source model by introducing the notion of "intrinsic conditional mutual information".

$$\mathbb{S}(X;Y||Z) \geq \max(\mathbb{I}(Y;X) - \mathbb{I}(Z;X), \mathbb{I}(X;Y) - \mathbb{I}(Z;Y)) \tag{1.7}$$

$$\mathbb{S}(X;Y||Z) \leq \min(\mathbb{I}(X;Y), \mathbb{I}(X;Y|Z)) \tag{1.8}$$

Ahlswede and Csiszàr [33] discusses the secret key capacities for the source and channel models without and with a wiretapper. If no wiretapper is present, the secret key capacity for the source model (i.e., $\mathbb{I}(X,Y)$) can be obtained with one-way public discussion, whereas the secret key capacity for the channel model (i.e., the capacity of the DMC) is achievable with no public channel. Then, the paper presents closed single-letter characterizations for the source and channel key capacities with forward (i.e., from Alice to Bob) public discussion. Lastly, it is outlined that the knowledge of the wiretapper's observations by one of the terminals can improve the secret key capacity.

Both studies point out the link between the source and the channel models: to a given DMC($p_{YZ|X}$), one can associate several DMS($p_{XYZ} = p_{YZ|X} \cdot p_X$) by varying p_X (or equivalently, a DMC can be viewed as a DMS in which one of the parties, namely the transmitter, can control the source). It can be inferred that the key capacity of the channel model is larger or equal to the supremum of the key capacities of the associated source models [33]. This link between the two information-theoretic models leads to several inquiries. What is the equivalent of these models in wireless physical layer security and what is the link between the corresponding source and channel models? Are there other more "practical" models for secret key generation exploiting the randomness of the wireless fading channel? Before classifying key generation models exploiting wireless channels, the next section presents the sequential key distillation procedure for the general source model.

1.2.2 Sequential key distillation for the source model

It has been shown that, unlike for the wiretap channel, the reliability and secrecy requirements for the source model can be implemented independently as a succession of several phases [18].

Randomness sharing. This first step corresponds to the observation of n random realizations of the source by Alice, Bob, and Eve.

Advantage distillation. In the unfortunate case in which Alice and Bob share less information than Alice and Eve for example ($\mathbb{I}(X;Y) \leq \mathbb{I}(X;Z)$), a public discussion phase is needed in order for Alice and Bob to obtain an advantage (e.g., by keeping only a subset of the realizations). Note that this step implies that Alice and Bob have some information about the statistics of Eve's realizations.

Information reconciliation. Even though Alice and Bob share an advantage with respect to Eve, they can still have differences in their observations. In order to correct them, Alice publicly sends partial information about her sequence to Bob, who uses it to correct the mismatches (i.e., one-way public discussion). Bob can also reply but in this case, the protocol is more difficult to analyze. The reconciliation information, which can be directly observed and employed by Eve, should be kept as small as possible. This step is theoretically equivalent to a source coding problem with side information: given the side information Y, the amount of information that Bob needs to recover Alice's observation X is $\mathbb{H}(X|Y)$. Since $\mathbb{H}(X|Y) = \mathbb{H}(X) - \mathbb{I}(X;Y)$, it is obvious that the more mutual information Alice and Bob have, the less information they need for reconciliation. Similarly to advantage distillation, Alice should preferably have statistical information about the number of mismatches between her sequence and Bob's in order to compute the needed reconciliation rate.

Privacy amplification. Since Eve's information about the shared advantage of Alice and Bob increases because of the information reconciliation step, the latter have to apply supplementary operations to their sequences (considered equal at this point). This can be a deterministic function (e.g., a hash function or a randomness extractor), whose role is to generate a uniformly distributed sequence. The result will be shorter but more secure (i.e., Eve will have negligible information about it). The present step also removes any correlation that might exist between consecutive observations, which is also an advantage for Eve.

Quantum key distribution. One of the early illustrations of the mentioned sequential procedure is QKD based on the laws of quantum physics [35] and implemented using single photons (or light pulses in practical scenarios), whose linear or circular polarities encode the key bits. For example, “1” can be encoded either as a vertically polarized photon or as a right-circularly polarized one. The linear and circular bases are conjugate bases: measuring the polarization in one of them randomizes the polarization in the other one (i.e., Heisenberg’s uncertainty principle). This quantum phenomenon allows the legitimate users to statistically detect an eavesdropper having measured the same photons using a different basis than the one used by the legitimate users (e.g., the BB84 protocol [16]).

Several QKD protocols based on different quantum phenomena have been developed [36] (BB84 and its variants, continuous-variable QKD, entanglement-based protocols, etc.). In order to illustrate the sequential distillation procedure, we briefly describe the BB84 protocol [16]. One may argue that the latter does not follow the source model but rather a channel model. However, given the particular physical properties of light transmissions, the sequential procedure can be applied as follows.

Randomness sharing. Alice sends a random sequence of bits, each one modulated in a randomly chosen basis. Bob measures the polarization of the received photons also in a randomly chosen basis and then, only keeps the bits measured with the correctly guessed bases. The two can also estimate the leaked information to an eavesdropper after a public discussion.

Information reconciliation. In order to find and correct possible errors, a block parity check algorithm is proposed, followed by a repeated parity check on randomly chosen subsets of bits. When an error is detected in a block, iterative parity check per sub-block is performed until the error is corrected.

Privacy amplification. In order to prevent information leakage from the public channel, for each parity bit disclosed on the public channel, Alice and Bob drop the last bit of the corresponding sequence. Furthermore, the leaked information during randomness sharing is suppressed by applying a hash function parameterized by the number of final bits, the estimate of the leaked information, and the desired level of security.

The aforementioned protocol has been proved to be perfectly secure under certain assumptions regarding its implementation, the equipment (photon sources, polarizers, beamsplitters, photon detectors) and the reliability of the quantum transmissions [35]. Commercial systems with dual key agreement (quantum and classical) are already available and current research in the QKD field focuses on subjects such as equipment imperfections that have not been yet considered in the proofs, undiscovered vulnerabilities, quantum key distribution networks, etc.[10]

Practical sequential key generation

In practical wireless scenarios, the classical key generation steps are [26]: channel probing and randomness extraction (e.g., measuring RSSI and extracting the small scale fading), quantization (conversion of real or complex values to bits), information reconciliation and privacy amplification. We can map the first three steps to the aforementioned "randomness sharing" phase.

In some cases, an additional public discussion phase before error correction is necessary, for example, if the measured sequences are not synchronized. In our studies, we will consider that the information reconciliation phase consists of: i) a preliminary public discussion for inducing coherence between Alices's and Bob's measurements (e.g., synchronization) ; ii) a conventional error correction scheme. Note that the former can be designed together with the quantization algorithm or separately.

1.2.3 Key generation models with wireless channels

Wireless channels are typically characterized by three main phenomena that affect transmissions: path-loss, shadowing and small-scale fading. Physical layer security is predominantly interested in the random time-varying component of the wireless channel, either in the form of small-scale fading in narrow-band systems or multipath components in UWB systems. In the following, we will refer to the "wireless channel" from a key generation perspective.

We identify several key generation models employing the wireless channel based on the aforementioned source and channel models. Essentially, the channel model consists

[10] For more details on QKD advantages and challenges one can refer to the *SECOQC White Paper on Quantum Key Distribution and Cryptography* and to *The black paper of quantum cryptography: real implementation problems* [36].

in transmission strategies that use the wireless channel as a medium in order to convey secret information, whereas the source model considers the wireless channel as a random source of information and not as a transmission support.

- Source model. This approach relies on the reciprocity and spatial decorrelation properties, which offer an inherent advantage to the legitimate users with respect to the eavesdropper. The legitimate users have correlated observations of the common source, the wireless fading channel, while an eavesdropper, located at a certain minimum distance of the legitimate users, has considerably lower chances of obtaining correlated observations. This key generation approach involves bidirectional training or channel probing and the phases described in Section 1.2.2. The majority of practical secret key generation algorithms presented in Section 1.2.4 fall under this category.

- Extended source model. The sender-excited model is a generalization of the source model with optimized probing signals [37]. In scenarios with long channel coherence times, the bidirectional entropy harvesting source model is not efficient. A possible solution would be the virtual channel approach where one of the parties induces controlled channel variations during the channel coherence time by using e.g., two antennas [38].

- Wiretap channel model. The channel model for key agreement can be implemented through classical wiretap coding strategies without a public channel and with various levels of channel state knowledge, as investigated in [39] (e.g., full main and wiretap CSI at the transmitter, main CSI only at the transmitter, main CSI only at the receiver). Usually, these strategies are not optimal because they do not make use of the public channel.

  If a public channel is also available, equiprobable dense parity codes can be employed to securely share a secret key in the case of a binary symmetric wiretap channel [40]. The key is distilled from the encoded sequences that are correctly received. The employed codes guarantee both a lower bound on the probability of detecting correctly received symbols and a lower bound on the equivocation rate, ensuring thus information theoretic security.

- Source-emulation channel model. An alternative implementation of the channel model for key agreement is the source-emulation: Alice generates a discrete memoryless source and sends it over the wireless channel in order to yield correlated observations for Bob. The final key is obtained by information reconciliation and privacy amplification like in the source model. Secret key capacities for Rayleigh fading channels with unknown CSI but known CSI statistics has been investigated in [41].

Based on the observations that fading is beneficial to secrecy capacity [23], practical secret sharing schemes [24] can be designed for AWGN Rayleigh fading channels when instantaneous CSI is available (i.e., complete main channel CSI and complete/partial wiretap CSI at the legitimate users and wiretap CSI at the eavesdropper). The proposed protocol consists in a one-way transmission of a random sequence during the time slots when the secrecy capacity is positive, followed by error correction by LDPC codes and privacy amplification with universal hash functions.

- Mixed models. Recent studies also consider mixed training-transmission key generation strategies with an initial phase of channel probing (i.e., source model) followed by a transmission phase using: i) wiretap coding with receiver CSI only [42]; ii) a source-emulation strategy based on the channel estimates from the first phase [43]. Secret keys are generated from both phases. In both cases, the results indicate that when the channel coherence time is long, the contribution from the source model vanishes and the channel model should be used to achieve high key rates.

- Reciprocity-based channel model.[11] Because the source model is highly dependent on the coherence time, static channels would not be adapted for key generation within the practical source model approach. Alternative schemes combining reciprocity/spatial decorrelation and user-generated randomness have been proposed [44] [45] [46]. For example in [45], Alice and Bob generate random phases ϕ_A and ϕ_B, respectively, and use them to modulate their transmitted signals. If both of them transmit the modulated signals during a period shorter than the channel coherence time, they will be able to obtain a secret key by quantizing the shared information $\phi_A + \phi_B + \phi_c$, where ϕ_c is the phase induced by the reciprocal bidirectional channel. Keys can also be generated from unilateral user-generated randomness if the channel effect is mitigated by inversion operations (e.g., conjugation [46]).

- Alternative model: unknown deterministic sources [47]. Recently, key generation has also been studied from a different perspective that combines the information-theoretic concept of "mutual information" with the notion of "unknown deterministic parameters" from detection and estimation theory. Let Alice and Bob observe a noisy deterministic quantity x. Alice "enciphers" a secret message W

[11] We note that this denomination is not part of the established terminology in the key generation state of the art.

with her quantized observation x_A^q as $W_e = f(W, x_A^q)$, where f is a deterministic function. Then, she sends W_e to Bob, who "deciphers" the intended message W using W_e and his quantized observation of the source x_B^q. The result is a new framework that can characterize the secret key rate, perceived as a secure communication rate achievable with secure mapping functions parametrized by the quantized observations of the source.

Secret key agreement with an active attacker (i.e., using a non-authenticated public channel) can be studied under various assumptions. If the attacker completely blocks the communication during the attack slots, the success of the key agreement depends on the joint initial distribution of the source p_{XYZ} and the suggested protocol consists in secret key agreement during the secure slots [48]-[49]. In the case of wireless fading channels, it is not possible to completely block the communication but it is commonly assumed that the attack is continuous. This scenario is examined in the context of a mixed source-channel model for fading channels and positive secret key capacity is shown to be possible [50].

Finally, we also note that secret key generation is not limited to wireless communications and can be also achieved, for example, in networked linear control systems (e.g., sensors and controllers) by exploiting common system state information modifiable by controllers [51].

1.2.4 Wireless channel characteristics for key generation

Key generation methods can be also classified based on the employed channel characteristics (e.g., phase, channel impulse response, magnitude) and their variation over various domains (e.g., time, frequency, space) [52]. Measurements are usually performed on absolute or relative scales and their variation is either inherent (i.e., the source model) or user-induced (i.e., the reciprocity-based channel model or intentional variation of transmission parameters such as antenna configurations).

In this taxonomy, we only focus on the source model with a few exceptions of references indexed by u, which represent scenarios of user-induced randomness. A review of some of the key generation methods involving received signal strength or phase information can be found in [53].

- Phase. Although it can be challenging to achieve coarse frequency synchronization, quantization of phase information has been investigated for both narrow-band [54] $[55]_u$ $[45]_u$ and wide-band channels $[44]_u$ [56] $[46]_u$.

- Received Signal Strength (RSS). RSS is a popular channel characteristic for bit extraction because of its wide availability in wireless devices [57] $[58]_u$ [59] [60] [27]. Subsequent studies on RSS concern key generation implementation for a network of mobile nodes and static anchors [61] or WBAN [62]. The trade-off between bit generation and resource consumption during channel probing is investigated in [63]. Despite the simple acquisition process, RSS is considered a low-entropy feature and key generation from RSS measurements usually requires highly dynamic environments to achieve large key lengths.

- Narrow-band CIR. The possibility of employing narrow-band CIR for key generation is investigated in comparison to RSS [60] or as a function of the sampling frequency and the channel coherence time [64] [65].

- Multiple channels. An example is to measure RSS over multiple frequency-selective channels [66]. OFDM-like channels have also proved to be a popular key extraction feature in comparison to RSS measurements [67], narrow-band coefficients [68] or when combined with: i) uncorrelated feature extraction [69]; ii) delay mitigation and mobility bias correction [70]; iii) user-generated randomness $[46]_u$. Experimental key generation from MIMO narrow-band RSS measurements is investigated in [71]. Regarding MIMO theoretical key generation rates, various limiting factors have been analyzed: measurement/estimation noise, spatial correlation between the legitimate and illegitimate channel [72] [73], and time correlation of the legitimate channel [73]. Moreover, key generation using beamformed multi-antennas can attain higher secret rates when the nodes have statistical knowledge about the MIMO channel [74].

- IR-UWB CIR. The delays and amplitudes of multipath components present in the IR-UWB CIR (see Section 1.3) represent an interesting source of information for symmetric key generation. After the first proof of concept for key generation using relative excess delays of the IR-UWB CIR [75], the same characteristic has been recently employed for IR-UWB key generation with channel probing considerations [76]. The upper-bound for the achievable secret key rates from CIR has been derived [77] [78]. Several experimental studies with metrological equipment confirm the reciprocity and spatial decorrelation properties of IR-UWB channels [29] [28] and investigate the properties of the generated keys [79] [28] [80]. Because of the temporal correlation of CIRs in typical indoor environments, generated keys can be highly predictable when perfect knowledge of the past channel realizations is available to an attacker [81]. The proposed solution [81] is to employ linear prediction in order to remove the predictable samples.

Other less conventional key extraction algorithms exploit characteristics such as relative node distance [82] or angle of arrival [83], special equipment like reconfigurable antennas $[84]_u$, jamming mechanisms [85], or multi-hop networks [86]. Moreover, recent studies investigate the negative impact of information leakage from antenna reflections at relatively high SNR values [87].

Although the main aim of a physical layer secret key generation procedure is symmetric key distribution, one can consider exploiting such schemes for security measures such as joint encryption-turbo coding [88].

1.3 Impulse Radio - Ultra Wideband

Although invented at the beginning of the 20^{th} by G. Marconi and employed in radar technology since the 1960's, UWB communications have gained increased popularity in the last decade [89]. According to the regulation defined by the US Federal Communications Commission (FCC) in 2002, UWB refers to all the emissions with a power spectral density lower than -41.3 dBm and with a bandwidth larger than 500 MHz or larger than 20% of the central frequency. This regulation was intended to define secondary usages of the 3.1 - 10.6 GHz band, which explains the low transmission levels corresponding to non-intentional emissions of electronic devices. European regulation defines a low band (3.4-4.8 GHz) and a high band (6-8.5 GHz) for UWB transmissions and imposes the implementation of additional mechanisms in order to protect the coexisting services: Detect And Avoid mainly for high data rate UWB transmissions and Low Duty Cycle aimed at limiting the emission time of low data rate UWB transmissions (e.g., to around 3%) [90].

Today, there are two approaches for UWB communications: the multi-band UWB based on OFDM for high data rate at short transmission ranges and the Impulse Radio UWB consisting in the transmission of short low duty cycle impulses of the order of a few nanoseconds. Because of its temporal characteristics, IR-UWB exhibits high temporal resolution capabilities, which makes it suitable for radio localization applications involving ranging through precise detection of the Time of Arrival (ToA) of transmitted packets. Moreover, because of the low duty cycle transmissions, IR-UWB is adapted to energy-efficient devices as well.

IR-UWB standardization

Two main IR-UWB standards have been proposed [90]:

- IEEE 802.15.4a for joint low data rate communications and peer-to-peer ranging. The IR-UWB technology, which is capable of achieving low data rates (from 100 kbit/s to 26 Mbit/s) over medium to long distances, has been the natural choice for the physical layer.
- IEEE 802.15.6 for high data rate solutions over short-distance communications in body area networks. The retained physical layer is based on a modified version of the initial 800.15.4a physical layer in order to integrate the particular topology and variable data rates of these networks.

IR-UWB signal

An IR-UWB signal consists in a repetition of pulses with a mean Pulse Repetition Period (PRP) and possibly scrambled by a Time Hopping code in order to smooth the corresponding spectrum. Popular pulse shapes are the Gaussian pulse, the Gaussian monocycle (the first derivative of the Gaussian pulse) and the second derivative of the Gaussian pulse. This is due to the optimal time-bandwidth product of Gaussian pulses (i.e., $B = 2/\tau$ where B is the bandwidth and τ is the useful temporal support of the pulse [90]), which achieves maximal time-rate resolution, but also to the facility of generating Gaussian pulses at the antenna level [89].

Classical modulation schemes include (Differential) Binary Phase Shift Keying (D)BPSK, Pulse-Position Modulation PPM (i.e., the temporal position of the pulse indicates the modulated information), On-Off Keying OOK (i.e., modulation by the presence or the absence of the signal) or Transmitted Reference TR (i.e., the information is modulated by the difference between a reference pulse and a secondary one). In the IEEE 802.15.4a standard, a combination of BPSK and/or PPM is considered for flexible modulation (i.e., depending on coherent/non-coherent Rx implementations) and a cascade of an RS encoder and a convolutional encoder is added for channel coding purposes.

IR-UWB channel model

The IEEE 802.15.4a statistical channel models for various type of environments (e.g., office, residential, outdoor, industrial, etc.) and two frequency regimes (above 3 GHz and below 1 GHz) [91] are based on a modified version of the Saleh-Valenzuela (S-V) indoor channel model developed in 1987 using measurements with 10 ns pulses [89]. The S-V small-scale fading model assumes that arrival times of multipath components are Poisson-distributed within clusters, which also follow a Poisson distribution. The CIR can be expressed as:

$$h(t) = \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} x_{l,k} e^{j\phi_{l,k}} \delta(t - T_l - \tau_{l,k}) \tag{1.9}$$

where T_l is the arrival time of the l-th cluster, $\tau_{l,k}$, $x_{l,k}$, and $\phi_{l,k}$ are the multipath arrival time, amplitude, and phase of the k-th multipath within the l-th cluster. The phases $\phi_{l,k}$ are uniformly distributed in $[0, 2\pi]$ and the amplitudes $x_{l,k}$ are Rayleigh random variables with an exponentially decaying power profile ($\mathbb{E}[x_{l,k}^2] = \mathbb{E}[x_{1,1}^2] e^{T_l/\Gamma} e^{\tau_{l,k}/\gamma}$ with Γ and γ the cluster and multipath decaying rates).

The standardized proposal [91] includes several modifications of the S-V model: i) the number of clusters is considered a Poisson random variable; ii) the multipath arrival times are modeled as a mixture of two Poisson processes and the cluster decaying rate Γ as a function of T_l; iii) the small-scale fading amplitudes $|x_{l,k}|$ follow a m-Nakagami distribution. Also, the IEEE 802.15.4a standard modifies the clustering model depending on the considered environment, includes a frequency-dependent term in the path gain model and proposes a new model for correlated shadowing for WBAN. In the following, we will consider this model for the performance evaluation and benchmark of some of our proposals.

Another approach to channel modeling can be achieved by deterministic propagation simulators, like ray-tracing tools that capture the dependency of the received signal with respect to Tx-Rx positions and to the environment. In Chapter 3, we will employ data generated with an IR-UWB ray-tracing tool [92] in order to investigate the effect of the spatial correlation of IR-UWB signals on an existing key generation protocol.

IR-UWB receivers

At the receiver, the high sampling rates imposed by the high signal frequencies and large bandwidth can be prohibitive given the consumption requirements for integrated devices. This poses challenges in terms of synchronization/timing acquisition, demodulation and channel estimation, which are usually designed to match a target level of performance depending on the application [90]. Generally, IR-UWB receiver architectures can be [89] [90]:

- Coherent. For example, coherent Rake receivers require *a priori* synchronization and channel estimation in order to separately process every received pulse and combine the results provided by the multipath diversity before making a demodulation decision.

- Semi-coherent. A system employing DPBSK or TR modulations can integrate a semi-coherent receiver, which demands only intra-symbol synchronization, i.e., the detection of the time of arrival of the first multipath component.
- Non-coherent. Energy detectors [93] and mixed architectures [94] can simplify a lot the receiver design. According to ED architectures (e.g., for n-PPM, OOK modulations), the energy of the received signal is integrated in small bins whose width is approximately equivalent to the unitary pulse duration, thus relaxing the traditional constraints on synchronization precision, relative clock drifts between Tx and Rx, and sampling rate, which becomes a function of the bin integration duration for non-overlapping bins [93]. In the case of semi-coherent DBPSK modulations, a prior stage of coherent integration of one-bit quantized signals can be added for SNR improvements [94].

Herein, we are solely interested in the channel estimation phase, irrespective of any modulation-related questions. A more detailed state of the art on channel estimators is provided in Section 2.4.1.

1.4 Motivations and contributions of the thesis

The present thesis explores several aspects of the secret key generation procedure involving the source model, which can be practically implemented with sequential methods (Section 1.2.2). Compared to the source emulation model, which can also be designed according to a sequential approach, the source model does not rely on a temporary or engineered advantage for achieving key sharing. On the contrary, the reciprocity and spatial decorrelation of radio signals are inherent characteristics of radio transmissions and the channel estimation phase is already employed for communication purposes in some systems. This means that the source model is relatively simpler to implement than the other key generation models. However, it is dependent on channel variability, which should be slow enough to allow bidirectional channel probing and fast enough in order to achieve an acceptable key generation rate. The study of channel variability is out of the scope of this work, which focuses on the exploitation of single channel realizations and assumes that the channel is sufficiently slow in order to fulfill the channel probing constraints.

Because of its rich multipath resolution capabilities and favored by its increasing popularity for localization-aware applications and potentially in various kinds of decentralized emerging networks (e.g., WSN, WPAN etc.), the IR-UWB physical layer has been chosen for our investigations. It will be therefore used for both single links and cooperative scenarios (i.e., in mesh networks). We focus on the first part of the sequential key generation chain (i.e., the gray-highlighted steps in Figure 1.6), prior to error-correction and privacy amplification, because of its close links to signal processing techniques.

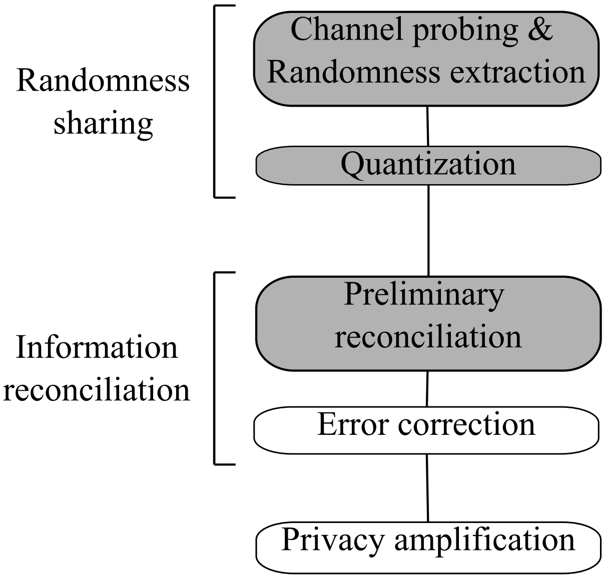

Randomness sharing
Channel probing & Randomness extraction
Quantization
Information reconciliation
Preliminary reconciliation
Error correction
Privacy amplification

Figure 1.6: Key generation steps within the source model (with concerned steps highlighted in gray)

Although joint design of, e.g., quantization and information reconciliation procedures could be a promising research direction, we choose to study them separately[12] because of: i) the simplicity of the approach; ii) the need to initially understand the specific structure and information that can be extracted from IR-UWB signals in single-link and cooperative contexts. Unlike RSS measurements, IR-UWB signals have a special vectorial structure with amplitude and temporal components (i.e., the attenuation and arrival time of multipath components) and the probing phase has an impact on the key generation performance depending on its parameters (e.g., sampling rate, estimation algorithm, etc.). Moreover, all the available physical layer characteristics should be exploited in order to obtain physical layer security solutions with a high potential of device integration. In the case of mobile IR-UWB devices, high-precision ranging provides a valuable source of reciprocal information even though the entropy of such measurements is lower than that of the CIRs. We also try to understand how far the original "physical layer" concept can be extended in the case of cooperative key generation without recurring to higher layer key distribution protocols, which may involve numerous packet exchanges.

This thesis provides guidelines concerning the aforementioned research axis. We start from existing IR-UWB key generation studies [95] [29] [79] [96], which include a key

[12] The obtained bit sequences will sometimes be called "keys" without any implication on their symmetric character.

generation protocol for directly sampled CIRs [79]. Regarding this signal model, we propose two extensions of the mentioned protocol: the first one improves the random character of the generated keys by a diversified bit encoding procedure [97] (Section 2.2) and the second one increases the immunity to eavesdropping by limiting or masking the public exchanges required for information reconciliation [98] (Chapter 3).

Then, we pursue our investigations in the context of synthetic channel estimates based on the IEEE 802.15.4a model in order to understand the impact of quantization on the reciprocity-randomness trade-off [99] (Section 2.3). In this context, we introduce the notions of "inter-key" and "intra-key diversity" to characterize randomness, show how to adapt the quantization thresholds in order to achieve a certain trade-off between reciprocity and inter-key diversity, and propose a scheme that favors intra-key diversity. Moreover, we extend the signal model to realistic channel estimates issued from various types of estimators (Section 2.4), some of which are designed to lower the prohibitive sampling rates that are usually needed for channel estimation. We examine the degradations in reciprocity incurred by these estimators and we propose a post-processing phase of the channel estimates that improves the reciprocity.

Finally, we develop a new signal-based method that extends the point-to-point source model to cooperative contexts with several nodes intending to establish a group key through the optimization of the channel probing signals [100] (Chapter 4). In this last chapter, the focus lies on the reciprocity at the signal level and on the design of the probing signals, which explains the simplistic ED-like approach that has been chosen for the quantization metric and the key performance evaluation.

Despite the fact that we employ the IR-UWB technology for all our simulations, the findings of Chapters 3 and 4 can be extended to other wireless technologies. Furthermore, our methods are not targeted or restricted to low-complexity networks such as WSN, but are rather general. On the one hand, in Section 2.4 we provide guidelines for quantizing channel estimations obtained with sub-Nyquist sampling. On the other hand, the cooperative key generation method developed in Chapter 4 requires terminals with evolved computational capabilities, such as multi-standard smartphones.

Each of the three main chapters also includes a review of related state of the art work and the concerned signal and system model.[13] The present document is concluded with a summary of the contributions and their limitations in Chapter 5 followed by a few personal thoughts about the societal impacts of information and communications technologies (Chapter 6). Other related subjects that have been investigated in parallel concern: i) quantization of ranging and device-dependent information for generating

[13] Throughout the document, the SNR is sometimes defined differently according to the context of the study.

pseudonyms and provide a security overlay to conventional authentication procedures (Appendix E); ii) quantization strategies for experimental IR-UWB channel estimates issued from low-complexity integrated devices [94] (Appendix F).

Chapter 2

Quantization of IR-UWB signals for secret key generation

This chapter addresses quantization issues in the context of symmetric key generation from IR-WB channels. After discussing the philosophy of key quantization and a few examples of algorithms (Section 2.1), we investigate several issues related to the quantization step.

- The random aspect of keys generated from directly sampled IR-UWB signals (Section 2.2): we propose a new bit encoding algorithm [97] that reduces the randomness defects compared to a previously proposed solution [79] [95].
- The inherent trade-off between reciprocity and randomness (Section 2.3): we introduce a new metric for quantifying the randomness of the quantization process, namely the *diversity* of the binary codewords. The mentioned trade-off is illustrated through a study on the optimization of the quantization thresholds for simulated IR-UWB channel estimates [99]. Also, a new diversity-aware quantization scheme is proposed for the same signal model.
- The impact of realistic IR-UWB channel estimates on the reciprocity performance of the channel estimates (Section 2.4): we evaluate three different estimators, namely a high sampling rate estimator and two sparse ones, and propose a pairing algorithm adapted to the obtained channel estimates in order to improve their reciprocity.

2.1 State of the art: quantization

Quantization for key generation

In the classical sense (e.g., for digital signal compression and/or reconstruction), the term "quantization" refers to the mapping of a continuous set of values to a discrete reproduction alphabet and its binary representation [101]. A memoryless quantizer can be represented by its three components: i) a lossy encoder α that maps the continuous value to an index; ii) a reproduction decoder β that maps the index to a codeword of the reproduction alphabet (i.e., the quantized values); iii) a lossless encoder γ that maps the index to a binary codeword. The quantization rule of an input value x is: $q(x) = \beta(\alpha(x))$.

The classical quantization problem is defined by the rate-distortion pair $(R(\alpha,\gamma), D(\alpha,\beta))$ from Eq. (2.1)-(2.2), which describes the trade-off between two conflicting goals : keeping the rate as small as possible and the distortion likewise (i.e., representing the input value as faithfully as possible using a minimum number of bits). We will call this scenario "classical quantization".

$$R(\alpha,\gamma) = \mathbb{E}_X[len(\gamma(\alpha(X)))] \tag{2.1}$$

$$D(\alpha,\beta) = \mathbb{E}_X[d(X,\beta(\alpha(X)))] \tag{2.2}$$

where d is a distance metric and len returns the number of bits of a binary sequence.

Classical quantizers can be classified as scalar or vector (depending on whether the lossy encoder α operates on scalars or on vectors), fixed or variable rate (depending on whether the reproduction decoder γ produces codewords of the same length or not), and memoryless or with memory (if α or γ depends on the previous quantization operations). Application of the quantization tools in the key generation context has to be adapted to its specific definitions and goals, as shown hereafter.

Let $y^u = [y^u_1, y^u_2, ...y^u_K]$ be K real samples issued from the channel probing/estimation phase at user $u \in \{A,B\}$. We consider the samples as realizations of K continuous independent RVs $[Y^u_1, Y^u_2, ...Y^u_K]$ denoted as Y^u. Further on, the term "key" ($\mathcal{K}^u$) will be used to denote the result of the quantization of y^u on each side of the link.

$$\mathcal{K}^u = q_{key}(y^u) = \gamma(\alpha(y^u)) \tag{2.3}$$

In most of the practical key generation protocols, there is no need for the reproduction alphabet or the reproduction decoder because we are only interested in the binary codewords. Moreover, the mapping between an eventual reproduction alphabet and the binary representation is a deterministic public function. Next, we choose between various quantizer types based on the goals of the key generation procedure and the metrics used to measure them.

First of all, an efficient quantization algorithm should be able to generate as many bits as possible from a single channel probe (i.e., high rate). In "classical quantization", vector quantizers are employed to achieve a lower bit rate given a fixed distortion [101]. This type of quantizers could be interesting from a joint quantization-reconciliation point of view. For simplicity, we will restrict the study to scalar quantizers. This means that a key is a concatenation of elementary codewords obtained from the quantization of the available samples of y^u :

$$\mathcal{K}^u = [q_{key}(y_1^u)||q_{key}(y_2^u)||...q_{key}(y_K^u)] \tag{2.4}$$

with || representing the concatenation operation. So, the length of the final key will be the sum of the lengths of the elementary binary codewords. Also, for simplicity, we only consider memoryless quantizers.

Secondly, a robust quantization algorithm should generate keys with good reciprocity, meaning low distortion between the keys generated at the two extremities of the link. The distortion in "classical quantization" implies a measure of the distance between the continuous input value x and its reconstructed quantized version $q(x)$. For key quantization, the distance has to be defined on a discrete space. One option would be the Hamming distance, which is defined for sequences of the same length.

Finally, key bits should have a random aspect, meaning, for example, that there should be diversity in the generated codewords of a single key or intra-key (i.e., over $k \in \{1 \dots K\}$), but also between the generated keys or inter-key (i.e., at fixed excess delay, over different channels). To illustrate the need of codeword diversity, we take as an example a quantization scheme in which the codewords are long (e.g., binary alphabet on 10 bits) and α is designed so that the first codeword is $q_{key}(y_1^u) = c_1$ for 99% of the input signals. The key rate is high (10 bits/codeword), the reciprocity is also high but the scheme performs very poorly from the security point of view because the first codeword of the key is highly predictable. Alternatively, an ideal quantizer from a diversity point of view is one that outputs equiprobable codewords. Even after key generation, randomness can be difficult to quantify. Studies usually use randomness tests for pseudo-random number generators that search for randomness flaws such as bit patterns or predictable oscillations [102].

Therefore, we identify three performance criteria for the evaluation of key quantization algorithms, namely the length or rate, the reciprocity and the random character of keys.

In this way, we incorporate into the quantization phase aspects related to reconciliation (i.e., reciprocity) or privacy amplification (i.e., randomness) in order to: i) control the desired trade-offs easier by trying to optimize only one operation; ii) eventually simplify the following steps of reconciliation and privacy amplification.

Quantization algorithms

In theory, the secret key rate is bounded by the mutual information of the observations of the legitimate users. In practice, quantization algorithms must be designed in order to "extract" the mutual information of the given measurements. For example, the mutual information between two correlated Gaussian random variables is compared to the secret achievable key rates with an equiprobable quantizer and a classical minimum-distortion quantizer, followed by an LDPC-based reconciliation procedure [103]. When using minimum-distortion quantization, the computed secret key rate must be adapted to take into account the entropy loss caused by the non-equiprobable quantization. Thus, the two quantizers have equivalent performances [103].

Despite the existence of general quantization models (e.g., equiprobable, minimum-distortion), quantization design depends on the type of input signal and its variations (e.g., RSS, CIR). As described in the following, it can also be concerned with the eventual public information needed to be exchanged between the two legitimate parties as a first stage of reconciliation (e.g., guard-band dropping or quantization map index [104]).

Quantization schemes usually fall into two categories: parallel quantization or asynchronous quantization.[1] Parallel quantization implies that samples are quantized independently by the two legitimate users according to a previously agreed upon algorithm and parameters. If needed, reconciliation information is exchanged after quantization. Asynchronous quantization is inspired from reconciliation procedures based on distributed source coding, namely source coding with side information. Accordingly, quantization is initially performed by one user and then, relevant quantization information is sent to the second user who can quantize his observation based on the received information. In the present thesis, we only consider parallel quantization because of its simple design and generality.

Parallel quantization

Several of the quantization algorithms that use RSS for bit extraction have been analyzed in terms of entropy, secret bit rate, and bit mismatch rate [27]. According to this

[1] We note that the denominations "parallel" or "asynchronous" as defined herein are not established terminology in the key generation state of the art.

study, the RSS quantizers fall into two categories: i) lossy or guard-band quantizers that drop measurements within a certain distance of the quantization thresholds (i.e., in the guard-bands) in order to achieve higher reciprocity [57] [58]$_r$ [60]; ii) lossless quantizers that do not drop samples but employ compulsory privacy amplification to remove the correlation between the generated bits [59]. The indexes of the dropped samples are publicly exchanged by the nodes.

The guard-band quantization algorithm from [60] is extended to an adaptive guard-band quantization algorithm, which divides the RSS measurements into blocks and applies quantization with block-dependent guard-bands [27]. Later, a guard-band quantization algorithm for IR-UWB CIRs [79] is developed: the corresponding guard-band thresholds are the same across a CIR but evolve during the iterative key extraction process starting from high values and decreasing (Section 2.2.2). In the context of experimental key generation based on IR-UWB CIRs, derivative-based and multi-bit uncensored quantization schemes are compared in terms of bit mismatch and bit frequency for both legitimate and illegitimate links [80].

Asynchronous quantization

An example of asynchronous quantization is given in [104] under the name of "Channel Quantization Alternating" (CQA). Instead of using guard-bands, the mismatch probability is reduced by adapting the quantization to each observed sample. For example, suppose that Alice quantizes her observation based on a one-dimensional quantization map. She then indicates Bob which side of the cell her observation fell in (e.g., the mapping bit can be "0" for left side and '"1" for right side). Before quantizing his observation, Bob shifts the initial quantization map so that the cells are centered on the sides of the initial cell corresponding to the mapping bit (i.e., if the mapping bit is "0" an initial quantization cell $(1, 2)$ becomes $(0.5, 1.5)$ for Bob).

Parallel studies propose other asynchronous quantizers: i) a mix of guard-band and asynchronous quantization including an authentication procedure intended for measurements with symmetric distributions [64]; ii) a more general syndrome-based scheme in which quantization and reconciliation are merged (i.e., Bob quantizes his observations based on the syndrome computed by Alice from her initial observations before quantization) [64]; iii) a multi-bit adaptive quantizer similar to CQA [104] preceded by an interpolation filter for reciprocity and a decorrelation transformation for randomness [30].

2.2 Keys from sampled waveforms: random aspect

Experimental studies [79] [95] show that relatively long keys can be obtained when using an IR-UWB metrological test-bed with trigger-based synchronization. The testbed consists of a pulse generator used to create a 2-9 GHz IR-UWB pulse, a digital oscilloscope with a time step of 50 ps at the reception side, and two omnidirectional dipole UWB antennas. Synchronization is achieved through a wired link used to send a trigger signal. These results have motivated supplementary evaluations of the key generation protocol on a variety of typical indoor IR-UWB signals with a more realistic synchronization method affected by asymmetric reception noise. The considered channels are thus simulated based on the IEEE 802.15.4a standard [91]. We aim to understand the effects of the previous quantization scheme (POS) on the random nature of the generated bits (i.e., intra-key randomness). The findings lead to the proposal of an extended quantization algorithm (HIST), which improves the random patterns of the keys, while maintaining an acceptable bit agreement ratio in noisy cases.

2.2.1 System model

The considered communication system includes two parties: Alice, denoted as A, and Bob, denoted as B. The key generation protocol consists of two phases : i) the half-duplex channel probing phase for signal acquisition and eventual processing and ii) the key extraction phase, which takes place in the application layer and deals with quantization and reconciliation. The signals received by Alice and Bob can be expressed as follows:

$$y_A(t) = (h_{BA} * p)(t) + w_A(t) \tag{2.5}$$

$$y_B(t) = (h_{AB} * p)(t) + w_B(t) \tag{2.6}$$

where $h_{AB}(t) = h_{BA}(t)$ are the reciprocal CIR, $p(t)$ is the transmitted pulse waveform with central frequency f_c and bandwidth B and $w_A(t)$, $w_B(t)$ are random processes of zero-mean Gaussian white noise with double-sided power spectral density $N_0/2$ and filtered in the band of the transmitted signal. The corresponding noise variance is therefore $\sigma_w^2 = N_0/2 \times 2B$. This model assumes that the pulse waveform is undistorted during propagation, irrespective of its associated multipath component.

The received signals are uniformly sampled at sampling frequency F_s and N samples are collected in an observation window whose overall time duration is $(N-1)/F_s$. The signal-to-noise ratio (SNR) is defined as the ratio between the mean power of the received noiseless signal and that of the noise:

$$\mathrm{SNR} = \frac{P_{sig}}{P_{noise}} = \frac{\frac{1}{N}\sum_{n=1}^{N}((h*x)[n])^2}{\sigma_w^2} \tag{2.7}$$

where $x[n]$ stands for the sampled version of $x(t)$.

Processing of received signals

An example of an IEEE 802.15.4a CM1 channel realization $h(t)$ and the analogue received signal are shown in Figure 2.1(a) - 2.1(b). On each side of the A-B link, the CIR can be estimated as $\hat{h}[n]$ after performing, for example, a straightforward frequency-domain deconvolution of the presumably known transmitted waveform $p[n]$ out of the received signal $y[n]$.

$$\hat{H}[l] = \frac{Y[l]}{P[l]} = \frac{H[l]P[l] + W[l]}{P[l]} \tag{2.8}$$

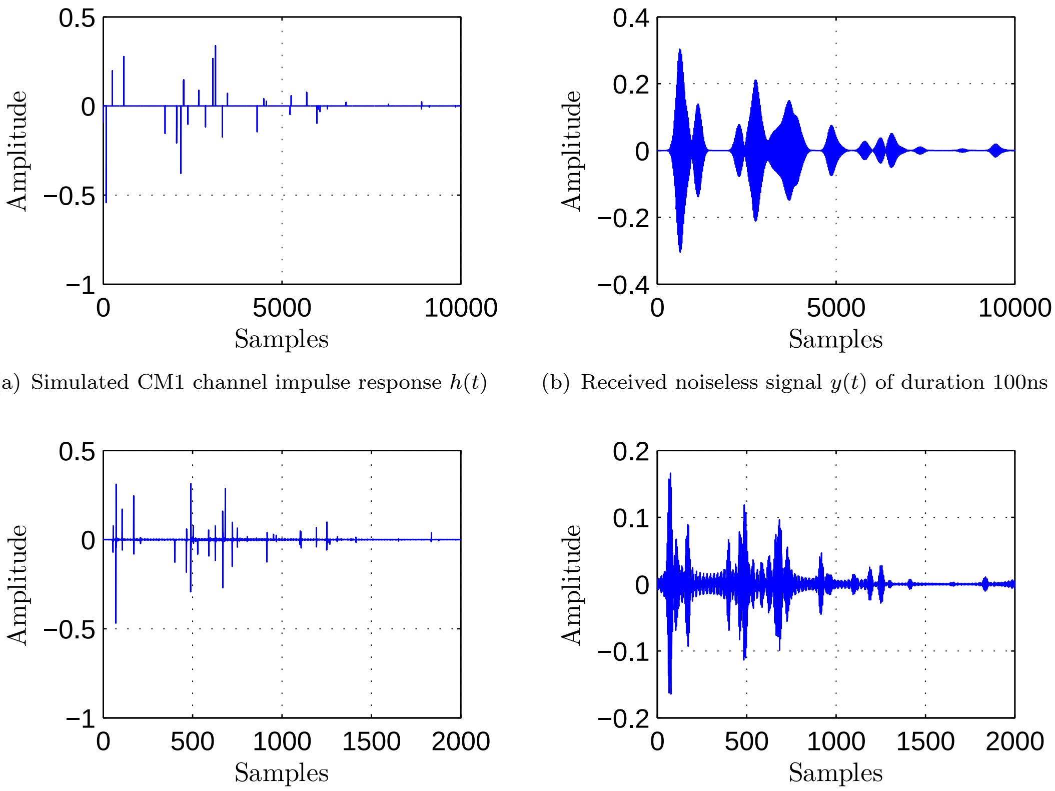
$$\hat{h}[n] = IDFT(\hat{H}[l]) \tag{2.9}$$

with $.[l]$ the Discrete Fourier Transform of $.[n]$.

In an ideal noiseless case, Tx deconvolution without filtering is possible and the channel estimation will tend towards a Dirac distribution like in Figure 2.1(c). In noisy cases, filtering is compulsory for Tx deconvolution [105]. This modifies the characteristics of the estimated signal: a simple brickwall filter equivalent to the pulse bandwidth would make the estimation $\hat{h}[n]$ tend to the form of a directly sampled signal (Figure 2.1(d)).

From the information-theoretic point of view, the bit extraction algorithms should use an estimated infinite-bandwidth CIR (i.e., pairs of times and amplitudes). Filtered signals like $y[n]$ involve correlated samples, which could lead to deterministic characteristics of the keys. However, channel estimation is a subject of research by itself and existing algorithms are constrained either from the sampling point of view (e.g., high resolution methods such as CLEAN [106]) or from the computational complexity point of view (e.g., compressed sensing algorithms [107]). Alternatives such as low-complexity algorithms, which can be implemented on low power devices [94], would be more limited in capturing the multipath richness (e.g., in terms of signal dynamics) and would be less suitable for long key generation. Channel estimation aspects are studied in more detail in Section 2.4.

In this study, the following input signals for quantization are considered: i) the estimated channel coefficients $\hat{h}[n]$ obtained by filtered frequency-domain deconvolution according to Eq. (2.9) (FD) and ii) the unprocessed samples $y[n]$ from direct sampling (DS). Both



(a) Simulated CM1 channel impulse response $h(t)$

(b) Received noiseless signal $y(t)$ of duration 100ns

(c) Noiseless deconvolution result without filtering

(d) Noiseless deconvolution result with filtering

FIGURE 2.1: Channel estimation by frequency deconvolution (noiseless case)

strategies allow the extraction of a sufficient number of bits because of the large number of available samples but they are also prone to degradations in randomness, which we will study next.

Synchronization considerations

The initial IR-UWB quantization algorithm [79] is evaluated on experimental traces acquired with a 2-9 GHz pulse generator and a 20 GHz real-time digital oscilloscope. The receiver's channel is synchronized with the pulse generator using a trigger signal. In this study, a more realistic synchronization method based on level-crossing (leading edge detection) will be used. This method aims at detecting the actual channel leading edge (first signal-carrying samples) and using it as a starting point for the temporal observation window. The used detection threshold (y_{th}) is set as a function of the noise variance. This means that the observation window on each side of the link will start at a time t_{start} defined by the level-crossing time ($t_{y(t)>y_{th}}$) and a fixed delay (Δ):

$$t_{start} = t_{y(t)>y_{th}} - \Delta \tag{2.10}$$

where Δ is employed in order to avoid missing significant samples from the signal.

