N° d'ordre : 254

Université de Lille 1

THÈSE

pour obtenir le grade de

Docteur

En

Spécialité : **Automatique, Génie informatique, Traitement du signal et images**

présentée et soutenue publiquement
par

Anh Thu PHAN HO

DOCTORAT DELIVRE PAR L'UNIVERSITÉ DE LILLE 1

Titre:

Théorie de l'information et méthodes statistiques pour l'analyse des systèmes d'authentification utilisant des codes graphiques

Information-theoretic and statistical approaches to the problem of authentication using graphical codes

Soutenue le 18/12/2014 devant le jury d'examen:

Jury

M. Igor Nikiforov, Prof. à l'Université Tech. de Troyes,	Président
M. Jean-Claude Belfiore, Prof. à Telecom ParisTech,	Rapporteur
M. Sviatoslav Voloshynovsjiy, Prof. à l'Université de Genève,	Rapporteur
Mme. Tanya Ignatenko, Dr à l'Université Tech. d'Eindhoven,	Examinateur
M. Yves Delignon, Prof. à Telecom-Lille,	Examinateur
M. Zbigniew Sagan, Ingénieur à Advanced Track and Trace,	Examinateur invité
M. Patrick Bas, Chercheur CNRS,	Directeur de thèse
M. Wadih Sawaya, MCF à Télecom Lille,	Co-Encadrant de thèse

Thèse préparée dans le Laboratoire d'automatique, génie informatique et signal **(LAGIS)**

dans le cadre de l'École Doctorale **SPI 072 (Lille I, Lille III, Artois, ULCO, UVHC, EC Lille)**

PRES Université Lille Nord-de-France

Information-theoretic and statistical approaches to the problem of authentication using graphical codes

Anh Thu PHAN HO

Contents

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my two advisors, Prof. Patrick BAS and Prof. Wadih SAWAYA for their thoughtful guidance, continuous support and encouragement during the years of my Ph.D process. I appreciate Prof. Patrick for many helpful discussions with pragmatic ideas, for teaching me scientific writing, presentation skills and especially for his constant help in every way possible. He even spent his weekends to help me rehearse my presentation. I thank Prof. Wadih for not only being my academic advisor but also my mentor and my best friend. When I was disappointed or felt like losing sight of my path, he was always available to inspire me, motivate me and to point the right direction for me.

My sincere thanks go to all the members of the Doctorate Committee: Prof. Jean-Claude Belfiore, Prof. Sviatoslav Voloshynovsjiy, Prof. Igor Nikiforov, Dr.Tanya Ignatenko, Prof. Yves Delignon and Dr. Zbigniew Sagan for spending their valuable time reading my thesis, attending my dissertation defense, and giving me helpful advices, comments, and suggestions.

I am also grateful to the secretary of the laboratory LAGIS, Ms. Christine Yvoz and the secretary of the Telecom Lille, Ms. Myriam Leprêtre for helping me with administrative procedures.

I could never forget to mention my labmates in Telecom Lille, Le Thu, Bao An, Thong, Noura, Mirabelle, Mehdi, Xia, Rim,.. and friends who made my time in France not only memorable but also enjoyable. I will always remember our long discussions with many different types of topics in Barrois with Tuan ca, Trang, An, Thong, Lam...

Last but not least, I would like to express my heartfelt gratitude to my beloved family: my parents PHAN Van Thong and HO Thi Dao and my sister Nam Uyen, my brother Nguyen Bao for their unconditional support, endless love and for sharing with me moments of disappointment, inspiration and achievement. Also I am deeply grateful to the dentists Mr. Huy-Lan BUI and Mrs. Céline BUI for their emotional support and care, for sharing their experience and teaching me many good things in life. I have lived far away my family for 12 years in many different places but it is their house where I can find the family atmosphere like at my home.

List of Figures

Abstract

In recent years, authentication theory has attracted a lot of attention in different applications. However, the theoretical analysis of authentication for printed graphical codes remains an open issue. In this thesis, the problem of authentication is investigated from an information theoretic security point of view. An authentication model is analyzed using two settings, namely non-channel coding based authentication and channel coding based authentication.

In the former, a reliable performance measurements of an authentication system relying on a Neyman–Pearson hypothesis test is provided. Specifically, an asymptotic expression using Sanov's theorem is first proposed to compute the probabilities of false alarm and non-detection, then a practical method based on MC simulations using importance sampling is given to estimate these very small probabilities. Thanks to these accurate computation of twos error probabilities, it is demonstrated that it is entirely possible to optimize the authentication performance when the model of the print and scan channel is known.

In the latter, the setup in which the authentication message is coded using the deterministic channel codes is studied. It is showed that using channel coding is possible to enhance the authentication performance. More precisely, it is demonstrated that finding codes making the probability of false alarm and non-detection arbitrarily small at the same time is possible. Such codes have rates between the capacity of main channel and the capacity of the opponent channel. It should be noted that the legitimate receiver does not know whether the observed message comes from the legitimate or from the opponent. Therefore it is the objective of the legitimate receiver to use a decoding rule matching with the distribution law of the main channel but mismatching with the opponent channel. Then the probability of non detection is concerned with mismatched decoding. Finally, a practical scheme using parallel concatenated codes with turbo decoding is proposed. The analysis of the EXIT chart is discussed to choose channel parameters so that the authentication performance is optimized.

Chapter 1

Introduction

The idea of authentication appeared centuries ago in Ali Baba and the Forty Thieves, the famous folk tale of the Middle Eastern. In this story, Ali Baba used the phrase “Open, Sesame” as a password to open the door of a magical cave.

Nowadays, with the advancement of information technology, internet and electronic commerce, authentication has become increasingly important. More specifically, there are numerous hackers, viruses and many other malicious adversaries who have posed a major threat to online trading. For example, the cost of global payment card fraud grew by 19% last year to reach $14 billion [4] . Therefore authentication technology is essentially needed in e-commerce to protect business transactions from eavesdroppers who can steal and modify the information in the transactions.

Authentication also plays an important role in restricting the access to a system. In particular, the access control system allows the entrance only to people knowing a password, owning a ID card or having determined physical or behavioral characteristics like faces, fingerprints, voice, irises, etc.

More remarkably, authentication has drawn attentions to counterfeits fighting. Because it is easier for adversaries today to make more accurate counterfeits at low-cost with the support of advent new technologies, they no longer restrict their activities to luxury goods but increasingly invest in physical products such as making illegal copied passports, banknotes, diplomas and drugs, etc. This not only has a negative impact on society and the global economy but also poses a major threat to public health and safety. For example, counterfeit pharmaceutical products have been known to cause serious illness, injury or even death. It is estimated that there are as high as 700,000 deaths globally each year [1]. The problem of authentication of physicals objects consequently is first and foremost a concern for everyone, especially for the manufacturers who produce the genuine products. In the next paragraphs we will give more details about what the authentication is and how it can be used in practice.

Authentication is the act of confirming the truth of an attribute of a single piece of data (datum) or entity [3]. In contrast with identification which refers to the act of stating or indicating a person or thing's identity, authentication is the process of actually verifying the validity of at least one form of identification. Authentication is applied in many different fields.

For example, authentication applying in biometric based on "something we are" which is the measurement of biological or behavioral characteristics such as fingerprints, voice, irises, etc [37]. These biometric characteristics are unique to each person and cannot easily be lost or stolen as physical tokens. However, the biometric system can never be exactly the same for each time extracting. Therefore it is first necessary to capture the biometric data of a user and store it as a template of that biometric. When a user attempts to authenticate, his biometric data is captured and processed again and then compared with his stored template. If his sample is similar enough to the template then he is positively authenticated. Traditionally, the accuracy of a biometric system is measured by two statistics, namely False Rejection Rate and False Acceptance Rate [71]. The former is the probability that an authorized individual is rejected by a biometric system and the latter is probability that an unauthorized individual is accepted by a biometric system. These performance measures can be used in other authentication systems.

Authentication is also applied to digital image security. There are two main techniques which have proposed to authenticate images such as labeling approaches [45] and watermarking approach [32]. In the former approaches, the authentication data are recorded in a separate file while in the latter, the authentication data is embedded into the image. It can be said that for image authentication, the objective is not to protect the image from being copied but to authenticate the image and assure the integrity of it.

Authentication for physical products is employed widely to distinguish genuine products from counterfeit ones. Authentication of physical products is generally done by using the stochastic structure of either the materials that composes the product or of a printed package associated to it. For instance, in [30], the authors use the optical detection based on the random feature of the object combine with digital signatures based on public key codes in order to protect banknotes against counterfeiting. More specifically, the fiber of banknotes is recorded and attached with a digital signature using the public key authentication. It is then encrypted into code image which is printed onto the banknotes and stored as the reference for the verification procedure. It is noted that the technique of digital signature based on asymmetric codes can guarantee that only an authorized person is able to produce the protected banknotes while everyone can verify them. To verify the banknotes, we take the image of the fiber then decode it and compare it to the stored reference. If the difference between them is over a certain threshold, it is declared as a counterfeit. However, such a system is practically heavy to deploy since each product needs to be linked to its high definition capture stored in a database.

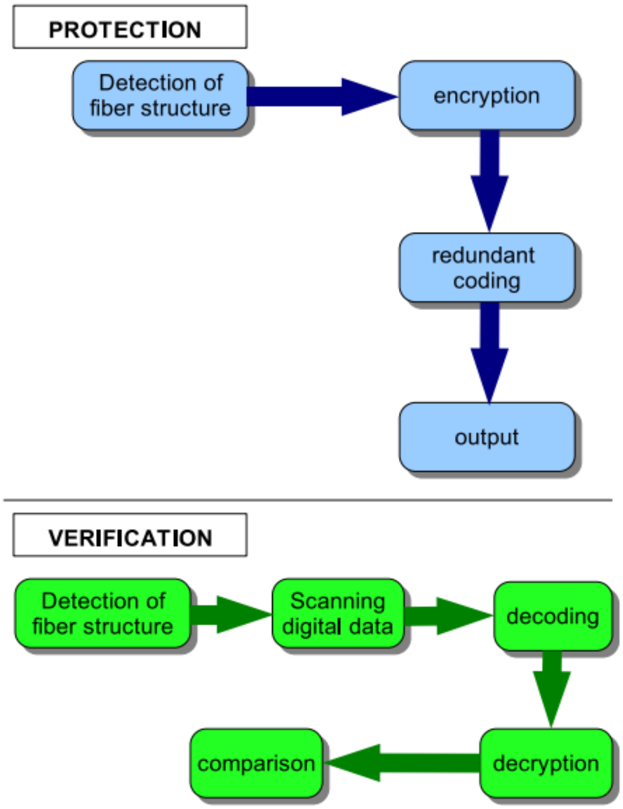


Figure 1.1: Protection and verification of banknote [30].

Another solution to do authentication is to rely on the degradation induced by the interaction between the product and a physical process such as printing, marking, embossing, carving ... Because of both the defaults of the physical process and the stochastic nature of the matter, this interaction can be considered as a Physically Unclonable Function (PUF) [65] that cannot be reproduced by the forger and can consequently be used to perform authentication. The PUF was first studied by Pappu [52] as a function mapping a set of challenges to a set of responses by an intractably complex physical system. It is noted that a PUF is similar to a one way function but it not really a function as most PUFs are noisy i.e. one challenge may result in several responses when it is embodied in different times. Properties of PUF is (a) easy to make but (b) hard to make a copy even using the same physical device and (c) unique for each physical object. Taking advantages of these properties, the authors in [65] use integrated circuits (ICs) as authentication devices to protect confidential information by exploiting PUF design. More specifically, the authors use PUF to generate directly a unique secret key from physical characteristics of ICs to achieve a low-cost authentication without using cryptographic.

We study in this dissertation the authentication for physical products by using graphical codes (GC) inserted on the package of products. Our authentication system is based on the fact that a printing process at very high resolution can be seen as a stochastic process due to the nature of different elements such as the paper fibers, the ink heterogeneity, or the dot addressability of the printer. The randomness of a paper is showed in Fig. 1.4. It turns out that the surface of a sheet of paper is not perfectly flat but is quite rough. In fact, it is like a surface of wood fibers which is highly random and difficult to reproduce. This authentication system has been proposed by Picard et al. [53] [54] and uses 2D pseudo random binary codes that are printed at the native resolution of the printer (2400 dpi on a standard offset printer or 812 dpi on digital HP Indigo printer). It can be said that the authentication based printing process shares the similarity with a PUF in the irreversibility property i.e. the opponent is not able to reproduce the original codes. The major difference is that this system is easier to deploy since the authentication process needs only a scan of the graphical code under scrutiny and the seed used to generate the original one and there is no need to store pairs of challenges and responses as for the PUF regime.

Fig. 1.2 shows an ink dot viewed under microscope.

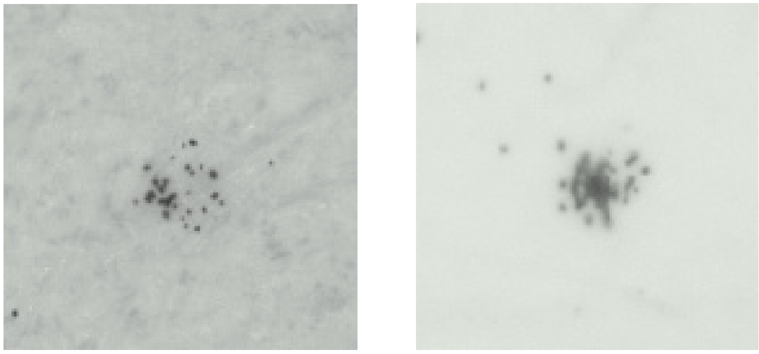


Figure 1.2: Left: Ink dot in uncoated paper printed in Laser printer (600dpi). Right: Ink dot in coated paper printed in Laser printer (600dpi).

The principle of the studied system can be depicted in Fig. 1.3.

- A secret message can either be mapped directly into a binary graphical code (GC) (step 1a) or being encoded, with some probably stochastic function, before this mapping (step 1b). In both cases, the resulting graphical code is printed on packages, thus we investigate in this dissertation two settings, namely non-channel coding based authentication and channel coding based authentication.
- Once printed on a package to be authenticated, the degraded code (greyscale) will be scanned (step 3) then processed (step 4) by an opponent (the forger). It should be mentioned that at this stage the processing is necessary because the industrial printers can only print dots, e.g. binary versions of the scanned code.
- The opponent produces a printed copy of the original code to manufacture his forgery (step 5).

- The receiver compares the scanned version (step 8a) (and potentially post-processed version (step 8b)) of the original code with the scanned version (step 7a) (and potentially post-processed version (step 7b)) of the copied code in order to perform authentication.

1.1 Security measure in authentication

Similar to secure transmission, there are two different ways to measure the security of authentication, namely computational security and information-theoretic security.

Systems relying on computational security are based on two assumptions:

- there are certain mathematical problems which are too difficult to solve,
- the opponent has limited computational power.

On the other hand, information-theoretic security does not depend on any assumption. When the system is information-theoretic secure, it disregards the computational power of the opponent.

In this dissertation, the problem of authentication using graphical codes is treated from the information theoretic security point of view. We consider a scenario where opponent can know everything except for the secret key or message and has unlimited computational power. He has to suffer a degraded channel with respect to the main channel. This degradedness is a stringent constraint to perform authentication. We measure essentially the security by two quantities: the probability of false alarm P_{FA} and the probability of non-detection P_{ND}. However, it should be noted that these two quantities do not guarantee perfect secrecy even they could be made very small. Perfect secrecy is defined in Shannon's paper [59]. More precisely, a system is said to achieve perfect secrecy if the mutual information

$$I\left(M;Y\right)=0,$$

where M represents the authentication message and Y represents what the opponent observes. However, the probability of false alarm and non detection are made arbitrarily small but they can hardly equal zero.

1.2 Estampille

The work presented in this dissertation is part of a project called Estampille. Estampille is supported by research national agency ANR under project number ANR-10-CORD-0019.

The goal of this project is to fight against forged printed documents and counterfeited goods. To this end, the project proposes to insert Graphical Codes (GC) as a

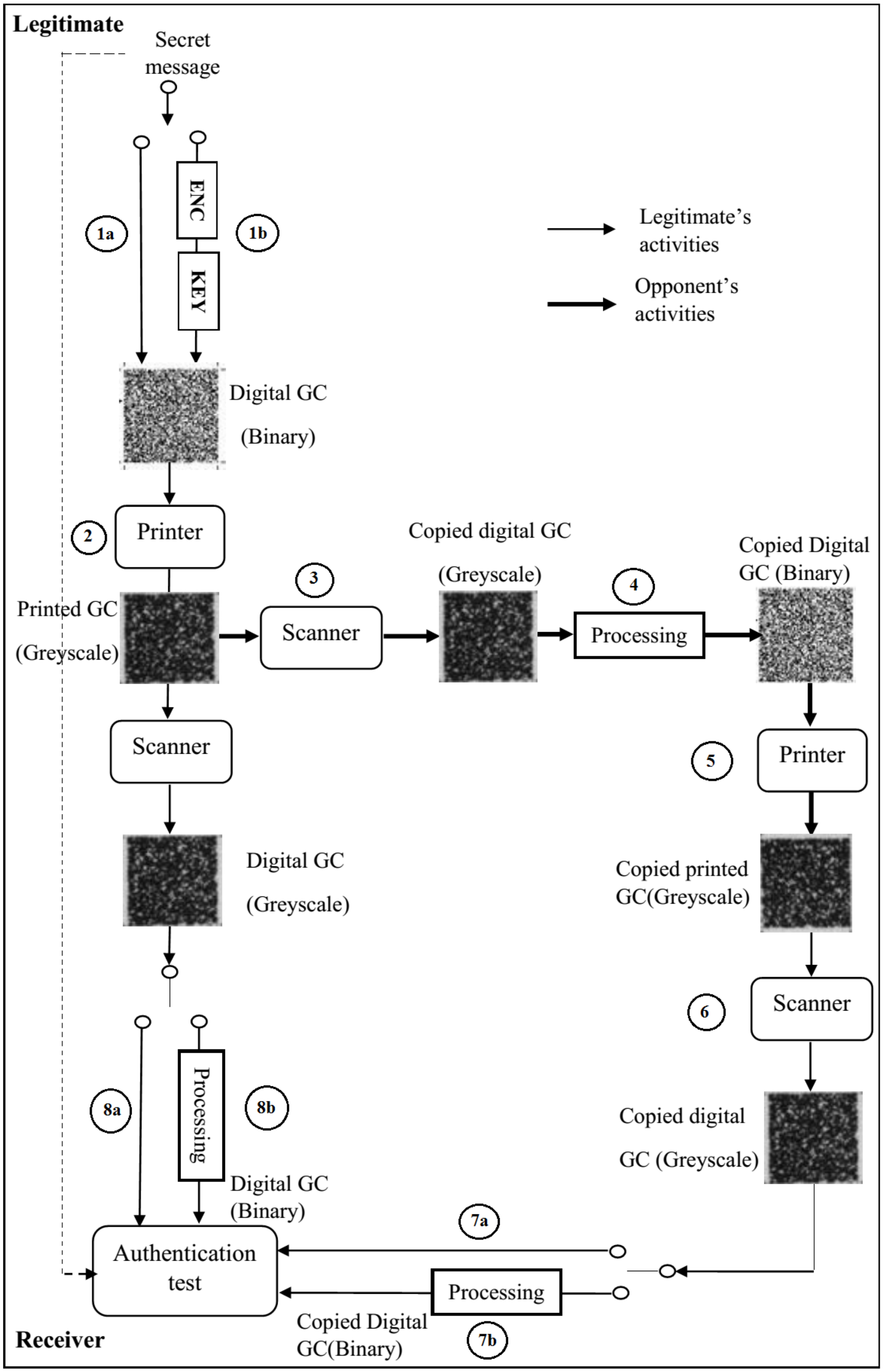


Figure 1.3: Principle of authentication using graphical codes (GC).

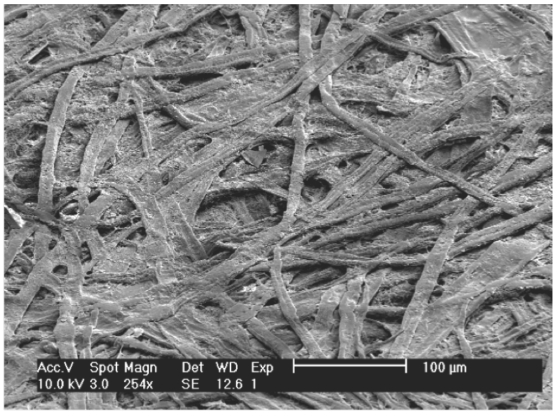


Figure 1.4: An ordinary piece of paper viewed under a microscope [15].

physical identifier on the document or the package of the good because GC have some the following advantages:

- Using GC enables to include integrity check of the printed document. Integrity check is possible by embedding a robust hash inside the graphical code which is not altered under some modifications such as small rotations, compression, scaling etc, and is sensitive to illegal operation like tampering.
- Using GC enables to perform authentication since a counterfeit of the original print will undergo a "scan and print" process that will yield to an additional noise. This noise will be evaluated and detected thanks to the analysis of the GC.
- GC has a wide range of applications. It can be used by custom services, by brand protection departments, on the assembly line, security authorities, etc.
- Using GC is easy to deploy on potentially several thousands of products.

An simulated GC is shown in Figure 1.5.

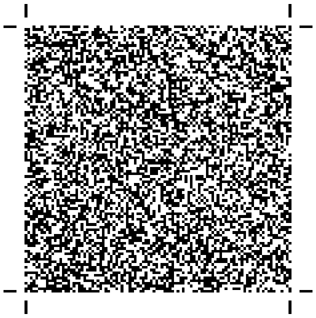


Figure 1.5: An simulated Graphical Codes in computer

The Estampille project has to address the problem of:

- (1) studying the physical process involved such as the stochastic behavior of the printing process, the roles of different parameters i.e. the type of printer, the resolution, the ink and the paper when a GC is printed,
- (2) using information theory approach for authentication,
- (3) building robust hashes to enable integrity check,
- (4) using technical solutions to bring forensics that can be used by a law court.

The Estampille project was formed as a joint project between six partners, i.e. the industrial company ATT (Advance Track and Trace), LAGIS (Laboratoire d'Automatique, Génie Informatique et Signal), GIPSA (Grenoble Images Parole Signal Automatique), LGP2 (Laboratoire Génie des Procédés Papetiers), CERDI (Centre d'Etudes et de Recherche en Droit de l'Immatériel) and the industrial company LATA.

- ATT is an industrial company working on authentication, protection against counterfeit and products tracking. In Estampille, ATT provides technical expertise in 2D graphical code authentication.
- LAGIS is a scientific laboratory in Ecole Centrale de Lille working on automatic systems, computer engineering and signal processing. LAGIS provides the expertise about authentication and stochastic modeling of printing processes.
- GIPSA is a joint laboratory between CNRS and university of Grenoble working on theoretical and applied research on signals and systems. GIPSA provides expertise about security analysis and integrity control.
- LGP2 is a laboratory in university of Grenoble working on intelligent processes, materials chemistry, solid mechanics, mechanics of materials and printing processes. LGP2 provides expertise about description and analysis of printing processes at the microscopic level.
- LATA is an industrial company working on printing technologies. LATA provides expertise and data from various printing processes.
- CERDI belongs to university Paris 11, and works on the juridical aspects of the information technologies. CERDI provides legal basis for the use of graphical code.

1.3 Dissertation structure

This dissertation aims at answering the second direction of the Estampille project. We approach an authentication system on a 2D GC from information theoretic point of view. The authentication model is studied using two settings, namely non-channel coding based authentication and channel coding based authentication. The probabilities of false alarm and non-detection which measures the security of the authentication system are analyzed in flavor of information theory in those two settings.

This dissertation is organized as follows.

- In Chapter 2, we introduces fundamental concepts of information-theoretic security, then we give a brief overview of several prior works related to this dissertation and we set the notation used in subsequent chapters.

- In Chapter 3, we consider a basic authentication model in which an authentication message is uncoded and shared secretly with the legitimate receiver via a main channel while it can be forged by the adversary via his own channel. The authentication performances are thus directly impacted by the discrimination between the two channels. Specifically, we use the Neyman–Pearson hypothesis test to perform the authentication and then to compute the probability of rejecting an authentic message and the probability of non-detecting an illegal copy. This is computed by using either Gaussian approximation or arguments relying on the Sanov's theorem.

- In chapter 4, we present the practical results for these two types of error probabilities by employing Monte-Carlo simulations. Regarding to the results of Monte-Carlo simulations, it is revealed that asymptotic expression relying on Sanov's theorem is accurate while Gaussian approximation is poor for small values of two types of error. More remarkably, importance sampling methods are studied and employed to practically estimate very small values of the non detection probability suggesting an optimized tilted distribution as a proposal. Moreover, by considering the expressions of the two types of error probabilities, we propose an optimization authentication performance in the case of using generalized Gaussian distribution as a model of the print and scan channel.

- Chapter 5 treats the authentication problem by using the channel deterministic codes. Without knowing whether the observed message comes from the legitimate or from the opponent, the legitimate receiver uses a decoding rule matching with the distribution law of the main channel and mismatching with the opponent channel. We establish the existence of the code with a rate between the capacity of main channel and the capacity of the opponent channel, which achieves arbitrarily small probabilities of false alarm and non-detection at the same time. We also propose a practical coding scheme using parallel concatenated codes with turbo decoding. The EXIT chart is analyzed to choose the channels' parameters so that the authentication performance is optimized.

- Finally, Chapter 6 summarizes our conclusions and outline some future research.

1.4 Related activities

During the time doing my Ph.D., I have conducted the redaction of two conference papers and one journal paper. I have also presented by work in two national workshops.

1. Poster: "Authentication using graphical codes" at workshop Journée Futur & Ruptures, Institut Telecom Paris, 2012.

2. Anh Thu Phan Ho, Bao An Mai Hoang, Wadih Sawaya, and Patrick Bas. Document authentication using graphical codes: impacts of the channel model. In Proceedings of the first ACM workshop on Information hiding and multimedia security, pages 8794. ACM, 2013.

3. Talk "Document authentication using graphical codes: Reliable performance analysis and channel optimization." at workshop of GdR ISIS, November 2013.

4. Anh Thu Phan Ho, Bao An Mai Hoang, Wadih Sawaya, and Patrick Bas. Document authentication using graphical codes: Reliable performance analysis and channel optimization. EURASIP Journal on Information Security, 2014.

5. Anh Thu Phan Ho, Bao An Mai Hoang, Wadih Sawaya, and Patrick Bas. Authentication using graphical codes:optimisation of the print and scan channels. In Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European. IEEE, 2014.

Chapter 2

Theoretical elements and Related works

In this chapter, we first set the notations which will be used throughout this dissertation. In the next section, we presents fundamentals in information theory needed to study our authentication system. More specifically, we recall first the basic measures of information, then we introduce the method of types, a powerful tool in information theory and statistics which plays an important role in many proofs and later developments. Next, we present the important results on channel coding and mismatched decoding. In the final section, we highlight several works on authentication and biometrics that are closely related to this dissertation.

2.1 Notations

In the rest of this dissertation, we designate sets by calligraphic font e.g. $\mathcal{X}$ and random variables (RV) ranging over these sets by the same italic capitals e.g. X. The cardinality of the set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The sequence of n variables $(X_1, X_2, ..., X_n)$ is denoted X^n or bold capital $\mathbf{X}$. We use x^n and $\mathbf{x}$ interchangeably to denote a sequence $(x_1, x_2, ..., x_n) \in \mathcal{X}^n$. $X \sim P_X(x)$ indicates that X is governed by the distribution law $P_X(x)$. The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$.

2.2 Fundamentals in Information Theory and Statistics

2.2.1 Basic definitions of information measures

First we recall some basic definitions on the information measures of random variables which will be used in the subsequent chapters. These definitions are restricted in the sense of discrete random variables i.e. random variables distributed over a finite alphabet. For more general discussion we refer the readers to the book by Gallager [28], Cover and Thomas [19], Csisar and Korner [22].

We describe a random variable X over a discrete alphabet $\mathcal{X}$ by its probability mass function $P_X(x) = \Pr\{X = x\}$, which is the probability that X takes on value $x \in \mathcal{X}$.

Definition 2.1. (Shannon Entropy). The Shannon entropy of a discrete random variable X taking values in a finite alphabet $\mathcal{X}$ with a probability mass function $P_X(x)$ is defined by

$$H(X) = -\sum_{x\in\mathcal{X}} P_X(x) \log P_X(x). \tag{2.1}$$

We use the convention that $0 \log 0 = 0$. The logarithms in 2.1 can be taken to any base but the most common base are 2 and e. For the base 2 logarithm, the Shannon entropy is measured in bits and for the base e logarithm, it is measured in nats. In this dissertation, the logarithms are taken to base e. Intuitively, the Shannon entropy $H(X)$ is a quantity measuring the uncertainty of the random variable X. In the rest of this dissertation, we call Shannon entropy just as entropy.

Similarly, we can make an extension of the definition to the pair of random variables (X, Y) as follows

Definition 2.2. (Joint Entropy). The joint entropy of a pair of discrete random variables (X, Y) taking values in finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ with a joint probability mass function $P_{XY}(x, y)$ is defined as

$$H(X, Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x, y) \log P_{XY}(x, y). \tag{2.2}$$

We have known that the entropy of a single random variable X measures the uncertainty about X. One might ask, how much the uncertainty of X is if we are fortunate to have an additional information about Y. Intuitively, this uncertainty will be reduced if Y reveals some information about X. Now we will introduce the conditional entropy as follows.

Definition 2.3. (Conditional Entropy). The conditional entropy of a pair of discrete random variables (X, Y) taking values in a finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ with a joint probability mass function $P_{XY}(x, y)$ and a conditional probability mass function $P_{X|Y}(x \mid y)$ is defined as

$$\begin{aligned} H(X \mid Y) &= \sum_{y\in\mathcal{Y}} P_Y(y) H(X \mid Y = y), \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x, y) \log P_{X|Y}(x \mid y). \end{aligned} \tag{2.3}$$

where $H(X \mid Y = y)$ is the entropy of X conditioned to the knowledge of the outcome $Y = y$.

The conditional entropy $H(X \mid Y)$ measures then the remaining uncertainty on X conditioned to the averaged over all observable outcomes in $\mathcal{Y}$. In a confidential communication system, if we suppose that X is a secret message and Y is a public message, the conditional entropy $H(X \mid Y)$ is then a relevant measure of the secrecy of the system. It is called the equivocation with respect to an eavesdropper (the opponent) who observes Y because it measures the average ambiguity of the observed signal. The system will achieve perfect secrecy if this equivocation is maximum, i.e. $H(X \mid Y) = H(X)$. It implies, in other words, that the public message Y and the secret message X are statistically independent.

We can also ask how much one random variable contains information about another. Here we turn to the mutual information, one important concept in information theory, leading also to higher skills in the state of the art of coding in communication.

Definition 2.4. (Mutual information). The mutual information between two discrete random variables X and Y is defined as

$$I(X;Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)\, P_Y(y)}. \tag{2.4}$$

The mutual information $I(X;Y)$ is a symmetric function and is a measure of the amount of information that X contains about Y and vice versa. In a communication system, mutual information can be understood as the average information flow through a channel. The mutual information $I(X;Y)$ can be expressed as the difference between $H(X)$ and $H(X \mid Y)$:

$$I(X;Y) = H(X) - H(X \mid Y). \tag{2.5}$$

As commented earlier, the conditional entropy is the remaining uncertainty on X given the knowledge of Y, and consequently mutual information is the reduction in the uncertainty of X due to the knowledge of Y. In a confidential communication system, where again X is a secret message and Y a public one, $I(X;Y)$ is referred as an information about the secret leaking to an eavesdropper. The perfect secrecy is the achieved when the information leakage $I(X;Y) = 0$, i.e. $H(X \mid Y) = H(X)$.

Mutual information is a special case of a more general quantity called relative entropy, which is a measure of the distance between two probability distributions.

Definition 2.5. (Relative Entropy). The relative entropy or Kullback Leibler divergence between two probability mass functions $P_X(x)$ and $Q_X(x)$ is defined as

$$D(P \parallel Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

In the above definition, we use the convention that $0 \log \frac{0}{Q} = 0$ and $P \log \frac{P}{0} = \infty$. The relative entropy $D\left(P \parallel Q\right)$ is always non-negative and is zero if and only if $P = Q$. It is important to note that the Kullback Leibler is not a true metric. In particular, it is not symmetric and the triangle inequality does not hold. The Kullback Leibler is also called as the discrimination between two distributions as in Blahut's book [13].

Definition 2.6. (Conditional Relative Entropy). The conditional relative entropy or conditional information divergence [22] $D\left(V \parallel W \mid P\right)$ is the average of the relative entropy between the rows of stochastic matrices V and W given an input of distribution P. More precisely,

$$\begin{aligned} D\left(V \parallel W \mid P\right) &= \sum_x P\left(x\right) D\left(V\left(. \mid x\right) \parallel W\left(. \mid x\right) \mid P\left(x\right)\right), \\ &= \sum_x P\left(x\right) \sum_y V\left(y \mid x\right) \log \frac{V\left(y \mid x\right)}{W\left(y \mid x\right)}. \end{aligned}$$

Now we will introduce a series of useful properties of information-theoretic quantities which will be used in the next sections. The proofs of these properties are easily found in the book of Thomas and Cover [19], Gallager [28], and Csiszar and Korner [22].

1. $H\left(X\right) \geq 0$.
2. (Conditioning reduces entropy) $H\left(X \mid Y\right) \leq H\left(X\right)$.
3. $I\left(X; Y\right) \geq 0$.
4. $I\left(X; Y\right) = H\left(X\right) - H\left(X \mid Y\right)$.
5. (Data processing inequality.) If $X \rightarrow Y \rightarrow Z$ is a Markov chain, i.e. $P\left(x, y, z\right) = P\left(x\right) P\left(y \mid x\right) P\left(z \mid y\right)$ then $I\left(X; Y\right) \geq I\left(X; Z\right)$.
6. $D\left(P \parallel Q\right)$ is convex in the pair $\left(P, Q\right)$, i.e. if $\left(P_1, Q_1\right)$ and $\left(P_2, Q_2\right)$ are two pairs of probability mass function, then for all $0 \leq \lambda \leq 1$:

$$D\left(\lambda P_1 + \left(1-\lambda\right) P_2 \parallel \lambda Q + \left(1-\lambda\right) Q_2\right) \leq \lambda D\left(P_1 \parallel Q_1\right) + \left(1-\lambda\right) D\left(P_2 \parallel Q_2\right). \quad (2.6)$$

2.2.2 The method of types and Sanov's theorem

In this subsection we give an overview of the method of types which is a powerful tool helping to extract precisely the exponential decay of the probability of rare events, and eases proofs related to channel coding. In our context, this method will be useful to compute accurately the probability of rejecting an authentic code and the probability of non-detecting an illegal copy in either channel coding based authentication system or authentication without channel coding. .

Throughout this dissertation, $\mathcal{P}(\mathcal{X})$ denotes the space of all probability measures on the alphabet $\mathcal{X}$. Here $\mathcal{P}(\mathcal{X})$ is identified with the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$, i.e., the set of all $|\mathcal{X}|$-dimensional real vectors with non-negative components that sum to 1. Therefore, the topology on $\mathcal{P}(\mathcal{X})$ is inherited as the subspace topology from the ordinary topology on $\mathbb{R}^{|\mathcal{X}|}$.

Definition 2.7. (Type.) The type $\hat{P}_{x^n}$ of a sequence $x^n = (x_1, ..., x_n) \in \mathcal{X}^n$ is its empirical distribution. More specifically, $\hat{P}_{x^n} = \left(\hat{P}_{x^n}(a_1), ..., \hat{P}_{x^n}\left(a_{|\mathcal{X}|}\right)\right)$ is an element of the set $\mathcal{P}(\mathcal{X})$ and

$$\hat{P}_{x^n}(a) = \frac{n(a \mid x^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i=a\}} \text{ for all } a \in \mathcal{X}.$$

where $n(a \mid x^n)$ is the number of times the symbol a occurs in the sequence $x^n \in \mathcal{X}^n$.

The set of all possible types on $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. It is therefore obvious that $\mathcal{P}_n(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$.

Lemma 2.8. *[19] (Type counting.)*

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}. \tag{2.7}$$

Proof. Note that there are $|\mathcal{X}|$ components in the vector that specifies a type $\hat{P}_{x^n}$. Moreover, every component of the type $\hat{P}_{x^n}$ takes values in the set $\left\{\frac{0}{n}, \frac{1}{n}, ..., \frac{n}{n}\right\}$ whose cardinality is $n+1$. Therefore, there are at most $(n+1)^{|\mathcal{X}|}$ choices for a type. □

Definition 2.9. For a given empirical distribution $\hat{P} \in \mathcal{P}_n(\mathcal{X})$, the type class $T_{X^n}(\hat{P})$ is the set of all sequences x^n in $\mathcal{X}^n$ having type $\hat{P}$:

$$T_{X^n}\left(\hat{P}\right) = \left\{x^n \in \mathcal{X}^n : \hat{P}_{x^n} = \hat{P}\right\}. \tag{2.8}$$

Note that a type class consists of all permutations of a given vector in this set.

Similarly for type of one sequence, we have the definitions of joint type and conditional type as follows.

Definition 2.10. (Joint type.) The joint type of two sequences $x^n = (x_1, ..., x_n) \in \mathcal{X}^n$ and $y^n = (y_1, ..., y_n) \in \mathcal{Y}^n$ is defined as

$$\hat{P}_{x^n y^n}(a,b) = \frac{n(a,b \mid x^n, y^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i=a, y_i=b\}} \text{ for all } (a,b) \in \mathcal{X} \times \mathcal{Y}. \tag{2.9}$$

where $n\left(a, b \mid x^n, y^n\right)$ is the number of times each pair (a, b) occurs in the pair of sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$.

The set of all possible types on $\mathcal{X}^n \times \mathcal{Y}^n$ is denoted by $\mathcal{P}_n\left(\mathcal{X} \times \mathcal{Y}\right)$. For a given $\hat{P} \in \mathcal{P}_n\left(\mathcal{X} \times \mathcal{Y}\right)$, the type class $T_{X^nY^n}(\hat{P})$ is the set of all sequences in $\mathcal{X}^n \times \mathcal{Y}^n$ having type $\hat{P}$:

$$T_{X^nY^n}\left(\hat{P}\right) = \left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \hat{P}_{x^ny^n} = \hat{P}\right\}. \tag{2.10}$$

Definition 2.11. (Conditional type.) The conditional type of y^n given x^n is a stochastic matrix whose elements are defined as

$$\hat{V}_{y^n|x^n}\left(b|a\right) = \frac{\hat{P}_{x^ny^n}\left(a, b\right)}{\hat{P}_{x^n}\left(a\right)} \text{ for all } (a, b) \in \mathcal{X} \times \mathcal{Y}. \tag{2.11}$$

Remark 2.12. Note that every $a \in \mathcal{X}$ must occur in the sequence x^n to ensure $\hat{P}_{x^n}\left(a\right) \neq 0$ for every a.

Given a sequence x^n and a stochastic matrix $V : \mathcal{X} \to \mathcal{Y}$, we denote by $T_{Y^n|x^n}(V)$ the conditional type class, i.e. the set of all sequences y^n in $\mathcal{Y}^n$ such that y^n has conditional type $\hat{V}$ given x^n,

$$T_{Y^n|x^n}\left(V\right) = \left\{y^n \in \mathcal{Y}^n : \hat{V}_{y^n|x^n} = V\right\}. \tag{2.12}$$

Given x^n, we denote $\mathcal{P}_n\left(\mathcal{Y} \mid \mathcal{X} \mid x^n\right)$ the set of all possible conditional type $\hat{V}$ such that for some $y^n \in \mathcal{Y}^n$, we have $\hat{P}_{x^ny^n}\left(a, b\right) = \hat{P}_{x^n}\left(a\right)\hat{V}_{y^n|x^n}\left(b|a\right)$.

A series of properties related to types are of great interest, but for the sake of concision we will not develop all of them. However, to give the reader some flavor of the power of this method, let us consider a multinomial distribution where X_1, X_2, ..., X_n are drawn i.i.d. according to distribution $Q\left(x\right)$. The probability $Q^n(x^n)$ of a sequence x^n is known exactly but its numerical computation may be subject to underflow on computers. One can naturally resolve this inconvenience by taking the logarithm or equivalently using the type of the sequence $\hat{P}_{x^n}$ and compute $Q^n\left(x^n\right) = e^{-n\left(H\left(\hat{P}_{x^n}\right)+D\left(\hat{P}_{x^n}\|Q\right)\right)}$ (see Theorem 11.1.2, [19]). On the other hand, computing the cardinality of a given type $\hat{P}$ is a simple combinatorial problem but numerical computation for large n is time consuming or subject to severe overflow. The method of types give an estimate to this size with simple exponential upper and lower bounds, and for large n we may write that $\frac{1}{n}\log\left|T_{X^n}\left(\hat{P}\right)\right| \to H(\hat{P})$ (see Theorem 11.1.3, [19]).

For more general discussion about the method of types, we refer the readers to the book of Csiszar and Körner [22], Thomas and Cover [19] and Dembo and Zeitouni [24].

Now since the sequence $X_1, ..., X_n$ is a random vector we denote $\hat{P}_{X^n}$ the random element associated to it, and taking values in the set $\mathcal{P}_n\left(\mathcal{X}\right)$:

$$\hat{P}_{X^n}(a) = \frac{n(a \mid X^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i=a\}} \text{ for all } a \in \mathcal{X}. \tag{2.13}$$

Consequently, given a distribution $P \in \mathcal{P}(\mathcal{X})$, we define the so-called random relative entropy $\mathbb{D}\left(\hat{P}_{X^n} \parallel P\right)$ which is a random element taking values in the following set

$$\left\{D\left(\hat{P}_{x^n} \parallel P\right) : \hat{P}_{x^n} \in \mathcal{P}_n(\mathcal{X})\right\}$$

Similarly, we define $\hat{P}_{X^nY^n}$ as the random element associated to the random vectors $X^n = (X_1, ..., X_n)$ and $Y^n = (Y_1, ..., Y_n)$ and taking values in the set $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$:

$$\hat{P}_{X^nY^n}(a,b) = \frac{n(a,b \mid X^n, Y^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i=a,Y_i=b\}} \text{ for all } (a,b) \in \mathcal{X} \times \mathcal{Y}. \tag{2.14}$$

And we finally define $\hat{V}_{Y^n|x^n}$ as the random element associated to the random vector $Y^n = (Y_1, ..., Y_n)$ given the sequence x^n and taking values stochastic matrices in the set $\mathcal{P}_n(\mathcal{Y} \mid \mathcal{X} \mid x^n)$:

$$\hat{V}_{Y^n|x^n}(b|a) = \frac{\hat{P}_{x^nY^n}(a,b)}{\hat{P}_{x^n}(a)} \text{ for all } (a,b) \in \mathcal{X} \times \mathcal{Y}. \tag{2.15}$$

In many applications related to Statistics like the hypothesis testing problem, the statistics (the functions) "will depend on the number of times each value $a \in \mathcal{X}$ appears in the observed sequence rather on this particular sequence" [13]. A decision region relative to some given constraints can then be defined as the set of types verifying this constraint. In addition, in the channel coding problem it is worth defining a typical set which will concentrate the most probable events in order to analyze the performance of an ensemble of codes. For these reasons, we will focus now on expressing the probability that a random empirical distribution as being defined in (2.13 to 2.15) belongs to a given region or a subset in $\mathcal{P}(\mathcal{X})$.

Let $X_1, X_2, ..., X_n$ be drawn i.i.d. according to $Q(x)$. By the weak law of large numbers we know that the empirical distribution $\hat{P}_{X^n}(x)$ converges to the true distribution $Q(x)$ in probability for all $x \in \mathcal{X}$, (Theorem 11.2.1, [19]). Let $E \subseteq \mathcal{P}(\mathcal{X})$ be a subset of the set of probability mass function on $\mathcal{X}$. Consequently, if $Q \in E$ the probability that the type $\hat{P}_{X^n}$ belongs to E, which is denoted $Q^n\left(\hat{P}_{X^n} \in E\right)$, will converge to 1. On the other hand, if $Q \notin E$ then $Q^n\left(\hat{P}_{X^n} \in E\right) \to 0$ exponentially fast . Sanov's theorem gives an accurate rate of the exponential decrease of the probability $Q^n\left(\hat{P}_{X^n} \in E\right)$. Before stating the theorem, we will need to introduce two important lemmas. The first one (Lemma 2.13) bounds the probability that the random element $\hat{P}_{X^n}$ equals a particular distribution in $\mathcal{P}_n(\mathcal{X})$, and the second one (Lemma 2.14) says that when n is large enough, the set $\mathcal{P}_n(\mathcal{X})$ approximates uniformly and arbitrarily well (in the sense of variational distance) any set in $\mathcal{P}(\mathcal{X})$ (see Lemma 2.14).

Lemma 2.13. *[24] . Let X_1, X_2, ..., X_n be a sequence drawn i.i.d. according to $Q(x)$. The probability that the random empirical distribution $\hat{P}_{X^n}$ associated to it equals a type $\hat{P} \in \mathcal{P}_n(\mathcal{X})$ is bounded as follows:*

$$\frac{1}{(n+1)^{|\mathcal{X}|}} e^{-nD\left(\hat{P}\|Q\right)} \leq Q^n\left(\hat{P}_{X^n} = \hat{P}\right) \leq e^{-nD\left(\hat{P}\|Q\right)}. \tag{2.16}$$

The probability $Q^n\left(\hat{P}_{X^n} = \hat{P}\right)$ is also recognized as the probability of a type class $T(\hat{P})$, i.e. $Q^n\left(\hat{P}_{X^n} = \hat{P}\right) = Q^n\left(T(\hat{P})\right)$.

Proof. see [22] or [19] or [24]. □

Lemma 2.14. *[24] For any probability vector $P \in \mathcal{P}(\mathcal{X})$,*

$$d_{TV}\left(P, \mathcal{P}_n(\mathcal{X})\right) \triangleq \inf_{\hat{P}\in\mathcal{P}_n(\mathcal{X})} d_{TV}\left(P, \hat{P}\right) \leq \frac{|\mathcal{X}|}{n}, \tag{2.17}$$

where $d_{TV}\left(P, \hat{P}\right) \triangleq \sum_{a\in\mathcal{X}} \left|P(a) - \hat{P}(a)\right|$ is the total variational distance between two distribution P and $\hat{P}$.

Proof. see [24]. □

Theorem 2.15. *(Sanov's theorem). Let $X_1, X_{2,\ldots,} X_n$ be i.i.d. $\sim Q(x)$. Let $E \subseteq \mathcal{P}(\mathcal{X})$ be a set of probability distributions. Then*

$$\begin{aligned} -\inf_{P\in\overset{\circ}{E}} D(P \parallel Q) &\leq \liminf_{n\to\infty} \tfrac{1}{n} \log Q^n\left(\hat{P}_{X^n} \in E\right) \\ &\leq \limsup_{n\to\infty} \tfrac{1}{n} \log Q^n\left(\hat{P}_{X^n} \in E\right) \leq -\inf_{P\in E} D(P \parallel Q) \end{aligned} \tag{2.18}$$

where $\overset{\circ}{E}$ is the interior of E.