2.2.2 Polarity-based quantization

The reference key generation method [79] [95] uses a dynamic threshold quantization for bit extraction and can be applied to input signals like $y[n]$ or $\hat{h}[n]$, denoted as $s[n]$ in the following. The algorithm, although adaptive with respect to the signal dynamics, does not depend on the temporal index (position) of the quantized sample and it is thus entitled POS for the rest of the document. Its phases are the following:

- estimation of the variance of the noise contained in the input quantization signal $s[n]$ (N_{lev}^2), which will be used to define a stopping rule for the bit extraction.
- bit extraction from $s[n]$:
  - compute the first thresholds ($i = 0$) for quantization: $L_0{}^+ = max(s[n]) > 0$ and $L_0{}^- = min(s[n]) < 0$.
  - at iteration i, apply the pos_i operator defined in Eq. (2.11) to the samples of $s[n]$ that cross the thresholds $L_i{}^-$ or $L_i{}^+$, memorize the extracted delay indexes n in a table $\mathcal{P}$ and the corresponding iteration step i in a table $\mathcal{I}$.

$$pos_i(s[n]) = \begin{cases} 1 & \text{if } s[n] \geq L_i^+ \\ 0 & \text{if } s[n] \leq L_i^- \end{cases} \tag{2.11}$$

  - update thresholds: $L_{i+1}^+ = L_i^+ - L_0{}^+/\delta$ and $L_{i+1}^- = L_i^- - L_0{}^-/\delta$, where δ is a scaling parameter of the protocol.
  - repeat the two last steps until the wanted length is reached or the noise level N_{lev} is approached within a guard interval depending on δ (e.g., $N_{lev}+L_0{}^+/\delta$). Even though, overall, several thresholds are used, the bits are extracted only from samples with amplitudes that cross the thresholds computed at the last iteration before stopping the extraction.
- public discussion involving the exchange of the index tables ($\mathcal{P}_A$ and $\mathcal{P}_B$), followed by the selection of the common indexes and their corresponding bits.
- key correction using a Reed-Solomon code to fix mismatching bits. For benchmarking purposes, we use the same code parameters as in [79]: total block length of 127 seven-bit symbols and encoded message length of 123 symbols. Therefore the maximum number of correctable symbols is 2.

The described bit extraction is adaptive with respect to the SNR. If the SNR is relatively high, a larger δ can be used; this corresponds to a lower stopping threshold and thus, to the extraction of more bits. Furthermore, given a desired key length, the bits corresponding to the most representative parts of the signal are given priority for extraction. Nonetheless, the quantization is based only on the polarity of the input signal with respect to the last threshold, independently for each sample/position. This simple polarity-based encoding applied to a deterministic waveform could possibly lead to regular patterns in the extracted bits. An extension of this bit extraction method, which takes into consideration the relationship between the amplitudes of different samples in order to break the regularity pattern, is presented in the next section.

2.2.3 Amplitude/delay-based quantization

The HIST extension of the POS algorithm concerns the quantization phase described in Section 2.2.2. The rest of the key generation protocol remains the same. The main idea is to exploit the absolute amplitude of the signal across the iterative extraction steps in order to improve the encoding process after applying pos_i. The observation window is split into N_b bins of a given sample length Δ_{bin}. These two parameters are considered public. The auxiliary bit operations of HIST, denoted by the operator $hist_b$ from Eq. (2.12), are performed independently for each bin $b \in \{1, 2, ..N_b\}$.

$$hist_b(pos_{i_k}(s[n])) = \begin{cases} pos_{i_k}(s[n]) & \text{if } k = odd \\ \neg pos_{i_k}(s[n]) & \text{if } k = even \end{cases}, i_k \in \mathcal{I}_b \tag{2.12}$$

where $\neg$ is the negation operator and $\mathcal{I}_b = \{i_1, i_2, ...\}$ is the set of ordered iteration steps (indexed by k) that correspond to the extracted bits in bin b. The result consists in a dynamic encoding of the samples depending on the threshold at which they are extracted.

The new algorithm is illustrated on an example with one arbitrary bin (see Table 2.1 and Figure 2.2). In this example, the indexes 21, 23, 24, 26 and 29 belong to the same bin if a bin contains 10 samples. The iteration steps corresponding to the bits extracted by POS are also memorized ("Iteration steps" row). Even though they can be available in both algorithms, the values of the iteration steps are not used for POS.

The result of applying this kind of post-processing is equivalent to a dynamic encoding of the negative and positive amplitudes as a function of the absolute value and the delay index or bin of the sample. With POS, the negative amplitudes are always encoded as "0" and the positive amplitudes as "1", but with HIST the encoding convention varies across the observation window. As the samples are processed following the decreasing

Table 2.1: Illustration of the bit extraction algorithms (HIST vs. POS)

Bits extracted by POS	1	**0**	0	1	**1**
Delay indexes of POS bits ($\mathcal{P}$)	21	23	24	26	29
Iteration steps # ($\mathcal{I}_b$)	10	12	10	17	12
Indices of the steps (k)	1	2	1	3	2
Final bits (HIST)	1	**1**	0	1	**0**

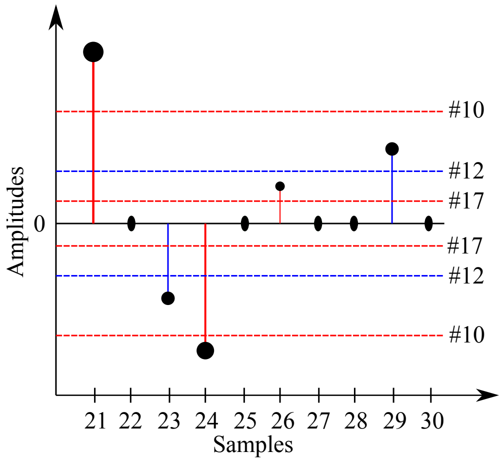


Figure 2.2: Illustration of HIST algorithm in one bin

order of their modules, nearby samples (i.e., belonging to the same bin) could have different extraction histories, meaning that they could have been extracted at different threshold levels. This type of information is transparent for POS but taken into account by HIST. Note that, asymptotically, HIST tends to POS when the bin length, Δ_{bin}, tends to the temporal resolution used for signal acquisition, meaning that each bin contains only one sample.

2.2.4 Performance evaluation

We assess the performance of the initial bit extraction algorithm (POS) and the proposed extension (HIST) with standard indoor channels. The advantages in terms of randomness of POS with respect to one-level quantization [58] or guard-band quantization [60] have already been observed using experimental traces [95]. Here, the data are generated from statistical channel models, which account for the high density of multipath profiles in indoor IR-UWB channels. We consider in particular CM1 and CM2, according to the IEEE802.15.4a standard [91]. The initial pulse characteristics (i.e., central frequency of 4.5 GHz and 1 GHz bandwidth) are set so as to comply with both the FCC spectral mask and one of the mandatory bands of the 802.15.4a band-plan, but also with recent low-power implementations of IR-UWB transceivers [94]. The simulated sampling frequency has the same order as that of the experimental tests [79] [95] (20 GHz[2]) and the observation window has a duration of 100 ns. The following performance metrics have been retained:

- *key agreement* ratio defined as the success rate of key establishment, i.e., when final bits (after reconciliation) on both sides of the link are identical (computed over different IR-UWB channel realizations); an extension of this metric is the *min. 64-bit key agreement* defined as key agreement ratio with an additional constraint on the minimum key length (64 bits).
- *mean key lengths* computed for successful keys only (i.e., after reconciliation).
- *mean bit agreement* ratio between the two parties after the extraction phase and before reconciliation.
- *mean number of different bits* between POS encoding and HIST encoding (before reconciliation).
- *random nature* of the extracted keys, tested with bit periodicity tests and with the NIST test suite [102](before reconciliation).

The algorithms are first of all evaluated in a noiseless scenario, in which generated bits at A and B are always identical (i.e., the key agreement ratio is 1). The goal is to evaluate their intrinsic randomness properties, which depend on the transmitted waveform and on the multipath channel. Next, the reciprocity properties are discussed as a function of the SNR.

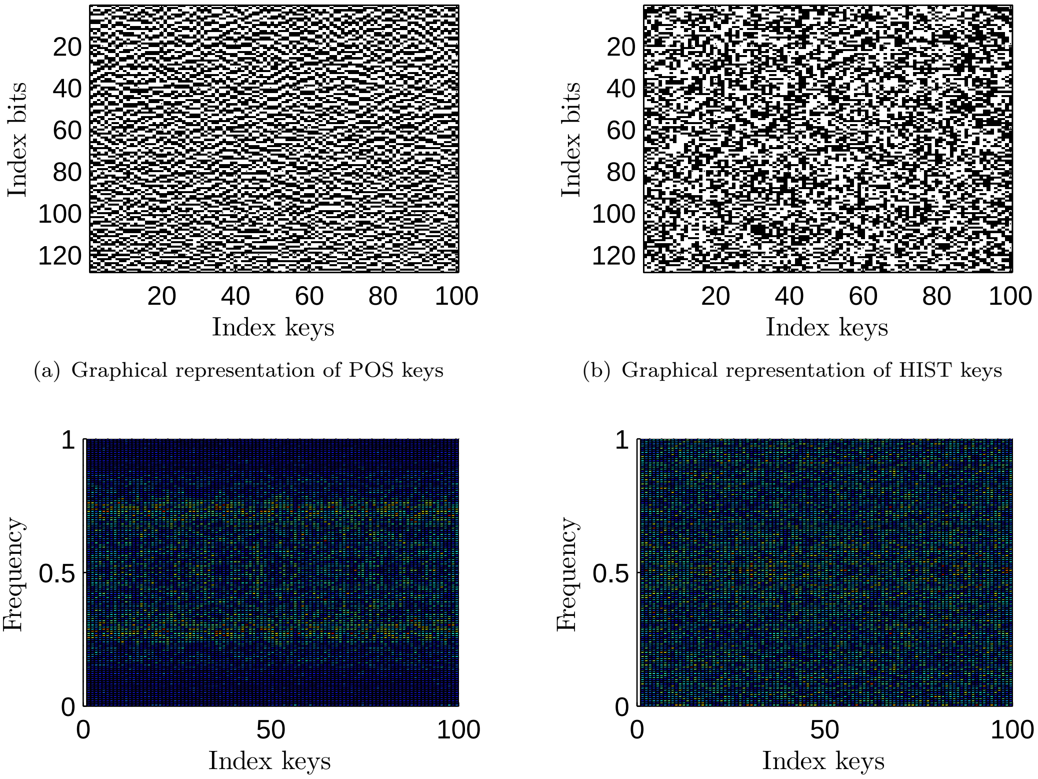
Random aspect evaluation with noiseless signals

Preliminary tests are performed on keys of minimum 128 bits generated with an initial scenario (CM1 channels, FD strategy). We analyze graphically the impact of the deterministic waveform on the generated bits and then, we apply the NIST tests. In the end, we show that the conclusions can be generalized to other scenarios.

First of all, the periodic nature of the extracted bits is tested using the Fast Fourier Transform (FFT) on each key. Figure 2.3 shows a regularity pattern for POS keys in both bit and frequency domains. The dominant frequency component at approximately 0.27 is equivalent to a bit pattern period (BPP) of $1/0.27 = 3.7$ bits. The quantization

[2] Although this sampling frequency is too restrictive for embedded devices, we maintain it for comparison purposes.

diversity of HIST seems to solve this issue because HIST keys do not show any particular dominant frequency.



(a) Graphical representation of POS keys

(b) Graphical representation of HIST keys

(c) Frequency representation of POS keys

(d) Frequency representation of HIST keys

FIGURE 2.3: Periodic nature of CM1 extracted keys using FD strategy. Parameters: $\delta = 10$ (POS and HIST) and $\Delta_{bin} = 10$ samples (HIST)

The regular bit pattern in POS is a consequence of the static encoding applied to the polarity of the waveform. The bits imitate the waveform's oscillations, which are determined by its central frequency. Therefore, when changing f_c from 4.5 GHz to 2.5 GHz, HIST presents the same approximately white frequency aspect as before while POS shows a dominant frequency of 0.158, i.e., a BPP of 6.33 bits (obtained by averaging the FFT representation over 500 keys). The BPP is actually proportional to the central frequency as shown in Eq. (2.13). Moreover, the BPP for POS approximately follows the analytical expression from Eq. (2.14) obtained by a geometrical computation on the sine wave and assuming that the extraction process stops at the last threshold ${L_0}^+/\delta$ for every key (Appendix A). This is actually an upper bound of the real BPP and its values are 4.2 bits for 4.5 GHz and 7.4 bits for 2.5 GHz.

$$\frac{6.33}{3.7} = 1.71 \approx 1.8 = \frac{4.5}{2.5} \tag{2.13}$$

$$uBPP = 2F_s(\frac{1}{2f_c} - 2\frac{\arcsin(1/\delta)}{2\pi f_c}) \tag{2.14}$$

Secondly, the use of the NIST statistical suite for pseudo-random number generators [102] allows the investigation of various characteristics of the keys such as: bit frequency (frequency test or block frequency test), random walks (cumulative sums test), frequency of oscillations between different bits (runs test and longest run within a block test), and frequency of all possible overlapping bit patterns of a certain length (serial test). Two of the tests (cumulative sums and serial) have 2 variants depending on the sense of the sequence processing: direct or reversed. It is important to keep in mind that such tests cannot tell whether a sequence is random. They can only show defects in the random nature by pointing out when certain keys are prone to a deterministic behavior. As they employ statistical methods, the tests have to be realized on a large number of fairly long keys.

The NIST suite is built upon the principles of statistical hypothesis testing, in which the null hypothesis is that a sequence is random, given the particular randomness feature. Each test computes a relevant randomness statistic and an associated “p-value” representing the probability that a perfect random number generator produced a sequence less random than the tested sequence [102]. A key is said to pass the test when its computed p-value is higher than 0.01 (a common value for the level of significance of the test, i.e., the probability of rejecting the null hypothesis when it is true). Nonetheless, a given set of keys pass a certain test if a very large proportion of the keys pass the test (pass rate) and if the p-values of the keys are uniformly distributed in the (0,1) interval. The need of p-value uniformity is explained with an example: at the limit, a constant p-value of 1 in the frequency test represents a 100% pass rate but also an equal proportion of ones and zeros, which means that from this point of view, the key is not entirely random.

For the considered initial scenario, POS passes the frequency and random walk tests, but fails the oscillations and pattern tests for almost all the keys. Moreover, it does not respect the uniformity criterion on the p-values for any of the tests. On the contrary, HIST shows an improvement in the uniformity of p-values over all tests and manages to pass them with a minimum proportion of 94%. These results reinforce the trends observed in the key-frequency diagrams from Figure 2.3, which give a complementary representation of the statistical pattern test.

Next, we present a synthesis of four randomness tests for the initial scenario (CM1, f_c = 4.5 GHz, B = 1 GHz, FD strategy) and for several extensions using various channel models, central frequencies and bandwidths, or with a different input signal (DS

Table 2.2: Statistics on NIST randomness tests for several scenarios: **A**(CM1/ [fc;B]=[4.5;1] GHz/FD) & **B**(**CM2**/ [fc;B]=[4.5;1] GHz/FD) & **C**(CM1/ [**fc**;B]=[2.5;1] GHz/FD) & **D**(CM1/ [fc;**B**]=[4.5;0.5] GHz/FD) & **E**(CM1/ [fc;B]=[4.5;1] GHz/**DS**)

Scenario		Metric	Block Freq.	Cum. Sums(1)	Runs	Serial(1)
A	POS	%	100%	100%	3%	0
		$10^3 Var$	3.33	10.93	NA	NA
	HIST	%	98%	98%	94%	94%
		$10^3 Var$	81.96	90.17	89.62	77.52
B	POS	%	100%	100%	2%	0
		$10^3 Var$	3.86	15.10	NA	NA
	HIST	%	97%	96%	93%	93%
		$10^3 Var$	87.45	96.59	96.23	83.99
C	POS	%	100%	100%	59%	0
		$10^3 Var$	5.93	15.84	15.26	NA
	HIST	%	99%	99%	98%	97%
		$10^3 Var$	83.53	86.85	86.91	83.62
D	POS	%	100%	100%	2%	0
		$10^3 Var$	3.51	11.58	NA	NA
	HIST	%	96%	95%	89%	91%
		$10^3 Var$	85.60	95.79	83.78	79.90
E	POS	%	100%	100%	1%	0
		$10^3 Var$	3.06	10.35	NA	NA
	HIST	%	98%	97%	90%	92%
		$10^3 Var$	80.06	89.18	90.41	78.69

strategy). The algorithm parameters, the key lengths and the number of tested keys are the same as for the previous tests. The condensed metrics are the pass rate (%) and the variance of the p-values used for representing the uniform character of the repartition (Var). If the variance is small, this would correspond to a non-uniform repartition of p-values meaning that the extraction algorithm is prone to produce keys within certain p-values, i.e., with a certain deterministic aspect, regardless of the channel.

The results are presented in Table 2.2. The conclusions of the initial scenario can be reported to all the other scenarios. HIST pass ratios in the runs and serial tests slightly decrease when using a narrower bandwidth pulse (scenario D). A supplementary test confirms that the pattern drawback found by the serial test (pass rate of 91%) is related to the bin size, which becomes comparable to the pulse duration in scenario D: by increasing the bin size to 20 samples, the serial(1) pass rate for HIST becomes 96%, meaning that the pattern defect is a limitation of the encoding procedure itself, which is, after all, a deterministic algorithm. Overall, from the point of view of these tests, the best scenario for both POS and HIST would be one using the FD strategy as well as a pulse with lower central frequency (i.e. with less oscillations), but large bandwidth.

In conclusion, the proposed extension, HIST, passes all the tests in most cases and solves

Metric name	POS	HIST
Key agreement	99.9%	24.5%
Mean key length	123	116
Min. 64-bit key agreement	98.2%	23.5%
Mean bit agreement	99.9%	93.4%
Mean number of changed bits	-	46.8

Table 2.3: Performance metrics for the extraction of keys of max. 128 bits from CM1 channels using the DS strategy at SNR=30 dB. Parameters: $\delta = 10$ (POS and HIST) and $\Delta_{bin} = 10$ samples (HIST).

the problem of the p-value repartition with the trade-off of a more complex quantization scheme that can become less effective in practical SNR conditions.

Reciprocity evaluations with noisy signals

In the case of noisy signals, the reciprocity is affected by both noise and synchronization errors (i.e., different starting points of the observation/quantization window). When the input signal for quantization is a noisy estimate of $h(t)$, the reciprocity becomes even lower because of the intermediate processing step (i.e., deconvolution). As the random character of the generated keys does not vary considerably between the FD and the DS strategies, we investigate the performance of the quantization of DS signals for different values of the algorithm parameters δ and Δ_{bin}. Ideally, the parameters' values should be as high as possible in order to generate many bits (large δ) with good random properties (large Δ_{bin}); specifically, a large Δ_{bin} implies many detected thresholds in a bin, which leads to diversified encoding according to HIST.

Preliminary tests using the initial scenario as in the previous section are performed for multiple 802.15.4a CM1 channel realizations. The results are summarized in Table 2.3. Even at large SNR, significantly lower key agreement rates are achieved for HIST (24.5 % in comparison to 99.9% for POS). This is mainly due to the more complex encoding of amplitude information, which is more sensitive to both noise and synchronization errors and leads to a lower bit agreement before reconciliation. A mean bit agreement of 93.4% cannot be entirely corrected by the chosen Reed-Solomon code; an alternative would be to increase the correction capacities of the Reed-Solomon code with the drawback of having a higher bit agreement rate with an attacker that would also benefit from the error correction scheme. However, if privacy amplification techniques were applied, this issue would be solved with the drawback of obtaining less bits.

Figure 2.4 shows the limitations of HIST in terms of generating successful 64-bit keys especially for high δ and Δ_{bin} values. The lower key agreement ratio for HIST is mainly due to the more complex encoding using the threshold information, but also to the key length constraint. POS does not experience the same phenomenon because it only encodes the polarity of the waveform, making it more robust to noise, but with a weaker random aspect.

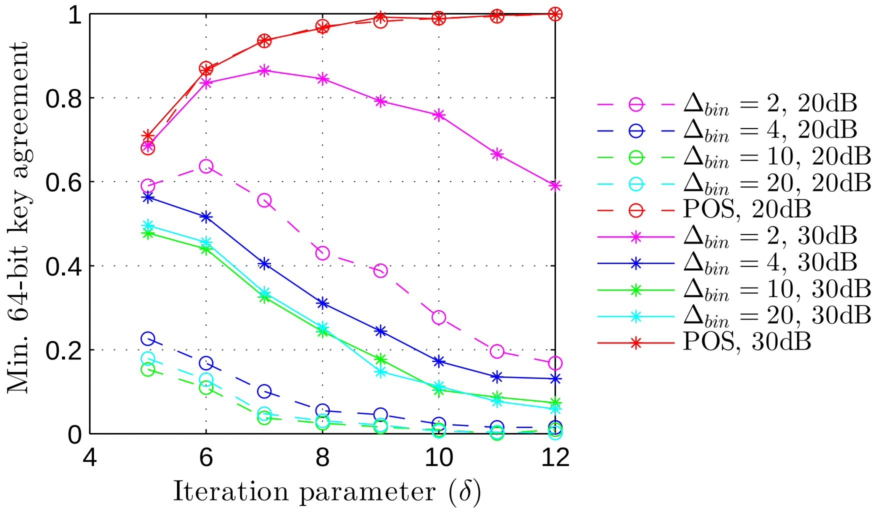


FIGURE 2.4: Key agreement for POS and several variants of HIST(Δ_{bin}(samples)), SNR=20 dB and SNR=30 dB

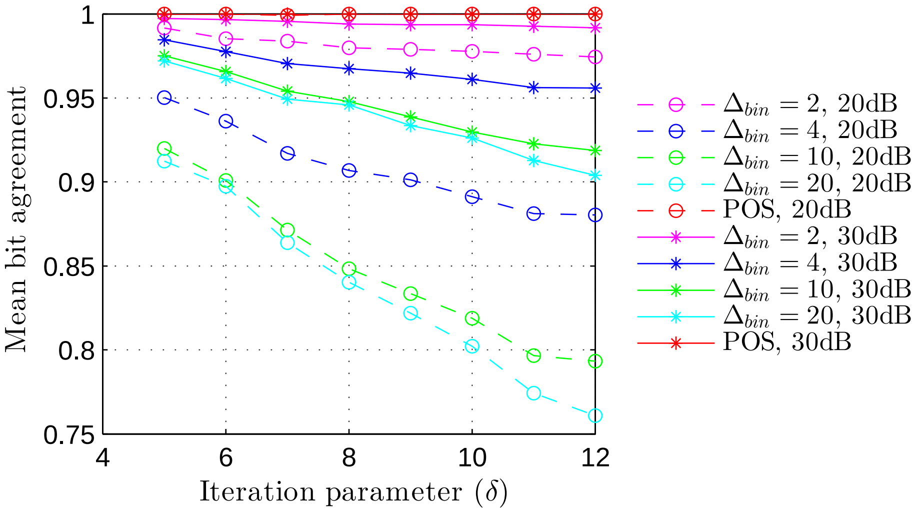


FIGURE 2.5: Bit agreement for POS and several variants of HIST(Δ_{bin}(samples)), SNR=20 dB and SNR=30 dB

However, HIST maintains a reasonably high bit agreement ratio (Figure 2.5) while providing a dynamic amplitude encoding. The number of differently encoded bits between POS and HIST, employed to measure the diversity offered by HIST, also remains relatively high even for small Δ_{bin} values (e.g., between 20 and 85 for $\Delta_{bin} = 4$ samples).

The findings about HIST can be extended to lower SNR values such as 10 dB or 5 dB, which are not directly exploitable because of the restrictive key agreement ratio, but remain, nonetheless, a good source of common bit extraction (e.g., at $\delta = 12$ and $\Delta_{bin} = 4$, the mean bit agreement ratio is 80% for 10 dB and 73% for 5 dB).

2.2.5 Summary

In this section, we have studied several randomness aspects of physical layer key generation using directly sampled IR-UWB signals or channel estimations obtained by frequency deconvolution. We suggest a new bit encoding algorithm (HIST) based on the delay and amplitude information of each sample (i.e., bins and thresholds) in order to improve the random patterns of the generated keys, irrespective of any reciprocity constraints.

HIST is an invertible overlay to the adaptive-threshold quantization algorithm developed in previous studies [79] [95]. Its main advantage is to improve the random patterns of the keys suffering from deterministic characteristics of the input signal, such as the pulse waveform that contributes to the correlation between samples and therefore between bits. Finally, we test the algorithms over multiple 802.15.4a channel realizations and find that an acceptable mean bit agreement ratio can be obtained for the proposed solution over a range of medium to high SNR values.

NIST tests results also show that a pulse with a lower central frequency and a larger bandwidth should be preferred for randomness considerations. The advantage of a larger bandwidth is obvious because it offers higher channel resolution capabilities and thus "more" information. Employing a lower central frequency or sampling directly the envelope of the signal means having to lower the sampling frequency as well, which will lead to shorter keys. These aspects are however common to both initial algorithm and the proposed solution. In this context, we argue that HIST provides an improved encoding mechanism for high-sampling scenarios in which key length is a priority.

2.3 Keys from CIRs: reciprocity-randomness trade-off

As explained in the previous section, secret keys are preferably extracted from the channel coefficients that should be estimated after the channel probing phase. Thus, based on simulations, we examine the behavior of various generic quantization schemes applied to noisy channel estimates of an IR-UWB channel.

While in high-resolution or rate-distortion theory [101] the "quantization" aims to reduce a certain distortion metric and the number of bits used to represent the quantized signal, in the key generation context a different trade-off arises. In former investigations on secret key quantization [30], a new extraction algorithm and a suitable reconciliation scheme have been proposed in order to find a good trade-off between key length and reciprocity (i.e., the number of matching bits between the two parties). Nevertheless, the problem can have an additional optimization axis : the probability of occurrence of each codeword of the codebook used for quantization. These probabilities can be set to their optimal values, which correspond to equally distributed codewords [30]. In this study, we will set the length of the codewords (relevant for, e.g., applications with fixed length constraints) and investigate the trade-off between reciprocity, as a measure of robustness, and codeword diversity, as a measure of randomness.

2.3.1 System model

The estimated CIR at each user, $\hat{y}^u(t)$, is a noisy version of the standard multipath model, considering noise on the amplitudes of the channel coefficients:

$$\hat{y}^u(t) = h(t) + w^u(t) \tag{2.15}$$

$$h(t) = \sum_{k=1}^{K} x_k \delta(t - \tau_k) \tag{2.16}$$

$$w^u(t) = \sum_{k=1}^{K} w_k^u \delta(t - \tau_k) \tag{2.17}$$

where $u \in \{A, B\}$, K is the number of estimated multipaths, x_k is the channel tap (or equivalently, the small-scale amplitude) of the k-th multipath, w_k^u is the estimation noise at user u for the k^{th} multipath and τ_k is the excess delay of the k^{th} multipath. The noise is generated from a synthetic noise model, meaning it does not correspond to the noise issued from a real channel estimator.

In order to generate a binary sequence, we apply a quantization algorithm to the K samples $y_k^u = x_k + w_k^u$ considered as realizations of the independent RVs $Y_k^u = X_k + W_k^u$, with X_k the RV modeling the channel taps at excess delay τ_k and $W_k^u \sim \mathcal{N}(0, \sigma_0^2)$ the RV modeling the hypothetical estimation noise at excess delay τ_k.

The employed guard-band quantization algorithm, which is allowed to vary as a function of the excess delay, is defined by :

- a binary Gray codebook of 2^b valid binary codewords and one invalid codeword corresponding to the guard-band interval : $\mathcal{C} = \{c_j \in \{0,1\}^* | len(c_j) = b, j \in \{-2^{b-1}, \ldots - 1, 1, \ldots, 2^{b-1}\}\} \cup \{c_0\}$.
- the respective inferior and superior quantization thresholds delimiting each quantization cell : $\theta^{inf} = \{\theta_j^{inf} | j \in \{-2^{b-1} \ldots 2^{b-1}\}, \theta_0^{inf} = -\sigma_0, \theta_{-2^{b-1}}^{inf} = -\infty\}$ and

$\theta^{sup} = \{\theta_j^{sup} | j \in \{-2^{b-1} \ldots 2^{b-1}\}, \theta_0^{sup} = \sigma_0, \theta_{2^{b-1}}^{sup} = \infty\}$. The quantization cells are considered adjacent ($\theta_{j-1}^{sup} = \theta_j^{inf}$). For simplicity, we use $\theta_j = \theta_j^{inf}$ for $j \geq 1$, $\theta_j = \theta_j^{sup}$ for $j \leq -1$ and $\theta = \theta^{inf} \cup \theta^{sup}$ for obvious symmetry reasons. This leads to the representation in Figure 2.6.

- a quantization rule : $q_{key}(y) = c_j$ if $y \in [\theta_j^{inf}, \theta_j^{sup})$.

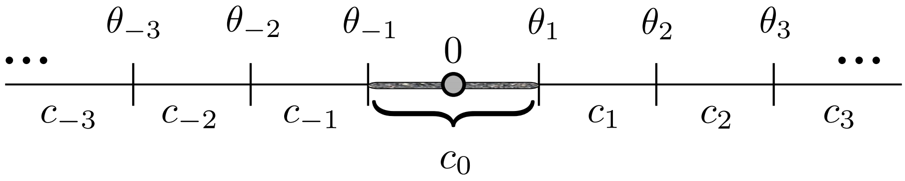


FIGURE 2.6: Quantization thresholds

In the following, we set the length of the codewords to a fixed value b and we define two types of reciprocity and diversity metrics: inter-key metrics (computed at a fixed excess delay τ by averaging over noise and channel realizations) and intra-key metrics (computed for one channel realization by averaging over the present channel taps). The inter-key metrics characterize the robustness and randomness properties of an "average" codeword generated at a given excess delay and are employed for the design of delay-adaptive quantization schemes (Sections 2.3.3). The intra-key metrics can measure the performance of a given quantization scheme applied to a single channel estimation (Section 2.3.4).

Inter-key reciprocity and diversity

The small scale amplitudes $|x_k|$ are modeled using a m-Nakagami distribution in the IEEE 802.15.4a standard [91]. In order to have tractable formulas and to take into account the sign of the channel tap, we adopt a Gaussian mixture approximation for the RV X_k. This model should represent the equally distributed positive and negative multipath components. As we are interested in the average behavior at a certain excess delay τ, we will denote by n the index of this uniformly sampled excess delay and we will focus on the typical channel tap at τ, namely X_n. According to this convention, k represents the index of the existing paths in one channel realization whereas n is the index corresponding to the sampled excess delay irrespective of any channel realization.

$$X_n \sim \frac{1}{2}\mathcal{N}(-\mu_n, \sigma_n^2) + \frac{1}{2}\mathcal{N}(\mu_n, \sigma_n^2) \tag{2.18}$$

with $\mu_n = \mathbb{E}[|X_n|]$ and σ_n, the standard deviation of $|X_n|$, representing channel parameters computed empirically from simulations using the 802.15.4a channel model CM1 and a reference acquisition window. The mean and the standard deviation of the channel taps are decreasing functions of the excess delay index n, similarly to the power delay profile (PDP).

The performance metrics per valid codeword at fixed excess delay n are:

- mean Hamming distance between two valid codewords C_n^A and C_n^B (to be minimized):
$$HD(n,\theta) = \mathbb{E}_{X_n}[\mathbb{E}_{W_n}[hd(C_n^A, C_n^B)|C_n^A, C_n^B \neq c_0]] \tag{2.19}$$

- spread of valid codewords (to be minimized):

$$\begin{aligned} CS(n,\theta) &= std(\{P_n^{u,1}, P_n^{u,2}, ..., P_n^{u,2^b}\}) \qquad (2.20)\\ P_n^{u,j} &= \mathbb{P}(C_n^u = c_j|C_n^u \neq c_0) \qquad (2.21)\end{aligned}$$

with C_n^u the codeword issued from the quantization on b_n bits of the n^{th} sample at user u, $hd(c_1, c_2) = |\{k|(c_1)_k \neq (c_2)_k\}|$, $(.)_k$ the k^{th} bit of a codeword.

We also define a tunable scalar cost function used to study the trade-off between reciprocity and diversity by varying the relative importance of the components using the weight λ according to Eq. (2.22). The reciprocity component is normalized with respect to a fixed number of bits $b_n = b$ and the diversity with respect to its maximum value (i.e., the standard deviation of a repartition with one codeword appearing with probability 1).
$$L(\lambda, n, \theta) = (1-\lambda)\frac{HD(n,\theta)}{b} + \lambda\frac{2^b \cdot CS(n,\theta)}{\sqrt{2^b - 1}} \tag{2.22}$$

The above-mentioned reciprocity (HD) and diversity (CS) metrics depend on the binary codebook, on the statistical model of Y^u as well as on the quantization thresholds. For our investigations, we arbitrarily choose a Gray codebook on $b_n = 3$ bits assuming that X_n follows the distribution in Eq. (2.18) and that it is independent of W_n. The details of the computation and the final expressions are provided in Appendix B. The inter-key reciprocity-randomness trade-off is studied in Sections 2.3.2-2.3.3.

Intra-key reciprocity and diversity

The same reciprocity and diversity metrics can be defined per channel realization by averaging over the existing paths in the given channel (Eq. (2.23)-(2.24)). The result can be subsequently averaged over channel realizations. These definitions are simulation-oriented and do not require an analytical model for the distribution of the channel taps.

$$
\begin{aligned}
HD_1(\theta) &= \mathbb{E}_k[hd(C_k^A, C_k^B)|C_k^A, C_k^B \neq c_0] && (2.23)\\
CS_1(\theta) &= std(\{P^1, P^2, ..., P^{2^b}\}) && (2.24)
\end{aligned}
$$

with P^j representing the probability of occurrence of codeword c_j within the codewords generated from one channel realization.

The intra-key reciprocity-randomness trade-off is investigated in Sections 2.3.4-2.3.5.

2.3.2 Quantization thresholds

This subsection presents the various threshold computation methods that have been employed in the present study.

Uniform and non-uniform quantization

Several quantization techniques with the same guard-band intervals (i.e., $\theta_0^{inf} = \theta_{-1} = -\sigma_0$ and $\theta_0^{sup} = \theta_1 = \sigma_0$) are considered:

- uniform quantization (UQ) : the quantization intervals for each codeword have a fixed width Δq.
- quantile-based quantization (QQ): the quantization thresholds are computed *a priori* in order to achieve a specific codeword repartition. A uniform repartition of all the codewords (*eq*: $P_n^{u,j} = 1/9$ for any j, n) and a non-uniform repartition (*ineq*: $P_n^{u,-1} = P_n^{u,1} = 40\%$ for any n) are given as examples.
- companding-based quantization (CQ): inspired by the work on companders [108], the initial samples are transformed with an exponential function and the output is quantized uniformly. The employed "expanding" function is the inverse of the μ-law function [108] with range $[-1, 1]$:

  $$exp(y_n) = sgn(y_n)\frac{((1+\mu)^{|y_n|} - 1)}{\mu} \qquad (2.25)$$

  with $sgn(.)$ the sign function and μ an arbitrary parameter. The result is a non-uniform quantization algorithm with gradually decreasing quantization intervals, which are also symmetric with respect to 0. Accordingly, the quantization intervals are expected to be larger for the values y_n with lower signal-to-noise ratio (SNR) and the reciprocity improved.

To sum up, the UQ and CQ methods have fixed quantization thresholds, while the QQ thresholds are dynamically computed as a function of the desired codeword distribution and the excess delay n.

Optimization of the quantization thresholds

The optimization of a multi-objective function $\{HD(\theta), CS(\theta)\}$ can be achieved using a scalarization technique (O_1) or a constraint-based method (O_2) [109].

$$\begin{aligned} (O_1): \quad \underset{\theta}{\text{minimize}} \quad L(\lambda,\theta) \quad \text{s.t} \quad & \theta_j \leq \theta_{j+1},\ j \in \{1,2,3\} \\ & \theta_j = -\theta_{-j},\ j \in \{2,3,4\} \end{aligned} \tag{2.26}$$

$$\begin{aligned} (O_2): \quad \underset{\theta}{\text{minimize}} \quad HD(\theta) \quad \text{s.t} \quad & \theta_j \leq \theta_{j+1},\ j \in \{1,2,3\} \\ & \theta_j = -\theta_{-j},\ j \in \{2,3,4\} \\ & CS(\theta) \leq c \end{aligned} \tag{2.27}$$

with c an *a priori* constraint parameter on the diversity metric.

2.3.3 Performance evaluation I

Reciprocity-diversity trade-off

We analyze the expected trade-off between reciprocity and codeword diversity computed using Eq. (2.19)-(2.20) and the channel tap model described in Section 2.3.1. The parameters of the Gaussian mixture model for the channel taps X_n are computed empirically from IEEE 802.15.4a simulations over a 50 ns acquisition window: $\mu_n \in [0.075, 0.44]$, $\sigma_n \in [0.05, 0.18]$. The estimation noise is fixed at a reference value of $\sigma_0 = 0.1$. This corresponds to a variation of the SNR from 13.5 dB to -0.9 dB between the beginning and the end of the acquisition window.

At fixed CS (QQ scheme), the mean Hamming distance increases almost linearly with the excess delay because of the degradation in the SNR (Figure 2.7). The UQ and CQ quantization algorithms achieve an optimal HD at a characteristic excess delay, which corresponds mainly to an asymptotically high value of CS (Figure 2.8). The UQ and CQ

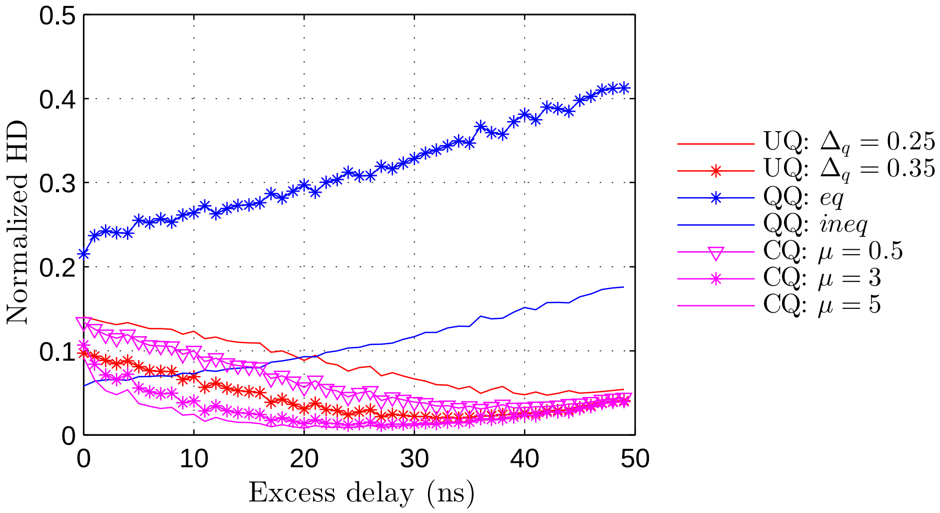


FIGURE 2.7: Inter-key reciprocity cost for various quantization schemes

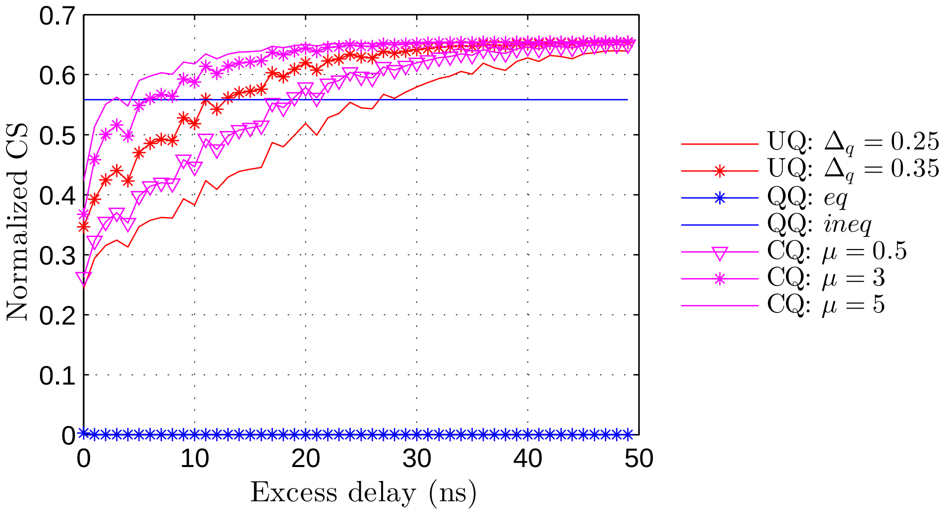


FIGURE 2.8: Inter-key diversity cost for various quantization schemes

behavior is linked to the actual values of the quantization thresholds and to the evolution of the signal dynamics with the excess delay (its mean and standard deviation).

In conclusion, on the one hand, fixing the codeword diversity like in the QQ case requires specifying the distribution of each codeword, which is unnecessary because, from the application point of view, the interest lies in the spread metric. On the other hand, using strategies like UQ and CQ gives no control on the final trade-off and no indication of the optimality of the scheme. To respond to these constraints, in the next section we present the results of an optimization study on the quantization thresholds.

Optimized quantization thresholds

Given the symmetry of the problem, the optimization can be performed on a three-dimensional variable $\theta = [\theta_2, \theta_3, \theta_4]$, which contains the positive thresholds corresponding to a quantization on 3 bits (i.e., 8 valid codewords in total including 4 codewords corresponding to positive values of y_n). The optimization routines are based on Eq. (2.19), (2.20), (2.22) and on the *a priori* measured channel statistics (μ_n, σ_n). Both problems are solved with a nonlinear constrained optimization function using the interior-point algorithm (*fmincon* solver from MATLAB®). In the case when the optimization results vary with the initialization values of the solver (e.g., for O_1), the final thresholds are the ones achieving the minimum cost over 100 trials starting from uniformly random initial values.

For illustration purposes, we use two SNR values corresponding to different excess delays in a real channel ($\text{SNR}_1 = 13.5$ dB and $\text{SNR}_2 = 7.3$ dB). The obtained thresholds with O_1 are shown in Figure 2.9.

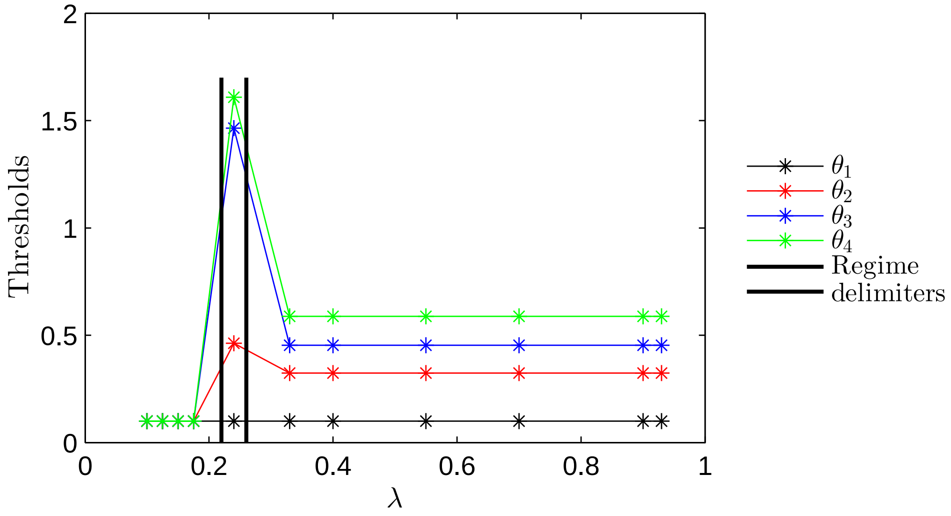


FIGURE 2.9: Illustration of the optimal thresholds O_1 for SNR_1

For both SNR values, we observe a sudden variation of the threshold values from small to large λ which corresponds to a bi-static behavior of the general cost function $L(\theta)$: at small λ values, the reciprocity is optimized (i.e., only one codeword is generated) while at high λ the codeword diversity is optimized (i.e., uniformly distributed codewords). For SNR_1, the optimization routine is able to find an intermediate point corresponding to the statistical domination of two codewords, but this is not the case for SNR_2. This means that O_1 is not efficient to look for a good compromise solution between reciprocity and diversity, especially since there is no direct *a priori* link between λ and the relative

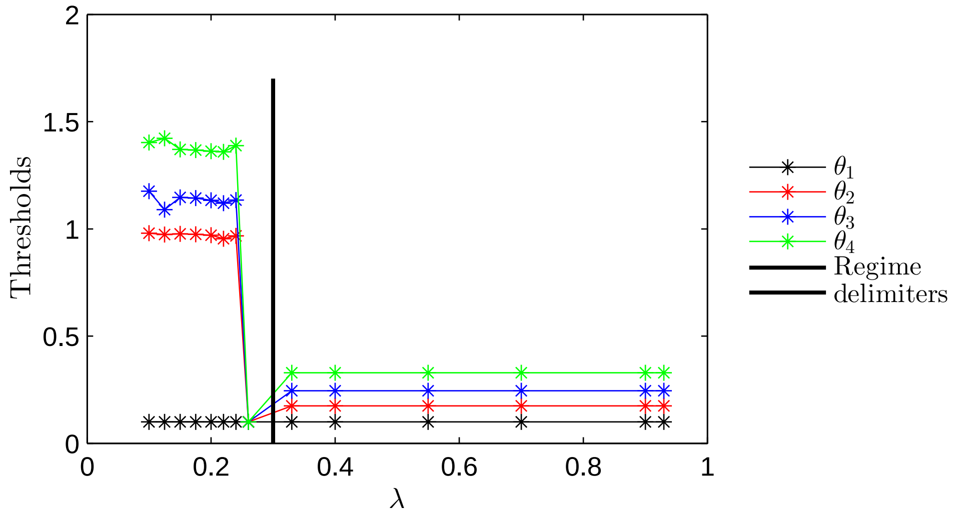


FIGURE 2.10: Illustration of the thresholds O_1 for SNR$_2$

importance of the objective functions [109], which makes it difficult to solve the multi-objective optimization problem.

The search for Pareto-optimal points for the multi-objective optimization problem can be solved by O_2, which will give access to weak Pareto-optimal solutions. In contrast to O_1, O_2 has the capability of finding points of the non-convex region of the Pareto-optimal front [109]. In Figure 2.11, we show the optimal points found by O_1 with different λ values and those found by O_2 with different constraint (c) value, as well as the costs achieved by the previous quantization schemes.