In addition, if the set E is a subset of the closure of its interior, i.e. $E \subseteq \overline{\overset{\circ}{E}}$ then the lower bound and upper bound are identical, i.e.

$$-\inf_{P\in\overset{\circ}{E}} D(P \parallel Q) = -\inf_{P\in E} D(P \parallel Q). \tag{2.19}$$

Therefore, the limit of $\frac{1}{n}\log Q^n\left(\hat{P}_{X^n}\in E\right)$ *exists and*

$$\lim_{n\to\infty}\frac{1}{n}\log Q^n\left(\hat{P}_{X^n}\in E\right)=-D\left(P^*\parallel Q\right),$$

or we can write

$$Q^n\left(\hat{P}_{X^n}\in E\right)\doteq e^{-nD(P^*\|Q)}. \tag{2.20}$$

where

$$D\left(P^*\parallel Q\right)=\inf_{P\in E}D\left(P\parallel Q\right) \tag{2.21}$$

$$P^*=\arg\min_{P\in\overline{\overset{\circ}{E}}}D\left(P\parallel Q\right), \tag{2.22}$$

and saying two sequences $a\left(n\right)\doteq b\left(n\right)$*, we means that*

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{a\left(n\right)}{b\left(n\right)}=0. \tag{2.23}$$

The proof of Sanov's theorem is presented in the appendix 2.6.1.

It might be a little hard to imagine how the set E looks like when saying it is a subset of the closure of its interior, i.e. $E\subseteq\overline{\overset{\circ}{E}}$. The following proposition gives a case of set E which is more familiar for non-mathematicians.

Proposition 2.16. *If* $E\subseteq\mathcal{P}\left(\mathcal{X}\right)$ *is a convex set of non-empty interior then* E *satisfies the property that* E *is a subset of the closure of its interior, i.e.* $E\subseteq\overline{\overset{\circ}{E}}$*. It turns out that the Sanov's theorem is still true when* E *is a convex set of non-empty interior.*

In chapter 3, we apply Sanov's theorem for convex sets whose interior is non-empty. The proof of this proposition is given in the appendix 2.6.2.

2.3 Fundamentals in channel coding theory

2.3.1 Channel coding

As mentioned in the introduction of this dissertation, the authentication model is analyzed using two settings, namely non-channel coding based authentication and channel coding based authentication. Chapter 4 treats the authentication problem by using the channel codes. Therefore, it is necessary to recall the literature on channel coding.

In his remarkable paper in 1948 [60], Claude E. Shannon established a conceptual basis for the problem of transmission of information. In his section dealing with transmission in discrete channels with noise, he advanced the necessity of encoding to achieve "as small a frequency of errors as desired". The communication system illustrated in Figure 2.1 will then be composed by a source of messages taking values in some finite alphabet, encoded into some sequence of channel symbols, which then produces the output sequence of the channel via a probabilistic mapping. The decoder attempts to convert the output sequence back to the transmitted message.

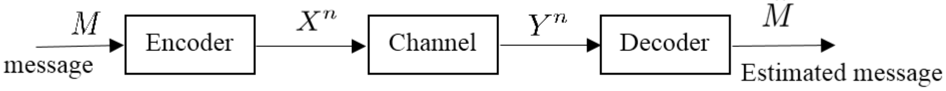


Figure 2.1: A communication system [19].

Shannon states that an encoder and a decoder can be designed to achieve a negligible probability of error as long as the transmitted information per channel use R is less than a quantity specifying the channel, the so-called channel capacity C. More specifically, the probability of error can be made arbitrarily small for any $R < C$ (achievability part), and conversely, the probability of error is bounded away from zero for any $R > C$ (converse part).

In this dissertation we focus on a discrete memoryless channel without feedback i.e. the input symbols do not depend on the past output symbols, which is defined as follow.

Definition 2.17. (Discrete Memoryless Channel). A discrete memoryless channel (DMC) denoted $(\mathcal{X}, W(y|x), \mathcal{Y})$ is a system consisting of a finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$ and a probability transition matrix $W(y \mid x)$ in which each output $y \in \mathcal{Y}$ is related to the corresponding input $x \in \mathcal{X}$ according to $W(y \mid x)$. For a memoryless channel, the transition function between an input and and output sequence $x^n = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ and $y^n = (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$ is given by:

$$W^n(y^n \mid x^n) = \prod_{i=1}^{n} W(y_i \mid x_i). \tag{2.24}$$

Definition 2.18. (Information Channel Capacity).The information channel capacity of a DMC $(\mathcal{X}, W(y \mid x), \mathcal{Y})$ is defined as

$$C(W) = \max_{P_X(x)} I(X;Y).$$

where the maximum is taken over all possible input distributions $P_X(x)$.

Definition 2.19. (Channel Codes). A $(\mathcal{M}_n, n)$ channel code $\mathcal{C}_n$ for a DMC $(\mathcal{X}, W(y \mid x), \mathcal{Y})$ consists of

- a message set $\mathcal{M}_n = \{1, 2, ..., \mid \mathcal{M}_n \mid\}$,
- an encoding function $\varphi : \mathcal{M}_n \to \mathcal{X}^n$, which maps a message m to a codeword x^n with n channel symbols,
- a decoding function $\psi : \mathcal{Y}^n \to \mathcal{M} \cup \{?\}$, which maps a channel output y^n to a message $\hat{m} \in \mathcal{M}_n$ or an error ?.

The set of codewords $\{\varphi(m) : m \in \mathcal{M}_n\}$ is called the codebook of $\mathcal{C}_n$. With a slight abuse of notation, we denote the codebook itself by $\mathcal{C}_n$ as well. Unless specified, messages are represented by a random variable M uniformly distributed in $\mathcal{M}_n$.

Definition 2.20. (Rate). A rate R of a channel code $(\mathcal{M}_n, n)$ is defined as

$$R = \frac{\log |\mathcal{M}_n|}{n} \text{ nats per channel use.}$$

Now we define the block (or word) error probability and the average block error probability as follows.

- The block error probability for a particular code $\mathcal{C}_n$ is defined as

$$Pe_B(\mathcal{C}_n) = \Pr\left\{\hat{M} \neq M \mid \mathcal{C}_n\right\} \tag{2.25}$$

Throughout this dissertation, we assume that all messages are equaly probable. Therefore the block error probability for the code $\mathcal{C}_n$ is alternatively expressed as follows

$$Pe_B(\mathcal{C}_n) = \frac{1}{|\mathcal{M}_\mathrm{n}|} \sum_{\mathrm{m}=1}^{|\mathcal{M}_\mathrm{n}|} \Pr\left\{\hat{M} \neq M \mid M = m, \mathcal{C}_n\right\}. \tag{2.26}$$

- The average block error probability over all possible random codebooks

$$\overline{Pe}_B = \sum_{\mathcal{C}_n} Pe_B(\mathcal{C}_n) \Pr\{\mathcal{C}_n\}. \tag{2.27}$$

When the message number m is represented as a binary vector $\mathbf{m}$ of length $K_n = \log_2 |\mathcal{M}_n|$ information bits and we can define the probability of bit error as the average of the probability that a bit m_k of $\mathbf{m}$ is different from its binary estimation $\hat{m}_k$,

$$Pe_b = \frac{1}{K} \sum_{k=1}^{K} \Pr\{\hat{m}_k \neq m_k\}. \tag{2.28}$$

Definition 2.21. (Achievable Rate). A rate R is called achievable for a DMC $(\mathcal{X}, W(y \mid x), \mathcal{Y})$ if there exists a sequence of $\left(\lceil e^{nR} \rceil, n\right)$ codes $\{\mathcal{C}_n\}_{n \geq 1}$ such that

$$\lim_{n \to \infty} Pe_B(\mathcal{C}_n) = 0.$$

To simplify the notation, we will write $\left(e^{nR}, n\right)$ codes to mean $\left(\lceil e^{nR} \rceil, n\right)$ in the rest of the dissertation.

Definition 2.22. (Operational Channel Capacity). The operational channel capacity of a DMC can be defined as the supremum of all achievable rates. In other words, it is the highest rate R at which information can be sent with arbitrarily low probability of error.

The channel coding theorem stated hereafter establishes that the information channel capacity is equal to the operational channel capacity. So from now on, unless specified, we use the terminology channel capacity to refer to both information capacity and operational capacity.

Theorem 2.23. *[19](Channel coding theorem). All rates below the channel capacity are achievable. Specifically, for every rate $R < C$, there exists a sequence of $\left(e^{nR}, n\right)$ codes $\{\mathcal{C}_n\}_{n \geq 1}$ such that $Pe_B(\mathcal{C}_n) \to 0$. Conversely, any sequence of $\left(e^{nR}, n\right)$ codes $\{\mathcal{C}_n\}_{n \geq 1}$ with $Pe_B(\mathcal{C}_n) \to 0$ must satisfy $R \leq C$.*

Proof. Please refer to chapter 7 in [19] . □

The proof of the direct part of the theorem (the existence of a code) is obtained with a random coding argument, thus calculating first the average probability of word error (2.27) over the ensemble of codebooks. This average probability of error is upper bounded by an exponentially vanishing term. One can argue then that since the average block error probability over the ensemble of codebooks can be arbitrary small, there exists at least one codebook such that the block error probability goes to zero. The error exponent specifying the rate at which the probability of error decays, depends on the decoding rule used to achieve the proof, namely the joint typical set decoding [19] or maximum likelihood decoding [28].

These decoding rules require the perfect knowledge of the channel. However, it should be noted that in our authentication problem, the legitimate receiver does not know whether the graphical code comes from the legitimate transmitter or the adversary. It is therefore natural for him to choose a decoding rule matched with the distribution law of the main channel. This implies that his decoder will be mismatched with respect to the opponent channel. We will present then in the next subsection a brief discussion about transmission in a channel with a mismatched decoder, and state a new version of the channel coding theorem for this scenario.

2.3.2 Mismatched decoding

Consider a memoryless channel with a transition law $W(y \mid x)$ mapping the input alphabet $\mathcal{X}$ to the output alphabet $\mathcal{Y}$. We study here the case where the decoder uses maximum likelihood decoding based on the additive metric $\log V(y \mid x)$ where $V(y \mid x)$ is different from the distribution law of the channel $W(y \mid x)$. Note that $V(y \mid x)$ is non-negative for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ but it is not necessary that $\sum_{y \in \mathcal{Y}} V(y \mid x) = 1$ for every $x \in \mathcal{X}$.

A rate R is called achievable for a DMC $W(y \mid x)$ and a mismatched decoding metric V if for every $\epsilon > 0$, there exists n large enough and a sequence of $\left(e^{nR}, n\right)$ codes such that the probability of error (2.26) when decoding with metric V is less than ϵ. The mismatched capacity of a DMC W with a mismatched decoding metric V, denoted as C_M, is defined as the highest achievable rate in the above definition.

The idea of decoding without the knowledge of the channel was first studied by Stiglitz. In [62], he dealt with the memoryless non stationary channel whose transition distribution is different for each channel use and chosen from a given set of transition distributions. In other words, the channel distribution varies from use to use and the receiver has no knowledge of which channel is selected. Later, Hui and Csiszar independently studied mismatched decoding for memoryless stationary channel. In [33], Hui used the random coding argument with the constraint that the empirical distribution of each codeword is close to a given distribution $P_X(x)$. He obtained a single letter expression for a lower bound on the mismatched capacity as the average block error probability (2.27) vanishes for large n. On the other hand, Csiszar and Körner [21] used graph decomposition technique to get the same lower bound, denoted C_{LM}, on mismatched capacity C_M.

Balakirsky in [8] succeeded in proving a converse coding theorem for mismatched decoding over binary input DMC channels, and proved that C_{LM} in this case is indeed the mismatched capacity C_M. The general converse coding theorem of mismatched decoding for arbitrary finite input alphabet is still an unsolved problem. Merhav et al. [51] established a converse with the random coding argument (averaging the probability of error over the ensemble of codebooks). They showed that C_{LM} is also an upper bound for the highest rate where the average error probability goes to zero.

So far, all results stated above are obtained with a deterministic coding strategy, i.e for a given codebook one codeword is assigned to one message. Merhav et al. [51] showed some forms of randomized encoding strategy capable of achieving rates higher than C_{LM}. By randomized encoding strategy we mean that, for a given codebook one codeword is chosen randomly from a given subset assigned to one message. For the latter example the achievability is expressed in terms of block error probability.

It is worth noting then that for transmission in a channel with a mismatched decoder, C_{LM} depends on the decoding metric, the strategy of encoding. i.e. deterministic or randomized encoding, and on the achievability criterion, i.e. block error probability or average error probability over the ensemble of codebooks. In this dissertation, unless specified, C_{LM} associates with deterministic encoding, maximum likelihood decoding and averaged block error probability (2.27).

In the sequel, we recall the theorem characterizing the lower bound C_{LM} of the mismatched capacity. We refer the readers to [42], [33] for relevant references.

Theorem 2.24. *[42] The mismatched capacity, $C_M\left(W,V\right)$, of a DMC $\left\{\mathcal{X},W\left(z\mid x\right),\mathcal{Z}\right\}$ in the presence of decoding mismatched according to the metric $d\left(x,z\right)=\log V\left(z\mid x\right)$ is lower bounded by $C_{LM}\left(W,V\right)$, which is given by*

$$C_{LM}\left(W,V\right)=\max_{P_X}I_{LM}\left(P_X\right), \tag{2.29}$$

where

$$I_{LM}\left(P_X,W,V\right)=\min_{f\in\mathcal{F}\left(P_X,W,V\right)}I_f\left(X;Z\right), \tag{2.30}$$

and the minimum is over all $f\left(x,z\right)\in\mathcal{F}\left(P_X,W,V\right)$ that satisfy

$$\mathcal{F}\left(P_X,W,V\right)=\left\{f\left(x,z\right)\in\mathcal{P}\left(\mathcal{X}\times\mathcal{Z}\right):\begin{array}{c}\sum\limits_{x\in\mathcal{X}}f\left(x,z\right)=P_Z\left(z\right),\text{ for all }z\in\mathcal{Z}\\ \sum\limits_{z\in\mathcal{Z}}f\left(x,z\right)=P_X\left(x\right),\text{ for all }x\in\mathcal{X}\\ \sum\limits_{x\in\mathcal{X}}\sum\limits_{z\in\mathcal{Z}}f\left(x,z\right)\log V\left(z\mid x\right)\geq -D\end{array}\right\} \tag{2.31}$$

where

$$\begin{aligned}P_X\left(x\right)&\overset{\triangle}{=}\sum_{z\in\mathcal{Z}}P_{XZ}\left(x,z\right)=\sum_{z\in\mathcal{Z}}P\left(x\right)W\left(z\mid x\right)\\ P_Z\left(z\right)&\overset{\triangle}{=}\sum_{x\in\mathcal{X}}P_{XZ}\left(x,z\right)=\sum_{x\in\mathcal{X}}P\left(x\right)W\left(z\mid x\right)\\ -D&\overset{\triangle}{=}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}}P\left(x\right)W\left(z\mid x\right)\log V\left(z\mid x\right)\end{aligned}$$

Remark 2.25. When the the decoding metric $d\left(x,z\right)=\log V\left(z\mid x\right)$,i.e. it mismatched with the channel $W\left(z\mid x\right)$, it is easy to check that $C_{LM}\left(W,V\right)\leq C\left(W\right)$. If the decoding metric $d\left(x,z\right)=\log W\left(z\mid x\right)$, i.e. it matches with the channel $W\left(z\mid x\right)$ then the decoding rule coincides with maximum-likelihood decoding and $C_{LM}\left(W,W\right)=C\left(W\right)$, the capacity of the channel W.

2.4 Related works on authentication

The idea of using information-theoretic security to analyze performance in authentication has been studied extensively. Without intending to be exhaustive, we will present the most significant works that may have directly or indirectly oriented many choices in this research. We focus on two major models concerned in authentication: message authentication, biometric systems and fingerprinting.

2.4.1 Message authentication

Message authentication is concerned in a secure communication system where the receiver of a message needs a sufficient amount of evidence to accept it as truly sent by the legitimate transmitter. It is then obvious that the need of authentication arises because of the presence of an opponent who sends fraudulent messages hoping that they will be accepted as valid by the receiver. Information theoretic analysis of message authentication models assume that the opponent has unlimited computing power. As for confidential communication, authentication can be achieved by encrypting the message where the sender and the receiver share secretly a key K. There are two types of attacks that may be launched by the opponent, namely impersonation and substitution attacks. In the impersonation attack the adversary introduces a fraudulent message to the receiver before the transmitter sends anything. In a substitution attack the adversary intercept the encrypted message sent by the transmitter and replace it by a fraudulent one. These attacks are considered successful if the receiver accepts the fraudulent message as valid.

In [61], Simmons provides information-theoretic lower bounds on the opponent's success probability in a noiseless communication system. Later, considering authentication as an hypothesis testing problem, Maurer [50] proposed a generalized treatment for multiple messages $M_1, ..., M_n$ authentication model. In case of single message authentication ($n = 1$), the opponent's success probabilities P_I (for an impersonation attack), and P_S (for a substitution attack), are lower bounded as follows:

$$P_I \geq 2^{-I(K;f(M,K))} \quad \text{and} \quad P_S \geq 2^{-H(K|f(M,K))} \ .$$

where M is the message and f is a cryptographic encoding function. In the impersonation attack the opponent sends his fraudulent message without observing the channel. Consequently the test will measure the degree of dependency between the key K and the received message. The cyphered message $C = f(M, K)$ sent by a legitimate source should then contains a sufficient amount of information about the key K in order to convince the authenticator that the transmitted message is valid. If not, the received message will be denied. Of course C should not leak all information about the key in this case because it will allow the adversary to guess K and to launch a substitution attack with a high probability of success. This tradeoff between P_I and P_S can easily be observed in the exponent terms as $I(K; f(M, K))$ should be large to reduce P_I, which unfortunately decreases $H(K \mid f(M, K))$. If the opponent can optimize his chance of success, i.e. by choosing between a substitution attack at time n or an impersonation attack at any time $1 \leq i \leq n$, Maurer [50] provides a lower bound on the opponent's success probability:

$$\max(P_{I_1}, P_{I_2}, ..., P_{I_n}, P_{S_n}) \geq 2^{-H(K)/(n+1)}.$$

Note that all these bounds suppose that the legitimate receiver never rejects a valid message. In case of one message authentication ($n = 1$), the key is then split in two parts where half of it is used to protect against impersonation attack and the second half is designed to protect against substitution attack.

2.4.2 Authentication using noisy channels

Unlike the model described above, channels in our authentication model are noisy channels. The receiver may then observe noisy versions of either a valid graphical code or a fraudulent one. The opponent's success probability will decrease as discrimination of the two noisy versions is increased. In chapter 3, we will express it then with a more general form since the error exponent involves the relative entropy rather than the mutual information. Moreover we will consider cases when the legitimate receiver can reject authentic codes.

In [41], L. Lai et al. studied the same message authentication model but over noisy channels. By exploiting an authentication counterpart of Wyner's wiretap channel [73], the authors showed that noise is not detrimental to authentication, but rather make the opponent's success probability smaller. More specifically, they aimed at designing a codebook in such a way that the distribution law of the key conditioned to the opponent's output is very close to uniform. This similarity is measured in term of total variational distance between the posterior distribution of the key and its prior one. Consequently, when observing a fraudulent message, the receiver will not be able to select accurately the exact key among all others as the uncertainty about it is maximal. With high probability the chosen key will not match to his database, and the access will be denied. Using a randomized coding strategy borrowed from the work of Csiszar [20] about a coloring function almost uniformly distributed when applied on a large space of events, and almost independent of another random variable, they conclude that there exists constants $c > 0$ and $\beta > 0$ so that:

$$2^{-H(K)} \leq \max\left\{P_I, P_S\right\} \leq 2^{-H(K)} + c\exp\left(-n\beta\right).$$

It means that all the key can be used to protect against substitution and impersonation attacks simultaneously. The authors also gave similar bounds on cheating probability in the case of sending multiple messages.

Discussion

In the model in [41], the opponent channel is a degraded version of the legitimate channel, where the characterization "degraded" follows the partial ordering of channels described in [40]. This is sufficient to prove the existence of an encoder and a decoder achieving secret and reliable transmission by using the wiretap model. In contrast, the opponent in our authentication model has a direct access to the printed graphical code in the same way as the legitimate receiver does. Thus the two channels[1] are statistically similar and the wiretap channel model is not relevant if we consider the opponent as an eavesdropper. To deal with the wiretap channel model, one must give artificially the role of the eavesdropper to the legitimate receiver as he should not be able to extract any information about the original message after observing a fraudulent graphical code. However, the bounds on cheating probability in [41] must be reconsidered because the decoder at the receiver stage is matched to the main channel and not to the channel involving the counterfeit process.

2.4.3 Biometric authentication

Another important field in the research of information theoretic authentication are biometric systems. Biometric systems are devoted to verify and authenticate user's identity using its biologic traits such as fingerprints or iris scanning. Biometric characteristics have the attractive advantage that they refer to an unique identifier. There are traditionally two phases in a biometric system: the enrollment and the authentication. In the enrollment phase, the biometric data of a user are extracted and saved as a biometric templates. Later, when an access is required, a fresh measurement of the same user's characteristic is observed, processed and compared to the reference template. If the observed sample is sufficiently similar to the database then the result is positively authenticated. One evident drawback of the system is that the measurements of a same biometric trait are different due to uncontrolled variability. Another important issue is the problem of secrecy and hence privacy, which arises if biometric data are stored in a clear form. The uniqueness of a biometric information leads paradoxically to an additional weakness. Once a database compromised, a biometric information cannot be revoked and it results in a "partially stolen identity" [34]. Biometric systems should then consider robustness along with secure storage to preserve privacy.

There has been an increasing interest in research to fulfill all these requirements for biometric systems.

Fuzzy extractors and secure sketch [26] are among the most rigorously analyzed approaches. Fuzzy extractor consists in extracting a secret key out from the biometric information along with a helper data during the enrollment phase. The secret key or a hash function of it may be stored in a central server, while the helper data is public. The role of the latter is to assist the reconstruction of the key from a noisy measurement when authentication is required.

In the enrollment phase of a secure sketch, a procedure having input data w will output a sketch s. In the authentication phase, if an observed noisy data w' is close to w it will be possible to recover w with the help of s. The sketch is secure because it does not reveal any information about w, but while the key generated out by the fuzzy extractor needs to be as uniform as possible, the secure sketch doesn't address uniformity. Fuzzy

[1] In chapter 3 we will give another definition to the opponent channel which will include the counterfeiter processing and printing.

commitment scheme introduced by Juels and Wattenberg [39] is one of the earlier approach of secure sketch. It is based on error-correcting codes and considers biometric data in the binary space with the hamming distance measure. Using results from the wiretap channel theory, initially devoted to confidential communication, Cohen and Zemor [18] implemented a fuzzy commitment scheme using a randomized strategy of coding with coset codes, to provide together secrecy and robustness (correctness property). Further implementations of secure sketch are applied on quantized real valued biometric templates in [43] where the authors study how the sketch affects the authentication performance and the probability that an adversary guesses the biometric data after observing the sketch.

In an enlightening framework on biometric systems with an information-theoretic point of view, Ignatenko and Willems [35] reconsidered the problem of security emphasizing on the privacy and secrecy aspects. Using a setup similar to the fuzzy extractor where a secret key is extracted out from biometric data, they present rather their biometric system as an instance of the secret sharing concept introduced by Maurer [49] and later by Ahlswede and Csiszár [5]. They adapted the source-type model in [5] where two terminals or sources X and Y communicate information over a noiseless public channel. Each source generates at the end of this exchange secret keys S and $\hat{S}$ respectively as functions of all exchanged information and source outputs. When $S = \hat{S}$ the two terminals deem to share a common secret key.

In the presented biometric system, only one message is exchanged and corresponds to the helper M, and the biometric data X^n at the enrollment and Y^n at the authentication phase are the source outputs. We say then that X and Y share a common secret key. Because the helper is public, it should not contain any information about the key to preserve secrecy and should leak a negligible amount of information about the source to prevent privacy theft. Moreover, the secret-key rate $\frac{1}{n}\log|\mathcal{S}|$ (where $|\mathcal{S}|$ is the size of the key's set) should be large enough to minimize the probability that it would be guessed by an unauthorized user. The authors also present other settings, particularly the setting where the secret key is generated independently from the biometric data (called "setting with chosen key"). Their results are stated in a theorem for each setting delimiting an achievable region of the pair (secret-key rate, privacy leakage). For instance, for the chosen key setting, achievability of a pair (R, L) is obtained if an encoder/decoder exists such that for any $\delta > 0$ we have:

$$\begin{aligned} \Pr\{S \neq \hat{S}\} &\leq \delta \\ \frac{1}{n}\log|\mathcal{S}| &\geq R - \delta \\ \frac{1}{n}I(M;S) &\leq \delta \\ \frac{1}{n}I(M;X^n) &\leq L + \delta \end{aligned}$$

The first inequality insures the property of correctness (the two terminals share a common secret), the second one insures that it is possible to generate a secret-key with a rate greater than the number R , the third inequality insures negligible secrecy leakage and finally the last one upper-bounds the privacy leakage by the number L. The maximal possible secret-key rate is the mutual information $I(X;Y)$, which is also a result from [5]. The authors proposed recently practical implementations for the chosen key setting, based on key-binding using BCH and convolutional codes [36]. They also studied in [71] upper-bounds on the false acceptance exponent terms in a scheme based on secret-key extraction. The computation of this exponent is based on the assumption that the biometric imposter sequence and the biometric sequence of an authorized individual have the same marginal distribution, but these two biometrics are independent.

Discussion

The problem of verifying if data extracted out from an observed measurement matches a locally stored secret message, arises naturally in our authentication scheme. The legitimate receiver will be deceived if he can recover the secret-message stored in his database from the observed measurement which comes from a counterfeiter. The probability that this event happens is enforced with the fact that the adversary launch a kind of informed attack as he can observe the printed and scanned code Y^n. Consequently, an available helper along with Y^n will be detrimental to our model. We conclude that the helper, if it exists, must be communicated in a secure channel. In addition, the fraudulent code obtained from the informed attack and the original code are depended random sequences, unlike the case of an impostor and the legitimate user in biometric system. False acceptance exponent will then be related to relative entropy rather than mutual information.

2.4.4 Authentication using fingerprinting

In [10], the authors analyzed an authentication model based on binary content fingerprint when an informed attack is launched. Informed attackers may indeed benefit from some leakage about the stored authentic fingerprints. In order to detect any counterfeit, the authors propose to test the similarity between the binary fingerprint obtained from the measurement of an item with a given claimed label and the corresponding authentic binary fingerprint stored in the database. The hamming distance between these two binary sequences is a sufficient statistic for this hypothesis testing. It is then compared to a threshold to infer a decision. The hamming distance has a binomial distribution and the authors use results from large deviation to upper-bound the probability of false rejection and the probability of false acceptance. Since there is always a trade-off between these two types of error, they optimize the authentication performance by finding the threshold that minimize the maximum value between them. This optimization is preferable to the Neyman-Pearson criterion only for strict technical and commercial requirements aiming at small values of both types of error.

2.4.5 Authentication using graphical codes

Combating counterfeit of commercial products or documents using printed graphical code authentication (see the work of Picard [53],[54]) shares many similarities with the setting in [10].

While in [10], the authors test the possible matching of binary sequences, this dissertation considers performance's analysis for an arbitrary discrete space of measurements in case of an informed counterfeiter. Bounds and accurate asymptotic expressions are then provided using the method of type and large deviation principle. As mentioned previously, a trade-off is necessary between the false rejection error and the false acceptance error and we just do with the Neyman-Pearson criterion. The use of channel encoder aims at breaking partially this trade-off by separating the objectives over each channel, namely the legitimate channel and the opponent channel. First, the channel coding theorem will sustain the idea of a possible arbitrarily small probability of false alarm (false rejection). Second, the converse of the mismatch decoding theorem will insure a very low probability of miss detection (false acceptance). However, in the case that channel coding is employed, secrecy leakage may be very large as the opponent may retrieve the message due to the structure of codes. Therefore, it is essential to provide some additional encryption function to conceal the secret message from the opponent. We will consider perfect secrecy (zero secrecy leakage) provided by an one-time-pad function.

2.5 Conclusions

In this chapter, we have presented the fundamentals in information theory which are needed in this dissertation such as basic information measures, the method of type and channel coding theorem for matched and mismatched decoding. We have also presented the Sanov's theorem which plays a very important role for our study in this dissertation. Its proof is given in details in the appendix 2.6. We have also introduced some related works using information-theoretic tools to analyze performance of authentication, for instance message authentication, biometric systems and fingerprinting.

2.6 Appendix

2.6.1 Proof of Sanov's theorem

First, we will show the upper bound of (2.18),

$$\limsup_{n\to\infty} \frac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E \right) \leq - \inf_{P \in E} D\left(P \parallel Q\right) \tag{2.32}$$

We have:

$$\begin{aligned} Q^n \left(\hat{P}_{X^n} \in E \right) &= \sum_{P \in E \cap \mathcal{P}_n(\mathcal{X})} Q^n \left(\hat{P}_{X^n} = P \right), \\ &\overset{(a)}{\leq} \sum_{P \in E \cap \mathcal{P}_n(\mathcal{X})} e^{-nD(P\|Q)}, \\ &\leq \sum_{P \in E \cap \mathcal{P}_n(\mathcal{X})} \sup_{P \in E \cap \mathcal{P}_n(\mathcal{X})} e^{-nD(P\|Q)}, \\ &= \sum_{P \in E \cap \mathcal{P}_n(\mathcal{X})} e^{-n \inf_{P \in E \cap \mathcal{P}_n(\mathcal{X})} D(P\|Q)}, \\ &\leq \sum_{P \in E \cap \mathcal{P}_n(\mathcal{X})} e^{-n \inf_{P \in E} D(P\|Q)}, \\ &= \left| E \cap \mathcal{P}_n\left(\mathcal{X}\right) \right| e^{-n \inf_{P \in E} D(P\|Q)}, \\ &\overset{(b)}{\leq} (n+1)^{|\mathcal{X}|} e^{-n \inf_{P \in E} D(P\|Q)}. \end{aligned} \tag{2.33}$$

where (a) comes from Lemma 2.13 and (b) comes from Lemma 2.8. Taking logarithm and limsup on both sides of (2.33), we have:

$$\begin{aligned} \limsup_{n\to\infty} \tfrac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E \right) &\leq \limsup_{n\to\infty} \tfrac{1}{n} \log \left(n+1\right)^{|\mathcal{X}|} \\ &\quad + \limsup_{n\to\infty} \left\{ - \inf\nolimits_{P \in E} D\left(P \parallel Q\right) \right\}, \\ &\overset{(c)}{=} \lim_{n\to\infty} \tfrac{1}{n} \log \left(n+1\right)^{|\mathcal{X}|} \\ &\quad - \liminf_{n\to\infty} \left\{ \inf\nolimits_{P \in E} D\left(P \parallel Q\right) \right\}, \\ &= - \inf\nolimits_{P \in E} D\left(P \parallel Q\right). \end{aligned} \tag{2.34}$$

(c) comes from the fact that $\limsup_{n\to\infty}(-x_n) = -\liminf_{n\to\infty} x_n$. So the upper bound is followed. Now we turn to prove the lower bound of (2.18),

$$-\inf_{P\in\overset{\circ}{E}} D(P \parallel Q) \leq \liminf_{n\to\infty} \frac{1}{n} \log Q^n\left(\hat{P}_{X^n} \in E\right). \tag{2.35}$$

Again we have:

$$\begin{aligned} Q^n\left(\hat{P}_{X^n} \in E\right) &= \sum_{P\in E\cap\mathcal{P}_n(\mathcal{X})} Q^n\left(\hat{P}_{X^n} = P\right) \\ &\overset{(a)}{\geq} \sum_{P\in E\cap\mathcal{P}_n(\mathcal{X})} (n+1)^{-|\mathcal{X}|}\, e^{-nD(P\|Q)} \\ &\geq (n+1)^{-|\mathcal{X}|} \sup_{P\in E\cap\mathcal{P}_n(\mathcal{X})} e^{-nD(P\|Q)} \\ &= (n+1)^{-|\mathcal{X}|}\, e^{-n\inf_{P\in E\cap\mathcal{P}_n(\mathcal{X})} D(P\|Q)}. \end{aligned} \tag{2.36}$$

where (a) comes from Lemma 2.13. Therefore,

$$\begin{aligned} \liminf_{n\to\infty} \tfrac{1}{n}\log Q^n\left(\hat{P}_{X^n} \in E\right) &\geq \liminf_{n\to\infty} \left\{-\inf_{P\in E\cap\mathcal{P}_n(\mathcal{X})} D(P \parallel Q)\right\}, \\ &= -\limsup_{n\to\infty} \left\{\inf_{P\in E\cap\mathcal{P}_n(\mathcal{X})} D(P \parallel Q)\right\}. \end{aligned} \tag{2.37}$$

We need now to show that for any $P \in \overset{\circ}{E}$, the right hand side of (2.37) is greater than $D(P \parallel Q)$.

Fix then an arbitrary point $P \in \overset{\circ}{E} \subseteq \mathcal{P}(\mathcal{X})$. By the definition of an interior point, there exists $\delta > 0$, small enough and an open ball centered on P with radius δ such that:

$$B_\delta = \left\{P' \in \mathcal{P}(\mathcal{X}) : d_{TV}\left(P, P'\right) < \delta\right\} \subset E.$$

The crucial point here is that as n becomes larger the open ball B_δ contains also elements of $\mathcal{P}_n(\mathcal{X})$. Indeed, by using Lemma 2.14 we know that $\mathcal{P}_n(\mathcal{X})$ approximates uniformly well the set $\mathcal{P}(\mathcal{X})$ in the sense of total variational distance. One can then find a sequence $\left\{\hat{P}_n\right\} \in \mathcal{P}_n(\mathcal{X})$ such that:

$$\lim_{n\to\infty} d_{TV}\left(P, \hat{P}_n\right) = \inf_{\hat{P'}_n\in\mathcal{P}_n(\mathcal{X})} d_{TV}\left(P, \hat{P'}_n\right) \leq \frac{|\mathcal{X}|}{n}. \tag{2.38}$$

Take now N_δ such that $\frac{|\mathcal{X}|}{n} < \delta$ for all $n > N_\delta$, thus $\lim_{n\to\infty} d_{TV}\left(P, \hat{P}_n\right) \leq \frac{|\mathcal{X}|}{n} < \delta$ and consequently $\hat{P}_n \in B_\delta$. Hence there exists a sequence $\hat{P}_n \in E\cap\mathcal{P}_n(\mathcal{X})$ such that $\hat{P}_n \to P$

as $n \to \infty$ in the sense of d_{TV}. Because of the continuity of the relative entropy we have:

$$\lim_{n\to\infty} D\left(\hat{P}_n \parallel Q\right) = D\left(P \parallel Q\right). \tag{2.39}$$

As $\hat{P}_n \in E \cap \mathcal{P}_n(\mathcal{X})$, we have:

$$\inf_{\hat{P}'\in E\cap\mathcal{P}_n(\mathcal{X})} D\left(\hat{P}' \parallel Q\right) \;\; \leq \;\; D\left(\hat{P}_n \parallel Q\right). \tag{2.40}$$

Taking limsup on two sides we have

$$\limsup_{n\to\infty}\left\{\inf_{\hat{P}'\in E\cap\mathcal{P}_n(\mathcal{X})} D\left(\hat{P}' \parallel Q\right)\right\} \leq \limsup_{n\to\infty} D\left(\hat{P}_n \parallel Q\right). \tag{2.41}$$

Moreover, because of the existence of the limit (2.39) then we have $\limsup\limits_{n\to\infty} D\left(\hat{P}_n \parallel Q\right) = \lim\limits_{n\to\infty} D\left(\hat{P}_n \parallel Q\right)$, and:

$$\begin{aligned} -\limsup_{n\to\infty}\left\{\inf_{P'\in E\cap\mathcal{P}_n(\mathcal{X})} D\left(P' \parallel Q\right)\right\} \;\; &\geq \;\; -\lim_{n\to\infty} D\left(\hat{P}_n \parallel Q\right), \\ &= \;\; = -D\left(P \parallel Q\right). \end{aligned} \tag{2.42}$$

The last equality comes from (2.39) and is true for any $P \in \overset{\circ}{E}$. Thus:

$$-\limsup_{n\to\infty}\left\{\inf_{P'\in E\cap\mathcal{P}_n(\mathcal{X})} D\left(P' \parallel Q\right)\right\} \geq \sup_{P\in\overset{\circ}{E}} - D\left(P \parallel Q\right) = -\inf_{P\in\overset{\circ}{E}} D\left(P \parallel Q\right). \tag{2.43}$$

So the lower bound of (2.18) is proved.

Now we show that if the set E is a subset of the closure of its interior, i.e. $E \subseteq \overline{\overset{\circ}{E}}$ then the lower bound and the upper bound are identical, i.e.

$$\inf_{P\in\overset{\circ}{E}} D\left(P \parallel Q\right) = \inf_{P\in E} D\left(P \parallel Q\right).$$

If the set E is a subset of the closure of its interior, then the resulting set ordering $\overset{\circ}{E} \subseteq E \subseteq \overline{\overset{\circ}{E}}$, we infer the following one:

$$\inf_{P\in\overset{\circ}{E}} D\left(P \parallel Q\right) \geq \inf_{P\in E} D\left(P \parallel Q\right) \geq \inf_{P\in\overline{\overset{\circ}{E}}} D\left(P \parallel Q\right). \tag{2.44}$$

Now Lemma 2.26 hereafter relates the first and the last terms in the ordered sequence (2.44).

Lemma 2.26. *For any open subset $S \subseteq \mathcal{P}(\mathcal{X})$ we have the following equality:*

$$\inf_{P \in S} f(P) = \inf_{P \in \overline{S}} f(P). \tag{2.45}$$

The proof of the lemma will be deferred to the end of this subsection. As a direct consequence of this lemma, one can easily establish the equality after taking $S = \overset{\circ}{E}$ and $f(P) = D(P \parallel Q)$:

$$\inf_{P \in \overset{\circ}{E}} D(P \parallel Q) = \inf_{P \in E} D(P \parallel Q) = \inf_{P \in \overline{\overset{\circ}{E}}} D(P \parallel Q).$$

This means that the lower and upper bounds are identical:

$$\begin{aligned} -\inf_{P \in \overset{\circ}{E}} D(P \parallel Q) &= \liminf_{n \to \infty} \tfrac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E\right) \\ &= \limsup_{n \to \infty} \tfrac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E\right) \quad = -\inf_{P \in E} D(P \parallel Q). \end{aligned} \tag{2.46}$$

In other words, limit of $\dfrac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E\right)$ exists and

$$\lim_{n \to \infty} \frac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E\right) = -\inf_{P \in E} D(P \parallel Q). \tag{2.47}$$

Moreover, $D(P \parallel Q)$is continuous on the compact set $\overline{\overset{\circ}{E}}$, thus it achieves minimum. So we can rewrite (2.47) as follows

$$\lim_{n \to \infty} \frac{1}{n} \log Q^n \left(\hat{P}_{X^n} \in E\right) = -D(P^* \parallel Q). \tag{2.48}$$

where

$$P^* = \arg\min_{P \in \overline{\overset{\circ}{E}}} D(P \parallel Q). \tag{2.49}$$

Proof of Lemma (2.26): Let $a = \inf_{P \in S} f(P)$ and $b = \inf_{P \in \overline{S}} f(P)$. Therefore from the property of infimum, there exists sequences $\{P_n\}_{n=1}^{\infty}$ and $\left\{P_n^{'}\right\}_{n=1}^{\infty}$ in S such that:

$$\lim_{n \to \infty} f(P_n) = a, \tag{2.50}$$

$$\lim_{n \to \infty} f\left(P_n^{'}\right) = b. \tag{2.51}$$

Then we have $f\left(P_n^{'}\right) \geq a$ because $P_n^{'} \in S$. Thus $b = \lim_{n \to \infty} f\left(P_n^{'}\right) \geq a$. But $a \geq b$ because $S \subseteq \overline{S}$. So $a = b$.

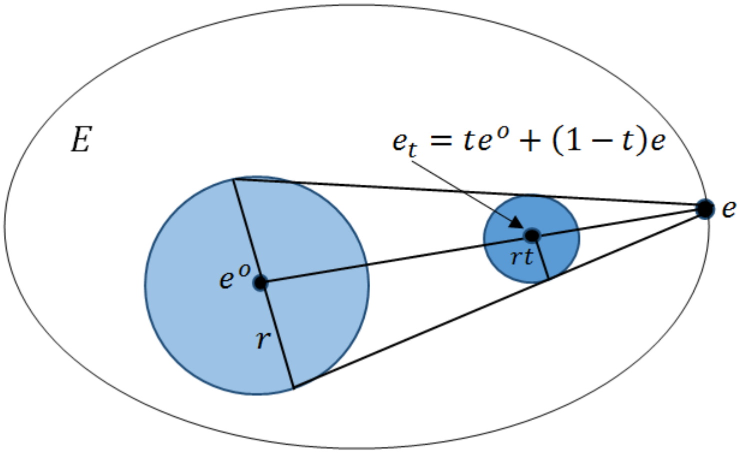


Figure 2.2: Geometry of C_δ.

2.6.2 Proof of Proposition 2.16

In this subsection, we will show that if E is a convex set whose interior is non-empty then $E \subseteq \overline{\mathring{E}}$. Assume that $e \in E$, we will show that $e \in \overline{\mathring{E}}$.

- If $e \in \mathring{E}$ then $e \in \overline{\mathring{E}}$ as desired.
- If $e \in \partial E$, the boundary of E, we will show that there exists a sequence $\{e_t\}_{t\in(0,1]} \subseteq \mathring{E}$ such that $e_t \to e$ as $t \to 0$. This means that $e \in \overline{\mathring{E}}$.

Now we present this second point in details. Take an arbitrary $e^\circ \in \mathring{E}$ and consider the line $e_t = te^\circ + (1-t)\,e$, with $t \in (0,1]$. We will show that e_t belongs to the set $\mathring{E}$ for all $t \in (0,1]$. When $t = 1$ we have $e_1 = e^\circ \in \mathring{E}$. So we have to prove that e_t belongs to the set $\mathring{E}$ for all $t \in (0,1)$. Because $e^\circ \in \mathring{E}$, then by definition of the interior point, there exists an $r > 0$ such that the ball $B\,(e^\circ, r) \subset E$. Let C_δ be the set of all lines connecting point e to all points $e^{'} \in B\,(e^\circ, r)$. This set C_δ can be understood as a cone if our set E is in $\mathbb{R}^3$. Particularly,

$$C_\delta = \left\{ e_t^{'} = te + (1-t)\,e^{'} \,:\, t \in (0,1)\,,\, e^{'} \in B\,(e^\circ, r) \right\} \tag{2.52}$$

We have then $C_\delta \subseteq E$ because of the convexity of E. Therefore $\mathring{C_\delta} \subseteq \mathring{E}$.

Moreover, $e_t \in \mathring{C_\delta}$ for all $t \in (0,1)$. Indeed, there exits a ball $B\,(e_t, {}^{rt}\!/_2) \subset \mathring{C_\delta}$, see Figure 2.2. Therefore, $e_t \in \mathring{C_\delta} \subseteq \mathring{E}$ for all $t \in (0,1)$. In addition, $e_t \to e$ as $t \to 0$. So $e \in \overline{\mathring{E}}$ as desired.

Chapitre 3

Authentication Without channel coding - Theoretical analysis

As stated in the first chapter, a secret message can either be mapped directly into a binary graphical code (GC) or being encoded with some probably stochastic function before this mapping. In both cases, the resulting graphical code is printed on packages and therefore two settings will be investigated for authentication in this dissertation, namely without and with channel encoding.

In the current chapter, we consider the setup in which a binary GC is printed on a package of a product without being encoded before. First, we describe the authentication setup and give a mathematical representation of the model. We then present two possible strategies that could be exploited by the receiver. The simplest one is naturally to convert the scanned and observed gray level code to a binary version in order to perform authentication, and the second one in contrast, is to process directly the observed gray level code. It is shown that the latter strategy offers better performance for authentication when Neyman-Pearson test is employed. These performances are given in terms of asymptotic expressions of the probability of false alarm and the probability of non-detection, based upon Sanov's theorem. For the sake of comparison, these two types of error probabilities are also computed using Gaussian approximation using CLT (the Central Limit Theorem).

3.1 Authentication problem formulation

3.1.1 Setup

Our authentication model without channel coding is summarized as follows. A secret message is generated uniformly randomly, then one-to-one mapped to a binary GC before being printed on the package of a product. Since the secret message is mapped using a one-to-one function to create the GC, our study will focus directly on the GC, thus considered as an authentication sequence x^n chosen at random from the message set $\mathcal{X}^n$, and shared secretly with the legitimate receiver. Throughout this dissertation, we assume that the authentication sequence is discrete, independent and identically distributed (i.i.d.). Once marked on a package, the printed code can be scanned directly by the legitimate receiver to test the authenticity of the product. It (the printed code) may also be scanned by an opponent who produces then a new printed copy that will mark the package of his forgery.

For the sake of simplicity, we combine devices like the printer and the scanner as one, which is depicted as "Printer & Scanner" in Figure 3.1. It is assumed also that the opponent uses the same quality of scanner as the legitimate receiver does. Therefore it can be considered that the grey level versions of the GC observed by the legitimate receiver and the one observed by the opponent have the same distribution law governing a random sequence Y^n.

The authentication model can then be set as follows: an authentication sequence x^n is published as a noisy random version Y^n taking values in the set of points $\mathcal{O}^n$. The legitimate receiver observes the outcome y^n, while the opponent processes his own observation of Y^n to print a fake code on a package of his counterfeit product. The scanned version of the latter is z^n taking values in the set of points $\mathcal{O}^n$. The receiver observes then a sequence o^n which can either be the scanned version of the original printed code y^n or the scanned version of the fake code z^n, and using his knowledge about the secret message, he establishes a statistical test in order to decide whether the observed sequence is genuine or not.

The authentication model may then be viewed as a secret communication problem involving two channels $\mathcal{X} \to (\mathcal{Y}, \mathcal{Z})$. We define the main channel as the channel between the legitimate parts (source and receiver), and the opponent channel as the channel between the legitimate source and the receiver but passing through the counterfeiter channel (see Figure 3.1). The two channels $\mathcal{X} \to (\mathcal{Y}, \mathcal{Z})$ are considered being discrete and memoryless with conditional probability distribution $P_{YZ|X}(y, z \mid x)$. The marginal channels $P_{Y|X}$ and $P_{Z||X}$ constitute the transition probability matrices of the main channel and the opponent channel respectively.

It is noted that the authentication sequence x^n is generated using a secure pseudo-random number generator (PRNG) having a sufficiently large key space to prevent brute-force attacks. The seed of the PRNG can be practically transmitted using both a secure lossless communication channel and via a key distribution system so that the receiver can generate x^n from the seed. The security of such a system is beyond the scope of this dissertation.

In the sequel, we aim at expressing the marginal distributions $P_{Y|X}$ and $P_{Z|X}$.