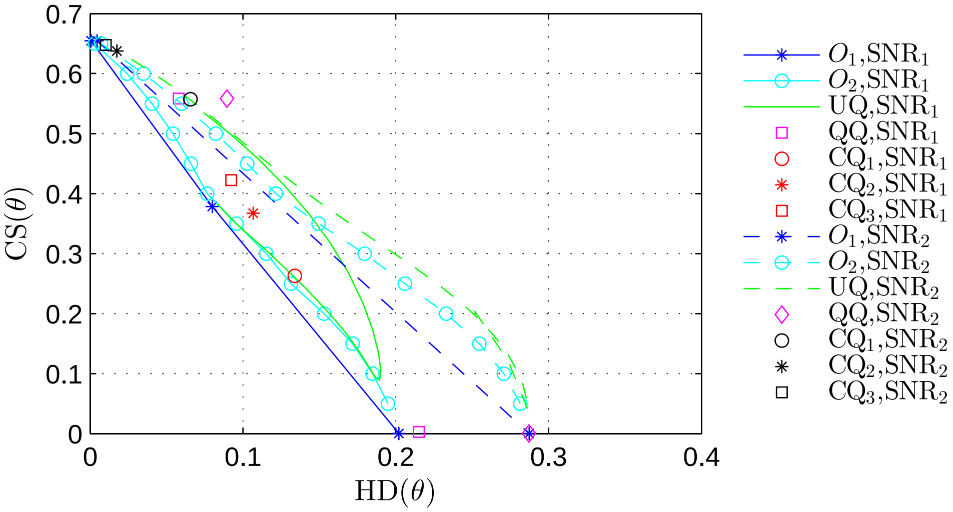


FIGURE 2.11: Inter-key costs for various quantization schemes

In the case of O_1, we can retrieve the points corresponding to the different regimes (3 for SNR_1 and 2 for SNR_2). O_2 finds points in the non-convex parts of the front, including one close to the intermediate regime of O_1, which visually corresponds to an inflexion point of the front (intersection of O_1 and O_2 for SNR_1).

All the other methods (CQ, QQ, UQ) achieve usually higher costs. The uniform quantization (UQ) parameterized with 36 different quantization steps between 0.05 and 0.4 presents an interesting behavior for SNR_1 : the performance for certain quantization widths ($\Delta q \geq 0.16$) is close to that of the optimal points found by O_2. This means that, for certain SNR and constraint values, we can simplify the problem and compute optimal uniform quantization steps, which would achieve the same performance as the resolution of O_2.

2.3.4 Diversity-aware quantization

After the study of inter-key reciprocity and diversity in Section 2.3.3, this section presents a quantization scheme, entitled DIV, that improves the intra-key diversity metric based on a circular dictionary rotation as a function of the excess delay.

The proposed method is an adaptation of the one-bit HIST algorithm for directly sampled signals (Section 2.2) to a general multi-bit quantization scheme while considering a simplified model for noisy estimated multipath components. We recall that HIST produces keys with better randomness properties because of its diversified encoding based on the amplitude and on the delay of the quantized samples. This can be considered as a form of intra-key diversity since it implies a change in the probabilities of occurrence of the one-bit binary codewords employed by the existing algorithm POS.

We consider the noisy channel estimate model from Eq. (2.15) with noise only on the multipath amplitudes and a uniform-threshold quantization scheme (UQ) on b bits of step Δ_q. Accordingly, each channel realization produces K noisy samples y_k^u for quantization. In this study, we consider the same quantization thresholds regardless of the excess delay τ_k of the quantized sample y_k^u.

The proposed DIV scheme operates as indicated in Figure 2.12. The excess delay axis is split into bins of width Δ_{bin} and the quantization dictionary varies from one bin to another by means of a rotation. The samples are quantized depending on their excess delay τ_k by using the dictionary corresponding to the bin in which they fall. We denote as FIX the alternative fixed-dictionary guard-band quantization in which the dictionary is the same irrespective of the excess delay.

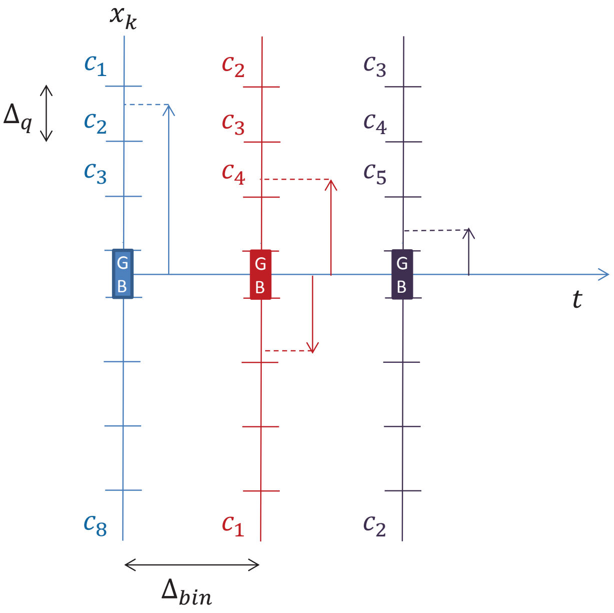


FIGURE 2.12: Illustration of the diversity-aware quantization method (DIV)

2.3.5 Performance evaluation II

In order to compare the two quantization approaches (FIX and DIV), we compute the intra-key reciprocity and diversity metrics according to Eq. (2.23)-(2.24) for various quantization steps Δ_q and bin widths Δ_{bin}. The noise variance is arbitrarily set at $\sigma_0 = 0.1$. A change in the value of σ_0 would only scale up or down the reciprocity-related values. Figures 2.13-2.14 show the averaged metrics over 1000 IEEE 802.15.4a channel realizations with a maximum excess delay of 50 ns and variable number of multipath components K. As expected, the reciprocity metric is not impacted by the proposed dictionary rotation because, in this model, we do not consider any errors on the excess delay estimations. The diversity measure is improved according to the bin size (smaller bin sizes are preferable) and is kept approximately constant with the quantization step. This means that DIV can help to recover some of the encoding diversity lost when increasing the quantization bin in order to obtain more reciprocal sequences.

In order to create intra-key codeword diversity, DIV relies on the information given by the temporal delays of the multipath components. When the corresponding estimates are also corrupted by noise, the performance of DIV in terms of reciprocity will be lower. The impact of realistic channel estimates on reciprocity is the focus of Section 2.4.

As already mentioned in Section 2.3.4, HIST also achieves intra-key diversity, which might seem contradictory at first sight because the uniformly sampled signals employed for testing HIST do not contain explicit delay information, such as excess delays. We

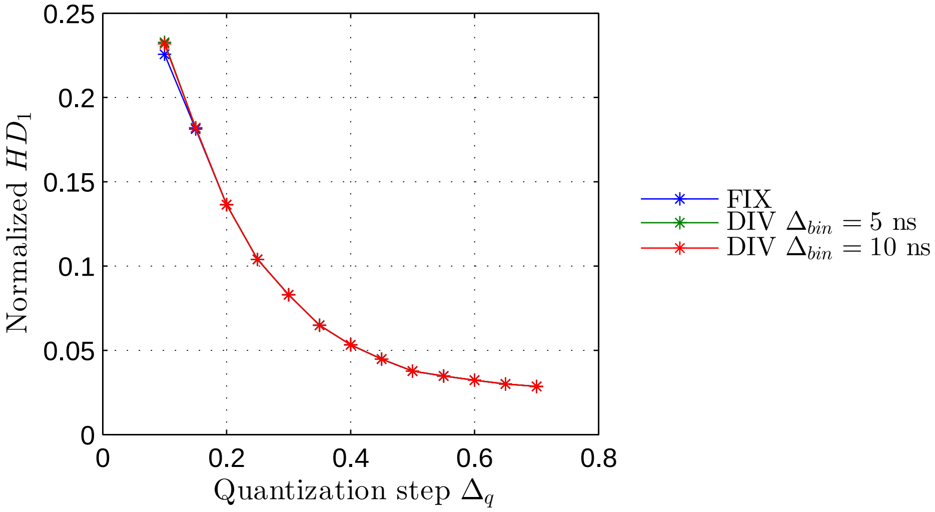


FIGURE 2.13: Mean intra-key reciprocity metric

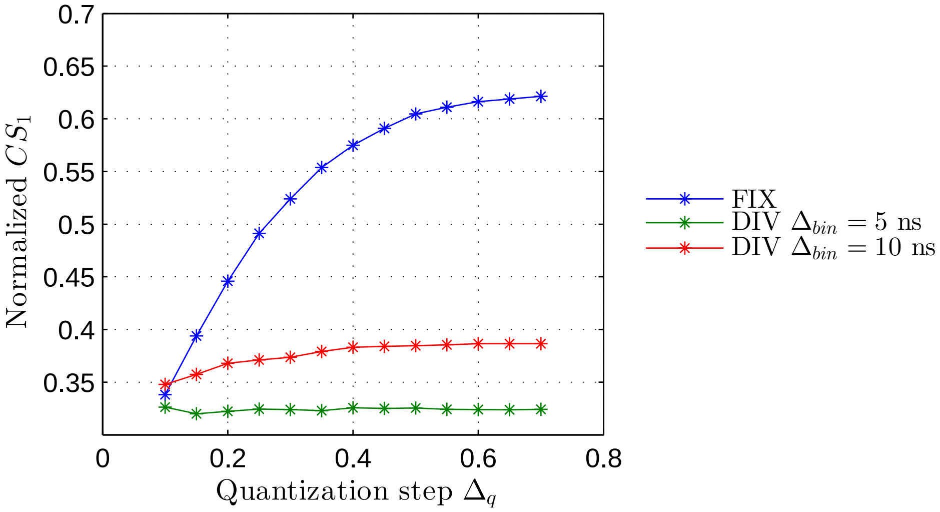


FIGURE 2.14: Mean intra-key diversity metric

argue that by combining the information given by the extraction threshold (or, equivalently, the amplitude) with the information provided by the bin affiliation (or, equivalently, the delay), HIST is a quantization scheme that indirectly depends on the alternations of the waveform with the excess delay, i.e., the multipath information. IR-UWB CIRs or signals are an interesting option for key generation first of all because of the fact that one measurement provides several values for quantization, but also because of the delay information contained in such signals. Schemes like HIST and DIV are examples of quantization schemes that consider both amplitude and delay information to improve the generated keys.

2.3.6 Summary

In this section, we focused on simulated IR-UWB channel estimates in order to investigate typical trade-offs that arise in the process of fixed-length quantization for key generation purposes. We have introduced a practical way of quantifying randomness based on the probabilities of occurrence of binary codewords at a fixed excess delay (inter-key diversity) or over an entire IR-UWB channel estimation (intra-key diversity). It should be noted that despite their similar flavor, the inter-key and intra-key metrics characterize different phenomena inherent to key generation using IR-UWB-like signals (i.e., amplitude-delay pairs).

In the first case, the amplitude quantization thresholds can be adapted as a function of the excess delay in order to achieve a desired trade-off between inter-key reciprocity and diversity. The threshold computation is performed through an optimization procedure based on the statistics of the IR-UWB channel and of the noise. It is therefore an *a priori* design method for "on average" optimal thresholds. Furthermore, we show that there exist uniform quantization steps achieving the same trade-offs as that of the optimized thresholds, which could potentially reduce the complexity of the quantization design step. In the second case, given the same uniform quantization step for the entire channel estimation, we illustrate an intra-key diversity-aware quantization scheme based on a delay-adaptive encoding dictionary. The proposed algorithm does not require identical quantization thresholds over the excess delay, so we can also use the computed thresholds in order to achieve the desired inter-key reciprocity and diversity trade-off. The mixed scheme would be characterized by the following:

- for each excess delay, we can choose beforehand the thresholds that achieve the desired inter-key reciprocity-diversity trade-off, meaning that, on average, some of the codewords will be likely to appear more often than others.
- when quantizing one particular channel estimate, the dictionary is varied circularly so that there are delay-dependent variations in the codewords that are more likely to appear.

The expected effect can be summarized as: decreasing diversity to achieve more reciprocity and then re-boosting diversity. However, this should be confirmed by further studies meant to investigate the interaction between the inter-key diversity and intra-key diversity and possibly extend them to incorporate more complex randomness aspects inspired, for example, by the NIST tests [102].

To conclude with, in this section, we have shown the interest of using statistical or absolute delay information for designing more robust quantization schemes. However, the findings may be limited by the necessity to operate with *a priori* known and reliable statistics in a given environment or by the channel estimation performance of real devices and multipath extraction algorithms. The latter will be investigated in the next section.

2.4 Keys from CIR estimates: impact on reciprocity

As already explained, an IR-UWB CIR is in theory a promising source of information for shared secret key generation [77] because of the fine temporal resolution capabilities of the IR-UWB technology (Section 1.3). The latter allows the extraction of accurate information such as amplitude and delay of the representative multipath components of a CIR. Not only does one CIR contain several scalar values, the latter are hard to predict by an attacker without knowing the exact positions of A and B or without employing complex ray-tracing tools [92] [96].

However, the main challenge is the way a receiver, usually limited in sampling frequency, representation dynamics or computational capabilities, can perceive the channel. This channel representation is the output of the channel probing phase and the input of the key quantization algorithm. Therefore, the reciprocal character of such signals is necessary for the study of key generation procedures using IR-UWB waveforms. In the present model, the degradations in the reciprocity of the inputs are caused by conventional additive noise on the received signal. However, a real system would also suffer from unbalanced radio frequency (RF) chains, antenna matching problems, non-linearities etc. on top of noise.

In this section, we start with an overview of channel estimation methods applicable to IR-UWB channel impulse responses. Then, we analyze the reciprocal character of IR-UWB channel estimates obtained with various types of estimators offering different trade-offs in terms of performance/complexity/hardware requirements: a high-resolution correlation-based estimator [110] and two sparse channel estimation methods [107] [111]. Finally, we infer a practical post-processing mechanism that can be applied to these signals before quantization in order to improve their reciprocity.

2.4.1 State of the art: channel estimation

Because of the high bandwidth occupied by IR-UWB communications, synchronization and channel estimation typically require high sampling rates of the order of tens of GHz. Moreover, in dense multipath environments (e.g., indoors), which we are mainly interested in, the number of parameters to estimate (i.e., delays and amplitudes) can be relatively large. The tutorial in Yang and Giannakis [89] provides several examples of timing acquisition techniques and channel estimation methods adapted to IR-UWB signals. Further, we will present an extended classification of channel estimators based on the required sampling frequencies employed for signal acquisition before the key generation stage (high, low, or sparse). The sparsity-aware estimators, which take into account the sparse structure of the CIR modeled as a Dirac stream, are divided into two categories based on their approach: Finite Rate of Innovation (FRI) estimators, which translate the channel estimation into an harmonic retrieval problem, and Compressed Sensing (CS) algorithms.

High-resolution methods

The first category includes the channel estimation algorithms that require sub-pulse sampling resolutions. Maximum-likelihood (ML) estimators for the amplitudes and delays are described for single-pulse channel sounding [112] and (non-) data-assisted [113] scenarios. These estimators can be applied in high-complexity Rake receivers.

Joint synchronization and channel estimation can be achieved by a least-squares estimator, which takes into account the clustered channel structure computed in advance with a subspace detection method [114]. This method assumes knowledge of the multipath order and an upper bound on the maximum excess delay.

Blind channel estimation for PPM transmissions is solved by exploiting the first-order cyclostationarity property of the received signal [115].[3] In this particular case, the mean over multiple equally distributed symbols of PPM transmissions is exploited to derive the CIR in a blind scenario. The proposed solution involves an FFT-based circular deconvolution.

More recent work on single-pulse channel estimation introduces the iterative detection of the multipaths delays and gains by splitting the ML estimate of the CIR [112] in two iterative steps: i) suboptimal search of the channel delays using a correlation-based method for resolvable paths; ii) computation of the channel gains based on the previously estimated delays. A variant of this approach is called the “search-subtract-readjust” algorithm introduced in [116] for characterizing the ranging precision of UWB localizers and exploited in [110] for evaluating the number of detectable multipaths in indoor channels. In our reciprocity evaluations, we will employ this channel estimation method, which, at each iteration step, readjusts all the previously estimated gains as a function of the newly estimated delay.

[3] Cyclostationary signals have statistical properties that vary circularly in time.

The aforementioned solutions consider the "sparse" channel model described in the standardization literature [91]. This model is extended to include "diffuse" components [117] modeled as Gaussian variables and adapted Bayesian channel estimation algorithms are introduced. Firstly, the nominal algorithm finds the parameters of the sparse components with an ML estimator consisting in an iterative Expectation-Maximization routine. Then, it estimates a complete hybrid channel estimation using MMSE or MAP estimators.

Low-resolution methods

In order to reduce sampling frequencies, an analog estimation of the aggregated CIR[4] is proposed for communication systems with a target demodulation sampling rate equal to the frame rate [118]. The resulting pilot waveform assisted modulation (PWAM) is inspired by the Transmitted Reference signaling.

Channel estimation for non-coherent energy detection (ED) receivers is translated into a Power Delay Profile (PDP) estimation problem [119]. The system employs multiple pulse transmissions and a phase offset between the sampling of consecutive pulse repetition periods in order to obtain a combined higher resolution image of the channel compared to typical ED resolutions. The image of the channel represents either a low-pass version of an aggregated PDP[5] or the estimated aggregated PDP by equalization of this low-pass version.

In low-complexity devices, such as the 500-MHz DBPSK transceiver [94] with 1Gbps direct sub-sampling and "1.5-bit" classical quantization, an image of the CIR can be obtained after differential digital correlations and accumulation of the correlation results over repetitions of a PN sequence on the received sequence.

These lower-complexity solutions will not be evaluated in this work because they either do not provide access to a digital CIR [118] or the channel estimation phase outputs signals with successive correlated samples [119]. In other cases, the resulting signals [94] have specific structure and dynamics requiring the design of adapted quantization metrics.[6]

[4] The so-called aggregated CIR is the result of the convolution between the transmitted pulse template and the CIR.

[5] An aggregated PDP can be expressed as $|p(t) * h(t) * p^*(-t)|^2$, where $p(t)$ is the transmitted pulse and $h(t)$ is the CIR.

[6] Note that we have recently initiated experimental studies regarding the achievable key rates exploiting this particular type of signals obtained in indoor mobile scenarios. For more details, the reader should refer to Appendix F.

Finite Rate of Innovation

Classical sampling and reconstruction theorems put forward by e.g., Shannon, Whittaker, Nyquist, state that a band-limited signal can be perfectly recovered by uniform sampling at a rate higher than two times the maximum frequency present in the signal. This is actually one possibility for recovering the information contained in a band-limited signal. Since then, generalized sampling theorems involving for example, band-pass sampling, derivative sampling or super-resolution sampling for image reconstruction, have been developed and have inspired the extension of sampling theorems to a larger class of signals: infinite-bandwidth signals with limited energy [120].

Similarly, perfect reconstruction of non-band-limited noiseless signals presenting a finite number of degrees of freedom per unit of time (i.e., the rate of innovation) is possible by employing a finite number of uniform samples [121] [122]. Examples of this type of signals include streams of Diracs or piecewise polynomials sampled with a sinc kernel at or above their rate of innovation. For example, in the case of a periodic stream of Diracs, the reconstruction is equivalent to the harmonic retrieval problem: estimation of the phases and amplitudes of a finite number of exponentials by sampling their sum. The main idea is to employ a tool from spectral theory, namely an annihilating filter[7] applied in the domain of Fourier series coefficients, in order to identify the time instants of the Diracs and then compute the Diracs' amplitudes.

To sum up, the reconstruction of h, a T-periodic stream of K Diracs can be achieved as follows: i) sampling the sparse signal h with a low-pass kernel to obtain M samples ($M \geq 2K+1$); ii) computing M Fourier series coefficients $H[m]$ of h by solving a linear system constructed with the M samples; iii) finding the coefficients $A[m]$ of the annihilating filter of $H[m]$; iv) finding the roots of the annihilating filter, each of them giving direct access to the time instants τ_k of the Diracs; v) computing the Diracs' amplitudes x_k by solving $H[m] = \sum_{k=1}^{K} x_k e^{-2\pi j m \tau_k / T}$.

The described method involves a root finding operation applied to the annihilating filter, which results in an ill-conditioned problem in the presence of noise. An alternative subspace method is proposed in order to apply sub-Nyquist FRI sampling to noisy signals [123] [122], such as received IR-UWB signals with the aim of estimating the CIR [111]. This solution exploits a matrix composed of the Fourier coefficients $X[m]$, the fact that its rank is K in the case of noiseless data and the shift-invariance property of particular subspaces of this matrix.[8] More details on the algorithm [111] implemented

[7] An filter $a[n]$ is said to annihilate the signal $s[n]$ if $(a * s)[n] = 0$.

[8] The FRI subspace method is similar to spectral theory algorithms like ESPRIT and MUSIC, which use the covariance matrix of the Fourier data instead. ESPRIT exploits the shift-invariance property for particular subspaces of this covariance matrix and MUSIC relies on the orthogonality between the signal and noise subspaces.

in our evaluations are given in Section 2.4.3. However, the frequency-domain processing may be questionable in low-complexity IR-UWB devices operating primarily in the time domain.

Compressed Sensing

The aforementioned FRI [121] solution for recovering sparse signals using a sub-Nyquist sampling rate has been initially developed for continuous non-band-limited signals. It is shown that these signals can be uniformly sampled using a low-pass or band-pass sampling kernel and reconstructed through a parametric estimation approach achieving provable performance lower bounds [124]. Compressed Sensing or Compressive Sampling is based on the paradigm that signals that are compressible in some basis (e.g., images in the wavelet basis) can be acquired with less samples since the beginning. Unlike FRI, CS methods have been designed for the discrete case and are adaptable to a larger class of signals, namely any signal that can be considered *sparse* in a certain orthonormal basis called the sparsifying basis. The sampling is usually realized with a non-adaptive random sampling/sensing kernel in a basis that is *incoherent* with the sparsifying basis and the information is reconstructed through nonlinear optimization algorithms minimizing the l_1-norm. Although both methods employ a lower number of samples than classical sampling methods, the theoretical performance of CS algorithms is probabilistic by nature and complex to analyze [124].

Compressed Sensing fundamentals Given an orthonormal basis $\Psi = \{\psi_n, n = 1, \dots, N\}$, a signal $\mathbf{x} \in \mathbb{R}^N$ is considered to have sparse coefficients $\theta_n = \langle \mathbf{x}, \psi_n \rangle$ if $||\theta||_p \leq R$, $0 < p < 2$, $R > 0$ [125].[9] In the context of CS, this definition is extended to the l_0 "norm", which represents the number of non-zero coefficients denoted as $K \ll N$, i.e., the orthodox connotation of "sparse". This condition represents the *sparsity* requirement [126], which in the IR-UWB case is equivalent to the aforementioned degrees of freedom of our signals, i.e., the cardinality of the multipath components.

The second necessary requirement is the *incoherence* between the sparsifying basis Ψ and the sensing basis $\Phi = \{\phi_m, m = 1, \dots, M\}$, which means that, unlike the signal of interest $\mathbf{x}$, the sensing kernels ϕ_m must have dense representations in Ψ. An example of incoherent basis pairs is the Dirac basis and the Fourier basis [126]. Random matrices are deemed to be largely incoherent with any fixed sparsifying matrix Ψ [126]. So by using a random sensing matrix, CS methods do not need to be adapted to the signal of interest.

[9] The l_p norm of θ is defined as $||\theta||_p = (\sum_n |\theta_n|^p)^{1/p}$.

Using $M < N$ samples $y_m = \langle \mathbf{x}, \phi_m \rangle$, the reconstructed signal is $\hat{\mathbf{x}} = \Psi \hat{\theta}$ where

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^N}{\operatorname{argmin}} \ ||\theta||_1 \text{ subject to } y_m = \langle \phi_m, \Psi \theta \rangle, \ m \in \{1, \ldots, M\}. \tag{2.28}$$

Note that the sparsity condition is translated into a l_1 minimization problem with linear constraints [126], which is easier to solve than the initial problem. In the noiseless case, if the signal $\mathbf{x}$ is K-sparse in Ψ and $M \geq c \cdot \mu^2(\Phi, \Psi) \cdot K \cdot \log(N/\delta)$, where c is a constant and the function μ is a measure of the incoherence between two basis, the solution to Eq. (2.28) is exact with probability $1 - \delta$. It can be found by the Basic Pursuit algorithm (BP) or greedy algorithms such as Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP).

In the noisy scenario or when the signal $\mathbf{x}$ is nearly sparse (i.e., the $N-K$ coefficients are small but not exactly zero), an additional condition is imposed: the Restricted Isometry Property (RIP). In order to be able to reconstruct K-sparse signals θ from the samples $\mathbf{y} = \Phi\mathbf{x} + \mathbf{z} = \Phi\Psi\theta + \mathbf{z}$, where $\mathbf{z}$ is the noise term, these vectors should not be in the null space of the matrix $\Phi\Psi$ [126]. Linear programming methods for the noisy case include Basic Pursuit Denoising (BPD), which adds a penalty term to the cost function in Eq. (2.28).

IR-UWB Compressed Sensing algorithms The received IR-UWB signal can be considered sparse in a “pulse” basis, i.e., a dictionary composed of delayed pulses [107]. By random time sampling at an equivalent $M/N = 0.36$ and MP reconstruction, both CS-based proposed receivers (a Rake-like receiver using CIR estimations and a correlator-based receiver using aggregated CIR estimations) outperform the correlator-based receiver described in [118] in terms of BER. The described CS algorithm for CIR estimation with a dictionary of delayed pulses will be used in our evaluations.

In a spread spectrum CS based approach for IR-UWB CIR estimation [127], the authors propose to modulate the received IR-UWB signal with a pseudo-random sequence in order to spread the spectrum. The result is then randomly sampled in the Fourier domain and the received signal is reconstructed using a dictionary of delayed pulses via BPD l_1 minimization. An additional frequency sensing technique for ToA and channel estimation and a new reconstruction algorithm inspired by OMP are proposed in [128] and compared to the aforementioned ones [107] [127]. This new CS solution achieves a trade-off between path detection accuracy and convergence speed.

The CS sensing framework can be combined with the noise statistics formulation of the ML estimator from [113] in order to achieve reliability performance in terms of BER close to that of the ML estimator but at a lower sampling rate [129]. The proposed

CS-ML estimator also shows better performance than the MP algorithm in terms of reconstruction error.

Sparse ultra-wideband channel estimation can be formulated within a Bayesian Compressed Sensing (BCS) context as well [130] [131]. The BCS method [132] replaces the traditional l_1 minimization with a MAP estimator, which takes into account the noise statistics within the likelihood function and models the sparsity information as, e.g., Laplacian or Gaussian priors on the searched CIR. BP l_1 minimization is compared to BCS for criteria like execution time and reconstruction error [130] and additionally for SNR regimes and sparsity structure of the channel [131].

It should be noted that studies like [133] [131] investigate the impact of channel sparsity on CS estimators concluding that reasonable performance can be achieved for CM1 or CM2 models. However, CS estimators could be unadapted to CM8-like channels (i.e., industrial NLOS) with denser multipath structure. Moreover, recent work on CS-based receivers investigates the effect of different types of noise (sky or amplifier noise) and CS sensing architectures (serial or parallel) on the achieved BER [134]. It is shown that CS signal detection with correlated noise is outperformed by the uncorrelated noise situation.

2.4.2 System model

We consider two communicating parties (A and B) and a received signal and channel models similar to the ones described in Section 2.2.1:

$$\begin{aligned} y_u(t) &= (h * p)(t) + w_u(t) && (2.29) \\ h(t) &= \sum_{k=1}^{K} x_k \delta(t - \tau_k) && (2.30) \end{aligned}$$

where $u \in \{A, B\}$, $p(t)$ is the transmitted pulse waveform with central frequency f_c and bandwidth B, $h(t)$ is the reciprocal CIR with K multipaths of amplitudes x_k and excess delays τ_k generated according to the IEEE 802.15.4a, $w_u(t)$ is a random processes of zero-mean Gaussian white noise with double-sided power spectral density $N_0/2$, which is filtered in the band B of the transmitted signal and has variance $\sigma_w^2 = N_0/2 \times 2B$. As before, we define the SNR as:

$$\mathrm{SNR} = \frac{\frac{1}{T}\int_0^T (h * p)^2(t)\,\mathrm{d}t}{\sigma_w^2} \qquad (2.31)$$

We denote the sampled version of $y(t)$ at sampling rate F_s as $\mathbf{y}$ and its length as M. After sampling, the signal is processed independently at A and B with three different estimators in order to generate CIR estimations. Next, we detail the considered channel estimators.

2.4.3 Channel estimators and pairing issues

Channel estimators

Matched Filter estimator (MF) Given a received signal and a pulse template, the MF algorithm [110] iteratively detects and substracts the most representative paths found in the signal. The operation is based on the output of the cross-correlation between the received signal and delayed versions of the pulse template, which gives access to the estimated path delay. The channel gains are jointly readjusted at each iteration.

Finite Rate of Innovation estimator (FRI) For FRI evaluation, we implement the subspace method from [111]. Based on an *a priori* targeted number of samples M, the received signal is filtered with a band-pass filter and sub-sampled. Then, the Fourier Transform is applied to $\mathbf{y}$ and the M FFT coefficients $\hat{H}$ of the searched signal $\hat{\mathbf{h}}$ are obtained after division by the FFT coefficients of the pulse. The resulting vector is arranged into a $P \times Q$ Hankel matrix $\mathbf{H}$[10] with $P + Q - 1 = M$, $P, Q \geq K$. The matrix $\mathbf{H}$, which has a a rank equal to K in the noiseless case, is decomposed with an SVD operation and divided into a signal subspace of dimension K indexed by s and a noise subspace indexed by n:

$$\mathbf{H} = \mathbf{U}_s \Lambda_s \mathbf{v}_s^T + \mathbf{U}_n \Lambda_n \mathbf{v}_n^T \tag{2.32}$$

The matrices $\mathbf{U}$ and $\mathbf{v}$ satisfy the shift-invariance subspace property, i.e., they have a Vandermonde structure[11] dependent on $e^{-2\pi j \tau_k / T}$. This allows the computation of $\hat{\tau}_k$ after matrix manipulations on $\mathbf{v}_s$ or $\mathbf{U}_s$. The path amplitudes $\hat{x}_k$ are then estimated from:

$$\hat{H}[m] = \sum_{k=1}^{K} \hat{x}_k e^{-2\pi j m \tau_k / T}, \ \forall m \in \{1, \ldots, M\} \tag{2.33}$$

Compressed Sensing estimator (CS) The CS estimator inspired from [107] uses a random standard Gaussian sensing matrix Φ in order to acquire M noisy samples $\mathbf{y} = \Phi y$. The sparsifying matrix Ψ contains the dictionary of delayed pulses with an arbitrary temporal resolution for the delay (e.g., 0.5 ns in our case). Note that this is a digital resolution and is independent of the ADC resolution used for signal acquisition. Then, a variant of the Matching Pursuit algorithm is implemented based on: $\mathbf{y} = \Phi \Psi \hat{\mathbf{h}}$.

[10] A Hankel matrix is a matrix with constant skew diagonals.

[11] A Vandermonde matrix is a matrix with the terms of a geometric progression in each row

This algorithm is similar to the MF estimator with the exception that the processing is realized in another domain and the amplitudes are not readjusted at every path detection. Also, the random sensing matrix should be the same for the bidirectional channel estimations.

Pairing of estimated multipath components

Based on our observations, an estimated CIR $\hat{\mathbf{h}}_u$ can contain misdetected paths or fail to detect paths that are detected on the other side of the link (Figure 2.17(a)). In order to avoid complete desynchronization of the sequences meant for quantization, we propose an heuristic pairing algorithm that involves the public exchange of the estimated excess delays associated with the estimated paths.[12] This means that the delay information cannot be used for quantization anymore and that schemes, such as HIST or DIV, cannot be applied after pairing. The pairing algorithm is detailed in Alg. 2, where $\hat{x}_u$ is the vector of estimated path amplitudes at u, $\hat{t}_u$ is the vector of estimated path delays, and a_u is the vector of path amplitudes after pairing. The number of estimated paths is denoted as K in the following.

```
**Data**: $\hat{x}_A, \hat{t}_A, \hat{t}_B$
**Result**: $a_A$
initialization pairing vectors $v_A$ and $v_B$ ;
**for** $i$ = 1, ..., $K$ **do**
    find $\hat{t}_B[j]$ closest to $\hat{t}_A[i]$;
    do $v_A[i] = j$;
    find $\hat{t}_A[j]$ closest to $\hat{t}_B[i]$;
    do $v_B[i] = j$;
**end**
find $\mathcal{I} = \{i | 1 \leq i \leq K, v_B[v_A[i]] = i\}$ ;
set $a_A = \hat{x}_A[\mathcal{I}]$ ;
```

Algorithm 2: Path pairing algorithm for CIR estimates (at A)

2.4.4 Performance evaluation

For our evaluations, we employ the following parameters and estimators:

[12] We do not use the traditional Hungarian algorithm for the complete assignment problem because we do not necessarily need to pair every single path, but rather we do not pair some of the paths if the cost of the resulting pair is too high.

- Signal and channel parameters: $f_c = 4.5$ GHz, $B = 1$ GHz, 1000 channel realizations according to the IEEE 802.15.4a CM1 model;
- Estimators: MF (a pseudo-analog matched filter estimator at $F_s = 100$ GHz), MF10 (an MF at $F_s = 10$ GHz), MF20 (an MF at $F_s = 20$ GHz), CS and FRI at the same $F_s \in (3.6, 3.7)$ GHz;[13]
- Quantization: an arbitrary two-bit quantization scheme with uniformly distributed codeword probabilities applied to the concatenated values issued from the CIR estimation phase (before or after pairing).

We evaluate the reciprocity before quantization through the inter-estimate RMSE metric from Eq. (2.34), where $\hat{x}_u$ can be replaced by $\hat{a}_u$ for the RMSE after pairing. This metric characterizes only the reciprocity errors regardless of the channel estimation performance (e.g., a similarly misdetected path on both sides of the link does not impact the key generation performance). For comparison purposes, the $\text{RMSE}_{A,B}$ for each estimator is normalized with respect to the $\text{RMSE}_{A,B}$ of the MF estimator. After quantization, we also compute the mean bit agreement ratio obtained after averaging the bit agreement of the quantized sequences for each channel realization.

$$\text{RMSE}_{A,B} = \mathbb{E}_{h(t)}[\text{RMSE}(\hat{x}_A, \hat{x}_B)] = \mathbb{E}_{h(t)}[\sqrt{\frac{\sum_{k=1}^{K} (\hat{x}_{k,A} - \hat{x}_{k,B})^2}{K}}] \tag{2.34}$$

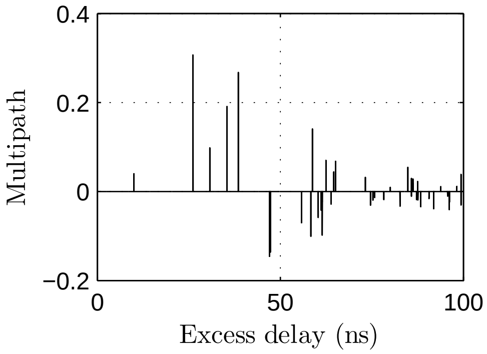
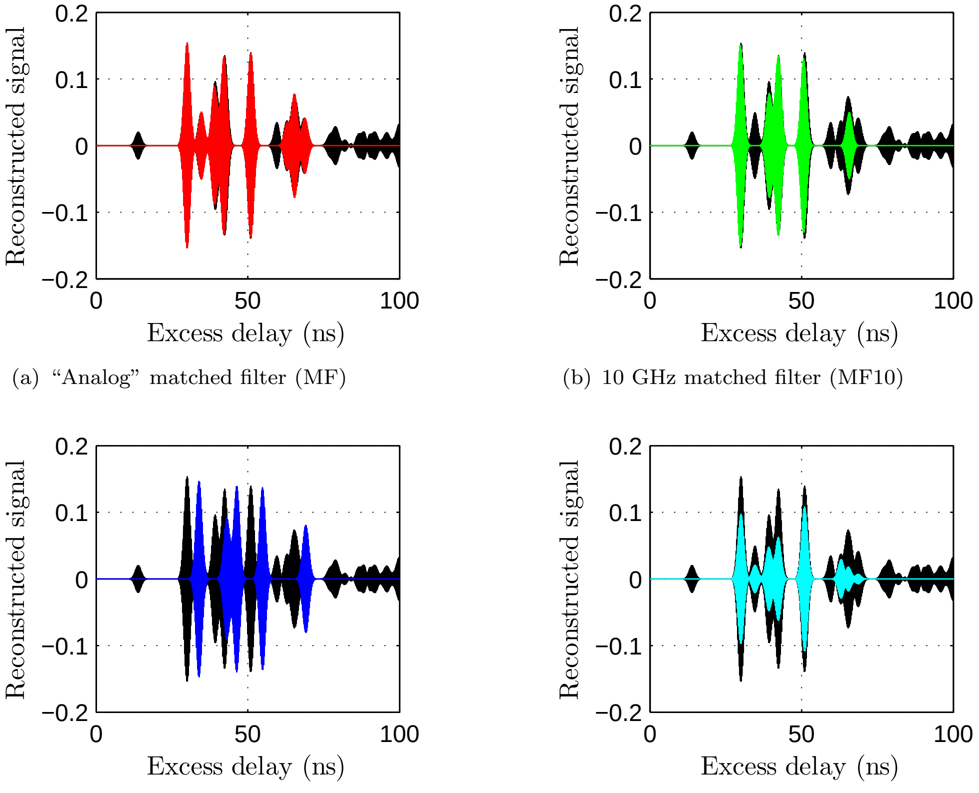


FIGURE 2.15: Illustration of an IEEE 802.15.4a true multipath channel

Figures 2.16-2.17 show examples of the reconstructed signals from various estimators and the corresponding estimated bidirectional CIRs for an arbitrary channel realization (Figure 2.15). The bidirectional CIRs are measured at A an B and are denoted by different colors. The number of estimated multipath components is $K = 8$ and the SNR is deliberately high for better visualization (SNR = 25 dB). The reconstructed signals (Figure 2.16) are shown only for one side of the link because there is no significant

[13] Note that this sampling rate corresponds to 5 times more the number of minimum samples needed for a noiseless FRI estimator using band-pass sampling.



(a) "Analog" matched filter (MF)

(b) 10 GHz matched filter (MF10)

(c) Compressed sensing estimator (CS)

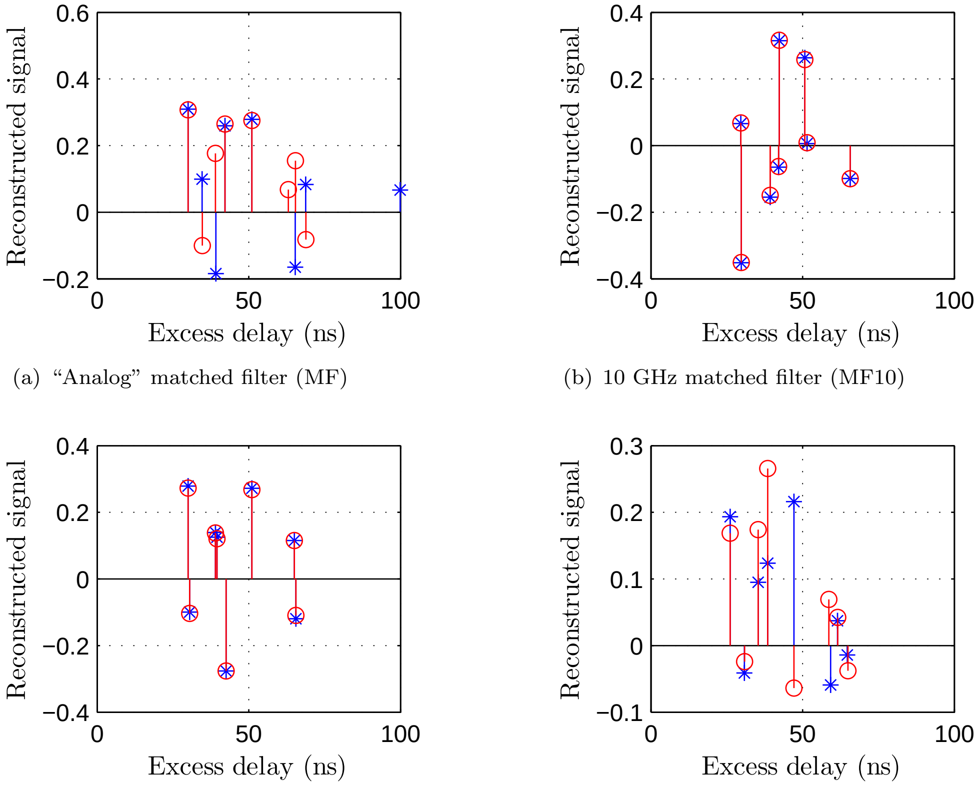
(d) Finite Rate of Innovation estimator (FRI)

FIGURE 2.16: Illustration of initial and reconstructed "analog" signals with various estimators

difference between the bidirectional received signals. The shift experienced in Figure 2.16(c) can be corrected based on the pulse template length and is therefore not a limitation of the CS method.

As expected, from the reconstruction point of view, the MF estimator performs the best (Figure 2.16) and exhibits the most "natural" CIR structure (Figure 2.17). However, it is also the most sensitive in terms of reciprocity of the paths (Figure 2.17(a)) because of several polarity inversions, which occur at the matched filter output, and one misdetected path. More details on the impact of the various sampling rates on the MF estimators are provided in the next section.

We also observe that the FRI estimator (Figure 2.17(d)) produces reciprocal excess delays but relatively noisy amplitude estimates. This behavior could be explained by the harmonic approach of initially retrieving the excess delays based on a subspace method, which seems favorable to reciprocity, and then computing the corresponding amplitudes. Although the order of the strongest FRI-sensed paths is reciprocal (SVD operation and signal space detection), sometimes small errors appear on the excess delays and are later translated to polarity inversions in the amplitudes computed in the time domain.



(a) "Analog" matched filter (MF)

(b) 10 GHz matched filter (MF10)

(c) Compressed sensing estimator (CS)

(d) Finite Rate of Innovation estimator (FRI)

FIGURE 2.17: Illustration of the estimated bidirectional CIRs with various estimators

Given the true dense channel from Figure 2.15, we conclude that all the estimators fail in absolute to retrieve the true multipath components because of the inherent limitations of the pulse resolution. However, they are able to recover different images of the channel. These images can be reciprocal and have different structures that are characteristic to each estimator and its parameters (e.g., sampling frequency, extraction algorithm, random sensing matrices for CS).

Reciprocity evaluation

In this subsection, we evaluate the $\mathrm{RMSE}_{A,B}$ for two configurations: fixed SNR and variable K and fixed K and variable SNR.

In Figures 2.18 and 2.19, we plot the normalized $\mathrm{RMSE}_{A,B}$ at fixed SNR values and variable number of estimated paths before and after pairing. The results for FRI are not shown on the graphics because of their extended range between 0.7 and 34 before pairing and 0.7 and 26 after pairing, which means that FRI estimates could be too non-reciprocal, at least in comparison with the reference MF estimator. All methods perform better at the highest SNR value and for a smaller number of paths, with CS having the

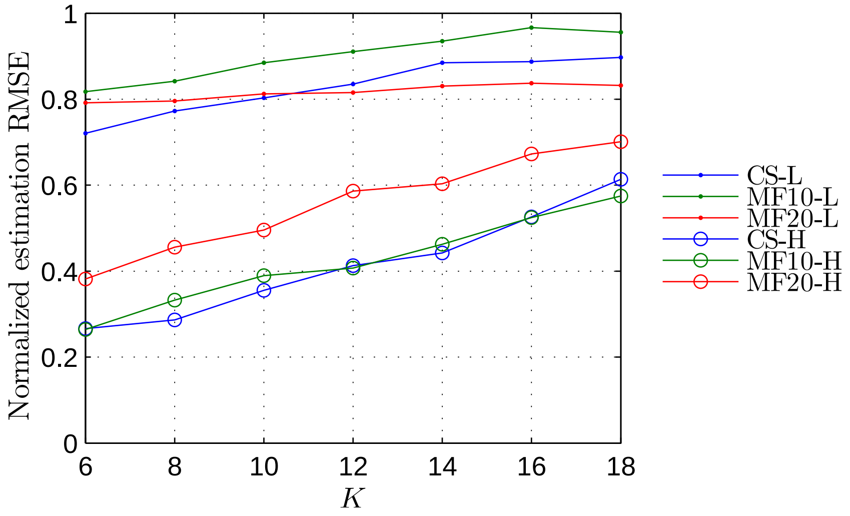


FIGURE 2.18: $\text{RMSE}_{A,B}$ before pairing at low and high SNR: 5 dB (L) and 30 dB (H)

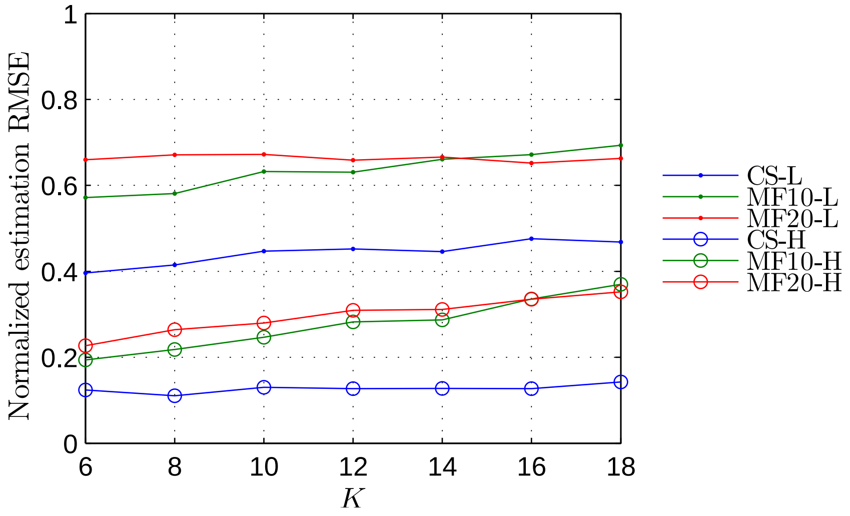


FIGURE 2.19: $\text{RMSE}_{A,B}$ after pairing at low and high SNR: 5 dB (L) and 30 dB (H)

best performance for most regimes. The proposed pairing algorithm contributes to the decrease in $\text{RMSE}_{A,B}$ for all regimes and to the beneficial uniformity of the $\text{RMSE}_{A,B}$ across K, while keeping a relatively high number of estimated paths (Figure 2.20).

Next, we fix K to the maximum number of detectable paths with the present sparse implementations (i.e., 18) and we look at the reciprocity evolution with the SNR (Figure 2.21). We conclude that before pairing the methods have equivalent performance, but after pairing an obvious gap is created between the CS and the MF estimators. Also, the value of the normalized $\text{RMSE}_{A,B}$ is systematically lower than 1, meaning that in terms of reciprocity the ranking of the methods is the following: CS, MF10 and MF20, and MF.

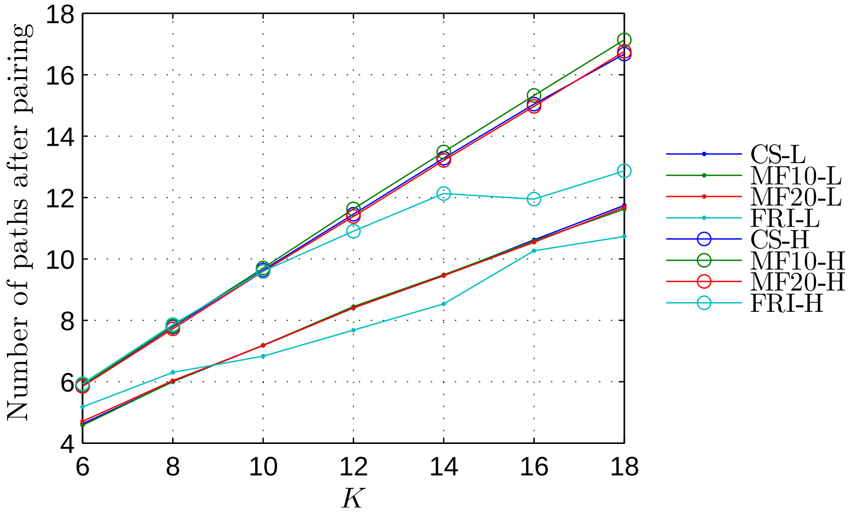


FIGURE 2.20: Mean number of paths after pairing: 5 dB (L) and 30 dB (H)

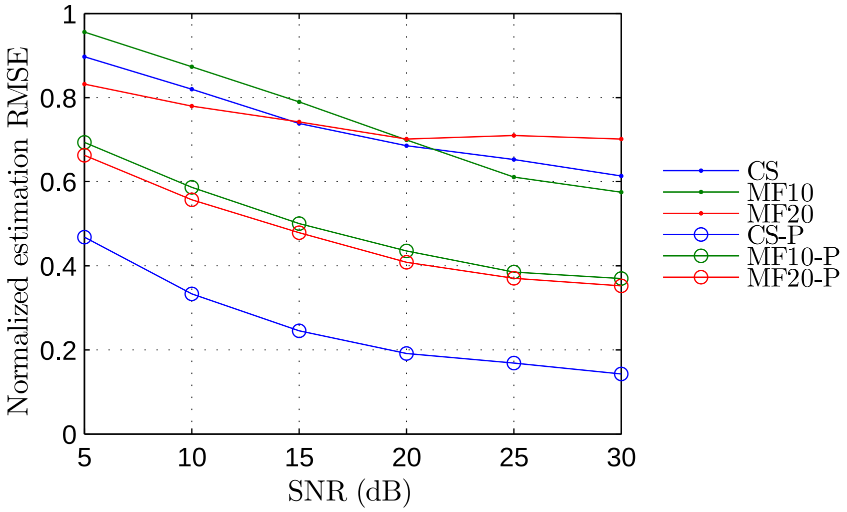


FIGURE 2.21: $\text{RMSE}_{A,B}$ before and after pairing (P) for $K = 18$

This seems counter-intuitive at first because higher sampling rates should mean better performances, but this result becomes more explicit when recalling the illustrative example from Figure 2.17. MF methods, especially the ones at higher F_s suffer from polarity inversions of certain path estimates. This is due to noise and it happens systematically when the error on the estimated delay is of the order of an odd multiple of $1/2f_c$. In other words, any inner oscillation of the unitary waveform (i.e., at the center frequency f_c) may be misinterpreted as the center of this waveform because of noise. MF methods with higher temporal resolutions are more sensitive to these phenomena since the maximum of correlation might not be detected in the same way on both sides of the link.

Nevertheless, the advantage of lower-sampling methods in terms of reciprocity does not come without a drawback. Regarding the visual structure of the estimated CIR from Figure 2.17, when the channel is represented less faithfully, its inherent randomness is less well captured, so sparse-sampling methods could suffer from randomness defects.

Bit agreement evaluation

In order to confirm our previous findings and verify the prohibitively low reciprocity of FRI estimates, we plot the average bit agreement ratios for all the estimators at $K = 18$ before pairing (Figure 2.22) and the respective bit agreement gains after pairing (Figure 2.23).

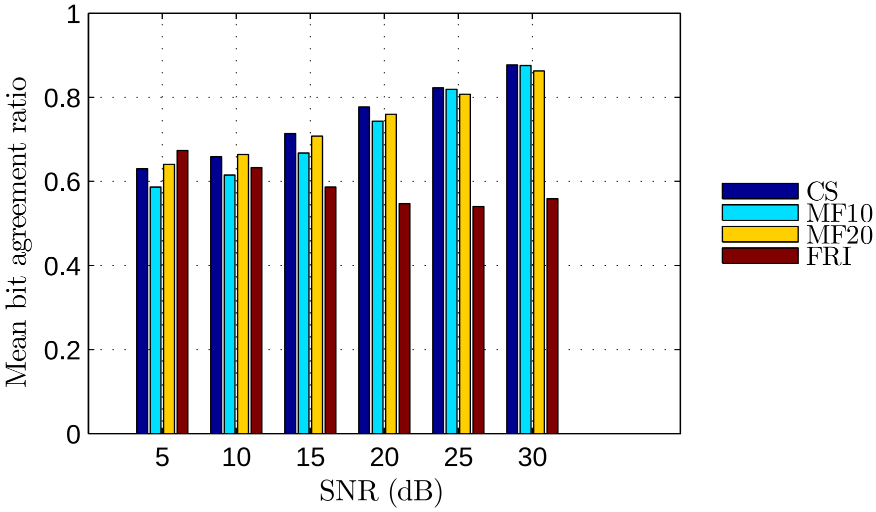


Figure 2.22: Mean bit agreement before pairing for $K = 18$

The FRI $\text{RMSE}_{A,B}$ results varying between 0.5 and 0.6 confirm that the implemented FRI estimator is less stable or too sensitive in terms of noise. Although the phenomena involved in the FRI estimation process are less transparent because it uses spectral methods, the presumed noise sensitivity indicates that temporal methods should be preferred from a reciprocity point of view for estimates of temporal signals.

Also, the estimator ranking observed in Figure 2.21 still holds for the bit agreement results as shown by the results of the gain after pairing (Figure 2.23). It can also be noted that the pairing has a powerful impact especially for CS at medium SNR values because at lower SNR pairing is not sufficient, while at high SNR it is less needed.

2.4.5 Summary

In this section, we examined the reciprocity degradation incurred by realistic CIR estimates. We compared three estimators: a high-resolution matched filter requiring

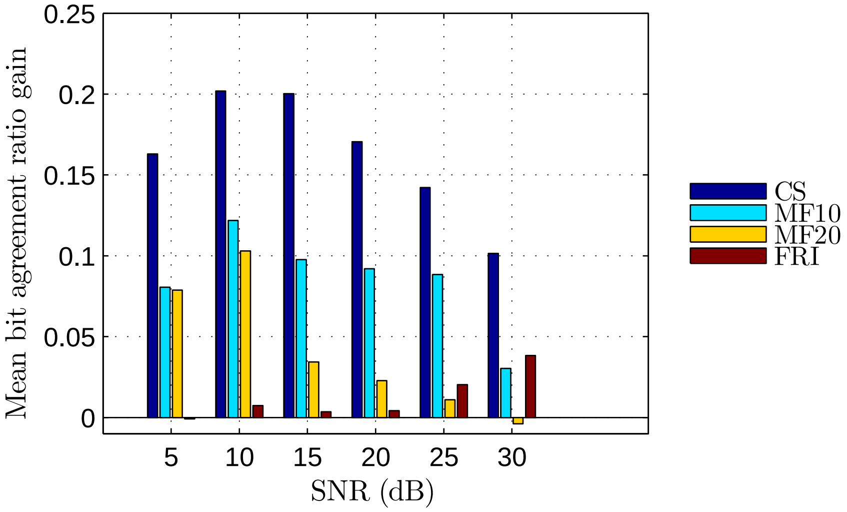


FIGURE 2.23: Mean bit agreement gain after pairing for $K = 18$

Nyquist or higher sampling rates, a compressed sensing algorithm, and a spectral method for sparse signal reconstruction. The MF methods are simple to implement, but require high sampling rates and the reciprocity is often deteriorated by misdetected paths. The FRI estimator has low performance in noisy cases and its Fourier-domain implementation along with the SVD operations require important computational capabilities. CS estimates achieve acceptable reciprocity but there are still open questions concerning the associated receiver architecture (e.g., implementation of the sensing part, choice and sharing of the random sensing matrix). Finally, we propose a pairing algorithm based on the exchange of estimated excess delays in order to avoid the mismatching caused by the misdetected paths.

This part of our work has been inspired by the study of the theoretical performance of unbiased IR-UWB channel estimators [135]. The authors derive the expressions of the Cramer-Rao Lower Bounds (CRLB) characterizing the accuracy of excess delay and amplitude estimates given the true multipath channel and a certain pulse used for probing. Initially, we aimed at creating a similar "CRLB model" meant to describe the theoretical reciprocity performance, equivalent of the estimation accuracy [135], and compare it to the performance achieved by various estimators. This would mean explaining only the noise effect on a realistic bidirectional channel estimate. Because of pulse interference, a true multipath component could be wrongly but similarly estimated by the two users, which means that it would still contribute to the mutual information for key generation. As an example, if there is pulse overlapping because two paths have very close time of arrivals (e.g., reflections on two close-by pieces of furniture), an estimator cannot "solve" both of them and there would certainly be an error with respect to the true channel. However, since both estimators will do the same error, the reciprocity of the estimates would not be affected.

This proved to be a difficult problem because of the assumption of realistic pulse interference, which was the core of the phenomenon we wanted to describe. Typically, this assumption is avoided in theoretical studies regarding the performance of unbiased estimators [135] or the mutual information between bidirectional IR-UWB channels for key generation [77]. Moreover, when considering the pulse interference phenomenon, it is admitted that, because of noise, the bidirectional estimates will potentially contain false alarms or missed paths. This leads to an assignment or pairing problem, which automatically implies additional reconciliation procedures. We conclude that the study of the reciprocity performance of channel estimates is dependent on the chosen reconciliation procedure and that the theoretical performance of joint probing-reconciliation schemes becomes even more complex than the initial aim of the "CRLB" reciprocity model.

Chapter 3

Public discussion strategies

In narrow-band communications, the radio signal perceived by an eavesdropper situated at more than a few wavelengths of one of the legitimate users is uncorrelated with the main channel signal [136]. However, in wide-band communications, notably in IR-UWB, the spatial decorrelation property of channels does not present the same behavior with the distance: it depends on the considered radio characteristic (e.g., CIR, averaged CIR, etc.) as well as on the propagation environment [29]. The mentioned study on the spatial correlation of IR-UWB channels [29] exploits experimental data from the same measurement campaign [95] [79] presented in Section 2.2.

The acquired radio signal of the eavesdropper and the information that transits on the public channel represent the leaked information impacting the immunity to eavesdropping attacks. The reconciliation phase from the previous key generation protocol [95] consists in the exchange of the position tables followed by Reed-Solomon error correction (see Section 2.2.2). The error correction scheme can be adapted to correct more or less mismatched bits in order to guarantee an acceptable compromise between bit agreement and security. On the contrary, the table exchange is a fixed protocol specification and the information transmitted on the public channel (i.e., the indexes of the extracted samples) represents the excess delay information of the significant channel coefficients.

In this chapter, we consider IR-UWB signals obtained by deterministic Ray-Tracing simulations, which, contrary to the statistical IEEE channel realizations, can model the location-dependent characteristics and therefore, the spatial correlation of the received IR-UWB signals. We aim to investigate the security impact of the public exchanges proposed in previous studies [95] when employing threshold-based synchronization and we propose alternative reconciliation strategies that limit or mask the publicly exchanged information [98]. Before discussing the system model and the main results, a short state of the art of practical reconciliation schemes is presented.

3.1 State of the art: information reconciliation

As already mentioned, we consider that a practical information reconciliation phase consists of two steps: i) a preliminary public discussion for sharing quantization-related information (e.g., indexes of dropped samples or information for asynchronous quantization); ii) error correction, after which the bits should be perfectly identical. We recall that all public exchanges are considered to be authenticated, which means that an attacker cannot tamper with the transmitted information or inject false information.

Preliminary public discussions are highly dependent on the key generation protocol, so we will focus on the one proposed in [95], which involves exchanging the indexes of the dropped samples after quantization. Whereas for consecutive uncorrelated RSS measurements, such exchanges do not offer any advantage to an eavesdropper, in the case of directly sampled IR-UWB CIRs, samples are usually correlated and this information can provide indications about the structure of the legitimate signal. It is therefore preferable to protect this information if possible.

Although in this work we only investigate the first step of reconciliation, a short classification of error correction schemes is also presented for completeness purposes. There are two main types of error correction: iterative and non-iterative error-correction techniques. The iterative Cascade protocol employed in [27] is similar to the information reconciliation step of QKD (Section 1.2.2). It involves random permutation of the bits, divisions in small blocks and iterative exchanges of block parity information. We classify the non-iterative error-correction techniques based on their “philosophy”.

- Distributed source coding approach or coding with side information. Typically, Alice maps her observations to codewords extracted from a codebook divided in cosets. The property of the coset organization is that the minimum distance between the elements of the cosets is maximized (i.e., the code”’words in a coset are “very” different). The indexes of the cosets are sent to Bob who uses them along with his observations to find the corresponding codewords. Possible implementations for this approach include: i) coset assignment through Reed-Muller block codes or trellis codes [77]; ii) LDPC syndrome computation by Alice and syndrome-based decoding by Bob before or after quantization [67] iii) classical quantization of the measurements and transmission of the difference between the measured and the quantized values over the public channel [137]. In the case of signal envelope quantization in IR-UWB [138], it is suggested to add a stage of bilateral independent LDPC decoding for better key agreement before reconciliation.

- Channel coding approach. After parallel quantization, Alice computes the parity check for her sequence with an error-correction code (e.g., Reed-Solomon codes [95]) and sends it to Bob over the public channel. If the number of errors is not too large comparing to the error correction capabilities of the employed code, Bob can now correct the mismatches of his own binary sequence using the received parity information. For benchmarking purposes, we will employ the same error-correction scheme as in [95].

Other more complex options for key reconciliation include compressive sensing algorithms, if we consider the difference between the users' quantized sequences as a sparse vector [139] or neural networks with specific training algorithms corresponding to typical bit errors [140]. The same study [140] proposes a second reconciliation scheme for key generation from signal fades of envelope magnitudes: generating a pseudo-random bit sequence from the average fade length and using it as a mask before applying error correction. Similarly to our proposed solution $\mathrm{POS}_{\mathrm{ToF}}$, this method also adds a supplementary layer of security against eavesdropping attacks. However, the mentioned work does not consider the impact of correlated channel observations of an attacker and it is specific to key generation procedures based on signal fades. Our public discussion strategy $\mathrm{POS}_{\mathrm{ToF}}$ does not operate at the error correction level, employs an external source of reciprocal information (i.e., not the quantized signal) and it can also be adapted to any key generation scheme involving public exchanges of sample indexes.

3.2 System model

We employ the same system model as in Section 2.2.1, to which we add an eavesdropper situated close to Bob (Eve, denoted as E). The signal received by E during the channel probing between A and B can be expressed as:

$$y_E(t) = (h_{AE} * p)(t) + w_E(t) \tag{3.1}$$

where the channel seen by E, h_{AE}, has a certain level of correlation (presumably small) with the legitimate channel.

Eve is assumed to know all the physical parameters of the transmission (synchronization method, pulse waveform, duration of observation window) as well as the public protocol parameters for key generation. The attacker performs the signal acquisition simultaneously with the legitimate parties, applies the same processing techniques and extracts her own key using the bit extraction algorithm described in Section 2.2.2. She will then eavesdrop on the public exchanges between the two parties and exploit the information as explained in Section 3.5. The missing information will be replaced with randomly generated bits.

This is a basic attack strategy which relies on the unmodified acquired signal and is straightforward to implement. Other more evolved strategies may rely on the received signal to infer the missing bits or even on deterministic ray-tracing predictions given partial knowledge of the physical environment and knowledge of the legitimate positions [96]. The mentioned study [96] concludes that an attacker equipped with a ray-tracing simulator is not able to entirely recreate the legitimate signal and that keys can be generated with the help of the diffuse components. Our focus herein is different in the sense that we use ray-tracing signals for the legitimate users also and we aim to investigate the immunity to a passive physical layer attack when employing only the “less” diffuse part of the signal, which would be a worse-case scenario from the signal point of view. Note that the ray-tracing simulator is used for data acquisition purposes only, not for the attack itself.

Round Trip - Time of Flight and synchronization

The Round Trip-ToF (RT-ToF or simply ToF) is also a reciprocal radio characteristic between A and B and it will be used to improve the secrecy of public exchanges during the reconciliation phase. This quantity is usually measured through cooperative handshake protocols, which are based on the time of arrival of exchanged packets [135]. For our simulations with ray-tracing signals, we will use noisy estimates of the ToF relying on the channel leading edge detection, defined by $t_{y(t)>y_{th}}$ from Eq. (2.10) as introduced in Section 2.2.1. The ToF measured by A and B (T_f^A and T_f^B) will be similar to the ToF measured by E using the A-E link (T_f^E) because we deliberately consider the worst case scenario when the attacker is close to B. For each participant, the measured ToF quantity can be transformed in number of samples :

$$u : N_{ToF}^u = T_f^u \cdot F_s \tag{3.2}$$

with u representing A, B or E.

The accuracy of the level-crossing synchronization depends on the SNR, which impacts both the key generation algorithm, by altering the reciprocity of the samples, and possible ToF-based reconciliation strategies, by altering the ToF estimation.

3.3 Limited public exchanges

We suggest, first of all, a bin table exchange (BIN) instead of the position (i.e., sample index) table exchange in order to reduce the public information. A bin is an interval of several samples that is defined beforehand for both sides.[1]

Instead of exchanging tables containing the indexes of the extracted samples, A and B exchange tables containing the occupancy of each bin, i.e., the number of samples above the final detection threshold in each bin. For illustration purposes, consider an observation window of $W_{obs} = 50$ samples, a bin length of $\Delta_{bin} = 10$ samples, and A's quantized samples with indexes [5, 6, 12, 35, 37, 39, 45] in W_{obs}. The resulting table transmitted by A over the public channel is [2 1 0 3 1]. After the exchange, A and B drop all the bits corresponding to the bins for which the occupancy is different.

3.4 Masked public exchanges

As the previous solution is expected to work in rather high SNR conditions, we suggest another approach for sending information over the public channel by masking or encoding it with a quantity (expressed in number of samples) equivalent to the independently measured ToF. The exchanged encoded tables are:

$$u : \mathcal{E}_u \equiv (\mathcal{P}_u + N^u_{ToF}) \bmod W_{obs} \tag{3.3}$$

with u representing A or B.

In order to recover the original index table of the other user, A and B try to decode $\mathcal{E}$ using their own N_{ToF} and $\mathcal{P}$. We consider two decoding strategies:

- POS^s_{ToF} where A and B assume perfectly symmetric ToF estimates (i.e. $N^A_{ToF} = N^B_{ToF}$) regardless of the actual quality/symmetry of their estimates and decode as follows:

  $$u : \mathcal{D}_v \equiv (\mathcal{E}_v - N^u_{ToF}) \bmod W_{obs} \tag{3.4}$$

  where $(u, v) \in \{(A, B), (B, A)\}$ and $\mathcal{D}_v$ stands for the positions extracted by v and decoded by u.

- POS^a_{ToF} where A and B try to infer the other's N_{ToF} by iteratively checking the values in the neighborhood of their own N_{ToF} in order to compensate for plausible asymmetric measurement biases. For each value, they decode in the

[1] Note that the bin definition is the same as in Section 2.2.3.

same way as before but choose the N_{ToF} that maximizes the number of common extracted indexes, assuming that similar ToF measurements (and thus leading edges' alignment) are likely to generate more common indexes :

$$u : \mathcal{D}_v \equiv (\mathcal{E}_v - \hat{N}^v_{ToF}) \bmod W_{obs} \tag{3.5}$$

where $\hat{N}^v_{ToF} = \underset{N_{ToF}}{\arg\max}\ |\mathcal{P}_u \cap \mathcal{D}_v(N_{ToF})|$.

After estimating each other's index table, A and B keep only the bits from the estimated common indexes. Contrary to POS, where the indexes are directly exchanged, this time there is no guarantee that the legitimate users estimate the common indexes correctly. However, the length of the final sequence is the same on both sides for POS^s_{ToF}, as it can be seen from Eq. (3.6), where $d = N^B_{ToF} - N^A_{ToF}$ and the second equality is due to sets' properties. The same is not always true for POS^a_{ToF} because of the asymmetrical nature in the errors made by A and B when using the estimation in Eq. (3.5). This makes POS^a_{ToF} vulnerable to key length issues even if, at first sight, it seems more robust to measurement noise.

$$|\mathcal{P}_A \cap \mathcal{D}_B| = |\mathcal{P}_A \cap (\mathcal{P}_B + d)| = |(\mathcal{P}_A - d) \cap \mathcal{P}_B| = |\mathcal{D}_A \cap \mathcal{P}_B| \tag{3.6}$$

3.5 Performance evaluation

Attacker strategy

The attacker E in the vicinity of B measures the channel between A and herself ($y_E(t)$) and the ToF on the same link (T^E_f). She obtains a bit sequence from the quantization of $y_E[n]$, which she uses along with the public information to guess the key generated between A and B. We assume that the attacker plays a passive role in the key generation procedure (i.e., she does not inject/modify packets, her final aim being the eavesdropping of the communication between A and B), but she has nonetheless the advantage to exchange packets with A in order to measure the ToF. These measurements along with the proximity to Bob would allow Eve to exploit spatial correlation effects.

In the case of POS, Eve will only keep the quantized bit values corresponding to the common indexes between A and B and guess the missing ones by randomly choosing between equally distributed “0” and “1”. For the BIN protocol, she will divide her observed signal samples into bins, select the bits corresponding to the common bin occupancy of A and B, and guess the missing ones. If her bins do not match the occupancy advertised by A and B, she will just add or delete bits at the end of the bins.

For POS_{ToF} variants, the attacker will either decode the position tables using her own N_{ToF} or try to infer them by maximizing the number of common positions between A and B based on Eq. (3.5). After decoding, she will continue with the same strategy as for POS. The attacker will have no clear indication about the number of bits conserved by A or B for POS_{ToF}.

Simulated signals

The used data are generated by a deterministic IR-UWB ray-tracing simulator using a reference indoor office environment [92]. The simulator gives access to $(x * h)(t)$ for 900 different Tx/Rx positions taken on a regular grid, which are then used to generate $y(t)$ by adding noise. The positions of A and B are chosen randomly on the grid and the position of E as close as possible to that of B (i.e., at 1 m). Ray-tracing signals can help quantifying the impact of spatial correlation on the generated keys because the generated signals depend on both the environment and the Tx/Rx locations. They are, however, more limited in the number of multipath components than statistical channel realizations (Section 2.2) because they mainly capture the most significant reflexion and diffraction phenomena. As expected, the diffuse components are captured less reliably because they usually derive from the details of the building layout or the complexity of the materials.

The sampling frequency of the available ray-tracing data is fixed at approximately 18 GHz and the pulse has a bandwidth of 2.5 GHz at a central frequency of 4.5 GHz. The retained metrics for evaluation are the key agreement ratio, mean successful key lengths, mean bit agreement between legitimate users (mean legal bit agreement), and mean bit agreement ratio with a potential attacker (illegal mean bit agreement), defined according to Section 2.2.4. If the keys have different lengths, the bit agreement ratio is considered to be 0. The standard deviations are computed for the key lengths and the legal/illegal bit agreement ratios.

Results

We compare the initial key generation protocol with position-based public discussion (POS) with the same quantization algorithm followed by the newly introduced variants of the public discussion phase (BIN, POS^a_{ToF} and POS^s_{ToF}). The first tests have been performed using typical algorithm parameters ($\delta = 10$ for the quantization process in all the algorithms and $\Delta_{bin} = 0.25$ ns for BIN[2]).

[2] Δ_{bin} (now expressed in ns for simplicity) corresponds to approximately half of the pulse duration in this case.

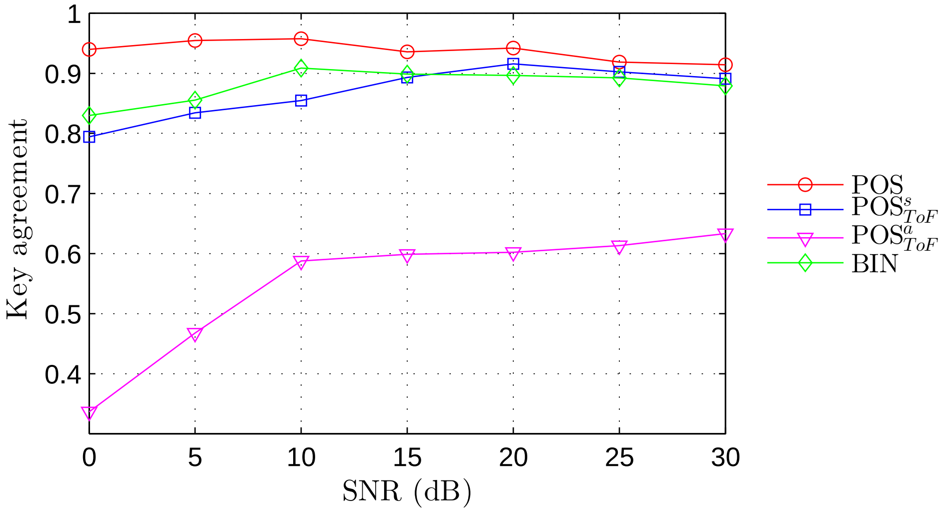


FIGURE 3.1: Key agreement. Parameters: $\delta = 10$ and $\Delta_{bin} = 0.25$ ns

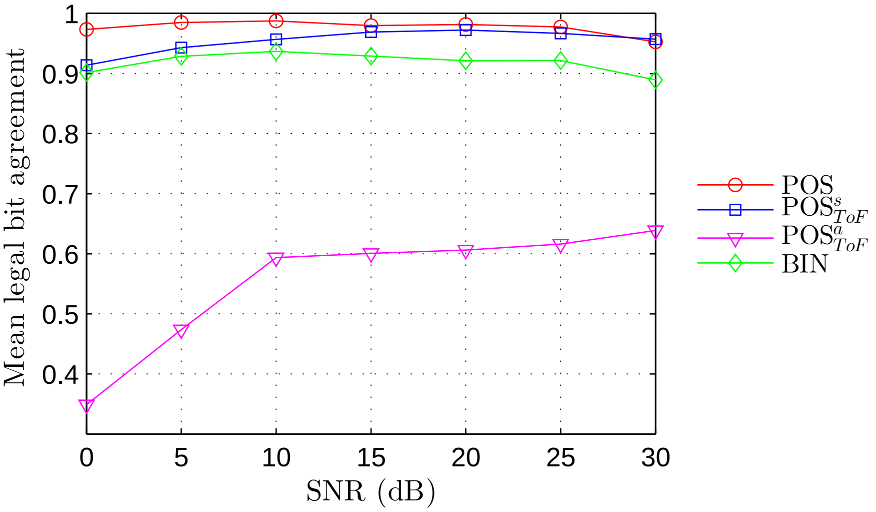


FIGURE 3.2: Mean legal bit agreement. Parameters: $\delta = 10$ and $\Delta_{bin} = 0.25$ ns

Figure 3.1 shows that POS^a_{ToF} is not a reliable solution for generating identical keys, which is later confirmed by the statistics on the legal bit agreement: a small mean value (Figure 3.2) and a large standard deviation with approximately the same order as the mean. This means that maximizing the number of common positions is not a good criterion to estimate the exact ToF of the other user because it leads to keys of different lengths, which are considered to have a bit agreement of 0. This explains the large variance on the legal bit agreement for POS^a_{ToF} while the standard deviations for all the other algorithms are between 0.06 and 0.22.

Both BIN and POS^s_{ToF} improve the mean illegal bit agreement (Figure 3.3 where the associated standard deviations are approximately 0.13 for all the algorithms). POS^s_{ToF} is better than POS and BIN over all SNR values and moreover, its mean value of 0.5 is equivalent to a blind guess of the attacker. It also yields a higher mean successful key length (Figure 3.4) because it does not drop bits by groups like BIN. For SNR values between 5 and 30 dB, the standard deviations of BIN key lengths span from 27 to 48 bits while those for POS^s_{ToF} are around 40-42 bits. Since the key length statistics (mean and standard deviation) are the same for POS and POS^s_{ToF}, we conclude that the key lengths depend more on the quantization algorithm and the channel itself than on the public discussion.

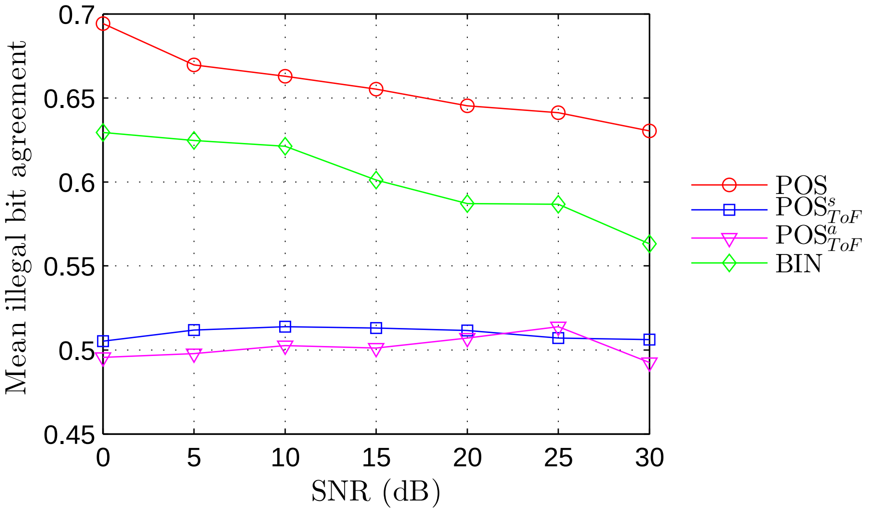


FIGURE 3.3: Mean illegal bit agreement. Parameters: $\delta = 10$ and $\Delta_{bin} = 0.25$ ns

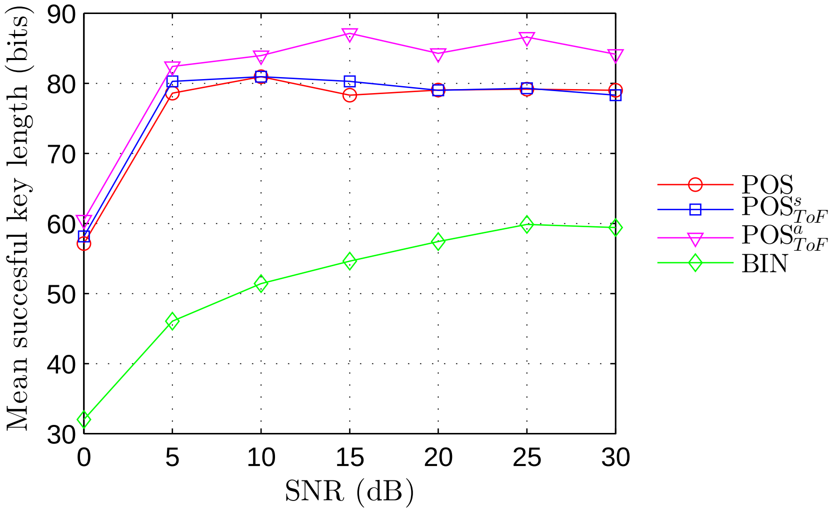


FIGURE 3.4: Mean successful key length. Parameters: $\delta = 10$ and $\Delta_{bin} = 0.25$ ns

Finally, we look at the effect of δ and Δ_{bin} on the performance of the reconciliation schemes at median SNR (15 dB). When increasing δ from 5 to 20, which corresponds to the extraction of more bits because of a lower final detection threshold, the key agreement ratio decreases almost linearly for POS, POS^s_{ToF} and BIN variants (Fig 3.5). When analyzing supplementary values of Δ_{bin} between 0.1 ns and 0.65 ns, we conclude that the key agreement ratio for BIN changes its behavior: it decreases almost linearly with δ at low Δ_{bin} and it presents an optimal δ for higher Δ_{bin}. The lower key agreement values at high Δ_{bin} and low δ are due to the empty-key phenomenon (i.e., after the public discussion phase, there are no common bins left).

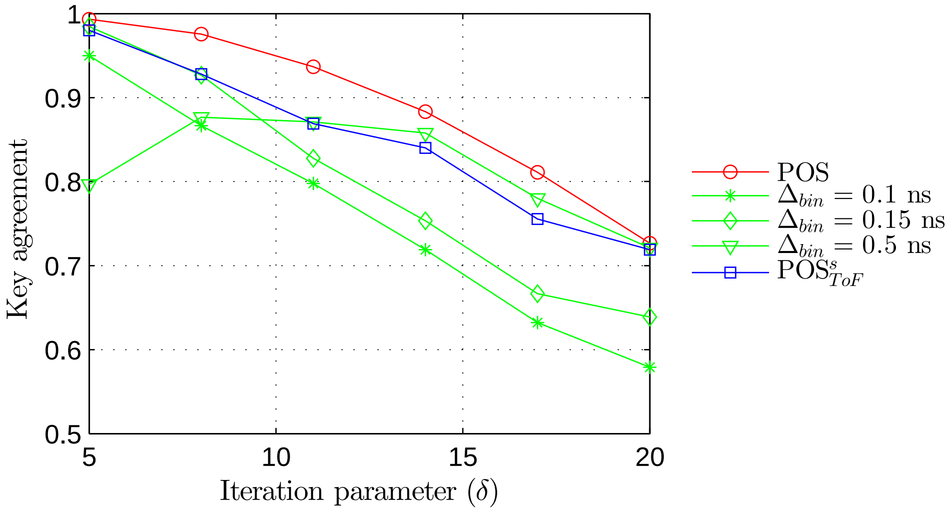


FIGURE 3.5: Key agreement at SNR=15 dB

The mean illegal bit agreement for POS and the BIN variants decreases with δ (Figure 3.6). The differences between POS and BIN are, however, smaller than the standard deviation of the samples which can be between 0.2 and 0.1. POS^s_{ToF} presents similar standard deviations but its mean is always approximately 0.5, which shows that it is a more efficient public exchange method for all δ values.

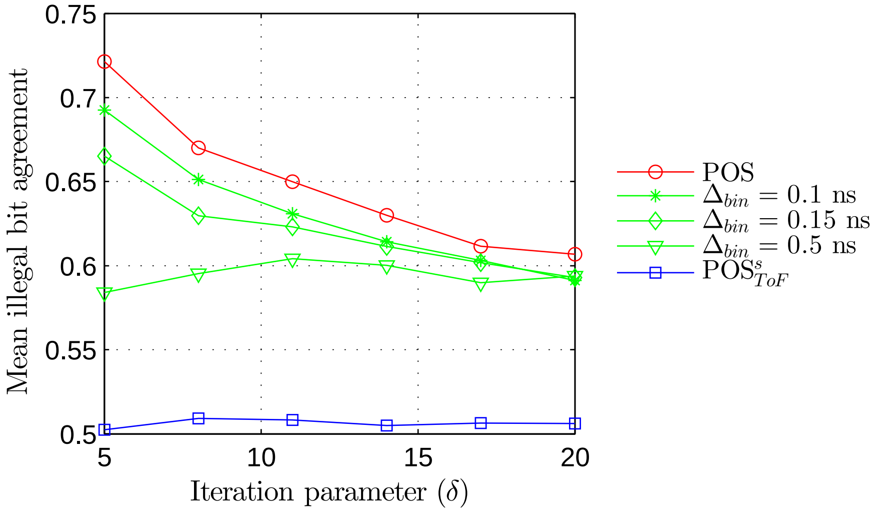


FIGURE 3.6: Mean illegal bit agreement at SNR=15 dB

3.6 Summary

In this section, we showed that signal spatial correlation and public information can help a passive attacker guess parts of the shared secret key between two legitimate users quantizing a directly sampled IR-UWB channel response according to [79] under more realistic synchronization assumptions. Among the proposed more discrete reconciliation methods, BIN can be easily adapted to the SNR conditions by varying the bin size, but it is less efficient at large bin sizes because it implies dropping bits by large groups. POS^{s}_{ToF} proves to be efficient (i.e., in terms of key length and also for achieving an optimal mean illegal bit agreement of 0.5) since it only hides the publicly exchanged information by using the ToF as a mask. Moreover, none of these methods increases the public discussion overhead in terms of exchanged packets. Since we use ray-tracing data for our simulations, our results would be valid for realistic IR-UWB signals from simple environments with limited multipath and solutions such as BIN or POS_{ToF} would be useful for adding a supplementary layer of protection against eavesdropping in these cases.

Chapter 4

Cooperative physical layer key generation

The previous chapters focused on the exploitation of the physical radio layer to generate symmetric keys over single point-to-point links. The channel can be also probed over successive transmissions for obtaining longer keys by concatenation but the performance is dependent on the channel coherence time. Therefore, an elementary issue for physical layer key generation is how to gather more entropy from channel measurements in a given static scenario. This leads to the idea of extending the key generation process to several nodes in order to exploit more physical links (cooperative/multi-link strategy) and/or to generate a group key (contrary to a point-to-point key). From a practical point of view, the solutions should be scalable, adapted to *ad hoc* scenarios and should avoid the high level complexity issues of classical key distribution techniques (e.g., key pre-distribution, numerous packet exchanges, etc.).

Some of the key generation models have been adapted to cooperative scenarios involving several nodes, either to reinforce the generated pairwise keys or to issue a common group key (i.e., shared by more than two nodes). In this chapter, we investigate key generation from IR-UWB multipath channels according to the source model. We propose a new method to generate group keys within cooperative scenarios, while exploiting all the available physical links in a full mesh topology and reducing over-the-air traffic. The main idea consists in adjusting the IR-UWB signals usually transmitted for channel probing so that a receiving node has access to non-observable channels corresponding to its non-adjacent links.[1] This operation leads to a deconvolution problem, for which we investigate various solutions.

[1] Non-adjacent links are links between other nodes in the network

The present chapter starts with an overview of the main cooperative key generation methods. Then, a general state of the art protocol named CKD (Cooperative Key Distribution) is presented for comparison purposes with the the proposed method entitled CKG (Cooperative Key Generation). For the latter, we detail the signal processing algorithms involved in computing the optimized transmitted signal. Firstly, we show that the maximum-likelihood (ML) approaches to the deconvolution problem are not stable with respect to imperfect channel estimates. Then, we introduce a parameterized maximum *a posteriori* (MAP) solution and we analyze two automatic methods for parameterization: Cross Validation (CV) and Expectation Maximization (EM). Finally, the protocol is analyzed in terms of traffic complexity and the key length gains are assessed through simulations.

4.1 State of the art: cooperative key generation

The secret key capacities for the source model with multiple terminals including a subset of helpers, various extents of an eavesdropper's knowledge, and unrestricted public discussion are characterized in [141]. First, the secret key capacity when the eavesdropper only observes the public exchanges without having any side information regarding the source is shown to be closely related to the multiterminal source coding problem with no secrecy constraints. Then, the expression of the secret key capacity when the eavesdropper wiretaps a subset of the helpers is also derived.

Similar studies [142] [143] extend the source emulation channel model to include a public channel eavesdropper and multiple users. The latter can be classified as: i) active users that want to generate a secret key per group [142] [143] or per sub-group [142]); ii) untrusted helpers, whose channel inputs and/or outputs can be wiretapped by the eavesdropper; iii) trusted helpers. The active users can be either senders or receivers [142] or both [143]. The results consist in lower and upper bounds for the secret key capacities.

Multiterminal or cooperative secret key generation has also been investigated for less complex systems with the aim to design practical protocols and to measure their performance.

Early work from [144] presents an extension of the source model to cooperative pairwise key agreement and group key generation in a pairwise independent network (i.e., a network in which the point-to-point channels are independent). Point-to-point keys are generated from each physical network link and the group keys (or extra secret bits for pairwise cooperative keys) are propagated through XOR-ing operations over a graph representation of the network.

In more recent studies [145] [146], the secret key between two user nodes is generated with the help of a relay. First, non-cooperative pairwise keys (from the main channel and the side channels, i.e., the channels between each node and the relay) are generated using a typical key distillation procedure based on channel gains [145] or phase estimations [146]. At this point, each user node holds the key from the main channel and a key from a side channel. After that, the relay, which holds the keys obtained from both side channels, publicly broadcasts the bitwise combination of those keys. The user nodes can therefore recover the key that was generated by the other user node and the relay and append one of these side channel keys to the main channel key. An equivalent approach based on pairwise keys is described in [147], where the received signal strength is quantized for non-cooperative pairwise key generation and a group key, generated by a root node, is securely distributed in the network using the pairwise keys.

If the wireless channels are static in time or too sparse in terms of multipath, there might not be enough information to harvest in order to generate a robust secret key. In order to deal with the issue of limited entropy of the source model, a recent study [148] extends the channel model for key generation to a cooperative scenario with a relay and an eavesdropper that is collocated with the relay. The authors derive the upper and lower bound for the secret key rate with a relay and propose a joint optimized design of the different key generation phases (advantage distillation, information reconciliation and privacy amplification) while focusing on the trade-off between security and protocol efficiency. Although it is shown that collocating with the relay is the worst case scenario for the secret key rate, this assumption also facilitates the advantage distillation phase because the legitimate users are informed about the quality of the eavesdropper's signal by the cooperative relay.

Overall, the previous contributions on the source model with cooperative pairwise keys or with group keys rely on the initial pairwise single-link key generation and subsequent key distribution. This involves extra-traffic and latency, while the length of the group key is limited by the shortest pairwise keys. Herein, we describe a novel method of physical layer group key generation that avoids the pairwise key generation before group key generation. An alternative way to deal with the entropy limitation in the case of the source model would be to use multiple links as input signals for quantization. The final group key is obtained after quantization and reconciliation of a concatenation of measurements from several links. The protocol involves several cooperative nodes and the obtained key is by construction known to all the participants, so it becomes a group key. This solution avoids extra public communication overhead but comes with an expense on the signal processing side where supplementary operations for channel probing are needed.

The "probing" of a non-adjacent channel is possible because of the cooperation between nodes, which will send specific signals (called *s-signals* in the following) in order to "induce" a certain channel state observed at the receiver. This operation depends on the type of measured signal; in the case of IR-UWB channel responses, which we will use for illustration purposes, the operation consists in a deconvolution operation, which is one of the scenarios with the highest complexity. This can be nonetheless acceptable for current personal devices such as smartphones and possibly even for next generation wireless sensors. The general concept and the protocol are however applicable to different technologies and channel measurements, which would require less complex signal processing capabilities.

Similar concepts (e.g., IR-UWB time-reversal [149]) have been put forward for improved communication robustness and intrinsic signal secrecy by spatial focusing of the signal energy. These methods rely on pre-filtering on the transmitter side and thus, enable location-dependent SNR gains on the receiver side. However, they are neither intended to provide secret material, such as keys, to higher layer cryptographic functions nor used in cooperative protocols as described in our method.

4.2 System model

We consider the simplest example of full mesh topology, consisting of three nodes (A, B, and C) with direct IR-UWB links between each pair. The received signal can be expressed according to Eq. (4.1)-(4.2).

$$y_{uv}(t) = (p * h_{uv})(t) + w_{uv}(t),\ w_{uv}(t) \sim \mathcal{N}(0, \sigma_w^2) \tag{4.1}$$

$$h_{uv}(t) = \sum_{k=1}^{K} x_k \delta(t - \tau_k) \tag{4.2}$$

where $y_{uv}(t)$ of duration T_w is the convolved noisy channel response between the transmitter $u \in \{A, B, C\}$ and the receiver $v \neq u, v \in \{A, B, C\}$, $p(t)$ is the transmitted pulse waveform of duration T_p, $h_{uv} = h_{vu}$ is the reciprocal channel impulse response between u and v, x_k and τ_k are respectively the amplitude and delay associated with the k^{th} multipath component ($k \in \{1, ..K\}$), and $w_{uv}(t)$ is the AWGN.

We arbitrarily define the SNR as the ratio between the power of the transmitted pulse and the noise power, assuming that all the CIRs are normalized ($\int_0^{T_w} h^2(t)\,\mathrm{d}t = 1$):[2]

$$\mathrm{SNR} = \frac{P_{pulse}}{P_{noise}} = \frac{\frac{1}{T_p}\int_0^{T_p} p^2(t)\,\mathrm{d}t}{\sigma_w^2} \tag{4.3}$$

The three channels (h_{AB}, h_{AC}, and h_{BC}) are considered independent two by two at a given channel probing time. Also, we assume that they vary from one channel probe to the next one (independently of the past channel realizations).

4.3 Cooperative key distribution

One way of extending the point-to-point source model for key generation to several links and to a group key is to generate pairwise keys on each link and then distribute a group key (similarly to [144]). CKD can be achieved in two phases.

- Each node generates a pairwise symmetric key with each of its neighbors based on the properties of the radio channel. The employed key generation method for the current evaluations will be described in Section 4.4.1. Contrary to CKG, the key generation for CKD is performed separately for each pair of nodes using only their corresponding channels.
- A group key, generated by a lead node using a random number generator, is propagated in the network by XOR-ing operations using the previous single-link keys. The security of the scheme relies therefore on the single-link keys.

It is not compulsory to have pairwise keys on all the links but at least on a sufficient number of links that can form a spanning tree over the given network. The length of the group key is limited by the length of the pairwise keys in the following manner: if we consider that there is a pairwise key between the lead node and each other node and we want to avoid additional traffic, the length of the group key is the minimum of the lengths of these keys. If not, the group key length is limited by the minimum key length in the most “advantageous” spanning tree (i.e., the spanning tree containing the links with the largest key lengths). In this case, additional traffic between nodes is needed in order for the lead node to be able to determine the group key length before initiating the distribution procedure.

[2] We adopt a new SNR definition based on the pulse duration because in this particular study we deal with various lengths for the observation windows: one for adjacent channel probing and another for non-adjacent channel acquisition. This aspect is however transparent in the protocol because of a supplementary windowing process.