3.1.2 Channel modeling

As mentioned previously, the device combining the principal physical elements, i.e. print and scan devices, is a stochastic process so we can model it by a stochastic

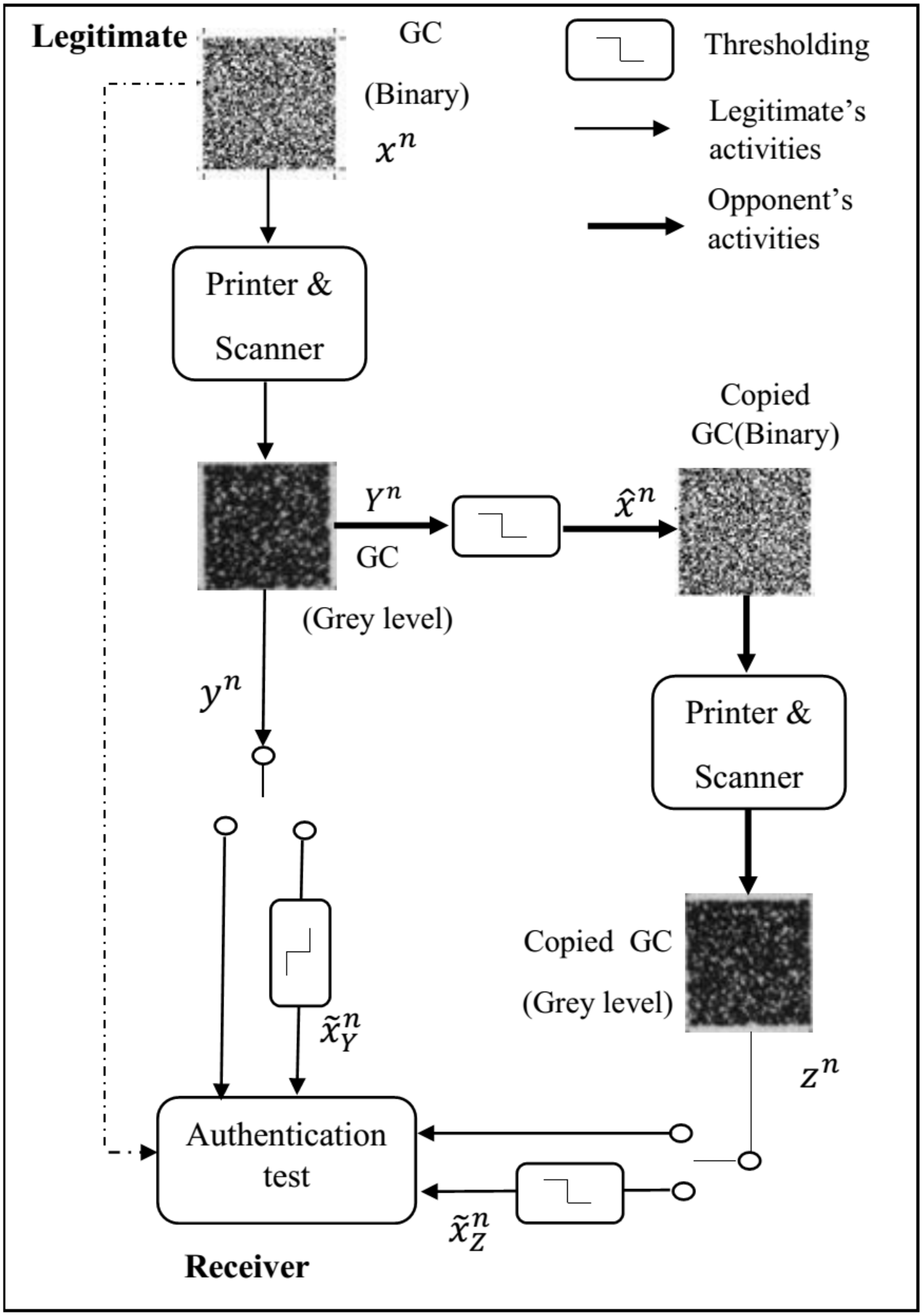


Figure 3.1: Non-channel coding based authentication model.

matrix relating its input to its output. Let T_m be the transition matrix modeling the combined devices involved in the main channel. The entries of this matrix are conditional probabilities $T_m(o|x)$ relating the binary input alphabet $\mathcal{X}$ and the gray level output alphabet $\mathcal{O}$. The transition matrix T_m may be any discrete distribution over the set $\mathcal{O}$. In practical and realistic situations, generalized Gaussian or Lognormal distributions are frequently used with an output set $\mathcal{O}$ of gray level values of cardinality $K = 256$.

The marginal distribution of the main channel $P_{Y|X}$ is equivalent to one print and scan process, and consequently we have:

$$P_{Y|X} = T_m. \tag{3.1}$$

On the other hand, $P_{Z|X}$ depends on the opponent processing as he has to retrieve the original sequence before reprinting it (Figure 3.1). As mentioned above, this processing is necessary to the opponent because the industrial printers can only print dots, e.g. binary versions of the scanned code. Before publishing his fraudulent sequence z^n, the opponent undergoes then inevitable errors in the estimated binary sequence $\hat{x}^n$ of the original code.

These errors are evaluated with probabilities $P_{e,1}$ when he confuses an original white dot ($X = 1$) with a black dot ($\hat{X} = 0$) and $P_{e,0}$ when an original black dot ($X = 0$) is decoded as a white dot ($\hat{X} = 1$). Recalling that the gray level sequence observed by the opponent is governed by the same distribution law of that of the main channel $P_{Y|X}$, the opponent's decoding error probabilities $P_{e,1} = P_{\hat{X}|X}(\hat{X} = 0 \mid X = 1)$ and $P_{e,0} = P_{\hat{X}|X}(\hat{X} = 1 \mid X = 0)$ are equal to:

$$P_{e,1} = \sum_{o \in \mathcal{D}_1^c} P_{Y|X}(o \mid X = 1), \tag{3.2}$$

$$P_{e,0} = \sum_{o \in \mathcal{D}_1} P_{Y|X}(o \mid X = 0). \tag{3.3}$$

where $\mathcal{D}_1$ and $\mathcal{D}_1^c$ are optimal decision regions for decoding white and black respectively, obtained after using classical maximum likelihood decoding at the output of $P_{Y|X}$:

$$\mathcal{D}_1 = \left\{ o \in \mathcal{O} : \; P_{Y|X}(o \mid X = 1) > P_{Y|X}(o \mid X = 0) \right\}. \tag{3.4}$$

We express now entries of $P_{Z|X}$ as marginals as follow:

$$P_{Z|X}(o \mid x) = \sum_{\hat{x}=0,1} P_{\hat{X}Z|X}(\hat{x},\, o \mid x). \tag{3.5}$$

As we can see in Figure 3.1, $X \rightarrow \hat{X} \rightarrow Z$ forms a Markov chain with the relation $P_{\hat{X}Z|X}(\hat{x}, o \mid x) = P_{\hat{X}|X}(\hat{x} \mid x)T_c(o \mid \hat{x})$, where T_c is the transition matrix of the counterfeit physical device. Entries of the marginal channel matrix $P_{Z|X}$ are then:

$$P_{Z|X}(o \mid x) = \sum_{\hat{x}=0,1} P_{\hat{X}|X}(\hat{x} \mid x)T_c(o \mid \hat{x}). \tag{3.6}$$

Finally, we have

$$\begin{aligned} P_{Z|X}(o \mid X=0) = &\ (1-P_{e,0})T_c(o \mid \hat{X}=0) \\ & +P_{e,0}T_c(o \mid \hat{X}=1), \end{aligned} \tag{3.7}$$

$$\begin{aligned} P_{Z|X}(o \mid X=1) = &\ (1-P_{e,1})T_c(o \mid \hat{X}=1) \\ & +P_{e,1}T_c(o \mid \hat{X}=0). \end{aligned} \tag{3.8}$$

3.1.3 Receiver's strategies

Two strategies are possible for the receiver.

Binary thresholding:

As a first strategy, the legitimate receiver first decodes the observed sequence o^n using a maximum likelihood criterion based on the knowledge of the main channel marginal distribution $P_{Y|X}$. He then restores a binary version $\tilde{x}^n$ of the original message x^n using the decision region defined by (3.4) and naturally undergoes errors.

- When $O^n = y^n$, let $\tilde{P}_{ye,1}$ be the error probability when confusing an original white dot ($X=1$) with a black dot ($\tilde{X}=0$) and $\tilde{P}_{ye,0}$ when confusing an original black dot ($X=0$) with a white dot ($\tilde{X}=1$). As noted before, the receiver uses a maximum likelihood criterion based on the main channel marginal distribution $P_{Y|X}$. Therefore, these error probabilities are exactly the same as (3.2) and (3.3). Particularly,

$$\begin{aligned} \tilde{P}_{ye,0} &= P_{e,0}, \\ \tilde{P}_{ye,1} &= P_{e,1}. \end{aligned} \tag{3.9}$$

Let us notice that in this case, the channel $X \rightarrow \tilde{X}$ can be modeled as a Binary Input Binary Output channel (BIBO) with transition probability matrix $P_{\tilde{X}|X}$:

$$\begin{bmatrix} P_{\tilde{X}|X}\left(\tilde{X}=0 \mid X=0\right) & P_{\tilde{X}|X}\left(\tilde{X}=1 \mid X=0\right) \\ P_{\tilde{X}|X}\left(\tilde{X}=0 \mid X=1\right) & P_{\tilde{X}|X}\left(\tilde{X}=1 \mid X=1\right) \end{bmatrix} = \begin{bmatrix} 1-P_{e,0} & P_{e,0} \\ P_{e,1} & 1-P_{e,1} \end{bmatrix}. \tag{3.10}$$

- When $O^n = z^n$, let $\tilde{P}_{ze,1}$ be the error probability when confusing an original white dot ($X=1$) with a black dot ($\tilde{X}=0$) and $\tilde{P}_{ze,0}$ when confusing an original black dot ($X=0$) with a white dot ($\tilde{X}=1$). The receiver is fixed and is optimized with respect to the marginal distribution $P_{Y|X}$. Thus, making uses of (3.7) and (3.8), we express $\tilde{P}_{ze,1}$ and $\tilde{P}_{ze,0}$ as follows:

$$\tilde{P}_{ze,1} = \sum_{o \in \mathcal{D}_1^c} P_{Z|X}(o \mid X = 1), \tag{3.11}$$

$$\begin{aligned} \tilde{P}_{ze,1} &= \sum_{o \in \mathcal{D}_1^c} (1 - P_{e,1}) T_c(o \mid \hat{X} = 1) \\ &+ P_{e,1} T_c(o \mid \hat{X} = 0). \end{aligned}$$

By setting

$$\begin{aligned} P'_{e,1} &= \sum_{o \in \mathcal{D}_1^c} T_c(o \mid \hat{X} = 1), \\ P'_{e,0} &= \sum_{o \in \mathcal{D}_1} T_c(o \mid \hat{X} = 0), \end{aligned} \tag{3.12}$$

we have

$$\tilde{P}_{ze,1} = (1 - P_{e,1}) P'_{e,1} + P_{e,1}(1 - P'_{e,0}). \tag{3.13}$$

The same development yields:

$$\tilde{P}_{ze,0} = (1 - P_{e,0}) P'_{e,0} + P_{e,0}(1 - P'_{e,1}). \tag{3.14}$$

For this first strategy, the channel $X \to \tilde{X}$ can be modeled as two BIBO channels, where the marginal transition probability matrix of the main BIBO channel is given by (3.10), and the marginal transition probability matrix of the opponent channel is given by the following cascaded BIBOs:

$$\begin{bmatrix} P_{\tilde{X}|X}\left(\tilde{X} = 0 \mid X = 0\right) & P_{\tilde{X}|X}\left(\tilde{X} = 1 \mid X = 0\right) \\ P_{\tilde{X}|X}\left(\tilde{X} = 0 \mid X = 1\right) & P_{\tilde{X}|X}\left(\tilde{X} = 1 \mid X = 1\right) \end{bmatrix} = \begin{bmatrix} 1 - \tilde{P}_{ze,0} & \tilde{P}_{ze,0} \\ \tilde{P}_{ze,1} & 1 - \tilde{P}_{ze,1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 - P_{e,0} & P_{e,0} \\ P_{e,1} & 1 - P_{e,1} \end{bmatrix} \times \begin{bmatrix} 1 - P'_{e,0} & P'_{e,0} \\ P'_{e,1} & 1 - P'_{e,1} \end{bmatrix}. \tag{3.15}$$

As we will see in the next section, the test that the receiver will perform to decide whether the observed decoded sequence $\tilde{x}^n$ comes from the legitimate source or not is tantamount to counting the number of errors.

Authentication based on gray level observations:

In the second strategy, the receiver performs his test directly on the received sequence o^n (gray level version) without any given decoding. We will see in the next section that this strategy performs better than the previous one for authentication.

3.2 Authentication using hypothesis testing

For a given input sequence $(x_1, ..., x_n)$, we consider here testing whether an observed sequence $(o_1, ..., o_n \mid x_1, ..., x_n)$ is generated from a given distribution $P_{Y|X}$ or if it comes from an alternative hypothesis associated to distribution $P_{Z|X}$, where o_i belongs to a discrete finite set $\mathcal{O}$. Practically, we are interested in performing authentication after observing a sequence of n samples $(o_i|x_i)$, attesting whereas this sequence comes from a legitimate source or from a counterfeiter. The receiver establishes then a decision based on a predefined test statistic, and assigns one of the two hypothesis H_0 or H_1 corresponding respectively to each of the aforementioned cases. According to this test, the space $\mathcal{O}^n$ will be partitioned into two regions $\mathcal{H}_0$ and $\mathcal{H}_1$. Accepting hypothesis H_0 while it is actually a fake (the observed n-samples sequence belongs to $\mathcal{H}_0$ while H_1 is true) leads to an error of type II having probability β_n. Rejecting hypothesis H_0 while actually the observed sequence comes from the legitimate source (the observed n-samples sequence belongs to $\mathcal{H}_1$ while H_0 is true) leads to an error of type I with probability α_n. It is desirable to find a test with a minimal probability β_n for a fixed or prescribed probability of type I. An optimal decision rule will be given by the Neyman-Pearson criterion. The eponymous theorem states that under the constraint $\alpha_n \leq \alpha^*$, β_n is minimized if only if the following log-likelihood test infers the choice of H_1:

$$L\left(o^n \mid x^n\right) = \log \frac{P^n(o^n \mid x^n, H_1)}{P^n(o^n \mid x^n, H_0)} \geq \gamma, \tag{3.16}$$

where γ is a threshold verifying the constraint $\alpha_n \leq \alpha^*$. In the next two subsections, the expressions of the log-likelihood ratio corresponding to the two possible strategies of the receiver will be given in detail.

3.2.1 Authentication via binary thresholding

In the first strategy, the final observed data is $\tilde{x}^n$ and the original sequence x^n is a side information containing two types of data ("0" and "1"). The conditional distribution laws of each random component $(\tilde{X}_i \mid x_i)$ are identical given type x. . Let $\mathcal{N}_0 = \{i : x_i = 0\}$ and $\mathcal{N}_1 = \{i : x_i = 1\}$ with $n_0 = |\mathcal{N}_0|$ and $n_1 = |\mathcal{N}_1|$. Because they are i.i.d. sequences, under hypothesis H_j, $j \in \{0, 1\}$ we have:

$$\begin{aligned} P^n(\tilde{x}^n \mid x^n,\, H_j) &= \prod_{i=1}^{n} P(\tilde{x}_i \mid x_i,\, H_j) \\ &= \prod_{i \in \mathcal{N}_0} P(\tilde{x}_i \mid X_i = 0,\, H_j) \times \prod_{i \in \mathcal{N}_1} P(\tilde{x}_i \mid X_i = 1,\, H_j). \end{aligned}$$

- Under hypothesis H_0 the channel $X \to \tilde{X}$ has a transition matrix given by (3.10):

$$P^n(\tilde{x}^n \mid x,\ H_0) = (P_{e,0})^{n_{e,0}}(1 - P_{e,0})^{n_0 - n_{e,0}} \times (P_{e,1})^{n_{e,1}}(1 - P_{e,1})^{n_1 - n_{e,1}},$$

where $n_{e,0}$ and $n_{e,1}$ are the number of errors ($\tilde{x}_i \neq x_i$) when black is decoded into white and when white is decoded into black respectively.

- Under hypothesis H_1, the channel $X \to \tilde{X}$ has a transition matrix given by (3.15) and we have:

$$P^n(\tilde{x}^n \mid x^n,\ H_1) = (\tilde{P}_{ze,0})^{n_{e,0}}(1 - \tilde{P}_{ze,0})^{n_0 - n_{e,0}} \times (\tilde{P}_{ze,1})^{n_{e,1}}(1 - \tilde{P}_{ze,1})^{n_1 - n_{e,1}}.$$

Applying now the Neyman-Pearson criterion, the log-likelihood ratio in (3.16) is rewritten as:

$$L_1\left(\tilde{x}^n \mid x^n\right) = \log \frac{P^n\left(\tilde{x}^n \mid x^n,\ H_1\right)}{P^n\left(\tilde{x}^n \mid x^n,\ H_0\right)} \underset{H0}{\overset{H1}{\gtrless}} \gamma, \tag{3.17}$$

$$L_1\left(\tilde{x}^n \mid x^n\right) = n_{e,0} \log \left(\frac{\tilde{P}_{ze,0}(1 - P_{e,0})}{P_{e,0}(1 - \tilde{P}_{ze,0})} \right) + n_{e,1} \log \left(\frac{\tilde{P}_{ze,1}(1 - P_{e,0})}{P_{e,0}(1 - \tilde{P}_{ze,1})} \right) \underset{H0}{\overset{H1}{\gtrless}} \lambda_1. \tag{3.18}$$

where $\lambda_1 = \gamma - n_{e,0} \log \left(\frac{1-\tilde{P}_{ze,0}}{1-P_{e,0}} \right) - n_{e,1} \log \left(\frac{1-\tilde{P}_{ze,1}}{1-P_{e,1}} \right)$ and L_1 stands for the log-likelihood ratio in the first strategy using binary thresholding.

It is remarkable that this expression of the test has the practical advantage to only count the number of errors in order to perform authentication. On can observe this more easily when channels are symmetric, as we will have $\tilde{P}_{ze,0} = \tilde{P}_{ze,1}$ and $P_{e,0} = P_{e,1}$. The test will resume in:

$$n_{e,0} + n_{e,1} \underset{H0}{\overset{H1}{\gtrless}} \eta \tag{3.19}$$

The simplicity of this test eases its implementation but at a cost of a loss of optimality. In the next subsection, we discuss the non thresholding strategy which offers better authentication performance.

3.2.2 Authentication via gray level observation

In the second strategy, the observed data is o^n. Here again, the conditional distributions of each random component $(O_i \mid x_i)$, $1 \leq i \leq n$, are the same for a given type x. The Neyman Pearson test is expressed as:

$$L_2\left(o^n \mid x^n\right) = \log \frac{P^n\left(o^n \mid x^n, H_1\right)}{P^n\left(o^n \mid x^n, H_0\right)} \underset{H0}{\overset{H1}{\gtrless}} \lambda_2, \tag{3.20}$$

Using 3.7 and 3.8 we have:

$$L_2\left(o^n \mid x^n\right)=\sum_{i\in\mathcal{N}_0}\log\frac{P_{Z|X}(o_i \mid X_i=0)}{P_{Y|X}(o_i \mid X_i=0)}+\sum_{i\in\mathcal{N}_1}\log\frac{P_{Z|X}(o_i \mid X_i=1)}{P_{Y|X}(o_i \mid X_i=1)}\overset{H1}{\underset{H0}{\gtrless}}\lambda_2, \tag{3.21}$$

$$\begin{aligned}
&= \sum_{i\in\mathcal{N}_0}\log\left((1-P_{e,1})\frac{T_c(o_i \mid \hat{X}_i=0)}{T_m(o_i \mid X_i=0)}+P_{e,1}\frac{T_c(o_i \mid \hat{X}_i=1)}{T_m(o_i \mid X_i=0)}\right)+\\
&\quad+\sum_{i\in\mathcal{N}_1}\log\left((1-P_{e,0})\frac{T_c(o_i \mid \hat{X}_i=1)}{T_m(o_i \mid X_i=1)}+P_{e,0}\frac{T_c(o_i \mid \hat{X}_i=0)}{T_m(o_i \mid X_i=1)}\right)\\
&\overset{H1}{\underset{H0}{\gtrless}}\ \lambda_2.
\end{aligned} \tag{3.22}$$

It is noted that the expressions of the transition matrices modeling the physical processes T_m and T_c are required in order to perform the optimal test. Knowing the distribution of the counterfeit channel may be doable as we may know several counterfeit products somehow and then use it to estimate the model printer of the opponent.

3.3 Reliable performance evaluation

In the previous section we have expressed the Neyman-Pearson test for the two proposed strategies resumed by (3.18) and (3.22). These tests may then be practically performed on the observed sequence in order to make a decision about its authenticity. We aim now at expressing the error probabilities of type I and II for each of the two possible strategies described previously. Let $m = 1,\ 2$ be the index denoting the strategy, a straightforward calculation gives

$$\alpha\left(m\right)=\sum_{l>\lambda_m}P_{L_m}(l \mid H_0), \tag{3.23}$$

$$\beta\left(m\right)=\sum_{l<\lambda_m}P_{L_m}(l \mid H_1). \tag{3.24}$$

where $P_{L_m}(l \mid H_j)$ is the distribution of the log-likelihood ratio L_m under hypothesis H_j.

3.3.1 The Gaussian approximation

As the length n of the sequence is generally large, we can use the central limit theorem to study the distributions P_{L_m}, $m = 1,\ 2$.

- For the binary thresholding strategy, $n_{e,1}$ and $n_{e,0}$ in (3.18) are binomial random variables with parameters depending on the source of the observed sequence. Let n_x stand for the number of data of type x (0 or 1) in the original code and $Q_{e,x}$ the cross over probabilities emerging from type x in the BIBO channels ($Q_{e,x} = P_{e,x}$ in (3.10)) or ($Q_{e,x} = \tilde{P}_{ze,x}$ in (3.15)). When n is large enough, the distribution of the random variable $\frac{n_{ex}-n_xQ_{e,x}}{\sqrt{n_xQ_{e,x}(1-Q_{e,x})}}$ can be approximated to the standard normal distribution according to the central limit theorem in his historical form stated by the De Moivre-Laplace theorem. We have then:

  $$n_{e,x} \sim \mathcal{N}(n_xQ_{e,x},\ n_xQ_{e,x}(1-Q_{e,x})). \tag{3.25}$$

  Combining with (3.18), L_1 is then a weighted sum of Gaussian random variables and one can obviously deduce the parameters of the normal approximation describing the log-likelihood L_1.

- For the second strategy, i.e. when the receiver tests directly the observed gray level sequence, the log-likelihood L_2 in (3.22) may be expressed as two sums of i.i.d. and becomes:

  $$\begin{aligned} L_2\left(o^n \mid x^n\right) &= \sum_{i\in\mathcal{N}_1}\ell(o_i \mid 1) + \sum_{i\in\mathcal{N}_0}\ell(o_i \mid 0) \underset{H0}{\overset{H1}{\gtrless}} \lambda_2, \\ &= L_2^1(o^n) + L_2^0(o^n). \end{aligned} \tag{3.26}$$

  where $\ell(o \mid x)$ is a function $\ell :\ \mathcal{O} \times \mathcal{X} \to\mathbb{R}$ having some distribution with mean and variance equal to:

  $$m_{x|j} = \mathbb{E}[\ell(O \mid x) \mid H_j] = \sum_{o\in\mathcal{O}}\ell(o \mid x)P_{O|X}(o \mid x, H_j), \tag{3.27}$$

  and

  $$\text{var}[\ell(O \mid x) \mid H_j] = \sum_{o\in\mathcal{O}}(\ell(o \mid x) - m_x)^2P_{O|X}(o \mid x, H_j), \tag{3.28}$$

  here $P_{O|X} = P_{Y|X}$ (respectively $P_{O|X} = P_{Z|X}$) for $j = 0$ (respectively 1). The central limit theorem is invoked again to state that the distribution of $\frac{L_2^x(o^n)-N_xm_x}{\sqrt{N_x\text{var}[\ell(O|x)|H_j]}}$ can be approximated to the standard normal distribution and thus L_2 to a Gaussian distribution whose parameters can easily be derived from (3.27) and (3.28) to compute type I and type II error probabilities.