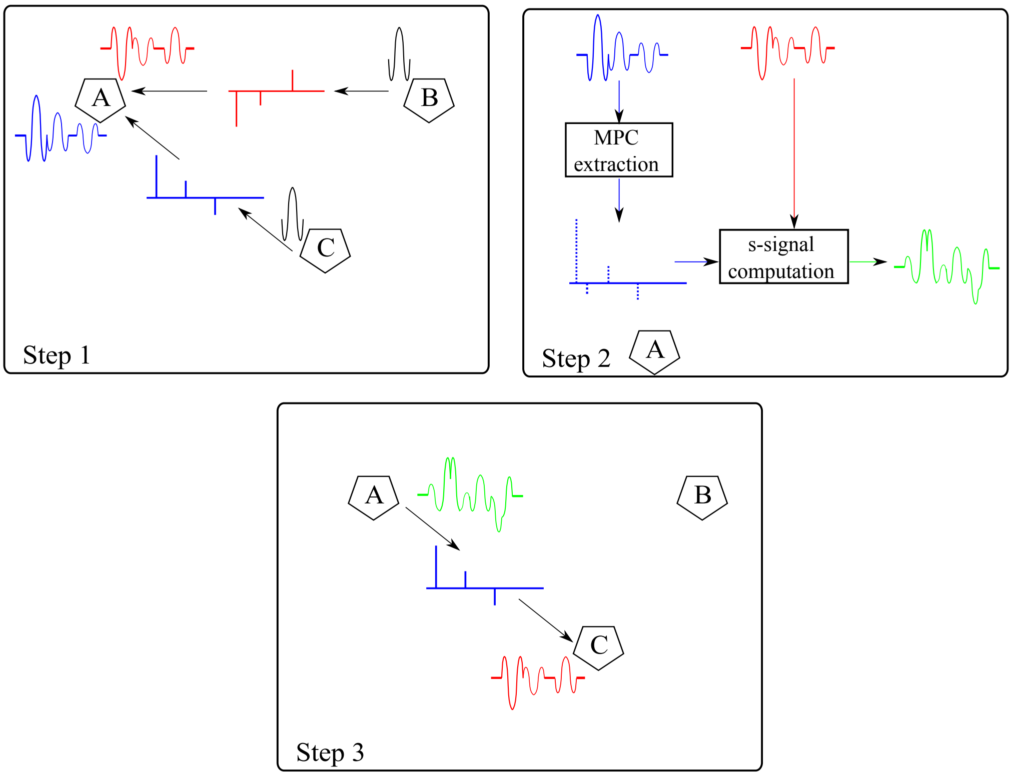


FIGURE 4.1: Physical layer cooperative key generation

4.4 Cooperative key generation

This section describes the proposed cooperative key generation method for IR-UWB signals [100] in three steps: description of the protocol sequence including a deconvolution operation (Section 4.4.1), focus on the deconvolution problem from a statistical estimation perspective and on state of the art approaches for statistical inference (Section 4.4.2), and performance evaluation of the selected methods for deconvolution (Section 4.4.3-4.4.4).

4.4.1 Protocol description

In the following description A, B, and C have different roles and the protocol must be repeated with interchanged roles in order to obtain a group key. We define A as the cooperator, B as the initiator and C as the generator. We also refer to an adjacent channel as the channel that can be directly probed by a node (e.g., channels $[B - A]$ and $[C - A]$ for node A). The CKG protocol consists of several steps:

- channel sounding using a pulse template signal to obtain adjacent channel responses ($y_{BA}(t)$ and $y_{CA}(t)$) for the cooperator A according to Eq. (4.1) (step 1 in Figure 4.1).

- CIR estimation (i.e., delay and amplitude extraction) at the cooperator A from the received y_{CA} (step 2 in Figure 4.1). This operation can be achieved by means of sampling at sampling frequency F_s and correlation with an *a priori* pulse template (corresponding to the expected unitary received pulse waveform $p(t)$). The used search-subtract-readjust algorithm [110] iteratively detects, estimates and then subtracts multipath components out of the acquired received signal y_{CA} to obtain $\hat{h}_{CA}$.

- computation of the so-called *s-signal* s_{AC} at the cooperator A (step 2 in Figure 4.1). The previous channel estimate $\hat{h}_{CA}$ is used by the cooperator A to compute s_{AC} to be sent to the generator C so that the latter deduces the non-adjacent channel $[B - A]$ represented by y_{BA}. The following problem needs to be solved:[3]

$$\text{Find } s_{AC} \text{ s.t. } (s_{AC} * \hat{h}_{CA})(t) = y_{BA}(t) \tag{4.4}$$

- transmission of the computed s-signal s_{AC} from A to C (step 3 in Figure 4.1). Accordingly, the employed transmitter must enable the programming of an arbitrary IR-UWB waveform, given an *a priori* occupied bandwidth (e.g., [94]), in order to generate s_{AC}. The generator C receives:

$$r_{BA}(t) = (s_{AC} * h_{AC})(t) + w_{AC}(t) \tag{4.5}$$

  Steps 1-3 can be performed by both A and B because they can both measure C's non-adjacent reciprocal channel $[B - A] \approx [A - B]$. If both of them send an s-signal, supplementary processing can be used at C to coherently take advantage of the two incoming signals r_{BA} and r_{AB} in order to obtain a more reliable version of the non-adjacent channel.

- processing of all the acquired signals at the generator C. At this point, C detains signals corresponding to all three channels. These signals are further processed (e.g., through windowing, signal squaring, low-pass filtering and down-sampling at frequency F_p, compatible with the multipath resolution capability allowed by the signal bandwidth). The generator C then concatenates the three signals to obtain the input quantization signal S_{CAB}. The concatenation is applied according to an arbitrary order without loss of generality.

- quantization of the input signal S_{CAB} using, e.g., uniform quantization with guard-bands.

[3] Although the equations are initially written in the continuous analog domain for more generality, the channel estimation and the computation of s-signals are solved in the discrete domain (See Section 4.4.2).

The steps 1-4 are repeated by all the nodes with interchanged roles so that all of them obtain the three channels.

- public discussion between the three nodes: i) sharing the indexes of the dropped samples falling into the guard-bands; ii) error correction using Reed-Solomon codes (a lead node, for example A, generates a syndrome representing its own bit sequence and sends it over the public channel to the other nodes, which will try to decode/correct their bit sequences to align them to A's sequence).

4.4.2 Parameterized s-signal computation

Searching for a solution to Eq. (4.4) is a non-trivial deconvolution problem. We will first explore straightforward solutions[4] and then, focus on the more general problem of s-signal estimation implying two levels of statistical inference (model fitting and model selection) separately or jointly.

Deconvolution solutions

As mentioned before, A needs to solve the following equation :

$$\mathbf{\hat{H}_{CA}s} = \mathbf{y_{BA}} \tag{4.6}$$

where $\mathbf{y_{BA}}$ is the $N \times 1$ sampled version of y_{BA}, $\mathbf{\hat{H}_{CA}}$ is the $N \times N_s$ matrix corresponding to the $N_h \times 1$ convolution kernel $\mathbf{\hat{h}_{CA}}$ such that $\mathbf{\hat{H}_{CA}s} = \mathbf{\hat{h}_{CA}} * \mathbf{s}$.

Considering the classical convolution definition, N_s should be $N - N_h + 1$. This leads to an overdetermined system of equations, which is consistent (i.e., has one or an infinity of solutions) only if a certain number of equations are linear combinations of the rest of the equations. As the coefficients from $\mathbf{\hat{H}_{CA}}$ are random by construction, the present system has high chances of being inconsistent. Consequently, we give the s-signal more degrees of freedom by imposing that only the valid part of the convolution[5] approaches $\mathbf{y_{BA}}$, which implies that $N_s = N + N_h - 1$. Eq. (4.6) becomes then an underdetermined system, which can have zero or an infinity of solutions.

Therefore, we replace the search for an exact solution to Eq. (4.6) with a least-squares (LS) minimization problem (Eq. (4.7)), whose solution will be denoted as ML (maximum

[4] We restrict our analysis to the natural temporal domain of IR-UWB signals in order to avoid supplementary processing incurred by the Fourier Transform, but also because of the richness of data processing techniques concerning deconvolution in similar domains (e.g., statistical spatial methods for image deconvolution).

[5] The samples of the central part of the convolution, where the two input signals overlap entirely. These samples are obtained from the summing of $\min(N_s, N_h)$ non-zero terms.

likelihood[6]). The LS problem is still ill-posed in the sense that it could have multiple solutions.

$$\mathbf{s}_{\mathbf{AC}}^{\mathbf{ML}} = \underset{\mathbf{s}}{\operatorname{argmin}} \; ||\hat{\mathbf{H}}_{\mathbf{CA}}\mathbf{s} - \mathbf{y}_{\mathbf{BA}}||^2 \tag{4.7}$$

We quantify the performance of the deconvolution operation with the residual error (RMSE) between the analog signals $(s_{AC} * h_{AC})(t)$ and $y_{BA}(t)$. Note that the receiver should perform additional windowing operations since the sent s-signal is longer than the needed observation window. This RMSE metric includes the degradation produced by the imperfect channel estimation $\hat{\mathbf{h}}_{\mathbf{CA}} \neq \mathbf{h}_{\mathbf{CA}}$. Additional artefacts are incurred by the additive noise from the $[A - C]$ channel. Even though these degradations are not directly considered for the computation of the s-signal, their effect is taken into account in the performance evaluation from Section 4.5.

In terms of implementation, Eq. (4.7) can be solved using solutions similar to MATLAB®'s linear least-squares solver *mldivide*, which employs QR decomposition on matrix $\hat{\mathbf{H}}_{\mathbf{CA}}$ and provides a solution $\mathbf{s}$ with the fewest possible non-zero components. After computing the s-signal using the aforementioned method on sampled signals at $F_s = 10$ GHz, we simulate the pseudo-analog s-waveform $s_{AC}(t)$ by sinc-interpolation.[7] The obtained waveform is next convolved with the true multipath channel $h_{AC}(t)$. In Figure 4.2, we show the noiseless received signal in C and the target signal to be deduced, $y_{BA}(t)$, for arbitrary realizations of CM1 channels (according to the IEEE 802.15.4a standard). All the received and target signals are normalized with respect to their maximum absolute value. We will refer to these particular channel realizations as "the basic channel configuration" hereafter. A value of SNR = 20 dB is employed for all the evaluations in the present section.

Therefore, we observe that the ML solution can be unstable with respect to the imprecision of channel estimations. This means that the criterion of fewest non-zero elements is not adapted to our problem and another type of least-square solution should be preferred. A simple preliminary test using MATLAB®'s Moore-Penrose pseudo-inverse function *pinv*, which produces a solution $\mathbf{s}$ with minimal l_2-norm, shows significant improvements at the expense of longer computation times. As it relies on singular value decomposition, the computation of the pseudo-inverse is deemed to add unnecessary complexity to the resolution of a linear system.

[6] LS and ML estimators are equivalent in the case of a noisy data model with Gaussian noise, which will be the case for our data model described in the subsequent subsections.

[7] The pseudo-analog waveforms are simulated at a simulation frequency of 100 GHz.

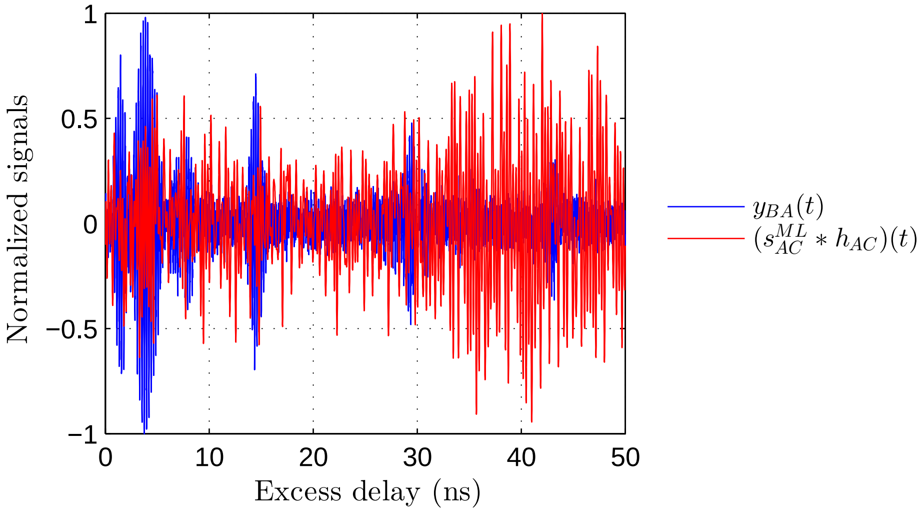


Figure 4.2: Example of deduced signal with the ML method (RMSE = 29.8%)

The instability issue of the ML solution can be efficiently addressed by adding a penalty term to Eq. (4.7) and solving it, for example, with *mldivide*-like methods. Note that if $\mathbf{P} = \mathbf{I_{N_s}}$, the solution of Eq. (4.8) is equivalent to the aforementioned *pinv* solution with minimal l_2-norm, where λ is a fixed parameter.

$$\mathbf{s_{AC}} = \underset{\mathbf{s}}{\text{argmin}} \; ||\mathbf{\hat{H}_{CA}s} - \mathbf{y_{BA}}||^2 + \lambda||\mathbf{Ps}||^2 \tag{4.8}$$

This operation is known as a Tikhonov regularization, where matrix $\mathbf{P}$ is chosen in order to constrain $\mathbf{s}$ and λ is a real scalar trade-off parameter. The penalty term, also called prior in a Bayesian setting, enforces the desired characteristics of the optimized s-signal (e.g., minimal l_2-norm, smoothness, etc.), while the first term keeps the result after convolution close to the data. Eq. (4.8) has a closed form solution, which we will denote as MAP (maximum *a priori*):

$$\mathbf{s^{MAP}_{AC}} = (\mathbf{\hat{H}^{\mathit{T}}_{CA}\hat{H}_{CA}} + \lambda\mathbf{P}^T\mathbf{P})^{-1}\mathbf{\hat{H}^{\mathit{T}}_{CA}y_{BA}} \tag{4.9}$$

Figure 4.3 shows the normalized deduced signal with the MAP solution for the computation of the s-signal (basic channel configuration). For this illustration, a value of $\lambda = -4$ dB is chosen. In the next subsections, we investigate methods for finding an adapted value for this parameter.

The signal instability observed for the ML solutions can be regarded as an overfitting issue: the s-signal is computed based on the imperfect channel estimation, $\mathbf{\hat{h}}$, but the final performance depends on the unknown real channel, $h(t)$. This means that even though the ML deconvolution solution is exact for the given data (channel estimates

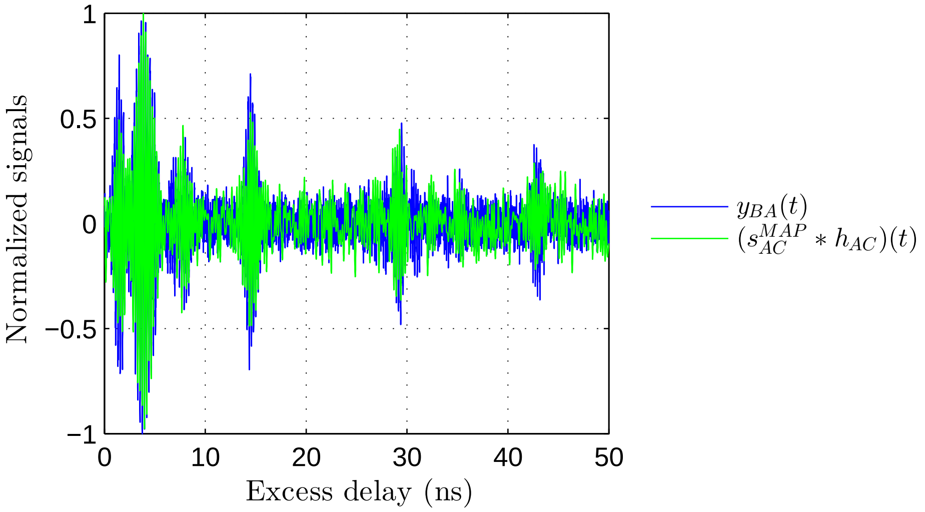


Figure 4.3: Example of deduced signal with the MAP method (RMSE = 9.8%)

and target channel), it could behave unpredictably for slightly different real channel conditions (equivalent to variations in the channel estimates). Adding the penalty term avoids the overfitting if its weight is properly chosen.

From deconvolution to inference

The deconvolution can be seen as a generalization of an interpolation problem (i.e., interpolation is a deconvolution with a kernel consisting of a Dirac delta function). The tutorial paper on Bayesian interpolation [150] uses the noisy data interpolation problem to illustrate the principles of Bayesian model selection, regularization and noise estimation and to compare them to classical (frequentist) misfit or cross-validation techniques. We first summarize the approach presented in the aforementioned work and then, describe a more recent alternative for statistical estimation problems with incomplete data, namely the Expectation Maximization algorithm.

Like other data modeling processes (e.g., pattern classification, clustering, detection, etc.), data interpolation or signal estimation[8] can consist in a statistical inference procedure with (minimum) two phases [150]:

- signal estimation based on a specified model with known parameters (also called first level of inference). This phase can be solved by Bayesian (e.g., MAP) or non-Bayesian (e.g., ML) estimators. In the Bayesian case, the fundamental relation

[8] Equivalent terminologies for interpolation include regression, curve-fitting, learning.

between the probability density functions is expressed as:

$$\begin{aligned} \text{Posterior} &= \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} && (4.10) \\ p(s|D,\mathbf{\Theta}) &= \frac{p(D|s,\mathbf{\Theta}) \times p(s|\mathbf{\Theta})}{p(D|\mathbf{\Theta})} && (4.11) \end{aligned}$$

where D is the known data, s is the searched signal and Θ represents the model parameters.

- model comparison/selection, which reduces to a parameter tuning phase if the aim is to find an optimal value for the parameters of a given model. Classical techniques include hypothesis testing, misfit criteria[9] or cross-validation (described in Section 4.4.3). The Bayesian "evidence maximization" method treats this step as a second level of interference, the parameters becoming the new quantity to be estimated (e.g., by ML or MAP estimators). Maximizing the evidence $p(D|\mathbf{\Theta})$ is equivalent to finding the parameters that can best "explain" the observed data with respect to the chosen model.

The two levels of inference are performed independently: first, estimation of the parameters based on evidence maximization and then, signal estimation with the obtained parameters.

Alternatively, the estimation of a signal using a model with unknown parameters can be solved by an Expectation Maximization (EM) algorithm. EM has been discovered and used independently in several domains ranging from genetics, statistics (estimation of parameters of mixture distributions) to maximum likelihood image reconstruction and speech recognition (estimation of parameters of Hidden Markov models) [151]. Often thought as an evolution of the maximum likelihood estimator, EM can be simply illustrated through the example of the maximum likelihood parameter estimation with incomplete data [152]. For simplicity, we will restrict to this example[10] and then, show how our deconvolution problem can be seen as an incomplete data model.

Given some data D sampled from a model with unknown parameters $\mathbf{\Theta}$, the ML estimator returns the parameters $\mathbf{\Theta}^{ML}$ that maximize the likelihood function $p(D|\mathbf{\Theta})$.[11] Most of the time, the model involves hidden variables S (e.g., $D = f(S, \mathbf{\Theta})$), which cannot be efficiently taken into account by the ML estimator on its own, but which can be naturally included in the EM framework. This algorithm alternates between solving

[9] The misfit metric represents the gap between the known data and the statistics of the predicted data with the unknown model parameters. The model parameters are derived from solving the equation that sets the misfit metric to a fixed value.

[10] A more detailed description of EM and its properties can be found in [151], [153], [154].

[11] This quantity represents the likelihood function with respect to the parameter estimation problem and also the evidence with respect to the signal estimation problem considered above.

for the hidden variables, knowing the latest parameter estimates and the observed data, and finding the optimal parameters, knowing the current hidden variable estimations and the observed data. Contrary to the level-based statistical inference, EM condenses the two inference levels in an iterative procedure with two steps for each iteration i:

- the Expectation step (E-step): computation of the conditional probability distribution $p(S|D, \mathbf{\Theta}_{i-1})$[12] and of the conditional expectation $\xi_i(\mathbf{\Theta})$.

$$\xi_i(\mathbf{\Theta}) = \mathbb{E}_{S|D,\mathbf{\Theta}_{i-1}}[\ln p(D, s|\mathbf{\Theta})] \tag{4.12}$$

- the Maximization step (M-step):

$$\mathbf{\Theta}_i = \arg\max_{\mathbf{\Theta}} \xi_i(\mathbf{\Theta}) \tag{4.13}$$

The convergence can be proved by showing that the algorithm increases the likelihood at each iteration [152].

The deconvolution problem presented in the previous subsection can be seen as a signal estimation problem with unknown model parameters or, equivalently, a parameter estimation problem (i.e., the regularization parameter or the weight of the prior) with hidden variables (i.e., the searched s-signal). In order to apply Bayesian model selection techniques, we create a statistical model corresponding to the regularized deconvolution equation Eq. (4.8).

Statistical deconvolution model

We identify the known data as the target signal ($\mathbf{y_{BA}}$) and the hidden data as the searched s-signal ($\mathbf{s}_{AC}$). The estimated channel ($\hat{\mathbf{H}}$) is a deterministic fixed quantity in this model. For simplicity and generalization purposes, we will drop the signal indexes representing the users for the rest of Section 4.4. The model consists of two equations: one for the data fit, where $\mathbf{e}$ represents the fitting error with independent samples of mean 0 and variance ϵ^2, and one for the signal prior, which offers an artificial representation of the resulting waveform as a noisy zero-mean process of sample variance γ^2.

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{s} + \mathbf{e},\ \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \epsilon^2\mathbf{I_N}) \tag{4.14}$$

$$\mathbf{Ps} \sim \mathcal{N}(\mathbf{0}, \gamma^2\mathbf{I_{N_s}}) \tag{4.15}$$

[12] Note that the E-step provides an estimation of the probability distribution over the hidden variables, so implicitly the mean and the associated uncertainties.

Therefore, we can write Bayes' rule for the given model:

$$p(\mathbf{s}|\mathbf{y}, \epsilon, \gamma) = \frac{p(\mathbf{y}|\mathbf{s}, \epsilon) \times p(\mathbf{s}|\gamma)}{p(\mathbf{y}|\epsilon, \gamma)} \tag{4.16}$$

By identification between Eq. (4.8) and the MAP estimator derived from the current model,[13] we have $\epsilon^2/\gamma^2 = \lambda$. Although the MAP solution expressed in Eq. (4.9) depends only on the value of λ, the present model provides a richer description of the underlying phenomena because the two model parameters (ϵ and γ) have a concrete meaning (i.e., ϵ represents the capacity of the model to fit the known data and γ can represent, for example, the energy of the searched signal).

We have therefore several solutions for our deconvolution problem:

- a MAP solution (solved by an LS minimization algorithm) with two choices of parameterization:
  - a weight parameter between the data fit and the prior (λ from Eq. (4.8)), which can be set using a cross-validation technique (Section 4.4.3).
  - Bayesian parameterization based on ϵ and γ, which can be estimated with an evidence ($p(\mathbf{y}, \hat{\mathbf{H}}|\epsilon, \gamma)$) maximization procedure [150]. We will not evaluate this option in our simulations because of the complexity of the evidence function in terms of optimization (see Appendix C).
- an EM solution, which provides a joint estimation of the parameters of the Bayesian model and of the s-signal (Section 4.4.4).

4.4.3 MAP solution with Cross Validation parameterization

The MAP solution depends on the choice of the trade-off parameter λ and of the prior matrix $\mathbf{P}$. Based on the empirical observations of the signal aspect at $F_s = 10$ GHz during the preliminary tests, we choose $\mathbf{P} = \mathbf{I_{N_s}}$,[14] corresponding to a minimization of the signal energy for regularization purposes. A parametric study on λ shows how the performance of the MAP solution depends on the trade-off parameter. In Figure 4.4, we represent the averaged RMSE (over 100 channel configurations) between the normalized noiseless received signal at C, $(s * h)(t)$, and the normalized target signal, $y(t)$. We also plot the standard deviations around the respective RMSE means with dotted curves.

[13] $\mathbf{s}^{\mathbf{MAP}} = \arg\max_{\mathbf{s}} p(\mathbf{y}|\mathbf{s}, \epsilon, \gamma) \times p(\mathbf{s}|\epsilon, \gamma)$

[14] A prior matrix corresponding to a differential kernel $[\mathbf{1}, -\mathbf{1}]^T$ gives similar results, but it is more adapted at higher sampling frequencies, where the signal can be considered smooth.

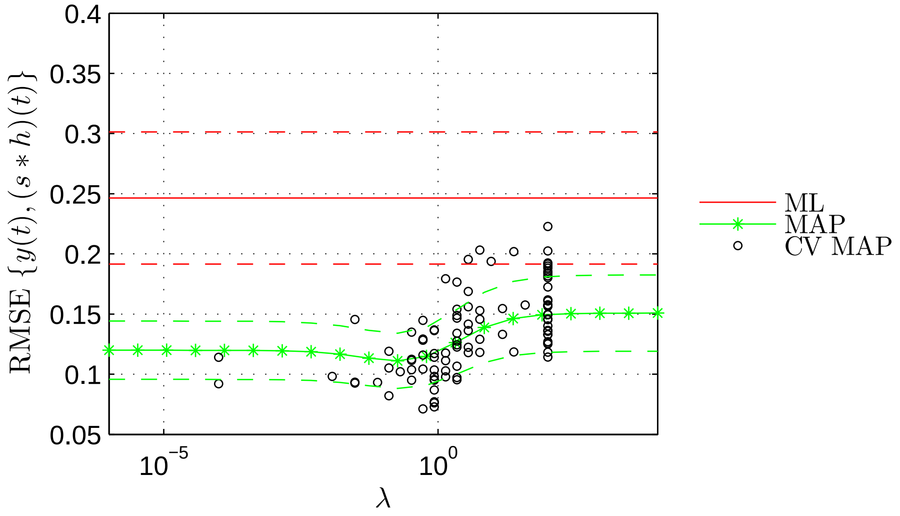


FIGURE 4.4: Non-adjacent signal reconstruction performance for 100 channel configurations (ML, MAP)

We conclude that at small λ values, the regularization works as intended, reaches an optimal value for λ values around 0 dB, but degrades after a certain upper threshold, when the prior obtains too much weight and "flattens" the signal in its initial form ($\mathbf{s} \to \mathbf{0}$). Although the estimated s-signal can be amplified or the received signal can be normalized, as it is the case, the received signal with unadapted λ contains more artefacts, as it can be observed for $\lambda = 2$ dB in Figure 4.5 (basic channel configuration).

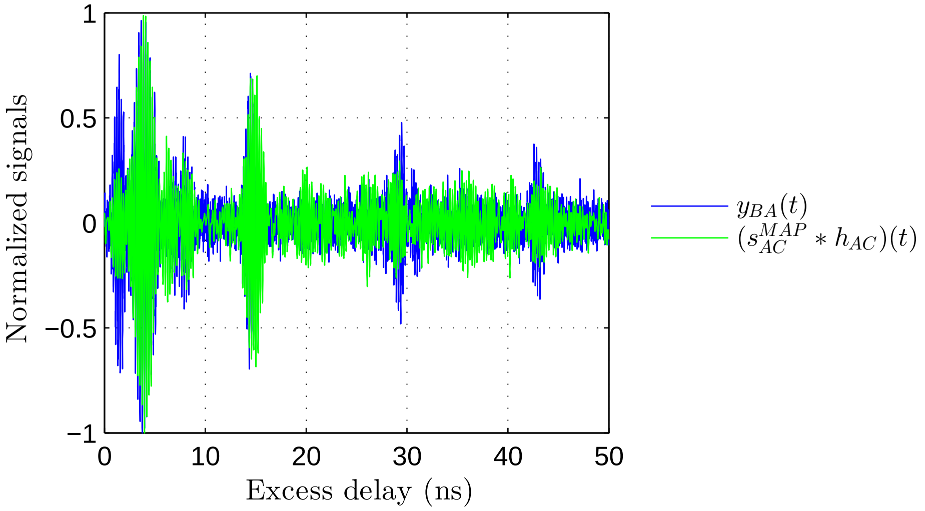


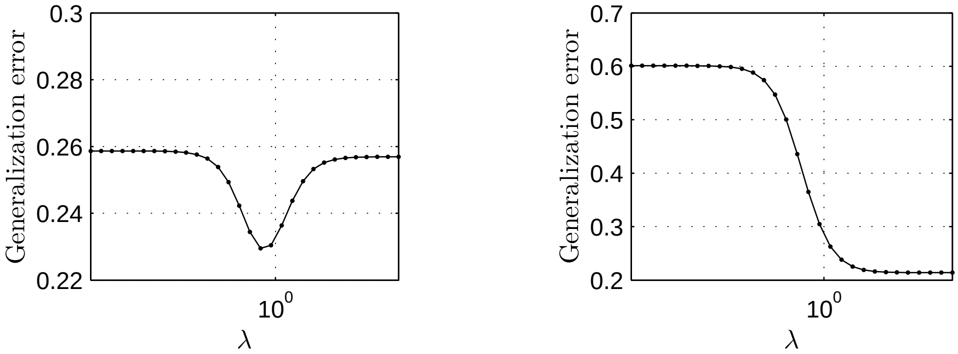
FIGURE 4.5: Example of deduced signal with the unadapted MAP method (RMSE = 11.6%)

Cross-validation (CV) methods are statistical tools for model validation, i.e., they are employed to evaluate how well a given model will generalize to unknown variations in the data set. Given a known data set (in our case, the channel estimates and the target channel), a basic cross-validation procedure splits it randomly into a training set and a validation set. The problem (in our case, Eq. (4.8)) is solved using only the training data set. Then, the found solution $\mathbf{s}(\lambda)$, which, herein, depends on the chosen λ value, is plugged into the given model for the validation set (Eq. (4.6)) and the performance is evaluated via the generalization error from Eq. (4.17). The optimal CV λ value is the one that minimizes the generalization error $\Delta_g(\lambda)$.

$$\Delta_g(\lambda) = ||\hat{\mathbf{H}}_{\mathbf{t}}\mathbf{s}(\lambda) - \mathbf{y}_{\mathbf{t}}|| \tag{4.17}$$

where the t-index represents the training data set.

For our tests, for each channel configuration, we employed 20 random partitions with 70% training data and 30% validation data. The RMSE for the same 100 channel configurations tested in Section 4.4.2 are also shown in Figure 4.4 together with the corresponding chosen λ values. It is observed that the performance of the CV technique depends strongly on the channel configuration. This can be visualized in Figure 4.6, where we show the shape of the generalization error (computed on the sampled un-normalized signals) for two different channel configurations. An RMSE of 9.5% corresponds to Figure 4.6(a) (the case of the basic channel configuration), whereas an RMSE of 18% corresponds to Figure 4.6(b).



(a) Example of convex generalization error

(b) Example of non-convex generalization error

FIGURE 4.6: Generalization error curves for two channel configurations

The generalization error does not always have a minimum and, therefore, model selection using CV does not have optimal performance for all channel configurations. This could be explained by the small cardinality of the data set on which cross-validation is applied (500 samples). In certain cases, the validation samples can be insufficient or unadapted to “generalize” the estimation of the s-signal to the unknown real channel conditions.

4.4.4 Expectation Maximization solution

The EM algorithm applied to our Bayesian data model consists in solving the following steps:[15]

$$\begin{aligned} \text{E-step: } \xi_i(\epsilon,\gamma) &= \mathbb{E}_{\mathbf{s}|\mathbf{y},\epsilon_{i-1},\gamma_{i-1}}[\ln p(\mathbf{y},\mathbf{s}|\epsilon,\gamma)] && (4.18) \\ \text{M-step: } (\epsilon_i,\gamma_i) &= \underset{(\epsilon,\gamma)}{\arg\max}\ \xi_i(\epsilon,\gamma) && (4.19) \end{aligned}$$

After the developments detailed in Appendix D, we obtain a mean signal $\mathbf{s}$, its associated covariance $\mathbf{\Sigma_s}$, and the model parameters ϵ and γ:

$$\begin{aligned} \mathbf{\Sigma_s} &= (\epsilon^{-2}\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \gamma^{-2}\mathbf{P}^T\mathbf{P})^{-1} && (4.20) \\ \mathbf{s} &= \epsilon^{-2}\mathbf{\Sigma_s}\hat{\mathbf{H}}^T\mathbf{y} && (4.21) \\ \epsilon &= \sqrt{\frac{\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\hat{\mathbf{H}}\mathbf{s} + \mathrm{Tr}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{\Sigma_s}) + \mathbf{s}^T\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{s}}{N}} && (4.22) \\ \gamma &= \sqrt{\frac{\mathrm{Tr}(\mathbf{P}^T\mathbf{P}\mathbf{\Sigma_s}) + \mathbf{s}^T\mathbf{P}^T\mathbf{P}\mathbf{s}}{N_s}} && (4.23) \end{aligned}$$

In order to evaluate the performance achieved by the EM signal estimation method, we measure the RMSE for the same previous 100 channel configurations and the s-signals obtained after 200 EM iterations starting from the same initial point λ as the MAP solution.[16] Representative parameter convergence curves are reported in Appendix D.

From Figure 4.7, it can be inferred that, at low λ values, the EM solution has lower performance compared to the MAP estimator with manually tuned parameter. However, it is a relatively "unfair" to compare a solution that gives an estimation of the signal and the parameters (EM) with one that takes as an input a suitable value of the parameter and provides only an estimation of the s-signal. Moreover, EM presents an RMSE advantage for relatively large values of λ, for which it manages to find optimal parameter values that, on average, incur less artefacts on the final deducted signal (e.g., Figure 4.8 for the basic channel configuration).

To sum up, the performance achieved by the EM solutions with the considered model are comparable or inferior to a manual tuning of the trade-off parameter (i.e., choosing a constant very small prior weight irrespective of the channel configuration). This suggests that the incurred complexity of the EM solution is unnecessary. However, the EM

[15] For reasons of consistency in the vector and matrix notations, we ignore the convention regarding the capitalization of the names of the random variables

[16] We actually set $\gamma_0 = 1$ and $\epsilon_0 = \gamma_0\sqrt{\lambda}$.

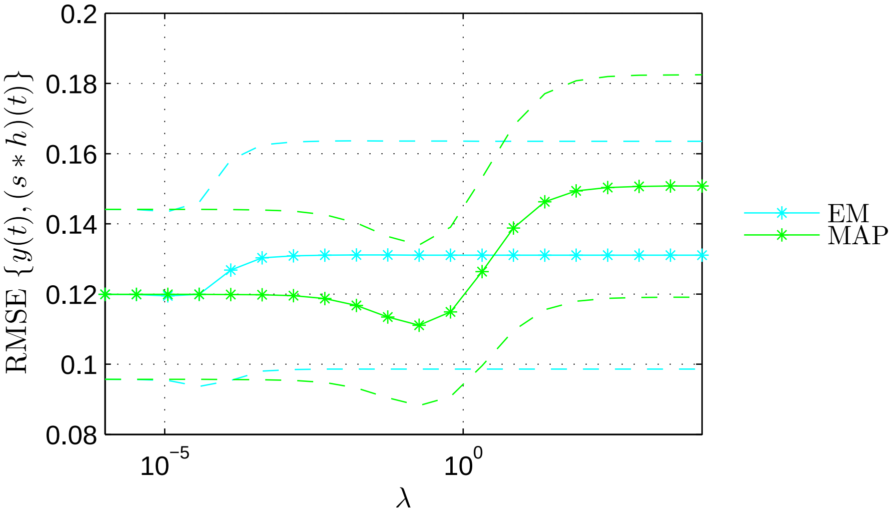


FIGURE 4.7: Non-adjacent signal reconstruction performance for 100 channel configurations (EM, MAP)

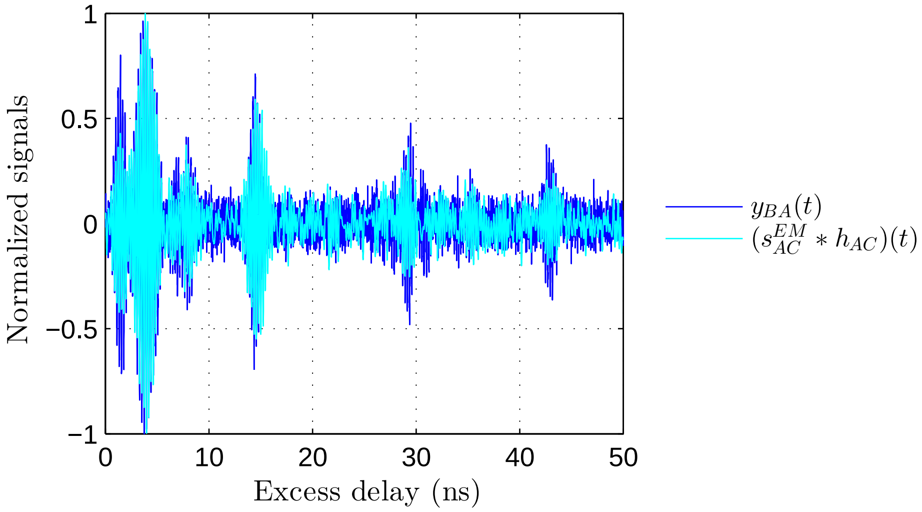


FIGURE 4.8: Example of deduced signal with EM (RMSE = 9.8%)

framework presents several advantages that cannot be achieved by any of the alternative solutions:

- computation of the uncertainty related to the s-signal estimation, potentially exploitable by the subsequent quantization layer.
- comparison between different priors (e.g., another example of prior could be the information about arbitrary regulatory emission masks compulsory for over-the-air transmissions).

- joint s-signal estimation and channel estimation adaptation if supplementary information about the channel estimation uncertainty is available. In this case, the hidden variable would be the couple $(\mathbf{s}, \mathbf{h})$.

4.5 Performance evaluation

Simulation framework

We now evaluate the two key generation protocols, namely CKD and CKG, in a full mesh network of three nodes considering only the MAP solution for CKG. We deem that the EM solution presents system-level advantages but its impact is less obvious in terms of final key performance. Our aim is to compare the two protocols in terms of number of generated bits after several rounds of channel probing for the same number of over-the-air packets. The quantization and reconciliation schemes are not analyzed in this section because they are not discriminant elements for the two protocols, which employ the same type of input signal for quantization. However, it should be noted that in a more exhaustive performance assessment, the random properties of generated keys, which depend on the input signal and the quantization algorithm, should be studied as well.

The reciprocal CIRs corresponding to the three links are generated independently using the IEEE 802.15.4a statistical model for LOS indoor environments (CM1). The transmitted pulse for initial channel probing $p(t)$ has a bandwidth of 1 GHz (defined at -10 dB of the Power Spectral Density) and a center frequency at 4.5 GHz. The duration of the observation window is set at $T_w = 50$ ns.

For CKG, the sampling frequency for channel impulse response estimation and for the computation of s-signals is set at $F_s = 10$ GHz. The s-signal is computed using the MAP solution in Eq. (4.9) with the regularization parameter set manually to $\lambda = -4$ dB and the obtained s-signal is then filtered to conform with the required signal bandwidth and central frequency.

The input signal for quantization has a sampling frequency of $F_p = 1/T_p$ and it is normalized with respect to the minimum and the maximum values in order to obtain signal samples between 0 and 1 for all nodes. An example of such a signal issued from link [A-C] and seen in A, B (based on the reception of an s-signal), and C is provided in Figure 4.9. For simplicity reasons, we choose a two-bit uniform quantization with corresponding cells {(0-0.25), (0.25-0.5), (0.5-0.75), (0.75-1)} and a Grey dictionary ({“00”, “01”, “11”, “10”}). The employed guard-bands (GB) vary between 0 and 0.1

around the borders of the quantization cells. As previously, reconciliation is implemented with a Reed Solomon code (i.e., 127 message codewords of 7 bits and 123 data codewords after padding the quantized binary sequence with dummy bits in order to obtain the needed block length for encoding).

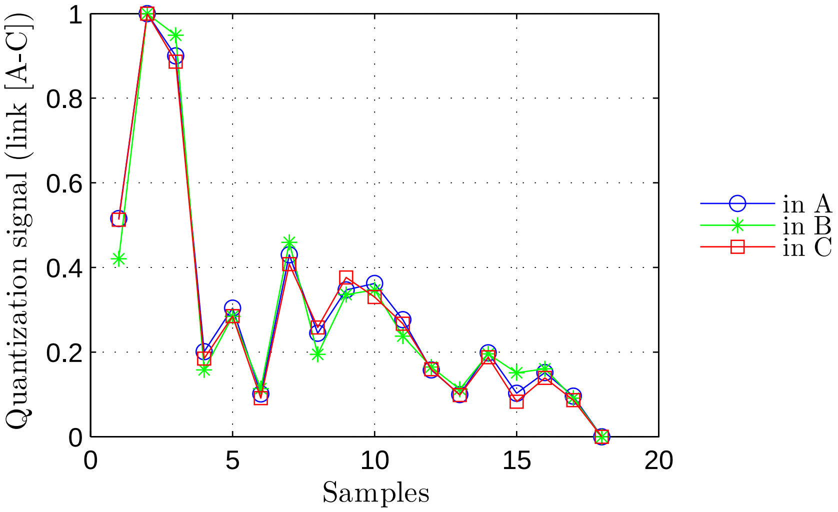


Figure 4.9: Typical input quantization signal seen by the three nodes (SNR= 20 dB)

Results

As a preliminary study, we compare the number of communication packets (including broadcasts when possible) needed for one round of key generation for the two methods (Table 4.1). We do not consider the packets exchanged in order to establish a lead node (for reconciliation in CKG and for distribution in CKD) because they are not discriminating. The cooperative channel probing for CKG can be achieved with 3 or 6 exchanges depending on the number of cooperators. The key distribution for CKD consists of one packet sent to the lead node to inform it of the length of the key on its non-adjacent channel and 2 packets for group key propagation. We conclude that, in any case, CKG requires less traffic than CKD to issue a group key.

Table 4.1: Exchanged packets for one group key generation/distribution

Phase/Method	CKD	CKG
Pairwise channel probing	3	3
Cooperative channel probing	0	3 or 6
Reconciliation (dropping)	6	3
Reconciliation (error-correction)	3	1
Key distribution	1 + 2	0
Total	15	10 or 13

Table 4.2 compares the number of packets needed for CKD and CKG in a full mesh network of N nodes when CKG inference is achieved with only one cooperator. Because of the cooperative channel probing, CKG is less scalable than CKD for $N > 6$. However, larger values of the number of nodes makes it difficult to obtain a full mesh scenario and to establish a key during the channel coherence time for both methods and they are, therefore, not practical.

Table 4.2: Exchanged packets for one group key generation/distribution in a full mesh of N nodes

Phase/Method	CKD	CKG
Pairwise channel probing	N	N
Cooperative channel probing	0	$[N(N-1)/2-(N-1)]\times N$
Reconciliation (dropping)	$N(N-1)$	N
Reconciliation (error-correction)	$N(N-1)/2$	1
Key distribution	$N(N-1)/2$	0
Total	$O(N^2)$	$O(N^3)$

The presented simulation results are obtained for a fixed-traffic scenario (30 packets corresponding to 2 rounds for CKD and 3 rounds for CKG). Keys that are not equal after reconciliation are considered to have length of 0. Averaging is performed over 500 channel configurations (each configuration comprising three channel realizations corresponding to the three pairwise links). The channels are independent over the three links and also from one round to another, meaning that bits can be concatenated for the final key.

From Figure 4.10, we conclude that CKG is more advantageous in terms of key length for higher SNR and larger guard-bands. This is due to the fact that before quantization, the CKG method does not have direct access to the non-adjacent channel measurements that are obtained after a deconvolution operation and another noisy transmission. This degrades the reciprocity of the final samples and leads to more keys that are not agreed upon for CKG at lower SNR or smaller guard-bands. However, it can be verified that CKG achieves a relatively high bit matching ratio between the three keys generated at the participating nodes (Figure 4.11). The bit matching ratios for the CKD method, although higher than those for CKG, do not present an interest in this study because they can only be computed for pairwise keys and represent solely the single-link reciprocity performance of the system.

4.6 Summary

In this chapter we have investigated an alternative method for generating secret group keys using the physical layer in IR-UWB systems without relying on classical pairwise

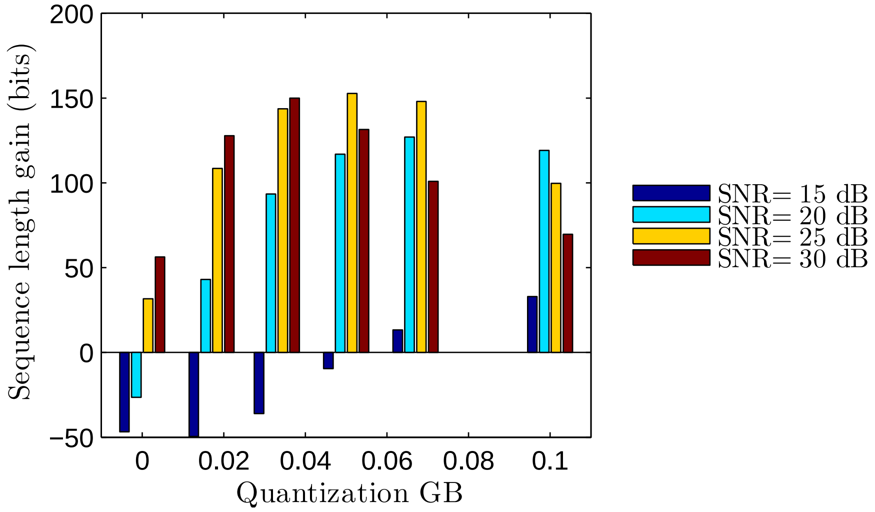


FIGURE 4.10: Average key length gain of CKG w.r.t. CKD

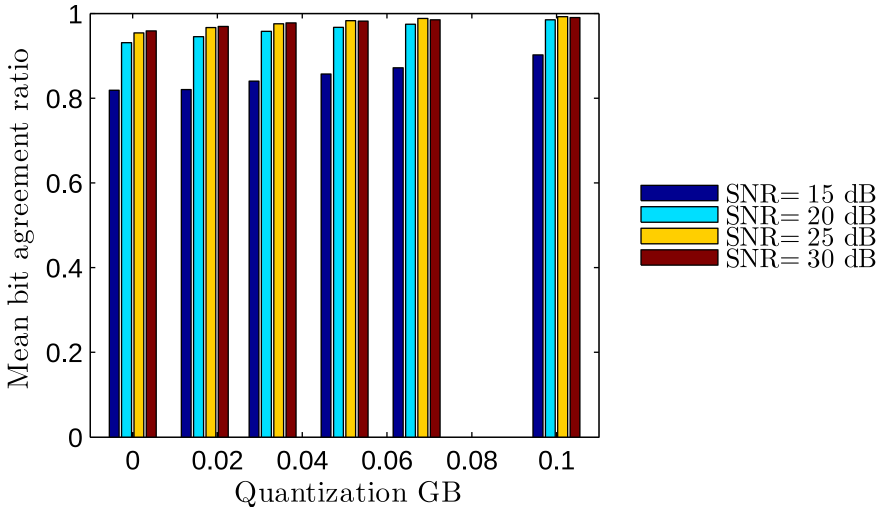


FIGURE 4.11: Average bit matching for CKG

key agreement. For each node, we exploit the concatenation of adjacent and non-adjacent links in a mesh network in order to gather more entropy before the quantization process. We show the advantages of our method in terms of traffic overhead as well as in terms of key length for certain parameter configurations.