3.3.2 Asymptotic expression

In the previous subsection, we have proposed Gaussian approximation to compute both probability of false alarm $\alpha(m)$ and probability of non-detection $\beta(m)$. It is unfortunate that the Gaussian approximation provides inaccurate error probabilities when the threshold λ_m in (3.23) and (3.24) is far from the mean of the random variable L_m. Sanov's theorem which is indeed the large deviation principle [24] is preferred in this context as very small error probabilities of type I and II may be desired. We aim now at computing asymptotically the error probabilities of type I and II. As the following arguments are the same for both strategies discussed above, we thus focus on computing these errors for the second strategy and drop the subscribe m which stands for the strategy for the sake of simplicity. As mentioned above, for the first strategy, when the channels are symmetric, the test will be resumed as (3.19) and we just simply count the number of errors to perform authentication. If one use the following arguments, he can have similar asymptotic bounds on the two probabilities of error which were discussed in the work of Beekhof et al. [10].

We recall that the log-likelihood test is

$$L\left(O^n \mid x^n\right) = \log \frac{P^n\left(O^n \mid x^n,\, H_1\right)}{P^n\left(O^n \mid x^n,\, H_0\right)} \underset{H0}{\overset{H1}{\gtrless}} \lambda, \tag{3.29}$$

and the two types of error probabilities are

$$\alpha_n = \Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right), \tag{3.30}$$

$$\beta_n = \Pr\left(L\left(O^n \mid x^n\right) \leq \lambda \mid H_1\right). \tag{3.31}$$

It is of our interest to approximate α_n and β_n by using arguments based on Sanov's theorem. Recalling from chapter 2 the definition of a type of an outcome of random sequence, the log-likelihood ratio L can be alternatively written as follows:

$$\begin{aligned}
L\left(O^n \mid x^n\right) &= \sum_{i=1}^{n} \log \tfrac{P_{Z|X}(O_i|x_i)}{P_{Y|X}(O_i|x_i)} \\
&= \sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}} n\left(o, a \mid O^n, x^n\right) \log \tfrac{P_{Z|X}(o|a)}{P_{Y|X}(o|a)} \\
&= \sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}} n\hat{P}_{O^n x^n}\left(o, a\right) \log \tfrac{P_{Z|X}(o|a)}{P_{Y|X}(o|a)} \\
&= \sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}} n\hat{P}_{O^n|x^n}\left(o \mid a\right) \hat{P}_{x^n}\left(a\right) \log \tfrac{P_{Z|X}(o|a)\hat{P}_{x^n}(a)}{P_{Y|X}(o|a)\hat{P}_{x^n}(a)} \\
&= \sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}} n\hat{P}_{O^n x^n}\left(o, a\right) \log \tfrac{\hat{P}_{o^n|x^n}(o|a)\hat{P}_{x^n}(a)}{P_{Y|X}(o|a)\hat{P}_{x^n}(a)} -
\end{aligned}$$

$$
\begin{aligned}
&-\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}} n\hat{P}_{O^n x^n}(o,a)\log\frac{\hat{P}_{o^n|x^n}(o|a)\hat{P}_{x^n}(a)}{P_{Z|X}(o|a)\hat{P}_{x^n}(a)}\\
&= n\left[\mathbb{D}\left(\hat{P}_{O^n x^n}\parallel P_{Y|X}\hat{P}_{x^n}\right)-\mathbb{D}\left(\hat{P}_{O^n x^n}\parallel P_{Z|X}\hat{P}_{x^n}\right)\right],
\end{aligned}
$$

It should be noted that $\mathbb{D}\left(\hat{P}_{O^n x^n}\parallel P_{Y|X}\hat{P}_{x^n}\right)$ is also a random element taking values in the following set:

$$
\left\{D\left(\hat{P}_{o^n x^n}\parallel P_{Y|X}\hat{P}_{x^n}\right):\ \hat{P}_{o^n x^n}\in\mathcal{P}_n\left(\mathcal{O}\times\mathcal{X}\right)\right\}.
$$

Then the log-likelihood test (3.29) becomes

$$
\mathbb{D}\left(\hat{P}_{O^n x^n}\parallel P_{Y|X}\hat{P}_{x^n}\right)-\mathbb{D}\left(\hat{P}_{O^n x^n}\parallel P_{Z|X}\hat{P}_{x^n}\right)\underset{H0}{\overset{H1}{\gtrless}}\frac{\lambda}{n}. \tag{3.32}
$$

Denote by $E^{'}_{x^n,H_1}$ the region on which the hypothesis H_1 is accepted when the transmitted message is x^n,

$$
E^{'}_{x^n,H_1}=\left\{P_{O|X}\hat{P}_{x^n}\in\mathcal{P}\left(\mathcal{O}\times\mathcal{X}\right):\ D\left(P_{O|X}\hat{P}_{x^n}\parallel P_{Y|X}\hat{P}_{x^n}\right)-D\left(P_{O|X}\hat{P}_{x^n}\parallel P_{Z|X}\hat{P}_{x^n}\right)\geq\frac{\lambda}{n}\right\}, \tag{3.33}
$$

where $\mathcal{P}\left(\mathcal{O}\times\mathcal{X}\right)$ is the set of all joint distributions.

The test (3.32) partitions the probability simplex into two regions $E^{'}_{x^n,H_1}$ and $E^{'c}_{x^n,H_1}$. When $\hat{P}_{o^n x^n}$, the empirical distribution of the observed sequence (o^n,x^n) belongs to the set $E^{'}_{x^n,H_1}$, we decide that the hypothesis H_1 is true, i.e. $\hat{P}_{o^n x^n}$ is governed by the distribution $P_{Z|X}\hat{P}_{x^n}$. Similarly, when $\hat{P}_{o^n x^n}$ belongs to the set $E^{'c}_{x^n,H_1}$, we decide that the hypothesis H_0 is true, i.e. $\hat{P}_{o^n x^n}$ is governed by the distribution $P_{Y|X}\hat{P}_{x^n}$. This means that $P_{Y|X}\hat{P}_{x^n}$ and $P_{Z|X}\hat{P}_{x^n}$ respectively belong to different sets $E^{'c}_{x^n,H_1}$ and $E^{'}_{x^n,H_1}$. Now Sanov's theorem will help us to compute the error probabilities of type I and II of the test (3.32). The probability of type I in (3.30) can then be alternatively expressed as follows

$$
\alpha_n=\Pr\left(L\left(O^n\mid x^n\right)\geq\lambda\mid H_0\right)=\Pr\left(\hat{P}_{O^n x^n}\in E^{'}_{x^n,H_1}\mid H_0\right). \tag{3.34}
$$

Now it is easy to check that $E^{'}_{x^n,H_1}$ is a close convex subset of the probability simplex $\mathcal{P}\left(\mathcal{O}\times\mathcal{X}\right)$ so that $E^{'}_{x^n,H_1}\subseteq\overline{\overset{\circ}{E}}'_{x^n,H_1}$ (see Proposition 2.16). Applying Sanov's theorem we have:

$$
\alpha_n\doteq\exp\left(-nD\left(P^{*}_{OX}\parallel P_{Y|X}\hat{P}_{x^n}\right)\right), \tag{3.35}
$$

where

$$P^*_{OX} = \arg\min_{P \in E'_{x^n,H_1}} D\left(P \parallel P_{Y|X}\hat{P}_{x^n}\right).$$

Let

$$E_{x^n,H_1} = \left\{P_{O|X} \in \mathcal{P}\left(\mathcal{O} \mid \mathcal{X}\right): \ D\left(P_{O|X} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_{O|X} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \geq \frac{\lambda}{n}\right\}, \tag{3.36}$$

where $\mathcal{P}\left(\mathcal{O} \mid \mathcal{X}\right)$ is the set of all stochastic matrices $P_{O|X} : \mathcal{X} \to \mathcal{O}$.

Then it is easy to check that (3.35) can be rewritten as follows

$$\alpha_n \doteq \exp\left(-nD\left(P^*_0 \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.37}$$

where

$$P^*_0 = \arg\min_{P \in E_{x^n,H_1}} D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right). \tag{3.38}$$

Similarly, we have

$$\beta_n \doteq \exp\left(-nD\left(P^*_1 \parallel P_{Z|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.39}$$

where $\hat{P}^*_1$ is the distribution such that

$$P^*_1 = \arg\min_{P \in \overline{E^c_{x^n,H_1}}} D\left(P \parallel P_{Z|X} \mid \hat{P}_{x^n}\right). \tag{3.40}$$

The following proposition provides the expression of P^*_0 and P^*_1 which are actually the same. In order to derive these expressions, we will solve a convex optimization problem with inequality constraints (3.33). However, thanks to specific constraints in the studied case, we show that the problem achieves minimum on the boundary. In the following, we give a proof with complimentary details which are omitted in the proof of Cover and Thomas [19]

Proposition 3.1. *The probabilities of type I and type II are asymptotically expressed as follows*

$$\alpha_n \doteq \exp\left(-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.41}$$

and

$$\beta_n \doteq \exp\left(-nD\left(P^*_s \parallel P_{Z|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.42}$$

where P^*_s is given in terms of parameter $1 \geq s \geq 0$ as

$$P^*_s\left(o \mid a\right) = \frac{P^{1-s}_{Y|X}\left(o \mid a\right) P^s_{Z|X}\left(o \mid a\right)}{\sum\limits_{o'} P^{1-s}_{Y|X}\left(o' \mid a\right) P^s_{Z|X}\left(o' \mid a\right)}, \tag{3.43}$$

and s is chosen so that $D\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_s^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) = \dfrac{\lambda}{n}$.

Proof. Instead of solving the problem $\min\limits_{P \in E_{x^n,H_1}} D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)$, we consider the following optimization problem

$$\begin{array}{ll} \text{minimize} & f(P), \\ & \\ \text{subject to} & g(P) \leq 0 \\ & h(P) = 0 \\ & P \in \mathcal{P}(\mathcal{O}|\mathcal{X}) \subseteq \mathbb{R}^{|\mathcal{X}|\times|\mathcal{O}|}, \end{array} \tag{3.44}$$

where

$$f(P) = D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)$$

$$g(P) = -D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) + D\left(P \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) + \frac{\lambda}{n} \tag{3.45}$$

$$h(P) = \sum_{a\in\mathcal{X}} \sum_{o\in\mathcal{O}} \hat{P}_{x^n}(a) P(o \mid a) - 1.$$

The *Lagrangian* $F : \mathbb{R}^{|\mathcal{X}|\times|\mathcal{O}|} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ associated with the problem (3.44) is defined as follows:

$$F(P, s, \eta) = f(P) + sg(P) + \eta h(P). \tag{3.46}$$

where $s \geq 0$. It is easy to check that $D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)$ and $g(P)$ are convex functions on $\mathcal{P}(\mathcal{O}|\mathcal{X})$ and $h(P)$ is linear. Therefore, KKT conditions are necessary and sufficient to find an optimum (see Appendix 3.6.3 and chapter 5 in [14] for more details). Hence P^* is a minimum of the problem (3.44) if and only if the set of unique points$\left(P^*, \quad s^*, \quad \eta^* \right)$ satisfy:

$$\begin{aligned} g(P^*) &\leq 0\ ;\ h(P^*) = 0 \\ s^* &\geq 0 \\ s^* g(P^*) &= 0 \\ \frac{\partial F}{\partial P}(P^*, s^*, \eta^*) &= 0 \end{aligned}$$

We now show that the problem (3.44) achieves minimum on the boundary, i.e. $g(P^*) = 0$. In order to do so, we will show that $s^* > 0$ by first assuming that $s^* = 0$, and then pointing out the contradiction. Because P^* satisfies the KKT conditions then we have:

$$\frac{\partial F\left(P^*\left(o\mid x\right),s^*,\eta^*\right)}{\partial P\left(o\mid x\right)} = 0$$

$$\frac{\partial f\left(P^*\left(o\mid x\right)\right)}{\partial P\left(o\mid x\right)}+s^*\frac{\partial g\left(P^*\left(o\mid x\right)\right)}{\partial P\left(o\mid x\right)}+\eta^*\frac{\partial h\left(P^*\left(o\mid x\right)\right)}{\partial P\left(o\mid x\right)}=0$$

Assuming that $s^*=0$ it comes:

$$\begin{aligned}\frac{\partial f\left(P^*\left(o\mid x\right)\right)}{\partial P\left(o\mid x\right)}+\eta^*\frac{\partial h\left(P^*\left(o\mid x\right)\right)}{\partial P\left(o\mid x\right)} &= 0\\ \log\frac{P^*\left(o\mid x\right)}{P_{Y|X}\left(o\mid x\right)}+1+\eta^* &= 0 \qquad (3.47)\end{aligned}$$

Solving the last equation (3.47) and using the fact that $\sum_{o\in\mathcal{O}}P\left(o\mid x\right)=1$ for all $x\in\mathcal{X}$, we find that $P^*\left(o\mid x\right)=P_{Y|X}\left(o\mid x\right).$

It turns out that $P_{Y|X}$ is a minimum of the problem (3.44). It means that $P_{Y|X}\in E_{x^n,H_1}$which is contrary to the aforementioned fact that $P_{Y|X}$ and $P_{Z|X}$ respectively belong to different sets $E^c_{x^n,H_1}$ and E_{x^n,H_1}. Hence, the problem (3.44) has its minimum on the boundary of the inequality constraint. In consequence we just need to consider the following problem:

$$\begin{aligned}&\text{minimize} && f\left(P\right)\\ &\text{subject to} && g\left(P\right)=0 \qquad (3.48)\\ &&& h\left(P\right)=0\\ &&& P\in\mathcal{P}\left(\mathcal{O}|\mathcal{X}\right)\subseteq\mathbb{R}^{|\mathcal{X}|\times|\mathcal{O}|}\end{aligned}$$

We develop now the Lagrangian as follows:

$$\begin{aligned}F\left(P,s,\eta\right) &= f\left(P\right)+sg\left(P\right)+\eta h\left(P\right)\\ &= \sum_{a\in\mathcal{X}}\hat{P}_{x^n}\left(a\right)\sum_{o\in\mathcal{O}}P\left(o\mid a\right)\log\tfrac{P(o|a)}{P_{Y|X}(o|x)}-\\ &\quad -s\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}P\left(o\mid a\right)\hat{P}_{x^n}\left(a\right)\log\tfrac{P_{Z|X}(o|a)}{P_{Y|X}(o|a)}+ \qquad (3.49)\\ &\quad +s\frac{\lambda}{n}+\eta\left(\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}\hat{P}_{x^n}\left(a\right)P\left(o\mid a\right)-1\right)\end{aligned}$$

Differentiating F with respect to $P\left(o\mid a\right)$ and setting to 0, we have

$$\log\frac{P\left(o\mid a\right)}{P_{Y|X}\left(o\mid a\right)}+1-s\log\frac{P_{Z|X}\left(o\mid a\right)}{P_{Y|X}\left(o\mid a\right)}+\eta=0, \qquad (3.50)$$

for all $o\in\mathcal{O}$, $a\in\mathcal{X}$.

Solving this set of equations we get the minimum of the problem (3.44) as follows:

$$P_0^* = P_s^*\left(o \mid a\right) = \frac{P_{Y|X}^{1-s}\left(o \mid a\right) P_{Z|X}^{s}\left(o \mid a\right)}{\sum\limits_{o'} P_{Y|X}^{1-s}\left(o' \mid a\right) P_{Z|X}^{s}\left(o' \mid a\right)}. \tag{3.51}$$

for all $o \in \mathcal{O}$, $a \in \mathcal{X}$ and s is chosen so that:

$$D\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_s^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) = \frac{\lambda}{n}. \tag{3.52}$$

We recall that, according to the Neyman-Pearson criterion, λ is chosen to satisfy a predefined level test $\alpha \leq \alpha^*$. Similarly, we also have

$$P_1^* = \frac{P_{Z|X}^{1-s'}\left(o \mid a\right) P_{Y|X}^{s'}\left(o \mid a\right)}{\sum\limits_{o'} P_{Z|X}^{1-s'}\left(o' \mid a\right) P_{Y|X}^{s'}\left(o' \mid a\right)}. \tag{3.53}$$

Let $t = 1 - s'$, we have:

$$P_1^* = P_t^*\left(o \mid a\right) = \frac{P_{Y|X}^{1-t}\left(o \mid a\right) P_{Z|X}^{t}\left(o \mid a\right)}{\sum\limits_{o'} P_{Y|X}^{1-t}\left(o' \mid a\right) P_{Z|X}^{t}\left(o' \mid a\right)}. \tag{3.54}$$

Because $P_t^*\left(o \mid a\right)$ and $P_s^*\left(o \mid a\right)$ satisfy both (3.52) and that the solution is unique, we have $P_0^* = P_1^* = P_s^*$, where $0 \leq s \leq 1$. □

It should be noted that when $s \to 1$, $P_s^* \to P_{Z|X}$ and as $s \to 0$, $P_s^* \to P_{Y|X}$. Moreover, from (3.52) the threshold λ may be expressed in term of s . Thus from these two previous remarks, λ may be tuned between

$$\left[-nD\left(P_{Y|X} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \quad nD\left(P_{Z|X} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right]$$

so that the problem of hypothesis testing is meaningful. Figure 3.2 illustrates the geometry of P_s^* in the space $\mathcal{P}\left(\mathcal{O} \mid \mathcal{X}\right)$.

Summarizing the results, for P_s^* of the form as in (3.51) it is therefore:

$$\begin{aligned} \alpha_n &\doteq \exp\left(-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \\ \beta_n &\doteq \exp\left(-nD\left(P_s^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right)\right). \end{aligned} \tag{3.55}$$

So far we have provided asymptotic expressions of the probabilities of type I and II errors for n sufficiently large. In the next subsection we establish tighter expressions for finite n which naturally coincide with (3.55) for n going to infinity.

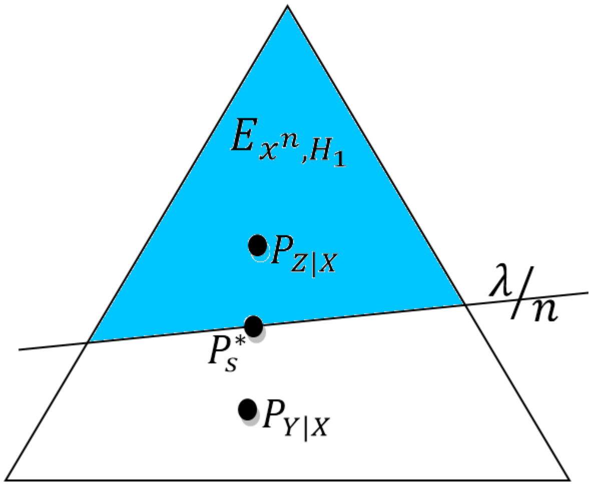


Figure 3.2: Geometry of P_s^*

3.3.3 Refinement of asymptotic expressions

The refinement suggested here is possible principally because the log-likelihood ratio $L\left(O^n \mid x^n\right)$ used as a test statistic in (3.29) is a sum of n i.i.d. random variables $l\left(O_i \mid x_i\right) \triangleq \log \frac{P_{Z|X}(O_i|x_i)}{P_{Y|X}(O_i|x_i)}$. When $O \mid x$ is governed by $P_{Y|X}$ the semi-invariant moment generating function of the random variable $l\left(O \mid x\right)$ is:

$$\mu_l\left(s;\, x, H_0\right) = \log \mathbb{E}_{P_{Y|X}}\left[e^{sl(O|x)}\right] = \log \mathbb{E}_{P_{Y|X}}\left[\frac{P^s_{Z|X}\left(O \mid x\right)}{P^s_{Y|X}\left(O \mid x\right)}\right]. \tag{3.56}$$

Then $\mu_L\left(s;\, x^n, H_0\right)$ can be expressed in terms of $\mu_l\left(s;\, a, H_0\right)$, with $a \in \mathcal{X}$ as follows:

$$\begin{aligned}
\mu_L\left(s;\, x^n, H_0\right) &= \log \mathbb{E}_{P^n_{Y|X}}\left[e^{sL(O^n|x^n)}\right],\\
&= \log \prod_{i=1}^{n} \mathbb{E}_{P_{Y|X}}\left[e^{sl(O_i|x_i)}\right],\\
&= \sum_{i=1}^{n} \log \mathbb{E}_{P_{Y|X}}\left[e^{sl(O_i|x_i)}\right],\\
&= \sum_{b\in\mathcal{O},a\in\mathcal{X}} n\left(b, a \mid o^n, x^n\right) \log \mathbb{E}_{P_{Y|X}}\left[e^{sl(O|a)}\right],\\
&= \sum_{b\in\mathcal{O},a\in\mathcal{X}} n\left(b, a \mid o^n, x^n\right) \mu_l\left(s;\, a, H_0\right),\\
&= \sum_{a\in\mathcal{X}} \mu_l\left(s;\, a, H_0\right) \sum_{b\in\mathcal{O}} n\left(b, a \mid o^n, x^n\right).
\end{aligned}$$

and finally:

$$\mu_L\left(s;\, x^n, H_0\right) \;=\; \sum_{a\in\mathcal{X}}\mu_l\left(s;\, a, H_0\right) n(a \mid x^n).$$

When the threshold is very far from the expected value of $L\left(O^n \mid x^n\right)$, i.e.:

$$\lambda \gg \mathbb{E}_{P_L}\left[L\left(O^n \mid x^n\right)\right]$$

the probability of the tail $\Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right)$ is very small thereby being difficult to compute. Using a tilted distribution $\widetilde{P}_s$ for L is a very useful and pertinent tool to evaluate tails of probabilities. However this must be done with the good choice of the parameter s, more specifically s is chosen such that $\mathbb{E}_{\widetilde{P}_{L_s}}\left[L\left(O^n \mid x^n\right)\right]$ is equal to the threshold λ. Let us firstly tilt the distribution of the channel and secondly extract the corresponding tilted distribution for L. We will naturally choose the distribution in (3.51) as a tilted channel distribution, with the parameter s chosen such that (3.52) is satisfied. Recalling that $l\left(o \mid x\right) \triangleq \log\frac{P_{Z|X}(o|x)}{P_{Y|X}(o|x)}$ we rewrite P_s^* as:

$$\begin{aligned}
P_s^*\left(o \mid x\right) &= \frac{P_{Y|X}\left(o \mid x\right)\exp\left[sl\left(o \mid x\right)\right]}{\sum\limits_{o^{'}\in\mathcal{O}} P_{Y|X}\left(o' \mid x\right)\exp\left[sl\left(o \mid x\right)\right]}, \\
&= P_{Y|X}\left(o \mid x\right)\exp\left[sl\left(o \mid x\right) - \mu_l\left(s;\, x, H_0\right)\right].
\end{aligned}$$

The tilted distribution $P_s^{*n}\left(o^n \mid x^n\right)$ will be of following form:

$$\begin{aligned}
P_s^{*n}\left(o^n \mid x^n\right) &= \prod_{i=1}^{n} P_s^*\left(o_i \mid x_i\right), \\
&= \prod_{b\in\mathcal{O}}\prod_{a\in\mathcal{X}}\left[P_s^*\left(b \mid x\right)\right]^{n(b,a|o^n,x^n)}, \\
&= \prod_{b\in\mathcal{O}}\prod_{a\in\mathcal{X}}\left\{P_{Y|X}\left(b \mid x\right)\exp\left[sl\left(b \mid x\right) - \mu_l\left(s;\, a, H_0\right)\right]\right\}^{n(b,a|o^n,x^n)}, \\
&= P_{Y|X}^n\left(o^n \mid x^n\right)\exp\left\{\sum_{b\in\mathcal{O},a\in\mathcal{X}} n\left(b, a \mid o^n, x^n\right)\left[sl\left(b \mid x\right) - \mu_l\left(s;\, a, H_0\right)\right]\right\}, \\
&= P_{Y|X}^n\left(o^n \mid x^n\right)\exp\left\{sL\left(o^n \mid x^n\right) - \sum_{a\in\mathcal{X}}\mu_l\left(s;\, a, H_0\right) n(a \mid x^n)\right\}, \\
&= P_{Y|X}^n\left(o^n \mid x^n\right)\exp\left\{sL\left(o^n \mid x^n\right) - \mu_L\left(s;\, x^n, H_0\right)\right\}.
\end{aligned}$$

Thus, $P_s^{*n}\left(o^n \mid x^n\right)$ has an equivalent form to $P_s^*\left(o \mid x\right)$. Now we turn to the tilted distribution associated to the likelihood test $L\left(O^n \mid x^n\right)$, sum of n i.i.d random variables $l\left(O_i \mid x_i\right)$. Let $\tilde{P}_s(l)$ be the resulting tilted distribution of the log-likelihood ratio $l(o \mid x)$ obtained at the output of the tilted channel. It can be expressed with respect to the original distribution $P(l)$ as:

$$\begin{aligned}
\tilde{P}_s(l) &= \sum_{o\in\mathcal{O}:l(o|x)=l} P_s^*\left(o \mid x\right), \\
&= \sum_{o\in\mathcal{O}:l(o|x)=l} P_{Y|X}\left(o \mid x\right)\exp\left[sl-\mu_l\left(s;\, x,H_0\right)\right], \\
&= \exp\left[sl-\mu_l\left(s;\, x,H_0\right)\right]\sum_{o\in\mathcal{O}:l(o|x)=l} P_{Y|X}\left(o \mid x\right), \\
&= \exp\left[sl-\mu_l\left(s;\, x,H_0\right)\right]P(l).
\end{aligned}$$

In the same way, the resulting tilted distribution of the sum $L\left(O^n \mid x^n\right)$ is then:

$$\begin{aligned}
\tilde{P}_s(L) &= \sum_{o^n\in\mathcal{O}^n:L\left(o^n|x^n\right)=L} P_s^{*n}\left(o^n \mid x^n\right), \\
&= \sum_{o^n\in\mathcal{O}^n:L\left(o^n|x^n\right)=L} P_{Y|X}^n\left(o^n \mid x^n\right)\exp\left\{sL-\mu_L\left(s;\, x^n,H_0\right)\right\}, \\
&= \exp\left\{sL-\mu_L\left(s;\, x^n,H_0\right)\right\}\sum_{o^n\in\mathcal{O}^n:L\left(o^n|x^n\right)=L} P_{Y|X}^n\left(o^n \mid x^n\right), \\
&= \exp\left\{sL-\mu_L\left(s;\, x^n,H_0\right)\right\}P(L). \qquad (3.57)
\end{aligned}$$

Proposition 3.2. *We derive now some properties related to the random variable L_s governed by the tilted distribution $\tilde{P}_s$.*

1. *The expected value of L_s according to $\tilde{P}_s$ is such that $\mathbb{E}_{\tilde{P}_S}\left[L_s\right] = \frac{d\mu_L(s;x^n,H_0)}{ds} = \mu_L'\left(s;\, x^n,H_0\right)$.*

2. *The value of s verifying (3.52) insure also that $\mu_L'\left(s;\, x^n,H_0\right)=\lambda$, where λ is the threshold of the test.*

3. *As a consequence of the two previous properties, the expected value of L_s coincide with the threshold λ.*

4. *Finally the variance of L_s can be easily expressed as the second derivative of the moment generating function:*

$$\mathrm{var}\left[L_s\right] = \frac{d\mu_L^2\left(s;\, x^n,H_0\right)}{ds^2} = \mu_L''\left(s;\, x^n,H_0\right) \qquad (3.58)$$

The probability of type I can now be computed as follows:

$$\begin{aligned}
\Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right) &= \sum_{L\geq\lambda} P(L), \\
&= \sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-sL + \mu_L\left(s;\, x^n, H_0\right)\right\}, \\
&= \exp\left\{\mu_L\left(s;\, x^n, H_0\right)\right\} \sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-sL\right\}, \\
&= \exp\left\{\mu_L\left(s;\, x^n, H_0\right) - s\mu_L'\left(s;\, x^n, H_0\right)\right\} \\
&\quad \times \sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-sL + s\mu_L'\left(s;\, x^n, H_0\right)\right\}, \\
&= \exp\left\{\mu_L\left(s;\, x^n, H_0\right) - s\lambda\right\} \\
&\quad \times \sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-s(L-\lambda)\right\}, \\
&\overset{(a)}{=} \exp\left\{-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right\} \qquad (3.59)\\
&\quad \times \sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-s(L-\lambda)\right\}.
\end{aligned}$$

(a) comes from the fact that the two exponent terms are equivalent (see appendix 3.6.2 for the proof):

$$-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) = \left(\mu_L\left(s;\, x^n, H_0\right) - s\lambda\right). \tag{3.60}$$

Since $s \geq 0$ and the sum in (3.59) holds for $L \geq \lambda$, the exponential term in this sum is ≤ 1 and consequently $\sum\limits_{L\geq\lambda} \tilde{P}(L) \exp\left\{-s(L-\lambda)\right\} \leq 1$. Hence one can express an information theoretic form of the Chernoff bound:

$$\Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right) \leq \exp\left\{-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right\}. \tag{3.61}$$

We turn to show that the probability of type I $\Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right)$ may be improved for moderate values of n with an appropriate approximation of the factor:

$$\sum_{L\geq\lambda} \tilde{P}_s(L) \exp\left\{-s(L-\lambda)\right\}. \tag{3.62}$$

Recalling property 3 in Proposition 3.2, the expected value of L_s coincide with the threshold λ. Applying the law of large number, outcomes of L_s are near λ with high probability, and the problem of large deviations changes to small deviations for L_s near its mean. The sum in (3.59) is then more significant within a small fraction of the standard deviation of L_s. Using arguments from [28] Appendix 5A, it is suggested to apply an appropriate version of the central limit theorem over this small deviations where the separation Δ_L between adjacent values of L_s becomes smaller and smaller as n gets larger. As a first approximation let $\tilde{P}_s(L) \approx q_s(L)\Delta_L(n)$. As $\Delta_L(n)$ goes to zero when n is increased, we may approximate the sum (3.62) by integrating it by parts. Taking the normalized version of the random variable L_s we have then:

$$\begin{aligned} &\sum_{L\geq\lambda} \tilde{P}_s(L) \exp\{-s(L-\lambda)\} \\ &\approx s\sqrt{\mu_L''(s;\, x^n, H_0)} \times \int_0^\infty [G(L_s') - G(0)] \exp(-s\sqrt{\mu_L''(s;\, x^n, H_0)}\, L_s') dL_s' \end{aligned} \tag{3.63}$$

where $\mu_L''(s;\, x^n, H_0) = \mathrm{var}(L_s')$ from property 4 of Proposition 3.2, $L_s' = \frac{L_s-\lambda}{\sqrt{\mu_L''(s;\, x^n, H_0)}}$ and $G(L_s')$ is the cumulative distribution of L_s' :

$$G(L_s') = \sum_{w\leq L_s'} \tilde{P}_s(w)$$

It remains now choosing a simple approximation for $G(L_s')-G(0)$ in regard to the central limit theorem where $\lim G(u) = \Phi(u)$ as n becomes larger, $\Phi(u)$ being the distribution function of the standard normal random variable. Because of small deviations near the mean, a first order of Taylor expansion may be sufficient here and we have:

$$G(L_s') - G(0) \approx \frac{L_s'}{\sqrt{2\pi}} \tag{3.64}$$

Plugging this approximation into (3.63) and completing integration, the error of type I may then be approximated by:

$$\Pr\left(L\left(O^n \mid x^n\right) \geq \lambda \mid H_0\right) \approx \frac{1}{s\sqrt{2\pi\mu_L''(s;\, x^n, H_0)}} \times \exp\left\{-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right\}. \tag{3.65}$$

3.4 Authentication with or without thresholding

In this setup and without loss of generality, we consider a Gaussian model for the physical devices T_m and T_c with variances σ^2 Figure 3.3 compares the Receiver Operating Characteristic (ROC) curves associated with the two different strategies. These error probabilities are computed using the results given in (3.65). We can notice that the gap between the two strategies is important. This is not surprising since the binary thresholding removes information about the gray level observation, yet this has a practical impact because one practitioner can be tempted to count the number of errors as an authentication score, as given in (3.18), for its easy implementation. An information theoretical analysis presented in appendix 3.6.1 shows that for a given level α_n, i.e. a given false alarm, the non-detection error exponent in case of gray level observation is greater than the corresponding exponent in case of thresholding. This result is in line with the remark of Blahut in *[13]* where in page 108 he writes that *" information is increased if a measurement is made more precise [...] (i.e. with a refinement of the set of measurement outcomes)."*

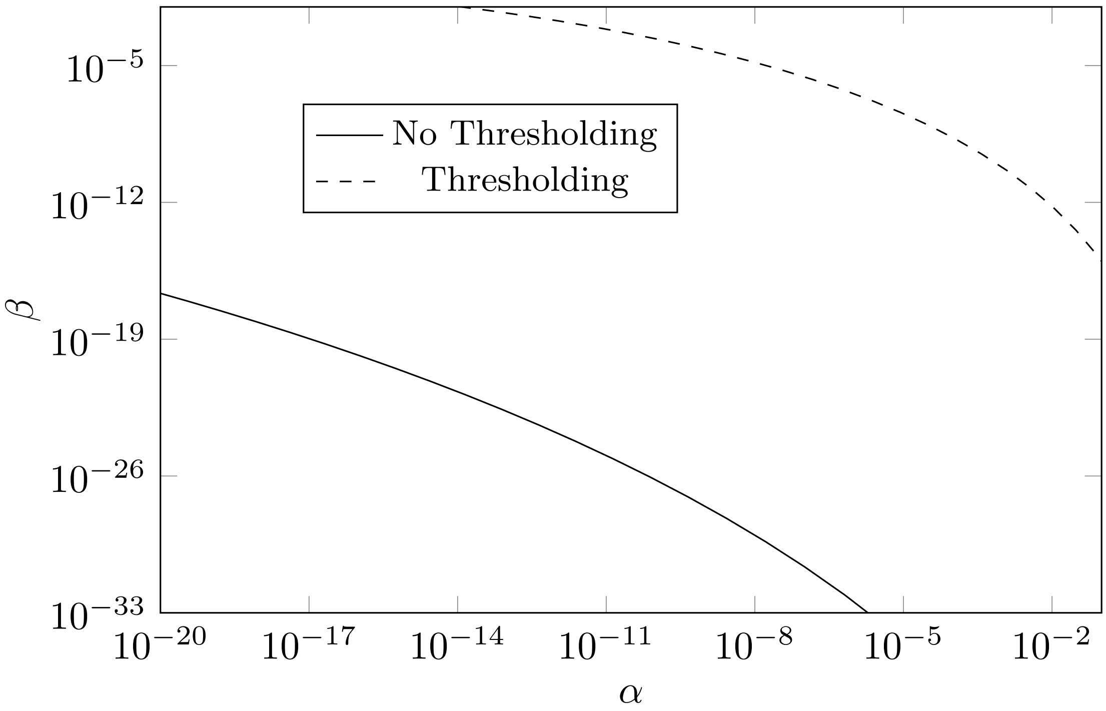


Figure 3.3: ROC curves for the two different strategies ($N_{\text{trials}} = 2000$, $\sigma = 52$). α is the probability of rejecting an authentic code and β is the probability of non-detecting an illegal copy.

3.5 Conclusions

In this chapter, we have presented a framework to analyze the performance of authentication without using channel coding. We have introduced two possible strategies for the receiver, namely binary thresholding and gray level observation. We have then used the optimal Neyman-Pearson test to perform authentication of the system. Particularly, we have computed the two types of error probabilities, false alarm and non-detection by using Gaussian approximation. More interestingly, we have also computed asymptotically these two types of error probabilities when they are small thanks to the Sanov's theorem. Relying on these asymptotic expressions, we have concluded that the gray level observation strategy offers a better performance for authentication than the former one. In the next chapter, we show how to practically compute these two types of error probabilities.

3.6 Appendix

3.6.1 Information-theoretic comparison between hypothesis testing with and without thresholding

Let us fix the false alarm probability for the two strategies and compare the two non-detection error exponents. In the gray level observation strategy, the final observation is $o^n \in \mathcal{O}^n$ which can be either y^n coming from the legitimate transmitter or z^n forged by the opponent, while in the binary thresholding strategy, the final observation is $\tilde{x}^n \in \mathcal{X}^n$ thereby being either $\tilde{x}^n_y$ from the legitimate transmitter or $\tilde{x}^n_z$ from the opponent.

- The grey level observation strategy

The hypothesis test (3.32) is recalled as

$$\mathbb{D}\left(\hat{P}_{O^n x^n} \parallel P_{Y|X}\hat{P}_{x^n}\right) - \mathbb{D}\left(\hat{P}_{O^n x^n} \parallel P_{Z|X}\hat{P}_{x^n}\right) \underset{H0}{\overset{H1}{\gtrless}} \frac{\lambda}{n}.$$

Similar to arguments in subsection 3.3.2 we have the probability of false alarm and non-detection of this strategy are asymptotically estimated as follows

$$\alpha_n \doteq \exp\left(-nD\left(P^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.66}$$

and

$$\beta_n \doteq \exp\left(-nD\left(P^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.67}$$

where

$$D\left(P^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) = \min_{P\in E_{x^n,H_1}} D\left(P \parallel P_{Y|X} \mid \hat{P}_{x^n}\right), \tag{3.68}$$

$$D\left(P^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) = \min_{P\in \overline{E^c_{x^n,H_1}}} D\left(P \parallel P_{Z|X} \mid \hat{P}_{x^n}\right), \tag{3.69}$$

and E_{x^n,H_1} the region on which the hypothesis H_1 is accepted when the transmitted message is x^n,

$$E_{x^n,H_1} = \left\{P_{O|X} \in \mathcal{P}\left(\mathcal{O} \mid \mathcal{X}\right) : \; D\left(P_{O|X} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_{O|X} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \geq \frac{\lambda}{n}\right\}.$$

- The binary thresholding strategy

Similarly, the test is

$$\mathbb{D}\left(\hat{P}_{\tilde{X}^n x^n} \parallel P_{\tilde{X}_Y|X}\hat{P}_{x^n}\right) - \mathbb{D}\left(\hat{P}_{\tilde{X}^n x^n} \parallel P_{\tilde{X}_Z|X}\hat{P}_{x^n}\right) \underset{H0}{\overset{H1}{\gtrless}} \frac{\lambda'}{n}. \tag{3.70}$$

Then we have the probability of false alarm and non-detection of this strategy are asymptotically estimated as follows

$$\alpha_n^{'} \doteq \exp\left(-nD\left(P'^{*} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.71}$$

and

$$\beta'_n \doteq \exp\left(-nD\left(P'^{*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.72}$$

where

$$D\left(P^{'*} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right) = \min_{P \in E'_{x^n,H_1}} D\left(P \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right), \tag{3.73}$$

$$D\left(P^{'*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right) = \min_{P \in \overline{E'^{c}_{x^n,H_1}}} D\left(P \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right), \tag{3.74}$$

and E'_{x^n,H_1} the region on which the hypothesis H_1 is accepted when the transmitted message is x^n,

$$E'_{x^n,H_1} = \left\{P_{\tilde{X}|X} : \; D\left(P_{\tilde{X}|X} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_{\tilde{X}|X} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right) \geq \frac{\lambda'}{n}\right\}.$$

Assume that $\alpha_n = \alpha'_n$ we will show that:

$$D\left(P^{'*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right) \leq D\left(P^{*} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right). \tag{3.75}$$

To prove (3.75), we need to define the following distribution

$$\begin{aligned} \overline{P^*}\left(1 \mid a\right) &= \sum_{o \in \mathcal{D}_1} P^*\left(o \mid a\right), \\ \overline{P^*}\left(0 \mid a\right) &= \sum_{o \in \mathcal{D}_1^c} P^*\left(o \mid a\right). \end{aligned} \tag{3.76}$$

where $a \in \mathcal{X}$ and $\mathcal{D}_1$ is defined in (3.4) which is

$$\mathcal{D}_1 = \left\{o \in \mathcal{O} : \; P_{Y|X}(o \mid X = 1) > P_{Y|X}(o \mid X = 0)\right\}.$$

Taking advantages of the distribution $\overline{P^*}$, we will show that

$$D\left(P^{*} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \geq D\left(\overline{P^*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right), \tag{3.77}$$

and

$$D\left(\overline{P^*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right) \geq D\left(P'^{*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right). \tag{3.78}$$

Lemma 3.3. *With the distribution* $\overline{P^*}$ *defined in (3.76), we have*

$$D\left(P^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \geq D\left(\overline{P^*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right). \tag{3.79}$$

Proof. We have

$$\begin{aligned}
&D\left(P^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) - D\left(\overline{P^*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right),\\
&= \hat{P}_{x^n}(0)\left\{\sum_{o\in\mathcal{O}} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)} - \sum_{\tilde{a}=0,1}\overline{P^*}(\tilde{a} \mid 0)\log\tfrac{\overline{P^*}(\tilde{a}|0)}{P_{\tilde{X}_Z|X}(\tilde{a}|0)}\right\} +\\
&+\hat{P}_{x^n}(1)\left\{\sum_{o\in\mathcal{O}} P^*(o \mid 1)\log\tfrac{P^*(o|1)}{P_{Z|X}(o|1)} - \sum_{\tilde{a}=0,1}\overline{P^*}(\tilde{a} \mid 1)\log\tfrac{\overline{P^*}(\tilde{a}|1)}{P_{\tilde{X}_Z|X}(\tilde{a}|1)}\right\},\\
&= \hat{P}_{x^n}(0) f(0) + \hat{P}_{x^n}(1) f(1).
\end{aligned}$$

We will show that $f(0) \geq 0$ and similarly $f(1) \geq 0$. We develop $f(0)$ as follows:

$$\begin{aligned}
&\sum_{o\in\mathcal{O}} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)} - \sum_{\tilde{a}=0,1}\overline{P^*}(\tilde{a} \mid 0)\log\tfrac{\overline{P^*}(\tilde{a}|0)}{P_{\tilde{X}_Z|X}(\tilde{a}|0)}\\
= \quad &\Bigg\{\sum_{o\in\mathcal{D}_1} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)} + \sum_{o\in\mathcal{D}_1^c} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)}\\
&-\overline{P^*}(0 \mid 0)\log\tfrac{\overline{P^*}(0|0)}{P_{\tilde{X}_Z|X}(0|0)} - \overline{P^*}(1 \mid 0)\log\tfrac{\overline{P^*}(1|0)}{P_{\tilde{X}_Z|X}(1|0)}\Bigg\},\\
= \quad &\Bigg\{\sum_{o\in\mathcal{D}_1} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)} + \sum_{o\in\mathcal{D}_1^c} P^*(o \mid 0)\log\tfrac{P^*(o|0)}{P_{Z|X}(o|0)}\\
&-\sum_{o\in\mathcal{D}_1^c} P^*(o \mid 0)\log\tfrac{\overline{P^*}(0|0)}{P_{\tilde{X}_Z|X}(0|0)} - \sum_{o\in\mathcal{D}_1} P^*(o \mid 0)\log\tfrac{\overline{P^*}(1|0)}{P_{\tilde{X}_Z|X}(1|0)}\Bigg\},\\
= \quad &\Bigg\{\sum_{o\in\mathcal{D}_1} P^*(o \mid 0)\log\tfrac{P^*(o|0)P_{\tilde{X}_Z|X}(1|0)}{P_{Z|X}(o|0)\overline{P^*}(1|0)}\\
&+\sum_{o\in\mathcal{D}_1^c} P^*(o \mid 0)\log\tfrac{P^*(o|0)P_{\tilde{X}_Z|X}(0|0)}{P_{Z|X}(o|0)\overline{P^*}(0|0)}\Bigg\},
\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{\geq} \sum_{o \in \mathcal{D}_1} P^* (o \mid 0) \left[1 - \frac{P_{Z|X} (o \mid 0) \overline{P^*} (1 \mid 0)}{P^* (o \mid 0) P_{\tilde{X}_Z|X} (1 \mid 0)} \right] \\
&+ \sum_{o \in \mathcal{D}_1^c} \hat{P}^* (o \mid 0) \left[1 - \frac{P_{Z|X} (o \mid 0) \overline{P^*} (0 \mid 0)}{P^* (o \mid 0) P_{\tilde{X}_Z|X} (0 \mid 0)} \right],
\end{aligned}$$

$$\begin{aligned}
&= \sum_{o \in \mathcal{D}_1} \left[P^* (o \mid 0) - \frac{P_{Z|X} (o \mid 0) \overline{P^*} (1 \mid 0)}{P_{\tilde{X}_Z|X} (1 \mid 0)} \right] \\
&+ \sum_{o \in \mathcal{D}_1^c} \left[P^* (o \mid 0) - \frac{P_{Z|X} (o \mid 0) \overline{P^*} (0 \mid 0)}{P_{\tilde{X}_Z|X} (0 \mid 0)} \right],
\end{aligned}$$

$$\begin{aligned}
&\overset{(b)}{=} \sum_{o \in \mathcal{D}_1} \left[P^* (o \mid 0) - \frac{P_{Z|X} (o \mid 0) \sum\limits_{o \in \mathcal{D}_1} P^* (o \mid 0)}{\sum\limits_{o \in \mathcal{D}_1} P_{Z|X} (o \mid 0)} \right] \\
&+ \sum_{o \in \mathcal{D}_1^c} \left[P^* (o \mid 0) - \frac{P_{Z|X} (o \mid 0) \sum\limits_{o \in \mathcal{D}_1^c} P^* (o \mid 0)}{\sum\limits_{o \in \mathcal{D}_1^c} P_{Z|X} (o \mid 0)} \right], \\
&= 0.
\end{aligned}$$

(a) comes from the fact that $\log x \geq 1 - \frac{1}{x}$

(b) comes from the fact that $P_{\tilde{X}_Z|X} (1 \mid 0) = \sum\limits_{o \in \mathcal{D}_1} P_{Z|X} (o \mid 0)$ and $P_{\tilde{X}_Z|X} (0 \mid 0) = \sum\limits_{o \in \mathcal{D}_1^c} P_{Z|X} (o \mid 0)$ and (3.76).

Similarly we have $f(1) \geq 0$. Therefore it follows that

$$D \left(P^* \parallel P_{Z|X} \mid \hat{P}_{x^n} \right) - D \left(\overline{P^*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n} \right) \geq 0.$$

□

With the same arguments we also have

$$D \left(P^* \parallel P_{Y|X} \mid \hat{P}_{x^n} \right) \geq D \left(\overline{P^*} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n} \right). \tag{3.80}$$

Moreover, as we assume that $\alpha_n = \alpha_n^{'}$, then

$$D \left(P^* \parallel P_{Y|X} \mid \hat{P}_{x^n} \right) = D \left(P'^* \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n} \right). \tag{3.81}$$

Thus

$$D\left(P'^{*} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right) \geq D\left(\overline{P^{*}} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right). \tag{3.82}$$

Because $D\left(P'^{*} \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right)$ is the minimum of $D\left(P \parallel P_{\tilde{X}_Y|X} \mid \hat{P}_{x^n}\right)$ over E'_{x^n,H_1}, then we can say $\overline{P^{*}}$ belongs to the set $\overline{E'^{\,c}_{x^n,H_1}}$. In other words, (3.78) follows. Finally, for a fixed false alarm, comparing the non-detection error exponent of the gray level observation strategy and the binary thresholding strategy we have:

$$D\left(P^{*} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \quad \geq \quad D\left(P'^{*} \parallel P_{\tilde{X}_Z|X} \mid \hat{P}_{x^n}\right).$$

.