Our proposal focuses on the parameterized estimation of specific s-signals obtained after a deconvolution process and emitted by cooperative nodes in order to "induce" non-adjacent channel information for their neighbors. We investigate the accuracy of the non-adjacent signal reconstruction phase at the reception side, which is an important factor for subsequent key generation. Consequently, we describe and test two methods for signal estimation and parameter specification. The first one is based on a cross-validation technique with the aim of choosing an optimal value for the maximum *a posteriori* deconvolution solution, whereas the second one applies the Expectation Maximization algorithm to obtain a joint estimation of the required signal and of the parameters of the employed statistical model. The two methods are iterative with different levels of complexity (i.e., iterative cost computations for CV and iterative matrix inversions for EM) but also different evolution opportunities (e.g., the EM model is more adapted for incorporating the errors on channel estimation).

Further studies should look into the complexity issues of the s-signal generation, potentially by considering a joint design between the s-signal generation and the rest of the key generation protocol (e.g., signal acquisition, estimation, quantization).

The proposed scheme can be adapted to other technologies (narrow-band, OFDM) by changing the cooperative signal generation method according to the most relevant channel features. Moreover, the concept of s-signal could be extended to cases where the target signals are not channel estimates but randomly generated information, which leads to a mixed key generation model (see Section 1.2.1). However, this complicates the overall key generation procedure because there would be no more direct acquisitions of channel information.

Chapter 5

Conclusion and perspectives

In this thesis we were interested in symmetric secret key generation methods using the physical layer. This approach to symmetric key distribution is part of the larger framework of physical layer security, which takes advantage of wireless propagation phenomena, such as fading and noise, in order to complement traditional confidentiality and authentication cryptographic solutions. Physical layer security solutions are situated at the crossroads of information theory and signal processing and are particularly appealing in decentralized wireless networks, where security becomes more challenging because of the lack of infrastructure. In this context, key agreement strategies based on public discussions have been proved simpler to implement than keyless secrecy achieving solutions, which require wiretap code design and/or physical advantage engineering (e.g., by jamming or beamforming). Similarly to the majority of secure communication solutions, physical layer key generation does not address initial authentication of the legitimate users.

The IR-UWB CIR is an interesting option for secret key generation according to the source model, first of all, because of the fact that one CIR measurement provides several values for quantization (i.e., several significant channel taps), but also because of the delay information contained in such signals (e.g., the excess delays of the significant channel taps). This is not the case for RSS measurements, for example, which are acquired at a uniform probing rate or over several parallel channels. As a general drawback of the source model, the wireless channel must provide temporal variability in order to generate sufficiently long keys or renew them. Since this issue is not directly linked to the signal structure and rather depends on the environment or on the network configuration (i.e., mobile or static), we did not address it in the main part of the present work (only partially in Appendix F).

We do not claim that IR-UWB CIRs are always more advantageous than other wireless technologies and channel characteristics because the design of key generation protocols, as any other security solution, depends on the context (i.e., application, deployment environment, type of network, hardware or software limitations, etc.). We believe that in order to achieve optimal integration of physical layer solutions, it is necessary to take advantage of all the opportunities provided by a certain radio technology. Therefore we provide guidelines for key generation using various types of IR-UWB temporal signals. By allowing further processing, IR-UWB signals could be exploited in the frequency domain as well with the advantage of achieving synchronization at the expense of less entropy (i.e., quantizing only the absolute value of the complex gain).

The work accomplished in this thesis is divided along three main research axis, corresponding to the three main chapters (Chapters 2-4): IR-UWB channel quantization issues, public discussion strategies for information reconciliation, and cooperative physical layer key generation. Although the first chapter is concerned with specific IR-UWB issues, the rest of the our findings can be adapted to other types of channel measurements. Two of our main proposals, namely HIST (Section 2.2) and POS$_{ToF}$ (Chapter 3), are extensions of an existing key generation protocol employing IR-UWB directly sampled signals [95] [79], whereas the rest of the work is built on new signal models (e.g., channel estimates) and has new aims (e.g., optimized quantization or cooperative techniques).

Quantization issues for IR-UWB signals

In this work, we considered the optimization of the quantization phase based on supplementary criteria related to reconciliation and privacy amplification. This resulted in a practical approach for key generation and for the study of the trade-offs between length, reciprocity and randomness. The studies of quantization algorithms applied to IR-UWB signals (Chapter 2) highlighted the potential of the delay information for achieving keys with better randomness properties (i.e., HIST scheme for directly sampled CIRs and DIV scheme for simulated noisy CIR estimates) or for improving reciprocity in the case of realistic channel estimates. We have also shown how to adapt the quantization thresholds in order to obtain a desired reciprocity-randomness trade-off for an “average” generated codeword given the expected average energy of the underlying multipath components. In this case, the randomness has been measured by the proposed inter-key diversity metric.

Once realistic CIR estimates are available for quantization, the delay information becomes useless for quantization if it is exchanged on the public channel for reciprocity purposes. Although the DIV scheme cannot be applied, the optimization of quantization thresholds or the search for equivalent uniform quantization thresholds at a fixed excess delay is still feasible. Nonetheless, since the optimized threshold computation is based on the statistics of the simplified signal model (i.e., with a Gaussian mixture model), the performance for individual channel realizations might be different. This can be caused by the fact that the considered statistics for the true channel might not be representative for the estimated channel tap, as it has been suggested by the structure of CIR estimates studied in Section 2.4. Therefore, a new model for realistic channel estimates should be investigated in order to be able to apply the proposed threshold optimization techniques. As already mentioned, the difficulty of such approaches resides in the dependence of the CIR structure on the employed estimator and, to the best of our knowledge, in the lack of theoretical models for pulse interference scenarios.

Nevertheless, the proposed methods and ideas remain valid for any type of signal that contains characteristic delay information. For example, we can imagine ED-based channel probing, similarly to the one used for quantization tests in Chapter 4. Although these signals have uniform measurement steps, the significant delay information can be extracted by a threshold operation on the amplitudes (e.g., all the samples lower than a given value are set to 0). In this way, in dense multipath environments, pairs of delays and amplitudes could be obtained and quantized according to the described proposals.

Discrete public discussion strategies

In Chapter 3, we presented two extensions (BIN and POS_{ToF}) of an existing public discussion method [95], which aim at achieving better immunity against eavesdropping attacks. BIN is an adaptable solution that limits the publicly disclosed information but is more sensitive to noise than POS_{ToF}, which only masks the public information.

We deliberately used ray-tracing data for our simulations because it gives direct insight into the advantage that an attacker could have by combining the location-dependent signal with the eavesdropped public information. The study is restricted because the available data did not allow to consider attackers that are closer than 1 m from one of the legitimate users. However, the obtained results suggest that the eavesdropping could become even more problematic at shorter distances for the considered type of signal. Further investigations should also analyze the case of realistic IR-UWB signals with denser and more diffuse multipath components, which are likely to favor the legitimate users.

We acknowledge the fact that although POS_{ToF} achieves good results in terms of bit agreement between the legitimate keys and the key generated by an eavesdropper, the performance is highly dependent on the attacker model. We considered a “naive” physical layer attacker that does not know the relative positions of the nodes and does not apply any data mining techniques, which would be a subject of research by itself. However, this is also a more “probable” attacker model.

Lastly, the masking procedure proposed by POS_{ToF} is not at all equivalent to an encryption operation and a simple ToF measurement is by far less powerful than a symmetric key. In order to achieve better secrecy performance, mobile nodes could use several consecutive high-precision ToF measurements (e.g., stored in memory) for masking one public exchange. Moreover, if enough ToF measurements are available, they could be employed similarly to one-time pads to exchange the excess delays of realistic channel estimates. This would therefore allow joint delay-based reconciliation and delay-dependent quantization for IR-UWB CIR estimates, which have been discussed in Chapter 2.

Cooperative physical layer key generation

Chapter 4 introduced a new idea for extending the key generation source model to several cooperative nodes in order to directly extract a group key. The proposed solution, a mix between a source model and a reciprocity-based channel model, relies on equalization-like operations and avoids over-the-air traffic needed for multiple pairwise key generation procedures and their associated reconciliation messages. Prior to quantization and reconciliation, nodes concatenate direct channel measurements from the adjacent links with measurements from cooperative transmissions of their neighbors. The cooperative transmissions are meant to assist distant nodes to “infer” their non-adjacent channels.

We illustrate the proposed idea in a scenario where devices can afford expensive computations such as deconvolution operations involving IR-UWB channels. We adopt a temporal Bayesian approach (i.e., MAP or EM solutions) for the estimation of the optimized signals that need to be transmitted in order to induce a certain desired signal at the receiver. The most important limitation of these methods, especially the iterative ones, is the advanced computational capabilities required for implementation and the computation time. This makes them more adapted to relatively static environments, which have been one of the motivations for this proposal, and to powerful devices, in the particular case of IR-UWB. However, the developed CKG protocol could be more easily implemented in narrow-band systems involving phase or channel gain information because of the simplification of channel equalization operations.

Chapter 6

Reflections

"We cannot solve our problems with the same thinking we used when we created them."
Albert Einstein

Progress has always been a part of human nature. The search for more, better and faster is something almost innate to people as a species and probably the engine that embarked us on an accelerated journey through technology.
From fire, agriculture and cities to medicine, Internet and autonomous cars ...
From survival and safety to comfort and pleasure ...
From local to global.

Technological progress is our response to the limitations we encounter and so, *telecommunications*[1] technology becomes the expression of our desire to share information despite the physical limitations of space. As any other invention, it carries within it the potential for good and bad, in the sense that it can be beneficial from a social or economic perspective, a drive for humanitarian actions, a tool of knowledge, a means of connection, but in the same time, it can be disruptive, misused, overused, detrimental. Although the binary "good vs. bad" model can be vastly debated, I believe that, at some point and for each specific context, lines have to be drawn and choices have to be made. As a technological researcher, it is part of my work to understand the impact that technology has on the people using it. Whenever possible, this understanding needs to be incorporated in the technological choices that are made regarding aspects such as centralization or security. As a daily technology user, I would like to be conscious about the way I employ it. The present discussion is a short overview of my current perception of sensitive telecommunications-related issues and the questions they rise.

[1] from the Greek "tele-", meaning "distant", and the Latin "communicare", meaning "to share"

You are now online. We own landline phones, computers, tablets, and smartphones that we can use for phone calls, SMS, MMS, online chat, online calls, video calls, personal and professional emails, social media, instant private or group messaging, conference calling. This makes a lot of options for contacting someone and for being contacted. If the advantages of this proliferation of communication means is obvious on a short-term basis (e.g., business models, personal life, etc.), the possible negative effects are usually long-term, more subtle, and dependent on the particular usage.

Some people can show signs of fatigue and decreased productivity when handling multitasking between offline work and abundant emails or phone calls. Others find it difficult to manage their personal life with constant access to professional emails or when owning a professional mobile phone. Teenagers or even adults can become so attached to their online identities and friends that they neglect or fear real social contact. Furthermore, this can lead to feelings of dissatisfaction and inferiority caused by the permanent comparison to the images that “others” project about them and ironically, to feelings of loneliness. As users, we should understand that being “always on” is both an opportunity and a trap, and that our physical rhythm or reality could sometimes have difficulties in coping with our instant accelerated online lives.

Based on the philosophical idea that organized refined writing, rather than oral communication, contributed to the development of modern society, it is argued that instant messaging and social media can potentially produce a regression in our language skills [155]. This could be caused by the widespread adoption of instant communication tools that replace the “thought-enhancing” written form of communication with a pseudo-oral written language. In my opinion, although scientific evidence for proving such claims can be difficult to obtain, the described scenario is nonetheless plausible given common-sense trade-off mechanisms between efficiency and creativity, for example.

The Internet of ... Everything! The Internet of Things is about connecting common objects to the Internet in order to optimize environmental monitoring, infrastructure management, energy consumption, medical services, transports, etc. or just make things easier and less time-consuming in our daily life. First of all, the aim of connecting billions of objects brings new tremendous challenges regarding the management, security, privacy or fail-safe requirements of such systems. Secondly, the promised positive impact should be carefully weighted against the input costs. For example, energy consumption monitoring systems could help reduce our domestic CO_2 footprint, but the overall environmental impact of the manufacturing, usage and disposal of electronic devices that are still difficult to recycle has yet to be completely understood.

Furthermore, at a psychological and social level, would an "easier" daily life or time gains necessarily make us happier or just more technology-dependent? Of course, the same could be argued about the washing machine that we now accept as an essential tool in our lives. However, all technology requires a learning process so in the beginning it can be perceived as rather disruptive than facilitative. Things become even more complex when considering the "quantified self" concept [156], which, on the one side, seems like a valuable tool of self-development, and on the other side, raises ethical concerns about dehumanization by promoting numerical, objective measurements of complex or subjective human states.

Regarding technological dependency, an interesting point is made by studies on slave-making ants (i.e., ants that "enslave" a part of their population), which present diminished survival capacities if the slave-ants are taken away from the colony [157].[2] In my opinion, this does not necessarily mean that we are becoming "enslaved" by our technology, but it is obvious that the way we use it shapes our skills and our expectations in terms of happiness, efficiency, and sometimes self-esteem.

Privacy or the right to accept the terms and conditions you do not understand. The accumulation of sensing data from various sources, especially personal-related ones ranging from localization and commercial preferences to household indicators, has become the "Graal" for a lot of ICT[3] companies. In a consumer-oriented society, big data brings companies big money through mechanisms such as targeted advertising or more generally, consumer profiling based on personal information that is initially not provided for these purposes. This is what allows companies to offer free online services while remaining highly profitable in some cases (e.g., Google, Facebook, etc.). Adverse reactions to this business model include views based on the slogan "If you're not paying for it, you are not the customer; you are the product being sold.". Objectively, postal mail, cellular services, and even libraries are not free, why would email be? However, the Internet situation is slightly different: the user already pays a fee to his Internet connection provider and historically speaking, we have been "educated" to expect free online services. In this context, any online service provider demanding even a small fee would be perceived as a bad deal.

It is probably that even with a full understanding of the underlying data processing mechanisms, some users would still accept the data disclosure in return for free Internet services. However, the advent of the Internet of Things and the fact that we will no

[2] The cited article references the mentioned ant study in the context of discussions about the 2014 IQ2 debate "We are becoming enslaved by our technology" (http://www.iq2oz.com/debates/we-are-becoming-enslaved-by-our-technology-/).

[3] Information and Communications Technology

longer entirely or easily control the information disclosed by our devices will possibly raise more interest in privacy-friendly solutions already available but less popular today. Users of a particular data gathering application should be able to transparently choose what happens to their data, with whom it is shared, and for what purposes. In this sense, applications should be provided with comprehensive privacy interfaces or even more, engineered according to social rather than purely commercial principles [158].

The main argument for dismissing privacy-related burdens can be resumed in a simple rhetorical question: "Why bother if I don't have anything to hide anyway?". Although it might seem legitimate at first, we should consider that privacy is not lost or gained in one shot, but rather incrementally [159]. If today we do not feel concerned about who can track our shopping preferences, tomorrow, our online profile might contain sensitive information (e.g., health-related) that allows discriminative profit-oriented polices to be implemented (e.g., health insurance fees). We should probably think harder before contributing to the fading of a right that we do not know when we might need. Fast-forwarding on the same slope, the paradigm could shift in the sense that disclosing information could become a requirement in order to have certain basic benefits (i.e., the unraveling effect [160]). From this hypothetical point and taking into account present surveillance capabilities and legislation of certain states (e.g., both in terms of data and meta-data), we could even imagine moving towards the dreaded orwellian[4] society.

This inevitably leads us to a more complex issue: government surveillance and the debated privacy-safety trade-off. Despite the fact that at the present moment we seem to be far from such orwellian scenarios, precautionary principles encourage us to raise awareness on privacy issues. On the contrary, a survey on the public perception of information gathering practices of the US government [161] shows an increase (1985-1996) followed by a decrease (1996-2006) of concerns regarding privacy threats. The decrease could be potentially explained by a "cultural lag": once people get accustomed to ICTs as part of their daily lives, the related fears diminish despite the fact that surveillance capabilities might increase. This might also be the case for biometrics in the following years. However, in this case, privacy issues regarding biometrics seem to be on the legislation table from the early beginning at least in the EU [162]. Although, the case of biometrics for border control or national identification is too complex and out of the scope of this discussion, the idea of using such intrusive identification for customer profiling is, to my mind, unnecessary for the simple reasons that we are not the things we buy and that our preferences and habits should evolve as freely as possible in order to conserve diversity.

[4] https://en.wikipedia.org/wiki/Orwellian

Despite its socio-economic and political connotations, privacy can mean two things: "the right to be left alone" and "the right to misrepresent oneself" [159]. Personally, I associate the former with a passive attitude and a personal legitimate choice of not disclosing ourselves. The latter is probably more complex since it can include active participation while escaping the responsibility of our sayings and doings. This type of duality characterizes implementations such as the "darknet".[5] Accessible to drug dealers and activists in the same time, the darknet is gaining increasing popularity even for mainstream activities because of its privacy features.[6]

*

Telecommunications tools bring solutions to the issues they were designed for. The way we use them can nonetheless cause new problems that could have hardly been anticipated at their creation. Similarly to the example of the rapid industrial development followed by the recent acknowledgment of the need for sustainability measures, ICTs have witnessed widespread adoption and it is now becoming necessary to think about *sustainable communication technologies*. This cannot be achieved by applying the same concepts that launched or promoted the improvement of ICTs, but rather by delving into the social, economic, and political contexts and defining the core principles that we want ICTs to follow. Since they are already so deeply rooted in our lives to the point that our online identities are becoming persistent, we should expect ICTs to follow the same principles that we use to define our ideal society: choice, awareness, security, non-discrimination, freedom of expression, solidarity, etc. As challenging as it may seem, we should try to translate aspects like these into technological solutions and integrate them in today's and tomorrow's systems.

[5] Wikipedia: "An overlay network that can only be accessed with specific software, configurations, or authorization, often using non-standard communications protocols and ports. Two typical darknet types are friend-to-friend networks (usually used for file sharing with a peer-to-peer connection) and anonymity networks such as Tor via an anonymized series of connections."

[6] http://www.ted.com/talks/jamie_bartlett_how_the_mysterious_dark_net_is_going_mainstream

Personal contributions

Acknowledgment: Part of this work has been carried out in the framework of the BUTLER project (FP7-ICT-287901) and the SocIoTal project (ICT-FP7609112), which are funded by the European Commission.

Proceedings of international conferences with peer reviews

I. Tunaru, B. Denis, R. Perrier, B. Uguen, “Cooperative Group Key Generation Using IR-UWB Multipath Channels”, IEEE International Conference on Ubiquitous Wireless Broadband 2015 (IEEE ICUWB’15), Montreal, Oct. 2015

I. Tunaru, B. Denis, B. Uguen, “Location-Based Pseudonyms for Identity Reinforcement in Wireless ad hoc Networks”, IEEE Vehicular Technology Conference-Spring 2015 (IEEE VTC-Spring’15), Glasgow, May 2015

I. Tunaru, B. Denis, B. Uguen, “Reciprocity-Diversity Trade-off in Quantization for Symmetric Key Generation”, IEEE International Symposium on Personal Indoor, and Mobile Radio Communications 2014 (IEEE PIMRC’14), Washington, Sep. 2014

I. Tunaru, B. Denis, B. Uguen, “Random Patterns of Secret Keys from Sampled IR-UWB Channel Responses”, IEEE International Conference on Ultra Wideband 2014 (IEEE ICUWB’14), Paris, Sep. 2014

I. Tunaru, B. Denis, B. Uguen, “Public Discussion Strategies for Secret Key Generation from Sampled IR-UWB Channel Responses”, Conference on Communications 2014 (COMM’14), Bucharest, May 2014

Patents

I. Tunaru, B. Denis, R. Perrier, “Procédé de géneration de clé secrète de groupe basée sur la couche physique radio et terminal sans fil associé”, filed patent application, Aug. 2015

Other communications

I. Tunaru, B. Denis, B. Uguen, talk "Randomness and Reciprocity in Quantization for Secret Key Generation from IR-UWB Channels", Journée "Sécurité au Niveau de la Couche Physique dans les Réseaux Sans Fil" du GDR ISIS, Télécom ParisTech, Paris, France, May 2014

I. Tunaru, B. Denis, B. Uguen, poster "Physical Layer Secret Key Generation for WSN", IEEE European School of Information Theory 2014, Tallin, Estonia, Apr. 2014

Contributions to research reports of EU projects

J. L. Hernández Ramos, et al., "Secure Group Communication", Deliverable D3.3 of SocIoTal, May 2015

F. Sottile, et al., "IoT Enabling Technologies and Future Developments", Deliverable D2.5 of BUTLER, Sep. 2014

Teaching and supervising activities

"Probabilities and Statistics" (FR, 10h seminar, 30 students), Level Masters' 1, Section Physics, Electronics, Telecommunications, Phelma, Institut National Polytechnique (INP), Grenoble, France, May 2015

"Integral Transforms" (FR, 20h course and seminar, 30 students), Level Masters' 1, Section Physics, Electronics, Telecommunications, Phelma, Institut National Polytechnique (INP), Grenoble, France, Oct. 2014

"Summer School in Mathematics" (FR, 28h, 15 students), Level Bachelors' 1, Section EURINSA, Institut National des Sciences Appliquées (INSA), Lyon, France, Aug. 2012

Internship of M. Bulenok on "Experimental key generation with low-complexity devices", CEA-Leti, Apr.-Sep. 2015

Appendix A

Bit pattern period

The upper bound on the bit pattern period ($uBPP$) is the number of samples of a sinusoidal function that cross the given thresholds ${L_0}^+/\delta$ and ${L_0}^-/\delta$. Figure A.1 shows that the $uBPP$ presents a symmetry with respect to 0. We remind that F_s is the sampling frequency and f_c is the central frequency of the signal.

From Figure A.1, we have:

$$\begin{aligned} {L_0}^+ \sin(2\pi f_c \alpha) &= {L_0}^+/\delta && \text{(A.1)} \\ \alpha &= \frac{\arcsin(1/\delta)}{2\pi f_c} && \text{(A.2)} \end{aligned}$$

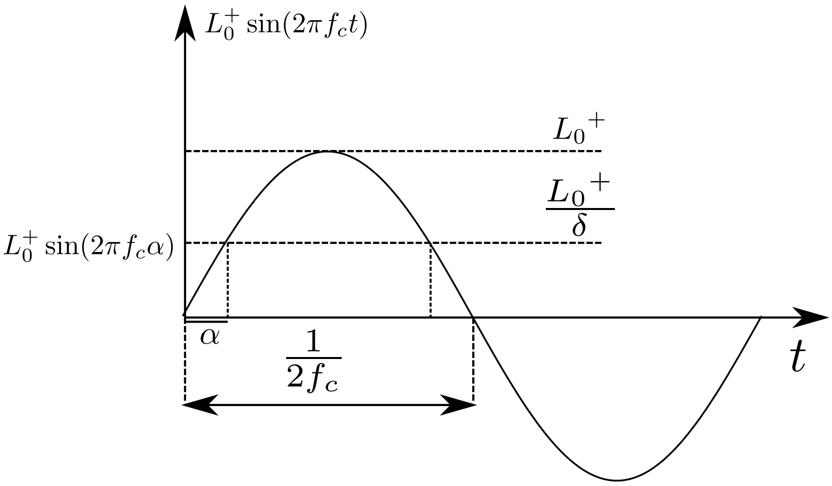


FIGURE A.1: Illustration of the computation of the $uBPP$

We define $T_{half} = 1/(2f_c)$ as the demi-period of a sine wave of frequency f_c and $\Delta_t = 1/F_s$ as the sampling resolution.

$$\begin{aligned} uBPP &= 2(\frac{T_{half}}{\Delta_t} - 2\frac{\alpha}{\Delta_t}) && \text{(A.3)} \\ &= 2(\frac{1/(2f_c)}{1/F_s} - 2F_s\frac{\arcsin(1/\delta)}{2\pi f_c}) && \text{(A.4)} \\ &= 2F_s(\frac{1}{2f_c} - 2\frac{\arcsin(1/\delta)}{2\pi f_c}) && \text{(A.5)} \end{aligned}$$

Appendix B

Reciprocity and diversity cost functions

According to the model in Section 2.3.1, the codewords C_n^A and C_n^B are independent given the noiseless value x_n. We define $Q(.)$ as the Q-function representing the Gaussian survival function for the estimation noise σ_0. Although the following expressions contain discrete variables C_n^A or C_n^B, we actually work with continuous probabilities because $\mathbb{P}(C_n^u = c_j) = \mathbb{P}(Y_n^u \in [\theta_j^{inf}, \theta_j^{sup}))$. Let $u \in \{A, B\}$.

The diversity cost $CS(n, \theta)$ (i.e., the spread of the probabilities of occurrence of the expected codewords) is detailed in the following equations:

$$\begin{aligned} CS(n,\theta) &= std(\{P_n^{u,1}, P_n^{u,2}, ..., P_n^{u,2^b}\}) && \text{(B.1)} \\ P_n^{u,j} &= \mathbb{P}(C_n^u = c_j | C_n^u \neq c_0) && \text{(B.2)} \\ &= \int_{-\infty}^{\infty} \mathbb{P}(C_n^u = c_j | C_n^u \neq c_0, x_n) \cdot f_{X_n}(x_n) \, \mathrm{d}x_n && \text{(B.3)} \end{aligned}$$

The term $P_{j,x_n}^u = \mathbb{P}(C_n^u = c_j | C_n^u \neq c_0, x_n)$ is given in Eq. (B.8).

The reciprocity cost $HD(n, \theta)$ (i.e., mean Hamming distance between two valid codewords) is computed through numerical integration according to the following equation:

$$\begin{aligned} HD(n,\theta) &= \mathbb{E}_{X_n}[\mathbb{E}_{W_n}[hd(C_n^A, C_n^B) | C_n^A, C_n^B \neq c_0]] && \text{(B.4)} \\ &= \int_{-\infty}^{\infty} \mathbb{E}_{W_n}[hd(C_n^A, C_n^B) | C_n^A, C_n^B \neq c_0, x_n] \cdot f_{X_n|C_n}(x_n | C_n^A, C_n^B \neq c_0) \, \mathrm{d}x_n \end{aligned}$$

The two terms of the integrand are detailed as follows. The conditional mean Hamming distance between two valid codewords:

$$\mathbb{E}_{W_n}[hd(C_n^A, C_n^B)|C_n^A, C_n^B \neq c_0, x_n] \quad = \quad \sum_{\substack{i,j=-4 \\ i,j\neq 0}}^{4} P_{i,x_n}^A \times P_{j,x_n}^B \times hd(c_i, c_j) \tag{B.5}$$

For $c_j \neq c_0$:

$$\begin{aligned} P_{j,x_n}^u &= \frac{\mathbb{P}(C_n^u = c_j \cap C_n^u \neq c_0|x_n)}{\mathbb{P}(C_n^u \neq c_0|x_n)} && \text{(B.6)} \\ &= \frac{\mathbb{P}(C_n^u = c_j|x_n)}{\mathbb{P}(C_n^u \neq c_0|x_n)} && \text{(B.7)} \\ &= \frac{Q(\frac{\theta_j^{inf}-x_n}{\sigma_0}) - Q(\frac{\theta_j^{sup}-x_n}{\sigma_0})}{1-(Q(\frac{\theta_0^{inf}-x_n}{\sigma_0}) - Q(\frac{\theta_0^{sup}-x_n}{\sigma_0}))} && \text{(B.8)} \end{aligned}$$

The probability distribution function of x_n conditioned on the validity of the codewords is expressed using Bayes' rule:

$$\begin{aligned} f_{X_n|C_n}(x_n|C_n^A, C_n^B \neq c_0) &= \frac{\mathbb{P}(C_n^A, C_n^B \neq c_0|x_n) \cdot f_{X_n}(x_n)}{\mathbb{P}(C_n^A, C_n^B \neq c_0)} && \text{(B.9)} \\ &= \frac{\mathbb{P}(C_n^A \neq c_0|x_n) \cdot \mathbb{P}(C_n^B \neq c_0|x_n) \cdot f_{X_n}(x_n)}{\mathbb{P}(C_n^A, C_n^B \neq c_0)} && \text{(B.10)} \end{aligned}$$

where

$$\begin{aligned} \mathbb{P}(C_n^u \neq c_0|x_n) &= 1-(Q(\frac{\theta_0^{inf}-x_n}{\sigma_0}) - Q(\frac{\theta_0^{sup}-x_n}{\sigma_0})) && \text{(B.11)} \\ \mathbb{P}(C_n^A, C_n^B \neq c_0) &= \int_{-\infty}^{\infty} \mathbb{P}(C_n^A \neq c_0|x_n) \cdot \mathbb{P}(C_n^B \neq c_0|x_n) \cdot f_{X_n}(x_n)\,\mathrm{d}x_n && \text{(B.12)} \end{aligned}$$

Appendix C

Parameterization with evidence maximization

The optimal parameters can be deducted by maximizing the evidence function:

$$
\begin{aligned}
(\epsilon, \gamma) &= \underset{\epsilon,\gamma}{\arg\max}\ \ln p(\mathbf{y}|\epsilon, \gamma) && \text{(C.1)} \\
p(\mathbf{y}|\epsilon, \gamma) &= \int_s p(\mathbf{y}|\mathbf{s}, \epsilon) \cdot p(\mathbf{s}|\gamma)\, \mathrm{d}\mathbf{s} && \text{(C.2)}
\end{aligned}
$$

In order to compute the integral analytically, we will transform it into a generalized Gauss integral. The integrand becomes:

$$
\begin{aligned}
& (2\pi\epsilon^2)^{N/2} \mathrm{e}^{-\frac{1}{2\epsilon^2}(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})^T(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})} \cdot (2\pi\gamma^2)^{N_s/2} \mathrm{e}^{-\frac{1}{2\gamma^2}(\mathbf{P}\mathbf{s})^T(\mathbf{P}\mathbf{s})} && \text{(C.3)} \\
= \; & (2\pi\epsilon^2)^{N/2} \cdot (2\pi\gamma^2)^{N_s/2} \cdot \mathrm{e}^{-\frac{1}{2\epsilon^2}\mathbf{y}^T\mathbf{y} - \frac{1}{2\epsilon^2}(-2\mathbf{y}^T\hat{\mathbf{H}})\mathbf{s} + \mathbf{s}^T(-\frac{1}{2\epsilon^2}\hat{\mathbf{H}}^T\hat{\mathbf{H}} - \frac{1}{2\gamma^2}\mathbf{P}^T\mathbf{P})\mathbf{s}} && \text{(C.4)} \\
= \; & (2\pi\epsilon^2)^{N/2} \cdot (2\pi\gamma^2)^{N_s/2} \cdot \mathrm{e}^{-\frac{1}{2\epsilon^2}\mathbf{y}^T\mathbf{y} + \frac{1}{2}\mathbf{m_s}^T\mathbf{m_s}} \cdot \mathrm{e}^{-\frac{1}{2}(\mathbf{s}-\mathbf{m_s})^T\mathbf{C_s}^{-1}(\mathbf{s}-\mathbf{m_s})} && \text{(C.5)}
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{C_s} &= (\epsilon^{-2}\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \gamma^{-2}\mathbf{P}^T\mathbf{P})^{-1} && \text{(C.6)} \\
\mathbf{m_s} &= \epsilon^{-2}\mathbf{C_s}\hat{\mathbf{H}}^T\mathbf{y} && \text{(C.7)}
\end{aligned}
$$

This leads to:

$$
\ln p(\mathbf{y}|\epsilon, \gamma) = -\frac{N}{2}\ln(2\pi\epsilon^2) - \frac{N_s}{2}\ln(2\pi\gamma^2) - \frac{1}{2\epsilon^2}\mathbf{y}^T\mathbf{y} + \frac{1}{2}\mathbf{m_s}^T\mathbf{m_s} + \frac{N_s}{2}\ln 2\pi + \frac{1}{2}\ln|\mathbf{C_s}| \qquad \text{(C.8)}
$$

The maximization of $\ln p(\mathbf{y}|\epsilon, \gamma)$ is usually time-consuming, because of the term $\frac{1}{2} \ln |\mathbf{C_s}|$, which involves computing the determinant of an $N_s \times N_s$ matrix.

Appendix D

S-signal estimation with Expectation Maximization

E-step and M-step

At iteration i:[1]

- identification of the parameters (mean $\mu_{\mathbf{s}}$, and covariance $\mathbf{\Sigma_s}$) of the searched signal $\mathbf{s}$ from the expression of the conditional probability density $p(\mathbf{s}|\mathbf{y}, \epsilon_{i-1}, \gamma_{i-1})$ by factorizing the terms in $\mathbf{s}$ and those in $\mathbf{s}^T\mathbf{s}$ from the exponential function.

$$p(\mathbf{s}|\mathbf{y}, \epsilon, \gamma) = (2\pi)^{-\frac{N_s}{2}} |\mathbf{\Sigma_s}|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{s}-\mu_{\mathbf{s}})^T \mathbf{\Sigma_s}^{-1}(\mathbf{s}-\mu_{\mathbf{s}})} \tag{D.1}$$

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}, \epsilon, \gamma) &\propto p(\mathbf{y}|\mathbf{s}, \epsilon) \times p(\mathbf{s}|\gamma) && \text{(D.2)} \\ p(\mathbf{y}|\mathbf{s}, \epsilon) &= (2\pi\epsilon^2)^{-\frac{N}{2}} \mathrm{e}^{-\frac{1}{2\epsilon^2}(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})^T(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})} && \text{(D.3)} \\ p(\mathbf{s}|\gamma) &= (2\pi\gamma^2)^{-\frac{N_s}{2}} \mathrm{e}^{-\frac{1}{2\gamma^2}(\mathbf{P}\mathbf{s})^T(\mathbf{P}\mathbf{s})} && \text{(D.4)} \end{aligned}$$

$$\mathbf{s}^T\mathbf{\Sigma_s}^{-1}\mathbf{s} - 2\mu_{\mathbf{s}}{}^T\mathbf{\Sigma_s}^{-1}\mathbf{s} + cte = \epsilon^{-2}(\mathbf{s}^T\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{s} - 2\mathbf{y}^T\hat{\mathbf{H}}\mathbf{s} + cte) + \gamma^{-2}\mathbf{s}^T\mathbf{P}^T\mathbf{P}\mathbf{s} \tag{D.5}$$

[1] We drop the index of the iteration in the first part of the demonstration for readability purposes: $\epsilon = \epsilon_{i-1}$ and $\gamma = \gamma_{i-1}$.

$$\begin{aligned}\mathbf{\Sigma_s} &= (\epsilon^{-2}\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \gamma^{-2}\mathbf{P}^T\mathbf{P})^{-1} && \text{(D.6)}\\ \mu_{\mathbf{s}} &= \epsilon^{-2}\mathbf{\Sigma_s}\hat{\mathbf{H}}^T\mathbf{y} && \text{(D.7)}\end{aligned}$$

- computation of the expectation based on $\mu_{\mathbf{s}}$ and $\mathbf{\Sigma_s}$ and ignoring the terms that do not depend on ϵ and γ because they do not influence the maximization step.[2]

$$\xi_i(\epsilon,\gamma) = \mathbb{E}_{\mathbf{s}|\mathbf{y},\epsilon_{i-1},\gamma_{i-1}}[\ln p(\mathbf{y},\mathbf{s}|\epsilon,\gamma)] \tag{D.8}$$

$$\begin{aligned}\xi_i(\epsilon,\gamma) &= cte + \mathbb{E}[\ln p(\mathbf{y}|\mathbf{s},\epsilon,\gamma)] + \mathbb{E}[\ln p(\mathbf{s}|\epsilon,\gamma)] && \text{(D.9)}\\ &= cte - \frac{1}{2\epsilon^2}\mathbb{E}[(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})^T(\mathbf{y}-\hat{\mathbf{H}}\mathbf{s})] - \frac{N}{2}\ln(2\pi\epsilon^2) && \text{(D.10)}\\ &\quad -\frac{1}{2\gamma^2}\mathbb{E}[(\mathbf{P}\mathbf{s})^T(\mathbf{P}\mathbf{s})] - \frac{N_s}{2}\ln(2\pi\gamma^2) && \text{(D.11)}\\ &= cte - \frac{1}{2\epsilon^2}[\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\hat{\mathbf{H}}\mu_{\mathbf{s}} + \mathrm{Tr}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{\Sigma_s}) + \mu_{\mathbf{s}}{}^T\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mu_{\mathbf{s}}] && \text{(D.12)}\\ &\quad -\frac{N}{2}\ln(2\pi\epsilon^2) - \frac{1}{2\gamma^2}[\mathrm{Tr}(\mathbf{P}^T\mathbf{P}\mathbf{\Sigma_s}) + \mu_{\mathbf{s}}{}^T\mathbf{P}^T\mathbf{P}\mu_{\mathbf{s}}] && \text{(D.13)}\\ &\quad -\frac{N_s}{2}\ln(2\pi\gamma^2) && \text{(D.14)}\end{aligned}$$

After re-arranging the terms, we have:

$$\xi_i(\epsilon,\gamma) = cte - \frac{1}{2\epsilon^2}T_1 - \frac{N}{2}\ln(2\pi\epsilon^2) - \frac{1}{2\gamma^2}T_2 - \frac{N_s}{2}\ln(2\pi\gamma^2) \tag{D.15}$$

- derivation with respect to ϵ and γ:

$$\begin{aligned}T_1\frac{2\epsilon}{2\epsilon^4} - \frac{N}{2}\frac{4\pi\epsilon}{2\pi\epsilon^2} &= 0 && \text{(D.16)}\\ T_2\frac{2\gamma}{2\gamma^4} - \frac{N_s}{2}\frac{4\pi\gamma}{2\pi\gamma^2} &= 0 && \text{(D.17)}\end{aligned}$$

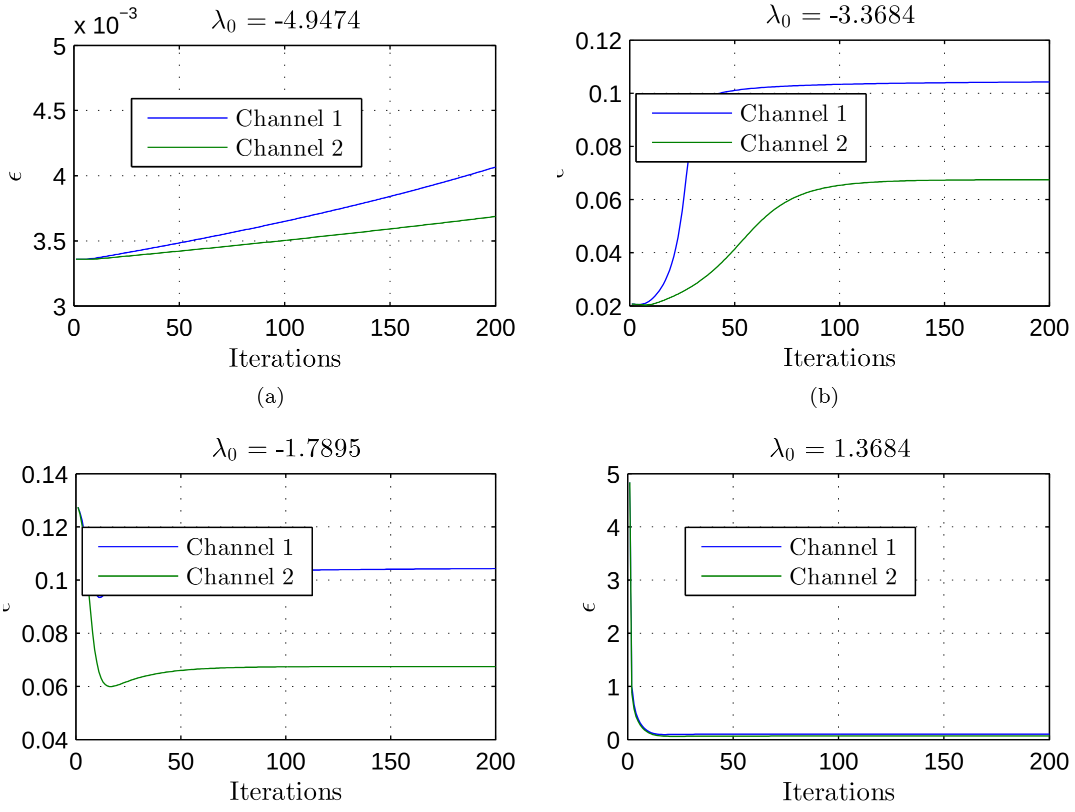
which leads to the final result:

$$\begin{aligned}\epsilon &= \sqrt{\frac{\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\hat{\mathbf{H}}\mu_{\mathbf{s}} + \mathrm{Tr}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{\Sigma_s}) + \mu_{\mathbf{s}}{}^T\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mu_{\mathbf{s}}}{N}} && \text{(D.18)}\\ \gamma &= \sqrt{\frac{\mathrm{Tr}(\mathbf{P}^T\mathbf{P}\mathbf{\Sigma_s}) + \mu_{\mathbf{s}}{}^T\mathbf{P}^T\mathbf{P}\mu_{\mathbf{s}}}{N_s}} && \text{(D.19)}\end{aligned}$$

[2] The index of $\mathbb{E}$ will be also neglected for readability purposes.

Convergence of the parameters

In this section, we show the behavior of the parameter values during 200 EM iterations starting from different initial conditions : $\gamma_0 = 1$ and $\epsilon_0 = \gamma_0\sqrt{\lambda_0}$. The convergence graphics are given for two different channel configurations. We note that, except for the very small λ_0 values, the algorithm manages to converge to stable parameter values.



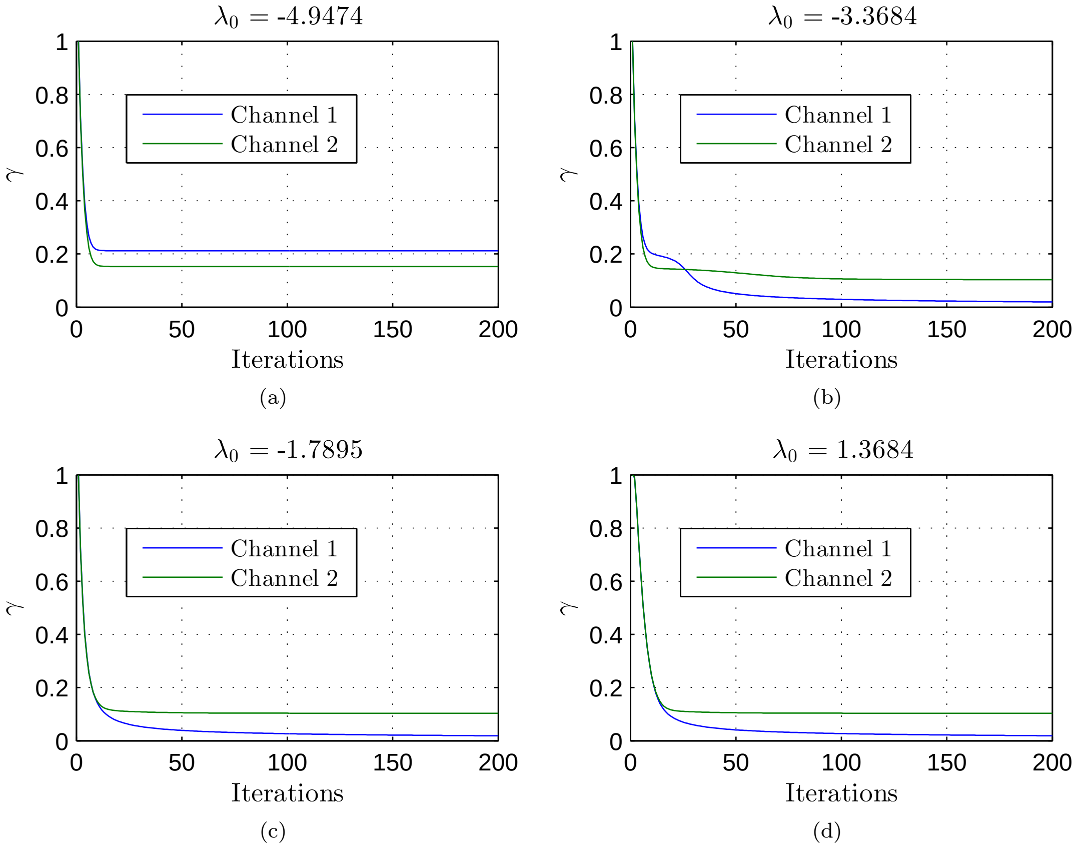
(a)

(b)

(c)

(d)

FIGURE D.1: Convergence behavior of parameter ϵ



(a)

(b)

(c)

(d)

Figure D.2: Convergence behavior of parameter γ

Appendix E

Location-based pseudonyms for identity reinforcement

I. Tunaru, B. Denis, B. Uguen (in Proc. IEEE VTC Spring'15)

In this paper we introduce methods that strengthen the identity of end-devices in order to prevent impersonation attacks in wireless ad hoc networks. Pseudonyms are locally generated from RSSI for narrow band IEEE 802.15.4 standard or Round Trip - Time of Flight (RT-ToF) measurements (optionally, along with relative clock drift estimates) for Impulse Radio - Ultra Wideband (IR-UWB) technology. These radiolocation features are converted into range measurements, quantized and then fed into hash functions to produce pseudonyms. For the two benchmarked radio technologies, practical trade-offs are illustrated depending on the input measurement accuracy under different channel assumptions. The evaluated solution enables to securely guess the pseudonyms of trusted neighbors with no information leakage. It also achieves advantageous low probability of successful attacks based on brute-force or statistics-aided strategies or compared to other impersonation detection strategies (e.g. RSSI history monitoring).

Introduction

In the emerging Internet of Things (IoT), location information may contribute to forge, reinforce or control the *ego* identity of independent devices and users for security or privacy purposes. It may be also useful to establish common identities within local groups or communities (sometimes referred to as IoT "Bubbles of Trust"). Such information reflects both the unique physical insertion of a device in its geographical environment (e.g., its absolute 2D position in a given building) and its interconnections with other collocated fellows or entities (e.g., its relative position with respect to neighbors). This "spatialized" identity could become a key feature in trust and reputation marks granted to mobile devices. Besides, various wireless localization technologies have emerged in the last past years, based on, e.g., fingerprinting with respect to WiFi access points, Bluetooth-Low Energy beacon-based cell attachment, range-based trilateration and co-operative positioning over Impulse Radio - Ultra Wideband (IR-UWB) or Zigbee links.

In this context, *ad hoc* location-based pseudonyms could be generated and used as a protection overlay to prevent impersonation attacks. Known uniquely by the legitimate devices (either through secure sharing or robust guess) and supposed secret with respect to attackers, they ease authentication procedures and protect data transactions. Secure procedures based on pseudonyms generated from stable connectivity or explicit location information rely on the fact that any malicious entity claiming a stolen public ID or address can neither experience nor generate the same physical awareness as that of the legitimate peers under attack. These schemes can also be extended to generate local secrets similarly to Physically Unclonable Functions (PUF) or to produce seeds for Pseudo Random Number Generators (PRNG).

This paper introduces simple methods to generate local pseudonyms out of various radiolocation features and technologies, which could be available in turn in most IoT devices, namely Received Signal Strength Indicator (RSSI) readings or measured Round Trip - Time of Flight (RT-ToF) of packets over the short-range peer-to-peer communication links. Particularly, we propose techniques that combine one-hop wireless connectivity, relative range information and optionally, device-dependent characteristics (e.g., estimated relative clock drifts, which are often required to mitigate RT-ToF measurement biases). The aim is not only to generate local pseudonyms but also to make possible the guess of these pseudonyms by trusted neighbors (secure legitimate inference). Simulation results show tangible gains in terms of : i) legitimate inference success rate, when integrating IR-UWB RT-ToF measurements instead of IEEE 802.15.4 RSSI readings as quantization inputs (because of lower noise dispersion in practical environments), ii) immunity against attackers with prior statistical knowledge about the expected range distributions, by coupling estimated relative clock drifts with range estimations, iii) prevention against impersonation attacks with the proposed range-based approaches as an alternative to direct RSSI monitoring solutions.

In Section E, we browse through related state-of-the-art contributions. Section E presents a generic scheme enabling location-based pseudonym generation, as well as a specific example based on peer-to-peer range information and optionally, relative clock drifts. In Section E, we evaluate the schemes as a function of system parameters such as the quantization grids. Section E concludes the paper.

State of the art

Available radiolocation technologies and modalities

Besides conventional GNSS means, alternative wireless localization solutions have been promoted recently [163], including low data rate IR-UWB and IEEE 802.15.4/Zigbee. Benefiting from low power consumption, these technologies also offer appealing peer-to-peer capabilities, which are suitable for mesh and cooperative connectivity in decentralized or versatile network contexts.

IR-UWB This technology enables RT-ToF and Time (Differences) of Arrival (T(D)oA) estimation with unprecedented timing accuracy, in the order of the nanosecond (i.e. within 30 cm spatial resolution) [135]. The RT-ToF gives direct access to the distance between two nodes. Side efforts have also been committed to design Medium Access Control (MAC) synergetic protocols with better support for both ranging and decentralized positioning functionalities. They typically rely on beacon-enabled Time Division Multiple Access (TDMA) superframe structures [164], estimating and compensating harmful relative clock drift effects through the use of cooperative n-way ranging transactions.

IEEE 802.15.4/Zigbee Various integrated solutions compatible with these two standards are currently available on the market. All of them can issue RSSI readings at the physical layer for each demodulated packet. Assuming a certain path loss model, such measurements can be exploited for parametric point-to-point range estimation [165]. However, in common environments, the expected precision of both ranging and positioning is hardly better than several meters [163].

Location-based identification and authentication

In order to prevent impersonation attacks, conventional cryptographic techniques may not be suitable in contexts like IoT, considering the massive deployment of low-cost and low-power entities with limited computational capabilities. Furthermore, cryptographic mechanisms usually need a centralized certified entity to distribute, refresh and revoke identity keys and signatures, which makes them more challenging to implement in decentralized networks with temporary *ad hoc* inter-connections.

Alternative non-cryptographic techniques relying on the lower layers of the communication protocol have thus emerged recently [166]. However, software-based methods (e.g., probe request behavior at the MAC level) or hardware-based solutions (e.g., radiometric fingerprinting, clock skewness or PUF) are still subject to practical limitations, requiring the exchange of numerous challenge-response messages or too specific hardware.

Additional lower layer techniques are based on the continuous monitoring of physical radio properties such as location-specific Channel State Information (CSI) or RSSI readings to detect identity-based attacks [167]-[168]. Channel sounding is then combined with hypothesis testing to determine if prior (authenticated/trusted) and current communications are issued by one unique user. The key challenge is to collect physical radio features over time so as to detect unexpected transients caused by impersonations. In [166], RSSI Similarity based Authentication (SA) and Temporal RSSI Variation Authentication (TRVA) methods are recalled. The SA technique aims at detecting large unexpected RSSI changes between consecutive frames at one receiver. In Section E, SA will be used as reference for benchmark purposes.

Finally, assuming error-free GNSS information in a WSN context [169], each sensor can be uniquely addressed by its 2D coordinates rather than an ID. Given a pre-loaded secret key (IK) generated from the initial ID and a system master key, each node securely receives an additional location-based key (LK) from one mobile entity. Node-to-node authentication and optional pairwise secret key establishment can then be applied using these location-based keys within a pairing-based crypto-system. Security lies in the secrecy of LK and one can verify that each node has the LK corresponding to its claimed position for authentication.

Overall, the previous concepts have not yet been extended to benefit from cooperation and heterogeneous radiolocation modalities for even better resilience against impersonations. Related recent work [170] explores the possibility of generating symmetric secret keys for cryptography out of relative location estimates in mobile networks.

Proposed scheme: location-based pseudonyms

Generic algorithm with heterogeneous inputs

The proposed approach consists in gradually using explicit or related location information to generate local pseudonyms, which complete or even temporarily substitute the IDs of wireless end-devices. Whenever multiple sources of location-dependent information are available at a given device, one can perform heterogeneous data integration, as follows:

$$PS_i = f(ID_i, [x_i, y_i], \{d_{ij}\}_{j\in Ne(i)}, \{\gamma_{ij}\}_{j\in Ne(i)} \cdots) \tag{E.1}$$

where ID_i is the public ID of node i, $[x_i, y_i]$ are the 2D absolute Euclidean coordinates of node i delivered by, e.g., GNSS or a WiFi-based localization system, $\{d_{ij}\}_{j \in Ne(i)}$ is the set of measured peer-to-peer distances (possibly drift-biased, with relative clock drifts $\{\gamma_{ij}\}_{j \in Ne(i)}$ w.r.t. neighbors $j \in Ne(i)$). Note that much simpler characteristics (still accounting for the "physical insertion" of the device) could be considered as well, such as node i's $(N - hop)$ connectivity w.r.t. to nodes $j \in Ne(i)$.

One straightforward implementation of $f(.)$ is a hash function. Hash functions map inputs of different lengths to a fixed-length output, while small changes in the input produce a completely different output. Moreover, cryptographic hash functions (e.g. SHA-1) are deemed to be collision-resistant, meaning that there is a small probability to find two different inputs that give the same output. This property is useful in order to avoid pseudonym collisions between legitimate nodes but also with respect to an attacker.

The pseudonyms between different legitimate nodes can be shared either through direct communication during a secure short time period or by inference (i.e., the neighbors of a given node try to guess its pseudonym to the best of their knowledge). In the inference case, the measurements feeding the hash function are quantized in order to provide better stability and reciprocity against measurement noise for legitimate nodes. The setting of the quantization step is a key parameter that can be adapted online according to empirical measurement statistics. It is obvious that inference is preferable for security reasons. However, the choice of the method depends on the available information used as input to the hash function. Overall we identify two preferred embodiments:

- One *global pseudonym* per node: the hash input can contain location information (e.g. position, relative distances w.r.t. all the neighbors) and/or connectivity information (e.g. adjacency information), and/or further device-dependent information (e.g. relative clock drifts w.r.t. all the neighbors). Inferring this type of pseudonym requires public exchanges, which, in the long term, could lead to the full disclosure of the network characteristics and can limit the advantage with respect to attackers. In this case, the pseudonyms must be shared securely.
- *Link-dependent pseudonyms*: the hash function outputs a new ID of the present node w.r.t. each neighbor using link-dependent information (e.g. relative distance or relative clock drift w.r.t. to the respective neighbor). In this case, the inference is facilitated thanks to the assumed reciprocity of the input data (See Figure E.1) and it can be achieved without any public exchange of information.

Therefore, pseudonym generation schemes rely mainly on non-complex digital operations (quantization and hash functions) and on the ranging capabilities increasingly present

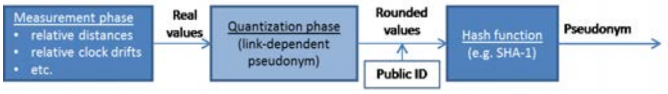


FIGURE E.1: Possible block diagram of a location-based pseudonym generator.

in a large number of communicating devices. The inference phase does not increase the complexity because it does not require any protocol overhead or any additional processing of the measured data.

We detail hereafter one example based on the quantization of peer-to-peer range information between two devices (from RSSI/RT-ToF measurements and optional relative clock drift estimates). The same pseudonym generation algorithm will be used for the evaluations in Section E.

Link-dependent pseudonyms from peer-to-peer ranges and relative clock drifts

In this particular scheme, each node generates N pseudonyms, with which it will be addressed by its N neighbours. As an example, the pseudonym (generated at A) that should be used by node B to address node A is:

$$PS^A(A) = hash([ID_A||qd_{AB}(\Delta_d))]) \tag{E.2}$$

with $hash$ the cryptographic hash function (SHA-1 in the present simulations), $||$ the concatenation function, ID_A the initial public ID of A, Δ_d the distance quantization step, qd_{AB} the quantized relative distance between A and B measured at A using: i) n-way ranging protocols and Time of Arrival (ToA) estimation to issue RT-ToF measurements (e.g., in IR-UWB, [135]), ii) RSSI-based ranging with a path loss model (e.g., in IEEE 802.15.4, [165]).

One step further, adding hardware device-dependent information (only for RT-ToF estimations), the new pseudonym is:

$$PS^A(A) = hash([ID_A||qd_{AB}(\Delta_d)||q\gamma_{AB}(\Delta_\gamma)]) \tag{E.3}$$

under the same notations as before, with $q\gamma_{AB}$ the quantized relative drift of A's clock with respect to B's clock [164] (by definition, only measured at node A using the same ranging procedures) and Δ_γ the related quantization step.

The inference of A's pseudonym without public information exchange is made possible because node B also estimates similar reciprocal characteristics: relative distance

($d_{BA} \cong d_{AB}$) and optionally, relative clock drift ($\gamma_{BA} \cong 1/\gamma_{AB}$), depending on the chosen technology and features. Asymmetries in the quantization inputs are thus mainly caused by measurement noise, which affects the unitary ToA estimates producing the RT-ToF/range estimations and by power fading for RSSI measurements. The legitimate peer B will then generate an inferred pseudonym for A:

$$PS^B(A) = hash([ID_A||qd_{BA}(\Delta_d)||q(1/\gamma_{BA})(\Delta_\gamma)]) \tag{E.4}$$

An attacker E (passively) eavesdropping or (actively) exchanging packets with nodes A and B can estimate neither the relative distance between A and B (because of its different location), nor the relative clock drift, which is inherent to both the link and the hardware characteristics of the legitimate peers. Moreover, if E compromises one node, it cannot find the pseudonyms that are used on the other non-adjacent links. Therefore, the attacker has to make a blind (or eventually statistics-assisted) guess on the generated pseudonym in order to obtain $PS^E(A)$.

Performance evaluation and discussion

In order to evaluate the robustness of our proposal with respect to noise over legitimate links, as well as its security in the presence of an attacker, we define two metrics:

- the probability of a successful inference of the pseudonym by a legitimate user : $P_l = P[PS^B(A) = PS^A(A)]$ evaluated through Monte Carlo simulations, considering 1000 distinct network realizations of 10 nodes each, with an average node degree between 7 and 8 neighbors and with uniformly distributed coordinates in a 20x20m area.
- the probability of a successful guess by an attacker: $P_{(p)BF} = P[PS^E(A) = PS^A(A)]$ evaluated by numerical integration over the internode distances in two cases: i) brute-force (BF) attack (i.e., the attacker makes a uniform random guess on the value of the internode distance and/or relative clock drift); ii) probabilistic brute-force (pBF) or statistics-aided guess (i.e., the attacker knows *a priori* the distribution of the true internode distances in the area and makes his best guess accordingly).

The successful attack probabilities (when using both range and drift information for quantization) are computed as follows:

$$P_{(p)BF} = \int_0^{R_{max}} p_R(r) P_{(p)BF|R}(r)\,\mathrm{d}r \tag{E.5}$$

$$
\begin{aligned}
P_{(p)BF|R}(r) &= Prob[PS^{E}(A) = PS^{A}(A)|r] && \text{(E.6)}\\
&= P^{d}_{(p)BF}(r) \times P^{\gamma}_{BF} && \text{(E.7)}\\
&= Prob[\tilde{qd}_{AB} = qd_{AB}|r] \times Prob[\tilde{q\gamma}_{AB} = q\gamma_{AB}] && \text{(E.8)}
\end{aligned}
$$

with R_{max} the maximum internode distance/range, $p_R(r)$ the *a priori* probability density function of the internode distance for uniformly distributed nodes, R the random variable representing the legitimate(A-B) internode distance, $\tilde{f}$ the guessed feature $f \in \{qd, q\gamma\}$ at the attacker, $P^{d}_{(p)BF}$ the probability of guessing the quantized distance based on: i) BF : $P^{d}_{BF}(r) = \frac{\Delta_d}{R_{max}}$ or ii) pBF: $P^{d}_{pBF}(r) = p_R(r)$, P^{γ}_{BF} the probability of guessing the quantized drift, independently of the internode distance: $P^{\gamma}_{BF} = \frac{\Delta_\gamma}{\frac{1+\delta}{1-\delta} - \frac{1-\delta}{1+\delta}}$ where δ is the maximum clock imprecision (specified by the manufacturer).

We perform three studies. Firstly, we investigate the differences between the quantization of the estimated distances issued from two different technologies, namely Narrowband (NB) IEEE 802.15.4 at 2.4 GHz and IR-UWB. The relative distances are computed from RSSI and RT-ToF measurements, respectively in NB and IR-UWB cases (both under typical radio assumptions and parameters). Secondly, we evaluate the performances of pseudonym generation schemes jointly based on distance and drift for the IR-UWB case only. Finally, we compare our impersonation prevention method to an RSSI-based monitoring method [166].

Pseudonyms from NB & IR-UWB range estimates

The received power measurements reflected by RSSI readings in NB are generally expressed (in dB) as a function of a log-normal distance-dependent path loss model [165], whose key parameters are PL_{ref}, the path loss reference at distance d_{ref}, α, the path loss exponent, and X_S, random centered Gaussian shadowing with standard deviation σ_s.

In our simulations, we consider assigning to all the feasible links some random channel configurations (i.e., Line of Sight (LOS), Non Line of Sight (NLoS) and severe NLoS (NLoS2)), along with their corresponding radio parameters (i.e., ToA standard deviation for RT-ToF vs. shadowing standard deviation, reference path loss and path loss exponent for RSSI), depending on the actual internode distance like in [171]. Once each link configuration has been allocated, one can use the associated conditional model parameters to estimate the range. For NB RSSI-based ranging, we consider the median estimator from [165], with $\alpha = [1.7, 3, 5]$ and $\sigma_s = [0.5, 3, 5]$ dB respectively in [LoS,

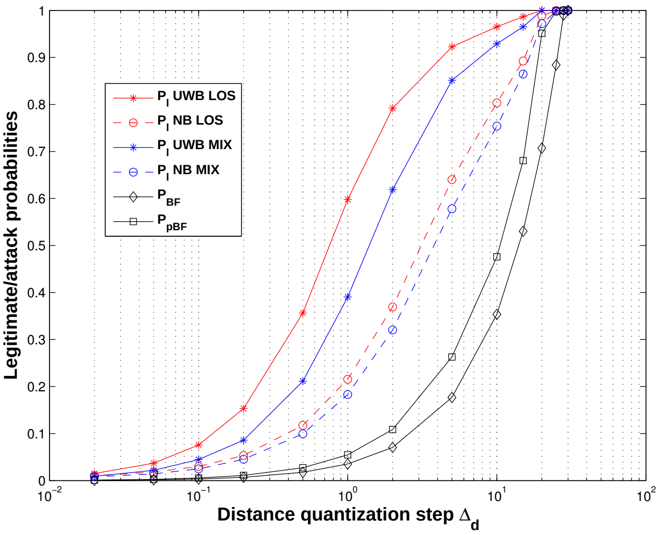


FIGURE E.2: Performance of range-based pseudonym generation : legitimate inference success probability (P_l) and attack success probability ($P_{(p)BF}$)

NLoS and NLoS2]. For IR-UWB RT-ToF estimates, we consider the standard deviation of ToA estimation noise as $[1, 2, 3]$ ns respectively in [LoS, NLoS, and NLoS2], according to empirical observations from [94]. Relative clock imprecisions are assumed to be bounded by $\pm\delta = 20$ ppm (worst case), which is representative for low-cost embedded oscillators.

In Figure E.2, we report the legitimate agreement probabilities for distance quantization from NB and IR-UWB estimates with two link configurations (only LoS and a mixture (MIX) of LOS, NLoS and NLoS2). Pseudonym generation from IR-UWB RT-ToF estimates is shown to be more robust to measurement noise for the same link configurations. This is due to the more advanced ranging capabilities of the IR-UWB technology compared to NB technologies (i.e., larger bandwidth implies more precision). The difference is significant in the LOS case where the detection of ToA of the first multipath component for IR-UWB is more robust than in the NLOS case. We also report the brute-force and the probabilistic brute-force successful attack probabilities as a function of the distance quantization step and identify advantageous quantization steps that maximize the distance between the legitimate agreement and the successful attack curves (e.g. $\Delta_d \in (1, 10)$ m).

Pseudonyms from IR-UWB range & clock drift estimates

Using the same framework as before, we incorporate the drift into the quantization procedure (we choose a fixed distance quantization step $\Delta_d = 10$m). When adding the

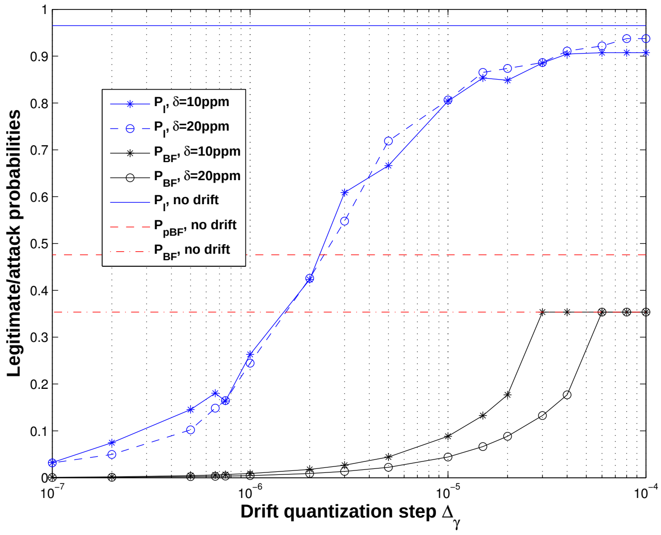


Figure E.3: Performance of pseudonym generation based on relative range vs. jointly relative range and clock drift measurements at fixed $\Delta_d = 10$ m

relative drift information, both the legitimate agreement and the successful attack probabilities decrease (Figure E.3) because both the legitimate inference and the attacker's guess become harder. Nevertheless, incorporating the drift presents an advantage with respect to the probabilistic brute-force strategy at relatively high values of Δ_γ: P_l with drift approaches the no-drift P_l while the new successful attack probability P_{BF} is always lower than the no-drift P_{pBF}. As the attacker has no indication on the possible value of the relative clock drift (hardware characteristic), the brute-force solution on the drift value is the only option, which makes the overall guess of the pseudonym more difficult. In conclusion, pseudonyms should be generated from reciprocal but uniformly distributed information in order to achieve both a satisfactory legitimate agreement probability and a low successful attack probability.

Comparison with RSSI monitoring authentication

The probability of a successful attack in RSSI monitoring methods (P_{sa}) depends on the variance of the shadowing (σ_s^2) that affects the RSSI readings and on the distance between the attacker and a legitimate node (d_{AE}) when averaging over the distances between the legitimate nodes (d_{AB}). The averaging is performed in the same way as in Eq. (E.5) by replacing $P_{(p)BF|R}$ with the probability of misdetection of an attacker as a legitimate node (conditioned on the range d_{AB}). We assume that the threshold for impersonation detection at the legitimate nodes is set at $\pm 3\sigma_s$ from the mean empirical RSSI value. Note that the mean RSSI value and d_{AB} are linked by a deterministic formula based on the propagation model mentioned in Section E. From Figure E.4, we can observe that even distance-based pseudonym generation with relatively large

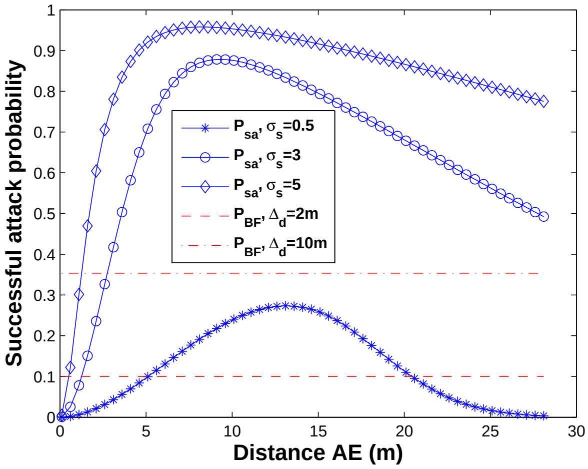


FIGURE E.4: Comparison of the successful attack probabilities in pseudonym generation vs. RSSI monitoring for different attacker-legitimate distances d_{AE}

Δ_d (e.g., 10m) outperform most of the RSSI-based authentication for medium to large shadowing variances.

Conclusion

In this paper we have investigated and discussed the possibility to generate, infer and/or share local pseudonyms out of various sources of radiolocation information for more secure IoT transactions and reinforced device identity. For the link dependent pseudonyms, we have shown that they could be securely guessed by legitimate neighbors, while achieving relatively low impersonation success rates. However, the absolute value of the attack success rates suggest that link dependent pseudonyms should be used as an authentication overlay.

Remaining challenges concern the degraded precision and the asymmetry of input measurements in typical environments. The performance of the proposed algorithms could be assessed on real integrated devices and field measurements. Mobility support is also challenging for the perennial tracking and management of new generated location-based IDs. It should also be noted that the pseudonyms provide a security overlay as long as the initial radiolocation measurements phase is not compromised by an active attacker. However, in mobility scenarios, pseudonyms can be easily refreshed in order to counter passive attacks (e.g., eavesdropping).

Appendix F

Experimental key generation with low-complexity devices

M. Bulenok, I. Tunaru, B. Denis (internship work of M. Bulenok)

In wireless decentralized networks security and privacy become more challenging because of the decentralized architecture and sometimes low-complexity profile. Confidentiality of the transmissions is usually achieved by symmetric cryptography, which implies a distribution of identical secret keys between the legitimate nodes. Classic solutions for symmetric key distribution can be too complex from a computational or infrastructural point of view. An alternative to these key management techniques is the physical layer key generation that exploits the radio channel as a source of common information. This paper is dedicated to the generation of secret keys from IR-UWB channels using low-power integrated devices. We investigate the possibility to generate identical secret keys in practical environments (e.g., typical indoor rooms, possibly occupied) and operating conditions (e.g., under mobility). The main challenge is to obtain reciprocal and random keys at an acceptable rate. Taking into account these three criteria (reciprocity, randomness, and key length), we propose an adapted quantization scheme for the measured CIRs acquired with embedded devices.

Introduction

Because of their broadcast nature, wireless networks can suffer from eavesdropping attacks. Therefore, one of the main issues for reliable communication is security, more specifically confidentiality of exchanges. This is usually achieved by symmetric-key cryptography, which relies on encryption and decryption operations using a secret shared key between the two legitimate users. So one of the main challenges is how to distribute this common key among legitimate users in a secure way.

Among the classical key distribution approaches there are the Key Distribution Center (KDC) [172], the Diffie Hellman (D-H) exchange protocol [7] and asymmetric cryptography. In the KDC method there is a third party who is responsible for the creation (or the storage) and distribution of keys among the users. This method is highly centralized, thus it is not adapted to ad-hoc/decentralized mobile networks. D-H exchange is a decentralized method but it includes expensive exponential operations, which should be avoided for resource constrained devices. Lastly, asymmetric cryptography relies on cumbersome certification operations for the public keys.

Physical layer security [18] includes an alternative approach for symmetric key distribution. It allows to extract a common secret key from the radio channel by measuring its features. The direct and reverse channels between two legitimate users (Alice and Bob) are theoretically reciprocal, but noise and other transmission artifacts can produce slightly different measurements. However, if an attacker, Eve, situated in a different position than Bob, will try to generate the same key, she will not succeed because of the spatial decorrelation of radio channels under certain distance and environmental conditions (i.e., the channel between Alice and Bob is uncorrelated with the channel between Alice and Eve).

The procedure of physical layer key generation can be separated in four steps [26]:

- Channel probing used to collect channel measurements by Alice and Bob. This step also includes post-processing in order to extract randomness from the measured signals.
- Quantization used to convert the extracted channel measurements into bits.
- Reconciliation used to correct random errors caused by imperfect reciprocity with the help of error correction codes such as Reed-Solomon or LDPC codes. During this procedure, the parity information is exchanged over a public channel and a certain amount of bit information will be revealed to Eve.
- Privacy amplification used to eliminate Eve's partial information about the key and the correlation among the bits.

Herein, we focus on the two first steps: post-processing of the channel acquisitions and quantization in order to produce appropriate raw keys.

Regarding quantization, one can apply several types of algorithms (Figure F.1). For example, *One-level* bit extraction means comparing the samples with a fixed threshold.

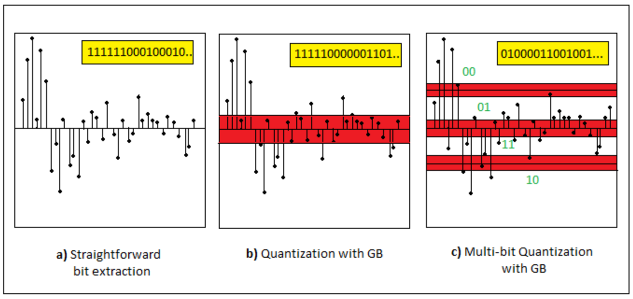


Figure F.1: Examples of types of quantization algorithms

If a sample is below the threshold the bit is set to “0”, otherwise it is set to “1”. It is a quite intuitive method, which guarantees a certain key generation rate (1 bit per sample) and a quite low mismatch rate, but keys might not be random enough. *Quantization with Guard-band* discards samples within a guard-band of the quantization threshold in order to obtain more reciprocal signals. It also implies that an additional sharing over a public channel of the discarded samples is needed. According to *Adaptive Quantization*, thresholds are adapted to each block of the received signal. Therefore every block is quantized individually, for example based on the mean value and standard deviation of its samples’ amplitudes. The main advantage of such schemes is the improved randomness. Instead of using guard-bands, [104] illustrates *“Channel Quantization Alternating”* (CQA) bit extraction scheme. In this method they combine quantization and reconciliation procedures. The main idea is that Alice quantizes her sequence and sends some information about it (a quantization map) to Bob through the public channel. After this, Bob quantizes his sequence using the obtained information. This helps Bob to reduce the error probability of his sequence.

This study is concerned with the generation of secret keys using Channel Impulse Responses obtained from IR-UWB low-complexity nodes developed at CEA-Leti [94]. The generated keys are evaluated in terms of reciprocity, randomness and length. Also, a new quantization method adapted to the signals issued by the employed nodes is proposed.

The remainder of the study is organized as follows. The first section introduces the measurement campaign, shows the environment conditions in which the measurements took place, and describes the main metrics employed in this study. In the second section, we show the limitations of amplitude quantization for our specific signals. The proposed solution, analysis and the evaluation of results are presented in the third section. Finally, we conclude the study and provide several perspectives for further work.

Experimental setup

Communication protocol

For obtaining a CIR pair we use the output of low-power integrated IR-UWB devices initially designed for localization purposes (i.e., TCR nodes [94]). Two sensor nodes have been used in our experiments. One of them is connected to a computer in order to transmit and save all the collected data by both of them. It is the main node and we call it Coordinator. The second one (Remote node) is not connected to the PC and it sends all its data to the Coordinator. Figure F.2 illustrates the structure of the communication protocol between these nodes. The time gap between the bidirectional measurements is 7.5 ms and the one between consecutive measurements 150.7 ms.

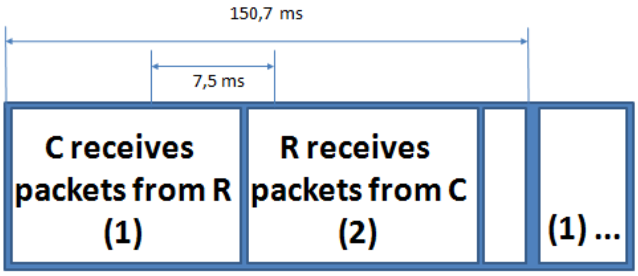


Figure F.2: Structure of the communication protocol

Experimental scenarios

The measurements took place in three separate rooms under different conditions of mobility, occupancy (with or without people) and geometry. We created three scenarios:

- The *static* scenario took place in a Laboratory with metallic equipment and furniture but without people. This is the simplest scenario in terms of reciprocity, because the nodes were not mobile and they always have LOS. The acquisition time was of approximately 2-3 min for every Tx-Rx position.
- The *occupied* scenario took place in the Coffee room. At any time there were at least a few people inside. The acquisition time was of 4-5 min for every Tx-Rx position.
- The *mobile* scenario took place in the Meeting room. As it can be seen from Figure F.3, the room is equipped with a U-shape table. There was a fixed position for the Coordinator at the end of the table. The Remote node was moving around table (dashed line) and did 3 circles.

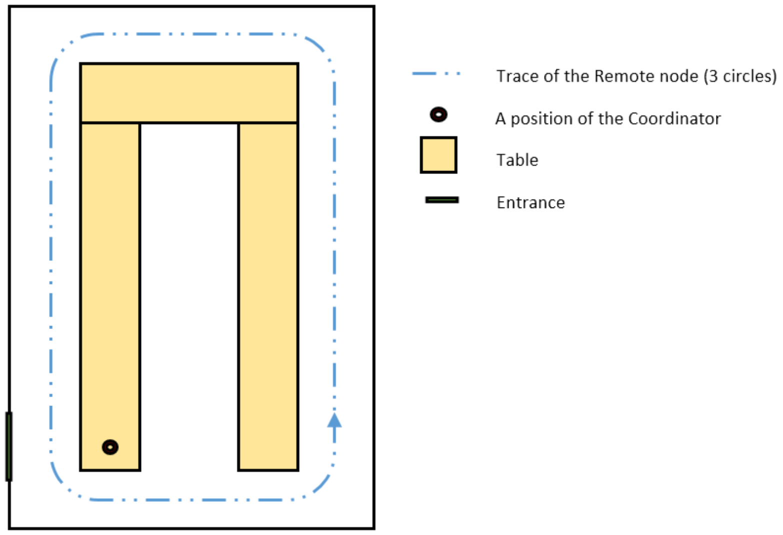

Trace of the Remote node (3 circles)
A position of the Coordinator
Table
Entrance

Figure F.3: Plan of the Meeting room

Signal preprocessing

There are two main problems caused by hardware and protocol limitations: signal desynchronization and predictive signal structure.

For solving the desynchronization problem between the bidirectional signals (see Figure F.5), we test two methods: MAXCORR and THRESHOLD.

The idea of MAXCORR is to find the shift (m^*), which maximizes the correlation between the fixed vector c (a CIR measured by the Coordinator) and the m-shifted vector b (a CIR measured by the Remote node).

$$m^* = \underset{m}{\operatorname{argmax}}\ f(m), \tag{F.1}$$

$$f(m) = \begin{cases} \frac{\sum_{n=1}^{N+m}(c_{n-m}-\mu_c)(r_n-\mu_r)}{N+m}, & m < 0; \\ \frac{\sum_{n=1}^{N-m}(c_n-\mu_c)(r_{n+m}-\mu_r)}{N-m}, & m \geqslant 0. \end{cases} \tag{F.2}$$

where N is the length of vectors b and c, m is a shift ($m < 0$ means that the vector is shifted to the right, $m > 0$ means that it is shifted to the left), $m \leq 0.35N$, μ_c is the mean value of the vector c, μ_b is the mean value of the vector b, c_n and b_n are the n^{th} elements of vectors c and b. This method is optimal but in a real scenario we cannot use it because one node is not aware about the measurements from the other node. This scheme will be used as a reference to estimate the optimal performance of key generation.

In the THRESHOLD scheme one sets a certain noise threshold and puts the first value above the threshold on the first position. This method can be applied to both nodes independently, thus it can be easily implemented in a realistic key generation scenario.

The reciprocity between the two bidirectional measurements after synchronization is measured using the Pearson linear correlation coefficients (i.e., the reciprocity coefficients). In order to compare the THRESHOLD synchronization scheme and the optimal MAXCORR synchronization, we show in Figure F.4 the histogram of the signed difference between the reciprocity coefficients (i.e., MAXCORR - THRESHOLD) after synchronization with these two methods. As expected, MAXCORR synchronization gives better results because on average MAXCORR reciprocity coefficients have a gain of 0.4.

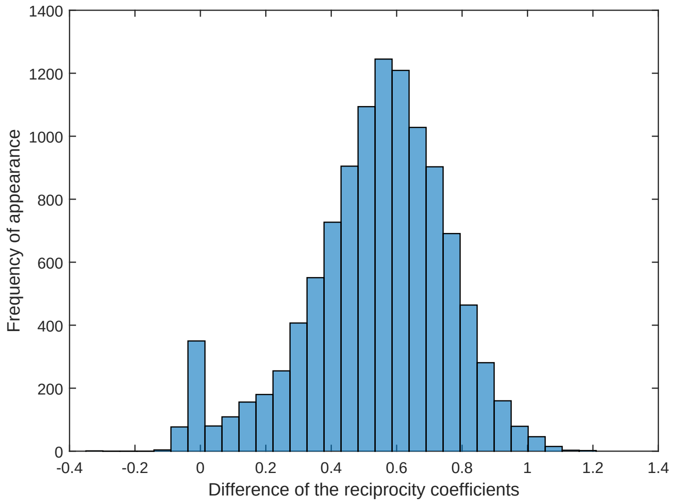


FIGURE F.4: Histogram of the difference between the reciprocity coefficients obtained with MAXCORR and THRESHOLD

Typical CIRs have some noisy parts at the beginning and at the end (see Figure F.5). In order to avoid these correlated samples, we set the noise threshold independently on each side and we crop the main part of the signal. Subsequently, we send through the public channel the sizes of the obtained sequences at each node and additionally crop the longer one in order to match the lengths.

Preliminary amplitude quantization

After channel probing both parties have to quantize the extracted channel measurements into bits. On the one hand, after applying the quantization scheme, keys should have small mismatches in order to be able to correct them with an error correction code without leaking too much information on the public channel. On the other hand, keys

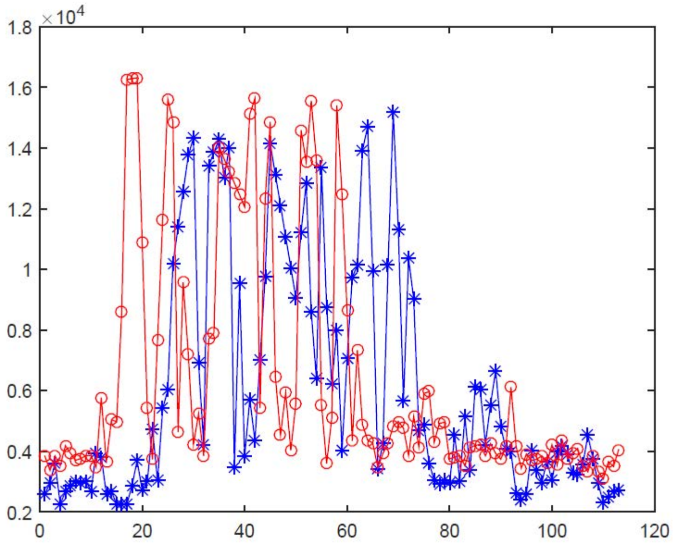


Figure F.5: Example of the typical bidirectional CIRs

should be random enough, e.g., they should pass the NIST tests [102]. Finally, they should also be as long as possible.

Firstly, we test several quantization approaches that exploit the signal amplitude of cropped or uncropped CIRs as the input data: 1-bit (or one-level), 2-bit and 2-bit with guard bands (GBs). The main parameters for each quantization algorithm are the thresholds that divide the amplitude space into cells. Our thresholds are defined independently for each node. They are computed offline for the whole set of measurements. Also, we concatenate 10 consecutive CIRs for obtaining sufficiently long keys. Table F.1 shows the mismatch rates for various quantization algorithms and scenarios.

Table F.1: Comparison of different quantization algorithms in Static (Mobile) scenarios, with (without) signal cropping

	Mismatch			
	Entire signal		Cropped signal	
Quantization algorithm	*Static*	*Mobile*	*Static*	*Mobile*
1-bit	0.14	0.25	0.29	0.33
2-bit	0.26	0.33	0.41	0.35
2-bit + GBs	0.22	0.24	0.26	0.29

Unfortunately, in our particular case, such kind of quantization algorithms give unacceptable high mismatch rate (i.e., always higher than 14%, even for static scenario and for the simplest one-level extraction algorithm). Moreover, it is not possible to use the entire uncropped signal for key generation because of the randomness defects of the obtained keys. One can clearly see this feature in Figure F.6 that compares keys obtained using 1-bit quantization on cropped and uncropped signals, where white corresponds to "1" and black corresponds to "0" bits).

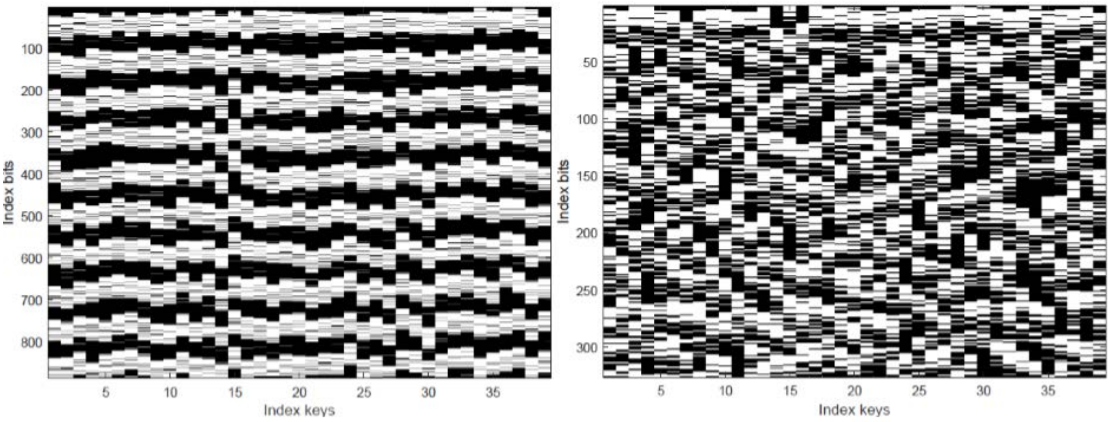


FIGURE F.6: The keys after 1-bit quantization for static scenario before and after cropping

In conclusion, concatenation of uncropped signals gives keys that can be predictable. The extraction of the main part of a signal (or cropping) helps to avoid such kind of problems but leads to higher mismatch rates. Thus a new quantization scheme is required.

Proposed solution

Before searching for a new quantization input, it is reasonable to refine the signal by passing it through a moving average filter. The filtering can help make the signals more reciprocal, because it makes them smoother. Figures F.7 and F.8 show the signals before and after filtering with a window size equal to 10. This choice is made so as to obtain a reciprocity coefficient of minimum 0.9.

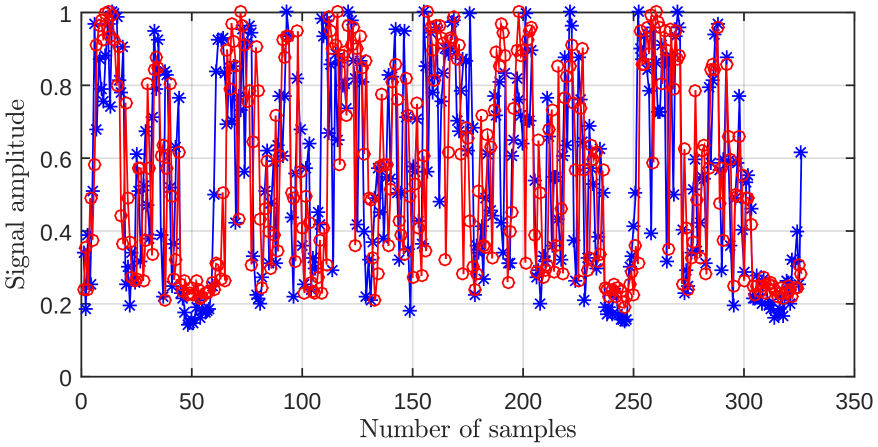


FIGURE F.7: Signals before filtering.

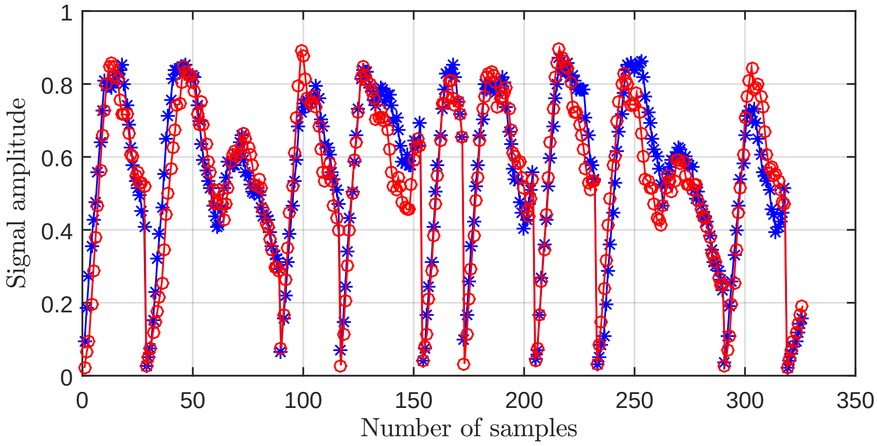


FIGURE F.8: Signals after filtering.

The filter helps decreasing the mismatch rate but we still have a problem with randomness. The samples of the signal become highly correlated and cannot be directly quantized. The idea of the new quantization scheme is to divide the signals in windows with a fixed number of samples inside, then to set a certain threshold and to quantize the number of samples above the threshold according to a one-level quantization scheme.

The proposed scheme gives a rather promising mismatch (less than 10%). In this method the length of the key becomes smaller and we need more time for obtaining a key of required length, which translates into a concatenation of more CIRs. For this reason the measurements in the meeting room and room with occupancy were repeated with longer acquisition time (15 minutes) and we concatenated 40 CIRs for obtaining one key. Therefore this scheme has a lower key rate ($\sim 18\frac{bits}{s}$) comparing to amplitude-based quantization, but it would still be higher than RSSI-based key generation algorithms employing the same communication protocol. Figure F.9 illustrates the keys after the moving average filtering and the new quantization.

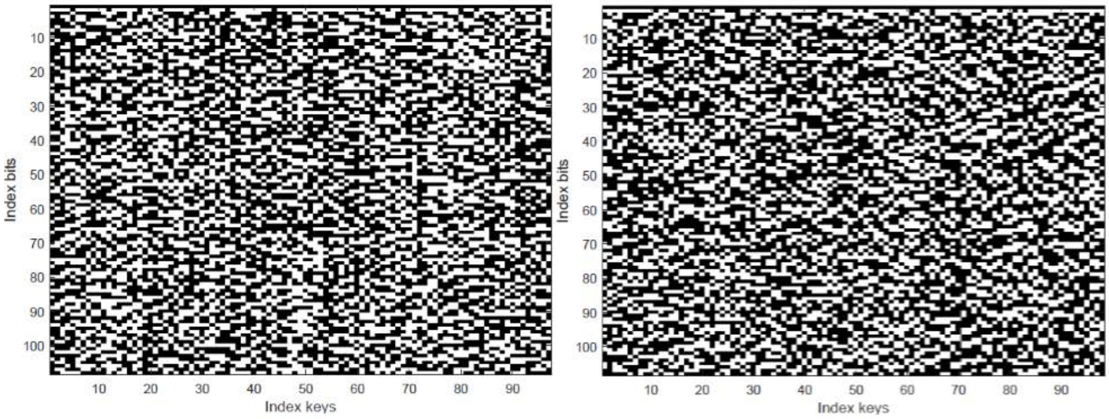


FIGURE F.9: The keys after the moving average filtering and the new quantization scheme applying.

Table F.2: Comparison of different quantization algorithms in Static (Mobile) scenarios, with (without) signal cropping.

Test	Occupancy scenario	Mobile scenario
The Frequency test	98/98	92/97
Frequency test in a block	98/98	95/97
The Cumulative Sums test	98/98	93/97
The Runs test	97/98	94/97
The Entropy test	98/98	87/97

In order to further investigate the random character of the keys after the new quantization scheme we applied the NIST tests. For the measurement in the Coffee room we obtain 98 keys of 108 bits and for the measurement in the Meeting room 97 keys of the same length. Table F.2 shows that a good percentage of keys pass the tests.

Conclusion

The generation of secret keys from IR-UWB channels, using low-complexity devices, has been investigated in this work. The measurements were performed in different environments. Estimated channel impulse responses were used as an input for key generation. After channel probing, CIRs were synchronized, cropped and quantized in order to obtain raw secret keys. Standard amplitude quantization methods are not suitable for these particular devices, thus we proposed an alternative technique based on filtering and quantization of the number of samples above a threshold. The final implementation achieves good performance in terms of reciprocity and randomness. Nevertheless it has a lower key rate than amplitude-based quantization.

Further work could continue the investigation of the secret key generation procedure in other environmental conditions. For example, it would be interesting to test a mix of mobile and occupied scenarios. The optimization of the final quantization parameters and realistic synchronization methods could also be studied. Finally, the impact of public information (e.g., cropping and quantization parameters) on the predictability of the keys should be considered as well.

Bibliography

[1] C. Shannon, “Communication theory of secrecy systems,” *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[2] C. Boyd and A. Mathuria, *Protocols for Authentication and Key Establishment.* Berlin, Germany: Springer, 2003.

[3] G. Stoneburner, C. Hayden, and A. Feringa, “Computer Security,” Information Technology Laboratory, NIST, Gaithersburg, Maryland, Tech. Rep., 2004.

[4] G. Boddapati, I. Matta, J. Day, and L. Chitkushev, “Assessing the Security of a Clean-Slate Internet Architecture Layer,” in *Seventh Workshop on Secure Network Protocols (NPSec)*, October 2012.

[5] “History of Cryptography,” 2013. [Online]. Available: http://book.itep.ru/depository/crypto/Cryptography_history.pdf

[6] J. Katz and Y. Lindell, *Introduction to Modern Cryptography.* CRC Press, 2014.

[7] W. Diffie and M. Hellman, “New directions in cryptography,” *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.

[8] Y. Fang, X. Zhu, and Y. Zhang, “Securing Resource-Constrained Wireless Ad Hoc Networks,” *IEEE Wireless Communications*, vol. 16, no. 2, pp. 24–30, Apr. 2009.

[9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A Survey on Sensor Networks,” *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[10] R. Watro, D. Kong, S.-f. Cuti, C. Gardiner, C. Lynn, and P. Kruus, “TinyPK: Securing Sensor Networks with Public Key Technology,” in *Proc. 2Nd ACM Workshop on Security of Ad Hoc and Sensor Networks*, ser. SASN ’04. New York, NY, USA: ACM, Oct. 2004, pp. 59–64.

[11] Y. Zhang, W. Liu, W. Lou, and Y. Fang, “Securing Mobile Ad Hoc Networks with Certificateless Public Keys,” *IEEE Transactions on Dependable and Secure Computing*, vol. 3, no. 4, pp. 386–399, 2006.

[12] ——, "Location-based Compromise-Tolerant Security Mechanisms for Wireless Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 247–260, Feb. 2006.

[13] K. Ren, W. Lou, and Y. Zhang, "LEDS: Providing Location-Aware End-to-End Data Security in Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 585–598, May 2008.

[14] Y. Zhou, Y. Fang, and Y. Zhang, "Securing Wireless Sensor Networks: a Survey," *IEEE Communications Surveys Tutorials*, vol. 10, no. 3, pp. 6–28, Mar. 2008.

[15] A. D. Wyner, "The Wire-tap Channel," *Bell Systems Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Jan. 1975.

[16] C. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental Quantum Cryptography," *Journal of Cryptology*, vol. 5, pp. 3–28, 1992.

[17] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley & Sons, 2001.

[18] M. Bloch and J. Barros, *Physical Layer Security.* Cambridge University Press, 2011.

[19] Y. Chen, "Wiretap Channel with Side Information," Ph.D. dissertation, Duisburg-Essen University, 2007.

[20] A. Mukherjee, S. Fakoorian, J. Huang, and A. Swindlehurst, "Principles of Physical Layer Security in Multiuser Wireless Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1550–1573, Apr. 2014.

[21] I. Csiszàr and J. Korner, "Broadcast Channels with Confidential Messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[22] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian Wire-tap Channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[23] J. Barros and M. Rodrigues, "Secrecy capacity of wireless channels," in *Proc. IEEE International Symposium on Information Theory, 2006*, July 2006, pp. 356–360.

[24] M. Bloch, J. Barros, M. Rodrigues, and S. McLaughlin, "Wireless information-theoretic security," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, June 2008.

[25] W. Trappe, "The challenges facing physical layer security," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 16–20, June 2015.

[26] K. Zeng, "Physical Layer Key Generation in Wireless Networks: Challenges and Opportunities," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 33–39, June 2015.

[27] S. Jana, S. N. Premnath, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, "On the Effectiveness of Secret Key Extraction from Wireless Signal Strength in Real Environments," in *Proc. ACM MobiCom'09*, Beijing, China, Sep. 2009.

[28] M. Madiseh, S. He, M. McGuire, S. Neville, and X. Dong, "Verification of Secret Key Generation from UWB Channel Observations," in *Proc. IEEE ICC'09*, Dresden, Germany, Jun. 2009, pp. 1–5.

[29] S. Tmar-Ben Hamida, J.-B. Pierrot, and C. Castelluccia, "Empirical Analysis of UWB Channel Characteristics for Secret Key Generation in Indoor Environments," in *Proc. IEEE PIMRC'10*, Istanbul, Turkey, Sep. 2010.

[30] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-Rate Uncorrelated Bit Extraction for Shared Secret Key Generation from Channel Measurements," *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 17–30, Jan. 2010.

[31] S. Tmar-Ben Hamida, J.-B. Pierrot, and C. Castelluccia, "On the Security of UWB Secret Key Generation Methods against Deterministic Channel Prediction Attacks," in *Proc. IEEE Vehicular Technology Conference (VTC2012-Fall)*, Quebec, Canada, Sep. 2012.

[32] U. Maurer, "Secret Key Agreement by Public Discussion from Common Information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[33] R. Ahlswede and I. Csiszar, "Common Randomness in Information Theory and Cryptography. I. Secret Sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[34] U. Maurer and S. Wolf, "Unconditionally Secure Key Agreement and the Intrinsic Conditional Information," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 499–514, Mar. 1999.

[35] S. Mjlosnes, "Chapter 5. Quantum Cryptography," in *A Multidisciplinary Introduction to Information Security*. CRC Press, 2011. [Online]. Available: http://arxiv.org/pdf/1108.1718.pdf

[36] D. Moskovich, "An Overview of the State of the Art for Practical Quantum Key Distribution," 2015. [Online]. Available: http://arxiv.org/abs/1504.05471

[37] T.-H. Chou, S. Draper, and A. Sayeed, "Key Generation Using External Source Excitation: Capacity, Reliability, and Secrecy Exponent," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2455–2474, Apr. 2012.

[38] P. Huang and X. Wang, "Fast Secret Key Generation in Static Wireless Networks: A Virtual Channel Approach," in *Proc. IEEE INFOCOM 2013*, April 2013, pp. 2292–2300.

[39] 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, Oct. 2008.

[40] S. Xiao, H. Pishro-Nik, and W. Gong, "Dense Parity Check Based Secrecy Sharing in Wireless Communications," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '07)*, Nov 2007, pp. 54–58.

[41] A. Agrawal, Z. Rezki, A. Khisti, and M.-S. Alouini, "Noncoherent Capacity of Secret-Key Agreement With Public Discussion," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 565–574, Sep. 2011.

[42] L. Lai, Y. Liang, and H. Poor, "A Unified Framework for Key Agreement Over Wireless Fading Channels," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 480–490, Dec. 2012.

[43] A. Khisti, "Secret-Key Agreement over Non-Coherent Block Fading Channels with Public Discussion," *Submitted to IEEE Transactions on Information Theory*, 2013.

[44] H. Koorapaty, A. Hassan, and S. Chennakeshu, "Secure Information Transmission for Mobile Radio," *IEEE Communications Letters*, vol. 4, no. 2, pp. 52–55, Feb 2000.

[45] S. Severi, G. Abreu, G. Pasolini, and D. Dardari, "A Secret Key Exchange Scheme for Near Field Communication," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Sep. 2014.

[46] D. Simmons, N. Bhargav, J. Coon, and S. Cotton, "Physical Layer Security Over OFDM-Based Links: Conjugate-and-Return," in *Proc. IEEE Vehicular Technology Conference*, May 2015.

[47] Y. Shen and M. Z. Win, "Intrinsic Information of Wideband Channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, Sep. 2013.

[48] U. Maurer and S. Wolf, "Secret-Key Agreement over Unauthenticated Public Channels-Part I. Definitions and a Completeness Result," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 822–831, Apr. 2003.

[49] ——, “Secret-key Agreement Over Unauthenticated Public Channels-Part II: Privacy Amplification,” *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 839–851, Apr. 2003.

[50] L. Lai, Y. Liang, and H. Poor, “Key Agreement over Wireless Fading Channels with an Active Attacker,” in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2010, pp. 1391–1396.

[51] H. Li, L. Lai, S. Djouadi, and X. Ma, “Key establishment via common state information in networked control systems,” in *Proc. American Control Conference (ACC)*, June 2011, pp. 2234–2239.

[52] M. Forman and D. Young, “A Generalized Scheme for the Creation of Shared Secret Keys through Uncorrelated Reciprocal Channels in Multiple Domains,” in *Proc. 18th International Conference on Computer Communications and Networks (ICCCN)*, Aug. 2009, pp. 1–8.

[53] K. Ren, H. Su, and Q. Wang, “Secret Key Generation Exploiting Channel Characteristics in Wireless Communications,” *IEEE Wireless Communications*, vol. 18, no. 4, pp. 6–12, August 2011.

[54] A. Hassan, W. Stark, J. Hershey, and S. Chennakeshu, “Cryptographic Key Agreement for Mobile Radio,” *Digital Signal Processing*, vol. 6, no. 4, pp. 207–212, Oct. 1996.

[55] Q. Wang, H. Su, K. Ren, and K. Kim, “Fast and Scalable Secret Key Generation Exploiting Channel Phase Randomness in Wireless Networks,” in *Proc. IEEE INFOCOM’11*, Shanghai, China, Apr. 2011, pp. 1422–1430.

[56] A. Sayeed and A. Perrig, “Secure Wireless Communications: Secret Keys Through Multipath,” in *Proc. IEEE ICASSP’08*, Las Vegas, NV, USA, Mar. 2008, pp. 3013–3016.

[57] M. Tope and J. McEachen, “Unconditionally Secure Communications over Fading Channels,” in *Proc. MILCOM’01*, Piscataway, NJ, USA, Oct. 2001, pp. 54–58.

[58] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, “Wireless Secret Key Generation Exploiting Reactance-Domain Scalar Response of Multipath Fading Channels,” *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 11, pp. 3776–3784, Nov. 2005.

[59] B. Azimi-Sadjadi, A. Mercado, A. Kiayias, and B. Yener, “Robust Key Generation from Signal Envelopes in Wireless Networks,” in *Proc. ACM CCS’07*, Alexandria, VA, USA, Oct. 2007.

[60] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-Telepathy: Extracting a Secret Key from an Unauthenticated Wireless Channel," in *Proc. ACM MobiCom'08*, Sep. 2008, pp. 128–139.

[61] P. Barsocchi, S. Chessa, I. Martinovic, and G. Oligeri, "AmbiSec: Securing Smart Spaces Using Entropy Harvesting," in *Proc. AmI'10*, Málaga, Spain, Nov. 2010, pp. 73–85.

[62] S. Ali, V. Sivaraman, and D. Ostry, "Secret Key Generation Rate vs. Reconciliation Cost Using Wireless Channel Characteristics in Body Area Networks," in *Proc. IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing (EUC)*, Dec 2010, pp. 644–650.

[63] Y. Wei, K. Zeng, and P. Mohapatra, "Adaptive Wireless Channel Probing for Shared Key Generation," in *Proc. IEEE INFOCOM'11*, Shanghai, China, Apr. 2011, pp. 2165–2173.

[64] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, "Information-Theoretically Secret Key Generation for Fading Wireless Channels," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 240–254, Jun. 2010.

[65] X. Wu, Y. Song, C. Zhao, and X. You, "Secrecy Extraction from Correlated Fading Channels: An Upper Bound," in *Proc. WCSP'09*, Nanjing, China, Nov. 2009, pp. 1–3.

[66] M. Wilhelm, I. Martinovic, and J. Schmitt, "Key Generation in Wireless Sensor Networks Based on Frequency-selective Channels: Design, Implementation, and Analysis," ArXiv.org, Tech. Rep., May 2010. [Online]. Available: http://arxiv.org/pdf/1005.0712v1

[67] Y. Liu, S. Draper, and A. Sayeed, "Exploiting Channel Diversity in Secret Key Generation From Multipath Fading Randomness," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1484–1497, Oct 2012.

[68] T. Mazloum, F. Mani, and A. Sibille, "Analysis of Secret Key Robustness in Indoor Radio Channel Measurements," in *Proc. IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[69] S. Yasukawa, H. Iwai, and H. Sasaoka, "Adaptive Key Generation in Secret Key Agreement Scheme based on the Channel Characteristics in OFDM," in *Proc. ISITA'08*, Auckland, New Zealand, Dec. 2008, pp. 1–6.

[70] Y. El Hajj Shehadeh, O. Alfandi, and D. Hogrefe, "Towards Robust Key Extraction from Multipath Wireless Channels," *Journal of Communications and Networks*, vol. 14, no. 4, pp. 385–395, Aug. 2012.

[71] K. Zeng, D. Wu, A. Chan, and P. Mohapatra, "Exploiting Multiple-Antenna Diversity for Shared Secret Key Generation in Wireless Networks," in *Proc. IEEE INFOCOM'10*, San Diego, CA, USA, Mar. 2010.

[72] J. Wallace, "Secure Physical Layer Key Generation Schemes: Performance and Information Theoretic Limits," in *Proc. IEEE ICC'09*, Dresden, Germany, Jun. 2009, pp. 1–5.

[73] G. Pasolini and D. Dardari, "Secret key generation in correlated multi-dimensional Gaussian channels," in *Proc. IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 2171–2177.

[74] B. Quist and M. Jensen, "Maximizing the Secret Key Rate for Informed Radios under Different Channel Conditions," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5146–5153, October 2013.

[75] A. Kitaura, T. Sumi, K. Tachibana, H. Iwai, and H. Sasaoka, "A Scheme of Private key Agreement based on Delay Profiles in UWB Systems," in *IEEE Sarnoff Symposium'06*, Princeton, NJ, USA, Mar. 2006, pp. 1–6.

[76] J. Huang and T. Jiang, "Dynamic Secret Key Generation Exploiting Ultra-Wideband Wireless Channel Characteristics," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 1701–1706.

[77] R. Wilson, D. Tse, and R. A. Scholtz, "Channel Identification: Secret Sharing Using Reciprocity in Ultrawideband Channels," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, Sep. 2007.

[78] M. Madiseh, M. McGuire, S. Neville, and A. Shirazi, "Secret Key Extraction in Ultra Wideband Channels for Unsynchronized Radios," in *Proc. CNSR'08*, Halifax, Canada, May 2008, pp. 88–95.

[79] S. Tmar-Ben Hamida, J.-B. Pierrot, and C. Castelluccia, "An Adaptive Quantization Algorithm for Secret Key Generation Using Radio Channel Measurements," in *Proc. NTMS'09*, Cairo, Egypt, Dec. 2009.

[80] F. Marino, E. Paolini, and M. Chiani, "Secret Key Extraction from a UWB Channel: Analysis in a Real Environment," in *Proc. IEEE International Conference on Ultra-Wideband (ICUWB)*, Sep. 2014.

[81] M. Madiseh, S. Neville, and M. McGuire, "Time Correlation Analysis of Secret Key Generation via UWB Channels," in *Proc. IEEE GLOBECOM'10*, Miami, Florida, USA, Dec. 2010, pp. 1–6.

[82] O. Gungor, F. Chen, and C. Koksal, "Secret Key Generation from Mobility," in *Proc. IEEE GLOBECOM Workshops'11*, Houston, TX, USA, Dec. 2011, pp. 874–878.

[83] A. Badawy, T. Khattab, T. El-Fouly, A. Mohamed, D. Trinchero, and C.-F. Chiasserini, "Secret Key Generation Based on AoA Estimation for Low SNR Conditions," in *IEEE Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–7.

[84] R. Mehmood, J. Wallace, and M. Jensen, "Key Establishment Employing Reconfigurable Antennas: Impact of Antenna Complexity," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6300–6310, Nov 2014.

[85] S. Gollakota and D. Katabi, "Physical Layer Wireless Security Made Fast and Channel Independent," in *Proc. IEEE INFOCOM*, April 2011, pp. 1125–1133.

[86] A. Limmanee and W. Henkel, "Secure Physical-Layer Key Generation Protocol and Key Encoding in Wireless Communications," in *Proc. IEEE GLOBECOM Workshop*, Dec 2010, pp. 94–98.

[87] H. Vogt and A. Sezgin, "Secret-key Generation from Wireless Channels: Mind the Reflections," in *Proc. IEEE International Conference on Communications Workshops (ICC)*, Jun. 2014, pp. 783–788.

[88] T. H. T. Nguyen and J.-P. Barbot, "Secret Key Management and Dynamic Security Coding System," in *Proc. IEEE Fifth International Conference on Communications and Electronics (ICCE)*, Jul. 2014, pp. 548–553.

[89] L. Yang and G. Giannakis, "Ultra-Wideband Communications: an Idea Whose Time Has Come," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[90] B. Denis, F. Dehmas, M. Pelissier, and L. Ouvry, "La Technologie UWB Radio Impulsionnelle: un Etat des Lieux des Solutions en Matière de Localisation Haute Précision et de Transfert de Données à Courte Portée," *Revue de l'Electricité et de l'Electronique*, vol. -, no. 5, pp. 62–74, Dec. 2013.

[91] A. Molisch, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. Schantz, K. Siwiak, and M. Win, "A Comprehensive Standardized

Model for Ultrawideband Propagation Channels," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3151–3166, Nov. 2006.

[92] N. Amiot, M. Laaraiedh, and B. Uguen, "PyLayers: An Open Source Dynamic Simulator for Indoor Propagation and Localization," in *Proc. IEEE ICC'13*, Budapest, Hungary, Jun. 2013.

[93] S. Dubouloz, B. Denis, S. de Rivaz, and L. Ouvry, "Performance analysis of ldr uwb non-coherent receivers in multipath environments," in *IEEE International Conference on Ultra-Wideband (ICU 2005)*, Sept 2005.

[94] M. Pezzin and D. Lachartre, "A low Power, Low Data Rate Impulse Radio Ultra Wide Band Transceiver," in *Proc. FUNEMS'10*, Florence, Italy, Jun. 2010.

[95] S. Tmar-Ben Hamida, "Signal-based Security in Wireless Networks," Ph.D. dissertation, Université de Grenoble, 2012.

[96] S. Tmar-Ben Hamida, J. Pierrot, B. Denis, C. Castelluccia, and B. Uguen, "On the Security of UWB Secret Key Generation Methods against Deterministic Channel Prediction Attacks," in *Proc. IEEE VTC Fall'12*, Quebec City, Canada, Sep. 2012.

[97] I. Tunaru, B. Denis, and B. Uguen, "Random Patterns of Secret Keys from Sampled IR-UWB Channel Responses," in *Proc. IEEE ICUWB'14*, Paris, France, Sep. 2014.

[98] ——, "Public Discussion Strategies for Secret Key Generation from Sampled IR-UWB Channel Responses," in *Proc. COMM'14*, Bucharest, Romania, May 2014.

[99] ——, "Reciprocity-Diversity Trade-off in Quantization for Symmetric Key Generation," in *Proc. PIMRC'14*, Washington DC, US, Sep. 2014.

[100] ——, "Cooperative Group Key Generation Using IR-UWB Multipath Channels," in *Proc. IEEE ICUWB'15*, Montreal, Canada, Oct. 2015.

[101] R. Gray and D. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[102] W. Burr, D. Dodson, and W. Polk, "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," Information Technology Laboratory, NIST, Gaithersburg, Maryland, Tech. Rep., 2010.

[103] C. Ye, A. Reznik, and Y. Shah, "Extracting Secrecy from Jointly Gaussian Random Variables," in *Proc. IEEE International Symposium on Information Theory, 2006*, July 2006, pp. 2593–2597.

[104] J. Wallace and R. Sharma, “Automatic Secret Keys From Reciprocal MIMO Wireless Channels: Measurement and Analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 381–392, Sept 2010.

[105] S. Riad, “The deconvolution Problem: An Overview,” *Proc. of the IEEE*, vol. 74, no. 1, pp. 82–85, Jan. 1986.

[106] R.-M. Cramer, R. Scholtz, and M. Win, “Evaluation of an Ultra-Wide-Band Propagation Channel,” *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 5, pp. 561–570, May 2002.

[107] J. Paredes, G. Arce, and Z. Wang, “Ultra-Wideband Compressed Sensing: Channel Estimation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[108] B. Smith, “Instantaneous Companding of Quantized Signals,” *Bell System Technical Journal*, vol. 36, no. 3, pp. 653–709, May 1957.

[109] M. Caramia and P. Dell’Olmo, *Multi-objective Management in Freight Logistics Increasing Capacity, Service Level and Safety with Optimization Algorithms.* Rome, Italy: Springer, 2008.

[110] W. Gifford, W.-L. Li, Y. Zhang, and M. Win, “Effect of Bandwidth on the Number of Multipath Components in Realistic Wireless Indoor Channels,” in *Proc. IEEE International Conference on Communications (ICC’11)*, Kyoto, Japan, Jun. 2011, pp. 1–6.

[111] I. Maravić, M. Vetterli, and K. Ramchandran, “Channel Estimation and Synchronization with Sub-Nyquist Sampling and Application to Ultra-Wideband Systems,” in *Proc. IEEE ISCAS’04*, vol. 5, Vancouver, Canada, May 2004.

[112] M. Win and R. Scholtz, “Characterization of Ultra-Wide Bandwidth Wireless Indoor Channels: a Communication-Theoretic View,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.

[113] V. Lottici, A. D’Andrea, and U. Mengali, “Channel Estimation for Ultra-Wideband Communications,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[114] C. Carbonelli, U. Mengali, and U. Mitra, “Synchronization and Channel estimation for UWB Signal,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM ’03)*, vol. 2, Dec. 2003, pp. 764–768 Vol.2.

[115] Z. Wang and X. Yang, "Ultra Wide-Band Communications with Blind Channel Estimation Based on First-Order Statistics," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 4, May 2004, pp. 529–532.

[116] C. Falsi, D. Dardari, L. Mucchi, and M. Z. Win, "Time of Arrival Estimation for UWB Localizers in Realistic Environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 032082, 2006.

[117] N. Michelusi, U. Mitra, and M. Zorzi, "Hybrid Sparse/Diffuse UWB Channel Estimation," in *Proc. IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications 2011 (SPAWC)*, Jun. 2011, pp. 201–205.

[118] L. Yang and G. B. Giannakis, "Optimal Pilot Waveform Assisted Modulation for Ultra-Wideband Communications," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1236–1249, 2002.

[119] B. Geiger, T. Gigl, and K. Witrisal, "Enhanced-Accuracy Channel Estimation and Ranging for IR-UWB Energy Detectors," in *Proc. IEEE International Conference on Ultra-Wideband 2010 (ICUWB)*, vol. 2, Sep. 2010, pp. 1–6.

[120] M. Unser and J. Zerubia, "A Generalized Sampling Theory without Band-Limiting Constraints," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 959–969, Aug. 1998.

[121] M. Vetterli, P. Marziliano, and T. Blu, "Sampling Signals with Finite Rate of Innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[122] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse Sampling of Signal Innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[123] I. Maravić and M. Vetterli, "Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2788–2805, 2005.

[124] M. Vetterli, "Sampling in the Age of Sparsity," 2009. [Online]. Available: http://lcav.epfl.ch/files/content/sites/lcav/files/Martin%20Vetterli/Talks/Sampling%20in%20the%20age%20of%20sparity.pdf

[125] D. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[126] E. Candes and M. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[127] F. Naini, R. Gribonval, L. Jacques, and P. Vandergheynst, "Compressive Sampling of Pulse Trains: Spread the Spectrum!" in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '09)*, Apr. 2009, pp. 2877–2880.

[128] E. Lagunas and M. Najar, "Sparse Channel Estimation Based on Compressed Sensing for Ultra Wideband Systems," in *Proc. IEEE International Conference on Ultra-Wideband (ICUWB)*, Sep. 2011, pp. 365–369.

[129] T.-K. Liu, X. Dong, and W.-S. Lu, "Compressed Sensing Maximum Likelihood Channel Estimation for Ultra-Wideband Impulse Radio," in *Proc. IEEE International Conference on Communications (ICC '09)*, Jun. 2009, pp. 1–5.

[130] L. Shi, Z. Zhou, L. Tang, H. Yao, and J. Zhang, "Ultra-Wideband Channel Estimation Based on Bayesian Compressive Sensing," in *Proc. International Symposium on Communications and Information Technologies (ISCIT)*, Oct. 2010, pp. 779–782.

[131] M. Ozgor, S. Erkucuk, and H. Cirpan, "Bayesian Compressive Sensing for Ultra-Wideband Channel Models," in *Proc. 35th International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2012, pp. 320–324.

[132] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[133] M. Basaran, S. Erkucuk, and H. Cirpan, "The Effect of Channel Models on Compressed Sensing Based UWB Channel Estimation," in *Proc. IEEE International Conference on Ultra-Wideband (ICUWB)*, Sep. 2011, pp. 375–379.

[134] Q. Zhou, Z. Zou, H. Tenhunen, and L.-R. Zheng, "Architectural Analysis of Compressed Sensing Based IR-UWB Receiver for Communication and Ranging," in *Proc. IEEE International Conference on Ultra-WideBand (ICUWB)*, September 2014, pp. 222–227.

[135] Z. Sahinoglu, S. Gezici, and I. Guvenc, *Ultra-Wideband Positioning Systems: Theoretical Limits, Ranging Algorithms and Protocols.* Cambridge University Press, 2008.

[136] G. Durgin, *Space-Time Wireless Channels.* Prentice Hall Professional, 2002.

[137] M. Wilhelm, I. Martinovic, and J. Schmitt, "On Key Agreement in Wireless Sensor Networks based on Radio Transmission Properties," in *Proc. IEEE Workshop NPSec'09*, Princeton, NJ, USA, Oct. 2009.

[138] M. Madiseh, M. McGuire, S. Neville, L. Cai, and M. Horie, "Secret Key Generation and Agreement in UWB Communication Channels," in *Proc. IEEE GLOBECOM'08*, New Orleans, LO, USA, Dec. 2008, pp. 1–5.

[139] G. Shirazi and L. Lampe, "A Compressive Sensing Approach for Secret Key Agreement based on UWB Channel Reciprocity," in *Proc. IEEE ICUWB'12*, Syracuse, NY, USA, Sep. 2012, pp. 135–139.

[140] D. Karas, G. Karagiannidis, and R. Schober, "Channel Level Crossing-based Security for Communications over Fading Channels," *IET Information Security*, vol. 7, no. 3, Sep. 2013.

[141] I. Csiszar and P. Narayan, "Secrecy Capacities for Multiple Terminals," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3047–3061, Dec 2004.

[142] ——, "Secrecy Generation for Multiaccess Channel Models," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 17–31, Jan. 2013.

[143] C. Chan and L. Zheng, "Multiterminal Secret Key Agreement," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3379–3412, Jun. 2014.

[144] C. Ye and A. Reznik, "Group Secret Key Generation Algorithms," in *IEEE International Symposium on Information Theory (ISIT 2007)*, Jun. 2007, pp. 2596–2600.

[145] 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, Sep. 2012.

[146] Q. Wang, K. Xu, and K. Ren, "Cooperative Secret Key Generation from Phase Estimation in Narrowband Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 9, pp. 1666–1674, Oct. 2012.

[147] Y. Wei, C. Zhu, and J. Ni, "Group Secret Key Generation Algorithm from Wireless Signal Strength," in *Sixth International Conference on Internet Computing for Science and Engineering (ICICSE)*, Apr. 2012, pp. 239–245.

[148] N. Wang, N. Zhang, and T. Gulliver, "Cooperative Key Agreement for Wireless Networking: Key Rates and Practical Protocol Design," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 272–284, Feb. 2014.

[149] L. De Nardis, J. Fiorina, D. Panaitopol, and M.-G. Di Benedetto, "Combining UWB with Time Reversal for Improved Communication and Positioning," *Telecommunication Systems*, vol. 52, no. 2, pp. 1145–1158, 2013.

[150] D. MacKay, "Bayesian Interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, May 1992.

[151] T. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[152] S. Borman, "The Expectation Maximization Algorithm. A Short Tutorial," 2009. [Online]. Available: http://www.cs.cmu.edu/~dgovinda/pdf/recog/EM_algorithm-1.pdf

[153] S. Roweis and Z. Ghahramani, "A Unifying Review of Linear Gaussian Models," *Neural Computation*, vol. 11, no. 2, pp. 305–345, Feb. 1999.

[154] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[155] L. Stillwaggon-Swan and L. Goldberg, "Are Social Media Making Us Stupid? [Opinion]," *IEEE Technology and Society Magazine*, vol. 34, no. 1, pp. 8–9, March 2015.

[156] D. Lupton, "Understanding the Human Machine [Commentary]," *IEEE Technology and Society Magazine*, vol. 32, no. 4, pp. 25–30, December 2013.

[157] J. Robbins, "Ruminations on the IQ2 Debate: We Are Becoming Enslaved by Our Technology? [Opinion]," *IEEE Technology and Society Magazine*, vol. 34, no. 1, pp. 6–7, March 2015.

[158] C. Nold and R. van Kranenburg, "The Internet of People for a Post-Oil World," 2011. [Online]. Available: http://situatedtechnologies.net/files/ST8_InternetOfPeople_web.pdf

[159] D. Geer, "The Right to Be Unobserved," *IEEE Security and Privacy*, vol. 13, no. 4, pp. 12–19, July 2015.

[160] S. Eldridge, "The Circle [Book Review]," *IEEE Technology and Society Magazine*, vol. 34, no. 1, pp. 4–5, March 2015.

[161] D. Anthony, T. Stablein, and E. Carian, "Big Brother in the Information Age: Concerns about Government Information Gathering over Time," *IEEE Security and Privacy*, vol. 13, no. 4, pp. 12–19, July 2015.

[162] J. Bustard, “The Impact of EU Privacy Legislation on Biometric System Deployment: Protecting citizens but constraining applications,” *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 101–108, Sept 2015.

[163] J. Nielsen, “Assessment of Cooperative and Heterogeneous Indoor Localization Algorithms with Real Radio Devices,” in *Proc. IEEE ICC’14, ANLN Workshop*, Sydney, Jun. 2014.

[164] M. Maman, B. Denis, M. Pezzin, B. Piaget, and L. Ouvry, “Synergetic MAC and Higher Layers Functionalities for UWB LDR-LT Wireless Networks,” in *Proc. IEEE ICUWB’08*, Hannover, Sep. 2008.

[165] M. Laaraiedh, S. Avrillon, , and B. Uguen, “Enhancing Positioning Accuracy through Direct Position Estimators Based on Hybrid RSS Data Fusion,” in *Proc. IEEE VTC-Spring’09*, Barcelona, Apr. 2009.

[166] K. Zeng, K. Govindan, and P. Mohapatra, “Non-Cryptographic Authentication and Identification in Wireless Networks,” *IEEE Wireless Communications*, vol. 17, no. 5, pp. 56–62, Oct. 2010.

[167] D. Faria and D. Cheriton, “Detecting Identity-Based Attacks in Wireless Networks Using Signalprints,” in *Proc. ACM WiSe’06*, Los Angeles, Sep. 2006.

[168] N. Patwari and S. K. Kasera, “Robust Location Distinction using Temporal Link Signatures,” in *Proc. ACM MobiCom’07*, Montreal, Sep. 2007.

[169] Y. Zhang, W. Liu, W. Lou, and Y. Fang, “Securing Sensor Networks with Location-Based Keys,” in *Proc. IEEE WCNC’05*, New Orleans, Mar. 2005.

[170] O. Gungor, F. Chen, and C. Koksal, “Secret Key Generation via Localization and Mobility,” *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–16, Jul. 2014.

[171] B. Denis, J.-B. Pierrot, and C. Abou-Rjeily, “Joint Distributed Time Synchronization and Positioning in UWB Ad Hoc Networks Using TOA,” *IEEE Transactions on MTT, Special Issue on Ultra Wideband*, vol. 54, no. 4, pp. 1896–1911, Apr. 2006.

[172] A. Menezes, P. Oorschot, and S. Vanstone, *Handbook of Applied Cryptography.* CRC Press, 1996.

Génération de clés secrétés avec la couche physique dans les réseaux sans fil décentralisés

Une quantité importante de données personnelles ou sensibles seront produites et échangées par de nouvelles applications concernant la surveillance environnementale, la domotique, les villes intelligentes, l'optimisation de la consommation énergétique, les paiements sans fil, etc. Ces applications seront supportées le plus souvent par des réseaux sans fil décentralisés de type réseaux de capteurs (WSN), réseaux mobiles ad hoc (MANET), réseaux véhiculaires (VANET) ou encore réseaux personnels (WPAN). Dans ce sens, de nouvelles technologies et standards sont en train d'être mis en place: WiFi Direct et l'option D2D dans les standards 5G en cours, IEEE 802.11p adapté aux VANET et technologies de communications à courte portée (Near Field Communications), Bluetooth-Low Energy, IEEE 802.15.4 ou Zigbee, IEEE 802.15.4a ou IEEE 802.15.6 à base de radio impulsionnelle ultra large bande (IR-UWB)).

Étant donné la nature de ces données et la vulnérabilité des réseaux de communication sans fil aux différents types d'attaques passives (par exemple, *eavesdropping*) ou actives (par exemple, corruption de l'intégrité des données), les systèmes de communication devraient être protégés par des mesures de sécurité adaptées en termes d'architecture, complexité de calcul, etc. La plupart des solutions cryptographiques assurant la confidentialité et l'intégrité des données reposent sur la cryptographie symétrique qui demande le partage *a priori* d'une clé secrète identique de part et d'autre du lien de communication et son renouvellement périodique. Le problème ainsi intitulé « distribution de clé » devient encore plus difficile dans les réseaux décentralisés, sans infrastructure ou avec une connectivité et une topologie dynamiques (typiquement, en situation de mobilité).

Classiquement, on peut utiliser la cryptographie asymétrique (basée sur un couple clé privée – clé publique) ou des protocoles à base du schéma Diffie-Hellman [Dif76] afin de distribuer une clé symétrique. Au-delà de la complexité calculatoire élevée de ces méthodes, la cryptographie asymétrique nécessite la mise en place d'un système de distribution des certificats et ne s'avère donc pas bien adaptée aux réseaux décentralisés. La distribution de clés à l'intermédiaire d'un centre de distribution de clés ou d'un serveur présente le même désavantage en termes d'architecture. Les solutions décentralisées [Zho08], comme par exemple la distribution des matériels cryptographiques avant déploiement afin de générer des clés symétriques après déploiement, ne peuvent pas assurer la flexibilité nécessaire au renouvellement de la clé symétrique. Les nouvelles recherches concernant la sécurité avec la couche physique proposent des méthodes alternatives pour la distribution des clés secrètes.

La sécurité avec la couche physique

Contrairement aux méthodes classiques de sécurisation qui reposent sur la difficulté calculatoire qu'un attaquant rencontrerait en essayant d'inverser les opérations cryptographiques utilisées par les algorithmes de chiffrage, la sécurité avec la couche physique de transmission radio utilise les avantages existants ou induits dans les canaux de communication sans fil. Des phénomènes comme le bruit ou les évanouissements (*fading*), considérés comme pénalisants pour la fonction de communication, sont utilisés afin de sécuriser les communications. En 1949, Shannon définit la sécurité au sens de la théorie de l'information à travers la notion d'ambiguïté qu'un message crypté induit pour un attaquant qui peut le lire parfaitement [Sha49]. En utilisant une clé sécrète de même longueur que le message et à l'aide d'opérations XOR, une communication peut être rendue parfaitement sécurisée indépendamment de la puissance de calcul détenue par l'attaquant. L'observation selon laquelle un message parvenant à un attaquant dans un canal de communication physique ne peut pas être récupéré sans aucune distorsion (*a fortiori* dans les canaux sans fil) a contribué à la définition du canal *wiretap*, ainsi qu'à l'extension de la définition de sécurité au régime asymptotique vis-à-vis de la longueur du message [Wyn75]. Ainsi, étant donné un encodage aléatoire des messages en émission, un attaquant disposant d'une copie imparfaite de ces messages peut avoir une ambiguïté maximale quand la longueur des messages tend vers infini. En 1993, Maurer et Ahlswede & Csiszàr étudient les capacités de génération d'un secret partagé à partir de deux modèles [Mau93] [Ahl93] : i) un canal *wiretap* auquel on ajoute un canal public authentifié utilisé pour assurer un retour/échange d'information entre les utilisateurs légitimes (modèle canal) ; ii) une source aléatoire observée par tous les utilisateurs et un canal publique (modèle source). Il se trouve que, grâce à la possibilité de traiter séparément les conditions de fiabilité et de confidentialité, le modèle source est plus facilement mis en œuvre.

On connaît ainsi des solutions qui consistent à générer des clés secrètes à partir des seules propriétés du lien de communication sans fil à protéger. Les terminaux des utilisateurs légitimes (Alice et Bob) viennent mesurer certaines métriques radio (par exemple, les réponses impulsionnelles du canal de communication reliant les terminaux ou des séquences de puissance reçue) afin d'extraire une clé secrète commune. Cette solution tire partie de la réciprocité bidirectionnelle entre le lien direct et la voie retour, et de la décorrélation "spatiale" des canaux de communication sans fil. Plus précisément, la réponse impulsionnelle du canal de communication entre Bob et Alice est théoriquement identique, au bruit près, à celle du canal de communication entre Alice et Bob. Alice et Bob peuvent donc séparément élaborer la même clé secrète à partir d'une estimation du canal de communication les reliant, réalisée de part et d'autre du lien. Par ailleurs, dès lors que le terminal de l'attaquant (Eve) est situé à plus de quelques longueurs d'onde de celui de Bob (dans le cas des communications bande-étroite), le canal de communication entre Alice et Eve (respectivement entre Bob et Eve) a des caractéristiques décorrélées de celui entre Alice et Bob (respectivement entre Bob et Alice). Il n'est donc pas aisé pour Eve de générer la même clé secrète en écoutant simplement ses propres canaux vis-à-vis d'Alice ou de Bob. Enfin, le canal de communication entre Alice et Bob est généralement sujet à des variations temporelles de ses caractéristiques, notamment lorsque l'une ou l'autre des parties se déplace. La clé secrète peut ainsi être renouvelée, ou sa longueur peut être augmentée.

Dans ce travail, on a employé le modèle source de partage de clé en l'appliquant aux communications impulsionnelles ultra large bande. Ces dernières permettent en effet de « capter » une quantité d'information mutuelle suffisamment riche de part et d'autre du lien, du fait du pouvoir de résolution multi-trajets conféré par la largeur de bande en réception. Typiquement, les méthodes de génération de clé à partir de la couche physique et du modèle source se décomposent en plusieurs étapes : i) acquisition du canal (mesures, estimations etc.) et extraction de la partie aléatoire ; ii) quantification des valeurs à travers un encodage binaire ; iii) réconciliation des séquences binaires légitimes en corrigeant les possibles différences à l'aide d'échanges sur le canal de communication public ; iv) amplification de *privacy*, utilisée pour diminuer la quantité d'information qu'un attaquant pourrait détenir ou gagner sur la clé finale. En la matière, on s'est particulièrement intéressé aux phases initiales d'estimation et de quantification, ainsi qu'aux discussions publiques à des fins de réconciliation. On a également proposé une nouvelle extension du modèle source à la génération coopérative (entre plusieurs nœuds) de clés de groupe.

Quantification des signaux IR-UWB pour la génération de clé en point-à-point

Le résultat de l'étape de quantification (ou la clé avant correction d'erreur) est évalué en fonction de trois critères de base : la longueur (préférablement grande), le nombre d'erreurs entre les séquences générées de part et d'autre du lien (ou inversement la réciprocité des séquences) et le caractère aléatoire, imprévisible. Ces trois critères constituent les axes d'un compromis caractérisant la génération de clé sur une liaison point-à-point. Le premier chapitre de cette thèse s'intéresse à plusieurs aspects de ce compromis en utilisant différents types de signaux IR-UWB.

Tout d'abord, à partir de travaux existants [Tma12] sur la génération de clé avec des signaux IR-UWB directement échantillonnés (à environ 20 GHz), on propose un nouvel algorithme d'encodage binaire qui favorise le caractère aléatoire des séquences finales tout en observant les impacts négatifs sur la réciprocité. Notre algorithme utilise un procédé d'embrouillement (*scrambling*) en fonction de l'amplitude et du retard du canal, qui « efface » les motifs binaires induits par la forme d'onde de l'impulsion de sondage du canal.

Ensuite, on s'intéresse au compromis entre réciprocité et caractère aléatoire des clés, en fonction de différentes stratégies de quantification et pour des réponses impulsionnelles (CIR) bruitées générées de manière synthétique à partir d'un modèle statistique du canal IR-UWB (IEEE 802.15.4a). A cette fin, on introduit une nouvelle métrique permettant d'évaluer un aspect du caractère aléatoire des clés : la diversité des mots de code binaires générés par l'algorithme de quantification. Le compromis mentionné est alors illustré par le biais d'une étude d'optimisation des seuils de quantification pour un nombre fixe de bits par échantillon, d'une part, et avec la proposition d'une nouvelle méthode de quantification favorisant la diversité des mots de code, d'autre part.

Finalement, l'impact d'estimations réalistes des CIR IR-UWB sur la réciprocité est mis en évidence pour trois estimateurs représentatifs : un estimateur à très haute résolution reposant sur un filtrage adapté à la forme d'onde reçue [Gif11] et deux estimateurs parcimonieux (*Compressed Sensing* [Par07] et *Finite Rate of Innovation* [Mar04]). Afin d'améliorer la réciprocité fortement dégradée par les fausses détections de trajets, on propose également à cette occasion un algorithme d'appariement.

Stratégies discrètes de discussion publique

Le signal radio directement acquis par l'attaquant et/ou les données transitant sur le canal public représentent une fuite d'information disponible pour l'attaquant. Dans les travaux précédents portant sur la génération de clé à partir des signaux expérimentaux IR-UWB, la réconciliation consiste à échanger les indices des échantillons choisis pour la quantification et à corriger les erreurs éventuelles à l'aide d'un code correcteur d'erreur du type Reed-Solomon. Tandis que le code correcteur peut être adapté en fonction du compromis entre les capacités de correction et la fuite d'information, l'échange des indices des échantillons en clair sur le canal public est une spécification fixée du protocole.

Dans le deuxième chapitre de cette thèse, on utilise des signaux IR-UWB obtenus par une méthode de modélisation déterministe du canal (*ray tracing*) pour mettre en évidence l'effet de la corrélation spatiale des signaux et de l'information publique sur la sécurité des clés obtenues avec une méthode existante [Tma12]. Ensuite, on présente deux méthodes plus discrètes de discussion publique : la première limite la quantité d'information divulguée et la deuxième la masque au moyen d'autres métriques réciproques caractérisant le lien légitime, telles que des mesures du temps de vol du signal (RT-ToF).

Génération de clés coopératives

Un des principaux défis pour le modèle source réside dans la collecte de mesures entropiques dans le cas de réseaux ou de liens peu dynamiques dans le temps. Pour pallier cette difficulté et arriver de surcroît à générer des clés partagées par plusieurs nœuds, on suggère d'étendre le modèle source point-à-point à des réseaux maillés de petite dimension en utilisant plusieurs liens physiques et des transmissions coopératives entre les utilisateurs.

L'idée principale du protocole proposé dans le troisième chapitre de cette thèse consiste à générer une clé de groupe à partir de mesures réalisées sur tous les liens d'un réseau maillé : des liens directs -ou adjacents- d'un nœud et des liens non-adjacents perçus par le biais de transmissions coopératives de la part de ses voisins. On aboutit alors à un problème d'égalisation, qui dans le cas de l'IR-UWB, prend la forme d'une déconvolution (ici, temporelle). Dans ce chapitre, on se propose donc de développer et de comparer différentes méthodes de calcul d'un signal porteur du canal non-adjacent pour son destinataire. Dans ce contexte, on s'intéresse à des méthodes d'estimation non-Bayésiennes de type maximisation de vraisemblance, qui s'avèrent assez peu efficaces, et à des méthodes Bayésiennes à base de validation croisée ou d'espérance-maximisation.

Finalement, on compare notre protocole à une solution de distribution utilisant des clés point-à-point générées à partir de la même couche physique IR-UWB. On montre ainsi que notre méthode apporte des gains significatifs en termes de trafic et de longueur de clé, au détriment d'une complexité calculatoire accrue en cas de déconvolution. Néanmoins, l'approche proposée demeure applicable à des cas plus simples d'égalisation reposant sur des mesures de phase ou de coefficients OFDM.

Conclusion

Après un récapitulatif des grands résultats obtenus, de leurs limitations, ainsi que des perspectives de recherche qui en découlent, en lien avec le sujet initial sur la sécurité, la thèse se conclut par quelques réflexions personnelles sur l'impact sociétal des nouvelles technologies de communication sur l'individu et sur les principes de conception qui pourraient être appliqués afin d'éviter certains effets ou dérives non désirables.

Enfin, des travaux techniques connexes et des résultats intermédiaires sont décrits en annexes : i) quantification des informations réciproques de localisation afin de générer des pseudonymes et de fournir une sur-couche supplémentaire d'authentification ; ii) stratégies de quantification s’appliquant à des mesures expérimentales issues des dispositifs IR-UWB intégrés.

Références

[Dif76] W. Diffie and M. Hellman, New directions in cryptography, IEEE Transactions on Information Theory, vol. 22, no. 6, pp. 644-654, Nov. 1976.

[Zho08] Y. Zhou, Y. Fang, and Y. Zhang, Securing Wireless Sensor Networks: a Survey, IEEE Communications Surveys Tutorials, vol. 10, no. 3, pp. 6-28, Mar. 2008.

[Sha49] C. Shannon, Communication theory of secrecy systems, The Bell System Technical Journal, vol. 28, no. 4, pp. 656-715, Oct. 1949.

[Wyn75] A. D. Wyner, The Wire-tap Channel, Bell Systems Technical Journal, vol. 54, no. 8, pp. 1355-1387, Jan. 1975.

[Mau93] U. Maurer, Secret Key Agreement by Public Discussion from Common Information, IEEE Transactions on Information Theory, vol. 39, no. 3, pp. 733-742, May 1993.

[Ahl93] R. Ahlswede and I. Csiszar, Common Randomness in Information Theory and Cryptography. I. Secret Sharing, IEEE Transactions on Information Theory, vol. 39, no. 4, pp. 1121-1132, Jul. 1993.

[Tma12] S. Tmar-Ben Hamida, Signal-based Security in Wireless Networks, Ph.D. Dissertation, Université de Grenoble, 2012.

[Gif11] W. Gifford, W.-L. Li, Y. Zhang, and M. Win, Effect of Bandwidth on the Number of Multipath Components in Realistic Wireless Indoor Channels, in Proc. IEEE International Conference on Communications (ICC'11), Kyoto, Japan, Jun. 2011, pp. 1-6.

[Par07] J. Paredes, G. Arce, and Z. Wang, Ultra-Wideband Compressed Sensing: Channel Estimation, IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 3, pp. 383-395, Oct. 2007.

[Mar04] I. Maravic, M. Vetterli, and K. Ramchandran, Channel Estimation and Synchronization with Sub-Nyquist Sampling and Application to Ultra-Wideband Systems, in Proc. IEEE ISCAS'04, vol. 5, Vancouver, Canada, May 2004.