3.6.2 The proof of 3.60

In this subsection we will show that

$$\exp\left[\mu_L\left(s;\, x^n, H_0\right) - s\lambda\right] = \exp\left(-nD\left(P_s^{*} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{3.83}$$

where P_s^{*} is given in terms of parameter $s \geq 0$ as:

$$P_s^{*}\left(o \mid a\right) = \frac{P_{Y|X}^{1-s}\left(o \mid a\right) P_{Z|X}^{s}\left(o \mid a\right)}{\sum\limits_{o'} P_{Y|X}^{1-s}\left(o' \mid a\right) P_{Z|X}^{s}\left(o' \mid a\right)},\ \forall o \in \mathcal{O}, a \in \mathcal{X}. \tag{3.84}$$

and s is chosen so that $D\left(P_s^{*} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_s^{*} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) = \dfrac{\lambda}{n}$.

Proof. We have:

$$P_s^{*}\left(o \mid a\right) = P_{Y|X}\left(o \mid a\right) \exp\left[sl\left(o \mid a\right) - \mu_l\left(s;\, a, H_0\right)\right]. \tag{3.85}$$

and

$$\log \frac{\hat{P}_s^{*}\left(o \mid a\right)}{P_{Y|X}\left(o \mid a\right)} = sl\left(o \mid a\right) - \mu_l\left(s;\, a, H_0\right). \tag{3.86}$$

where $l\left(o \mid a\right) = \log \frac{P_{Z|X}(o|a)}{P_{Y|X}(o|a)}$ and $\mu_l\left(s;\, a, H_0\right) = \log \mathbb{E}_{P_{Y|X}}\left[e^{sl(O|x)}\right]$.Then the RHS of (3.83) is developed as follows

$$
\begin{aligned}
&\exp\left[-n\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_s^*(o\mid a)\log\tfrac{P_s^*(o|a)}{P_{Y|X}(o|a)}\right]\\
&=\exp\left[-n\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_{Y|X}(o\mid a)\exp\left[sl(o\mid a)-\mu_l(s;\,a,H_0)\right]\times\right.\\
&\left.\times\left(sl(o\mid a)-\mu_l(s;\,a,H_0)\right)\right]\\
&=\exp\left[n\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_{Y|X}(o\mid a)\,\mu_l(s;\,a,H_0)\right]\exp\left[-n\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_{Y|X}(o\mid a)\times\right.\\
&\left.\times\exp\left(sl(o\mid a)-\mu_l(s;\,a,H_0)\right)sl(o\mid a)\right]\\
&=\exp\left[\sum_{o\in\mathcal{O},a\in\mathcal{X}}n\hat{P}_{o^n,x^n}(o,a)\,\mu_l(s;\,a,H_0)\right]\exp\left[-n\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_s^*(o\mid a)\,sl(o\mid a)\right]\\
&=\exp\left[\sum_{o\in\mathcal{O},a\in\mathcal{X}}n\left(o,a\mid o^n,x^n\right)\mu_l(s;\,a,H_0)\right]\exp\left[-sn\sum_{o\in\mathcal{O},a\in\mathcal{X}}\hat{P}_{x^n}(a)\,P_s^*(o\mid a)\log\tfrac{P_{Z|X}(o|a)}{P_{Y|X}(o|a)}\right]\\
&=\exp\left[\mu_L(s;\,x^n,H_0)\right]\exp\left\{-sn\left[D\left(P_s^*\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)-D\left(P_s^*\parallel P_{Z|X}\mid\hat{P}_{x^n}\right)\right]\right\}\\
&=\exp\left[\mu_L(s;\,x^n,H_0)\right]\exp\left\{-sn\frac{\lambda}{n}\right\}\\
&=\exp\left[\mu_L(s;\,x^n,H_0)-s\lambda\right].
\end{aligned}
\tag{3.87}
$$

□

3.6.3 KKT optimality conditions

We consider the following optimization problem

$$
\begin{array}{ll}
\text{minimize} & f(x)\\
\\
\text{subject to} & g(x)\leq 0\\
 & h(x)=0\\
 & x\in D\subseteq\mathbb{R}^m.
\end{array}
\tag{3.88}
$$

The *Lagrangian* $F:\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ associated with the problem (3.88) is defined as follows:

$$F(x,s,\eta) = f(x) + sg(x) + \eta h(x). \tag{3.89}$$

Let x^* be the optimal point of (3.88), then the KKT conditions say that there exists a unique s^* and η^* such that

$$\begin{aligned} g(x^*) &\leq 0 \\ h(x^*) &= 0 \\ s^* &\geq 0 \\ s^* g(x^*) &= 0 \\ \frac{\partial F}{\partial x}(x^*, s^*, \eta^*) &= 0. \end{aligned}$$

If the problem 3.88 is convex i.e. f and g are convex and h is affine, then KKT conditions are necessary and sufficient to find a minimum. For more general discussion, we refer the readers to the book of Stephen Boyd [14].

Chapitre 4

Authentication Without channel coding - Practical results

In the previous chapter, we have showed the expressions of the false alarm and non-detection probabilities by using either the Gaussian approximation or the asymptotic expressions based on the Sanov theorem. In this chapter, we present the practical results for these two types of error probabilities. Numerical simulations using Monte-Carlo estimates of the error probabilities show the good accuracy of the asymptotic expression while Gaussian approximation is poor. More remarkably, importance sampling methods are studied and employed to practically estimate very small values of the non detection probability suggesting an optimized tilted distribution as a proposal. Moreover, by considering the expressions of the two types of error probabilities, we propose to optimize the authentication performance when using generalized Gaussian distributions as a model of the print and scan channel.

4.1 Numerical computation of α and β via importance sampling

This section addresses the problem of estimating numerically type I and II error probabilities, i.e. α_n and β_n. Monte Carlo (MC) simulation methods [31] give accurate solutions since these probabilities can be expressed as expectations of a function of a random variable governed by a probability distribution. We have:

$$\alpha_n = \sum_{o^n \in \mathcal{H}_1} P^n\left(o^n \mid x^n,\, H_0\right) = \sum_{o^n \in \mathcal{O}^n} P^n\left(o^n \mid x^n,\, H_0\right) \phi\left(o^n \mid x^n; \mathcal{H}_1\right), \tag{4.1}$$

where $\phi\left(o^n \mid x^n; \mathcal{H}_1\right) = 1$ whenever $o^n \in \mathcal{H}_1$ and zero if not. The probability of type I error is then expressed as the expectation of $\phi\left(o^n \mid x^n; \mathcal{H}_1\right)$ under distribution $P^n\left(o^n \mid x^n,\, H_0\right)$. In the same way, type II error probability β is the expectation of $\phi\left(o^n \mid x^n; \mathcal{H}_0\right)$ under distribution $P^n\left(o^n \mid x^n,\, H_1\right)$. In the sequel, we denote $P^n\left(o^n \mid x^n,\, H_0\right) = P^n_{Y|X}\left(o^n \mid x^n\right)$ and $P^n\left(o^n \mid x^n,\, H_1\right) = P^n_{Z|X}\left(o^n \mid x^n\right)$ and we have:

$$\alpha_n = \mathbb{E}_{P^n_{Y|X}} \left[\phi \left(O^n \mid x^n; \mathcal{H}_1\right)\right], \tag{4.2}$$

$$\beta_n = \mathbb{E}_{P^n_{Z|X}} \left[\phi \left(O^n \mid x^n; \mathcal{H}_0\right)\right]. \tag{4.3}$$

MC methods make use of the law of large number to infer an estimation for α_n and β_n by computing numerically an empirical mean for $\phi\left(o^n \mid x^n; \mathcal{H}_0\right)$ and $\phi\left(o^n \mid x^n; \mathcal{H}_1\right)$ respectively. Clearly, the computer runs N_{trials}, each one generating an i.i.d. vector o^n, where samples $\left(o^n\right)^i$ are driven from distributions $P^n_{Y|X}$ or $P^n_{Z|X}$ respectively, which gives the following estimates:

$$\hat{\alpha}_n = \frac{1}{N_{\text{trials}}} \sum_{i=1}^{N_{\text{trials}}} \phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_1\right), \text{ where } \left(o^n\right)^i \text{ is generated from } P^n_{Y|X}, \tag{4.4}$$

$$\hat{\beta}_n = \frac{1}{N_{\text{trials}}} \sum_{i=1}^{N_{\text{trials}}} \phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_0\right), \text{ where } \left(o^n\right)^i \text{ is generated from } P^n_{Z|X}. \tag{4.5}$$

The MC estimator is unbiased ($\hat{\alpha}_n \to \alpha_n$ and $\hat{\beta}_n \to \beta_n$ almost surely when $N_{\text{trials}} \to \infty$) and the rate of convergence is $N_{\text{trials}}^{-1/2}$. Recalling that for a zero mean and unit variance Gaussian random variable U, $P(\mid U \mid \leq 1.96) = 0.95$, the confidence interval at 0.95 obtained from each estimation is

$$\left[\hat{\alpha}_n - \frac{1.96\sigma_{\alpha_n}}{\sqrt{N_{\text{trials}}}}, \hat{\alpha}_n + \frac{1.96\sigma_{\alpha_n}}{\sqrt{N_{\text{trials}}}}\right], \tag{4.6}$$

$$\left[\hat{\beta}_n - \frac{1.96\sigma_{\beta_n}}{\sqrt{N_{\text{trials}}}}, \hat{\beta}_n + \frac{1.96\sigma_{\beta_n}}{\sqrt{N_{\text{trials}}}}\right], \tag{4.7}$$

where σ_{α_n} (resp. σ_{β_n}) is the standard deviation of $\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_1\right)$ (resp. $\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_0\right)$).

As $\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_1\right)$and $\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_0\right)$ are Bernoulli random variables with parameters α and β, respectively, their variances are easily deduced, e.g. $\sigma^2_{\alpha_n} = \alpha_n - \alpha^2_n$ and $\beta^2_{\alpha_n} = \beta_n - \beta^2_n$. When α_n and β_n are very small, i.e. $\sigma^2_{\alpha_n} = \alpha_n$ or $\beta^2_{\alpha_n} = \beta_n$, accurate estimations are then difficult to achieve with realistic number of trails. Roughly speaking the number of trials needed is $N_{\text{trials}} > 10^3/\alpha$ (or $N_{\text{trials}} > 10^3/\beta$) when the desired confidence interval at 0.95 is constrained to be about the tenth of the expected value of α_n and β_n. When we need to evaluate numerically very small values of α_n and β_n to draw the curve $\beta\left(\alpha_n\right)$, the required number of trials fails to be realistic.

We propose here to use the importance sampling methods [31] which enable us to generate rare events and thus reduce considerably the required number of trials. Let us consider distributions $Q_{Y|X}$ and $Q_{Z|X}$ over the set $\mathcal{O}$ such that $Q_{Y|X}$ and $Q_{Z|X}$ are positive and (4.2), (4.3) are rewritten as:

$$\alpha_n = \mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right)\right] = \mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right)\frac{Q^n_{Y|X}}{Q^n_{Y|X}}\right], \tag{4.8}$$

$$\beta_n = \mathbb{E}_{P^n_{Z|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_0\right)\right] = \mathbb{E}_{P^n_{Z|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_0\right)\frac{Q^n_{Z|X}}{Q^n_{Z|X}}\right]. \tag{4.9}$$

One can alternatively express type I and type II error probabilities as

$$\alpha_n = \mathbb{E}_{Q^n_{Y|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right)\frac{P^n_{Y|X}}{Q^n_{Y|X}}\right], \tag{4.10}$$

$$\beta_n = \mathbb{E}_{Q^n_{Z|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_0\right)\frac{P^n_{Z|X}}{Q^n_{Z|X}}\right]. \tag{4.11}$$

MC simulations with importance sampling methods give the two following estimates:

$$\begin{aligned}\tilde{\alpha}_n &= \frac{1}{N_{\text{trials}}}\sum_{i=1}^{N_{\text{trials}}}\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_1\right)\times\left[\frac{P^n_{Y|X}\left(\left(o^n\right)^i \mid x^n\right)}{Q^n_{Y|X}\left(\left(o^n\right)^i \mid x^n\right)}\right], \\ & \quad \text{each } (o)^i \text{ is generated from } Q_{Y|X},\end{aligned} \tag{4.12}$$

$$\begin{aligned}\tilde{\beta}_n &= \frac{1}{N_{\text{trials}}}\sum_{i=1}^{N_{\text{trials}}}\phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_0\right)\times\left[\frac{P^n_{Z|X}\left(\left(o^n\right)^i \mid x^n\right)}{Q^n_{Z|X}\left(\left(o^n\right)^i \mid x^n\right)}\right], \\ & \quad \text{each } (o)^i \text{ is generated from } Q_{Z|X}.\end{aligned} \tag{4.13}$$

The problem of importance sampling is to choose an adequate proposal function $Q_{O|X}$ such that the variance of the estimated probabilities in (4.12) and (4.13) are very small. The number of trials will be considerably reduced and accurate estimations of very low values of α_n and β_n is then possible.

Let

$$Q_{Z|X}\left(o \mid x\right) = Q_{Y|X}\left(o \mid x\right) = P^*_s\left(o \mid x\right), \tag{4.14}$$

where P^*_s is the distribution such that

$$P^*_s = \underset{P_{o^n|x^n}\in E_{x^n,H_1}}{\arg\min}\ D\left(P_{o^n|x^n} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right), \tag{4.15}$$

$$E_{x^n,H_1} = \left\{P_{O|X} \in \mathcal{P}\left(\mathcal{O} \mid \mathcal{X}\right):\ D\left(P_{O|X} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_{O|X} \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) \geq \frac{\lambda}{n}\right\}.$$

With the aforementioned choice of the proposal functions $Q_{Y|X}$ and $Q_{Z|X}$, we now show that when the importance sampling is employed, the variances of the estimators $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ are lower than that of estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ using the classical MC sampling. Moreover, we prove that when n, the length of the code, is sufficiently large, the variances of $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ converge to 0 in probability even though N_{trials} is not very large.

Proposition 4.1. *Let $Q_{Y|X}$ be defined as in (4.14). Then* $\operatorname{var}(\tilde{\alpha}_n) \leq \operatorname{var}(\hat{\alpha}_n)$.

Proof. First, we will compute $\operatorname{var}(\hat{\alpha}_n)$. Recall that

$$\hat{\alpha}_n = \frac{1}{N_{\text{trials}}} \sum_{i=1}^{N_{\text{trials}}} \phi\left(\left(o^n\right)^i \mid x^n; \mathcal{H}_1\right), \text{ where } \left(o^n\right)^i \text{ is generated from } P_{Y|X}. \tag{4.16}$$

Then

$$\begin{aligned} \operatorname{var}(\hat{\alpha}_n) &= \frac{1}{N_{\text{trials}}^2} \sum_{i=1}^{N_{\text{trials}}} \operatorname{var}\left[\phi\left(\left(O^n\right)^i \mid x^n; \mathcal{H}_1\right)\right] \\ &= \frac{1}{N_{\text{trials}}} \operatorname{var}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right)\right] \\ &= \frac{1}{N_{\text{trials}}} \left\{\mathbb{E}_{P^n_{Y|X}}\left[\phi^2\left(O^n \mid x^n; \mathcal{H}_1\right)\right] - \left[\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right)\right]\right]^2\right\} \\ &= \frac{1}{N_{\text{trials}}} \left(\alpha_n - \alpha_n^2\right). \end{aligned} \tag{4.17}$$

Now we compute $\operatorname{var}(\tilde{\alpha}_n)$

$$\begin{aligned} \operatorname{var}(\tilde{\alpha}_n) &= \operatorname{var}\left[\frac{1}{N_{\text{trials}}} \sum_{i=1}^{N_{\text{trials}}} \phi\left(\left(O^n\right)^i \mid x^n; \mathcal{H}_1\right) \times \left(\frac{P^n_{Y|X}\left(\left(O^n\right)^i | x^n\right)}{Q^n_{Y|X}\left(\left(O^n\right)^i | x^n\right)}\right)\right] \\ &= \frac{1}{N_{\text{trials}}^2} \sum_{i=1}^{N_{\text{trials}}} \operatorname{var}\left[\phi\left(\left(O^n\right)^i \mid x^n; \mathcal{H}_1\right) \times \left(\frac{P^n_{Y|X}\left(\left(O^n\right)^i | x^n\right)}{Q^n_{Y|X}\left(\left(O^n\right)^i | x^n\right)}\right)\right] \\ &= \frac{1}{N_{\text{trials}}} \operatorname{var}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right) \times \left(\frac{P^n_{Y|X}(O^n | x^n)}{Q^n_{Y|X}(O^n | x^n)}\right)\right] \end{aligned} \tag{4.18}$$

$$\begin{aligned}
\mathrm{var}\left(\tilde{\alpha}_n\right) &= \frac{1}{N_{\text{trials}}}\left\{\mathbb{E}_{Q^n_{Y|X}}\left[\phi^2\left(O^n \mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}\left(O^n \mid x^n\right)}{Q^n_{Y|X}\left(O^n \mid x^n\right)}\right)^2\right]\right.\\
&\left.\quad-\left[\mathbb{E}_{Q^n_{Y|X}}\left[\phi\left(O^n \mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}\left(O^n \mid x^n\right)}{Q^n_{Y|X}\left(O^n \mid x^n\right)}\right)\right]\right]^2\right\}\\
&= \frac{1}{N_{\text{trials}}}\mathbb{E}_{P^n_{Y|X}}\left[\phi^2\left(O^n \mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}\left(O^n \mid x^n\right)}{Q^n_{Y|X}\left(O^n \mid x^n\right)}\right)\right]\\
&\quad-\left[\mathbb{E}_{P^n_{Y|X}}\phi\left(O^n \mid x^n;\mathcal{H}_1\right)\right]^2\Bigg\}\\
&\overset{(a)}{=} \frac{1}{N_{\text{trials}}}\left\{\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}\left(O^n \mid x^n\right)}{Q^n_{Y|X}\left(O^n \mid x^n\right)}\right)\right]-\alpha_n^2\right\},
\end{aligned}$$

where (a) is from the fact that $\alpha_n = \mathbb{E}_{P^n_{Y|X}}\phi\left(O^n \mid x^n;\mathcal{H}_1\right)$ and $\phi\left(O^n \mid x^n;\mathcal{H}_1\right) = \phi^2\left(O^n \mid x^n;\mathcal{H}_1\right)$.

Now we will develop $\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)}\right)\right]$.

We have

$$\begin{aligned}
\frac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)} &= \prod_{i=1}^{n}\frac{P_{Y|X}(O_i|x_i)}{Q_{Y|X}(O_i|x_i)}\\
&= \prod_{a\in\mathcal{X}}\prod_{o\in\mathcal{O}}\left(\frac{P_{Y|X}(o|a)}{Q_{Y|X}(o|a)}\right)^{n(a,o|x^n,O^n)}\\
&= \exp\left[n\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}\frac{n(a,o|x^n,O^n)}{n}\log\frac{P_{Y|X}(o|a)}{Q_{Y|X}(o|a)}\right]\\
&= \exp\left[n\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}\hat{P}_{O^n x^n}\left(o,a\right)\log\frac{\hat{P}_{O^n|x^n}(o|a)}{Q_{Y|X}(o|a)}\frac{P_{Y|X}(o|a)}{\hat{P}_{O^n|x^n}(o|a)}\right]
\end{aligned} \tag{4.19}$$

$$\begin{aligned}
&= \exp\left[n\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}\hat{P}_{x^n}(a)\,\hat{P}_{O^n|x^n}(o\mid a)\log\frac{\hat{P}_{O^n|x^n}(o|a)}{Q_{Y|X}(o|a)}\frac{P_{Y|X}(o|a)}{\hat{P}_{O^n|x^n}(o|a)}\right]\\
&= \exp\left[n\sum_{a\in\mathcal{X}}\sum_{o\in\mathcal{O}}\hat{P}_{x^n}(a)\,\hat{P}_{O^n|x^n}(o\mid a)\left(\log\frac{\hat{P}_{O^n|x^n}(o|a)}{Q_{Y|X}(o|a)}-\log\frac{\hat{P}_{O^n|x^n}(o|a)}{P_{Y|X}(o|a)}\right)\right]\\
&= \exp\left\{n\left[\mathbb{D}\left(\hat{P}_{O^n|x^n}\parallel Q_{Y|X}\mid\hat{P}_{x^n}\right)-\mathbb{D}\left(\hat{P}_{O^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\right]\right\}.
\end{aligned}\tag{4.20}$$

Plugging (4.20) into $\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n\mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)}\right)\right]$, we have

$$\begin{aligned}
&\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n\mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)}\right)\right]\\
&=\mathbb{E}_{P^n_{Y|X}}\left\{\phi\left(O^n\mid x^n;\mathcal{H}_1\right)\times\exp\left[n\left(D\left(\hat{P}_{O^n|x^n}\parallel Q_{Y|X}\mid\hat{P}_{x^n}\right)-D\left(\hat{P}_{O^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\right)\right]\right\}\\
&=\sum_{o^n\in\mathcal{O}^n}P^n_{Y|X}\left(o^n\mid x^n\right)\phi\left(o^n\mid x^n;\mathcal{H}_1\right)\exp\left[n\left(D\left(\hat{P}_{o^n|x^n}\parallel Q_{Y|X}\mid\hat{P}_{x^n}\right)-D\left(\hat{P}_{o^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\right)\right]\\
&=\sum_{o^n\in\mathcal{H}_1}P^n_{Y|X}\left(o^n\mid x^n\right)\exp\left[n\left(D\left(\hat{P}_{o^n|x^n}\parallel Q_{Y|X}\mid\hat{P}_{x^n}\right)-D\left(\hat{P}_{o^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\right)\right].
\end{aligned}\tag{4.21}$$

Plugging $Q_{Y|X}\left(o\mid x\right)=P^*_s\left(o\mid x\right)$ as in (4.14) into (4.21), we have

$$\begin{aligned}
&\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n\mid x^n;\mathcal{H}_1\right)\times\left(\frac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)}\right)\right]\\
&=\sum_{o^n\in\mathcal{H}_1}P^n_{Y|X}\left(o^n\mid x^n\right)\exp\left\{n\left[D\left(\hat{P}_{o^n|x^n}\parallel P^*_s\mid\hat{P}_{x^n}\right)-D\left(\hat{P}_{o^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\right]\right\}.
\end{aligned}\tag{4.22}$$

Using a Pythagorean like theorem (cf. Theorem 11.6.1, [19]), we have

$$D\left(\hat{P}_{o^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\geq D\left(\hat{P}_{o^n|x^n}\parallel P^*_s\mid\hat{P}_{x^n}\right)+D\left(P^*_s\parallel P_{Y|X}\mid\hat{P}_{x^n}\right).$$

Equivalently,

$$-D\left(P^*_s\parallel P_{Y|X}\mid\hat{P}_{x^n}\right)\geq D\left(\hat{P}_{o^n|x^n}\parallel P^*_s\mid\hat{P}_{x^n}\right)-D\left(\hat{P}_{o^n|x^n}\parallel P_{Y|X}\mid\hat{P}_{x^n}\right).\tag{4.23}$$

From (4.22) and (4.23) we have

$$\begin{aligned}
&\mathbb{E}_{P^n_{Y|X}}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right) \times \left(\tfrac{P^n_{Y|X}(O^n|x^n)}{Q^n_{Y|X}(O^n|x^n)}\right)\right] \\
&\leq \sum_{o^n \in \mathcal{H}_1} P^n_{Y|X}\left(o^n \mid x^n\right) \exp\left\{-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right\} \qquad (4.24) \\
&\leq \sum_{o^n \in \mathcal{H}_1} P^n_{Y|X}\left(o^n \mid x^n\right) = \alpha_n.
\end{aligned}$$

The last inequality is easy to see because the relative entropy is non negative so $\exp\left\{-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right\} \leq 1$. Therefore, from (4.18) and (4.24) we have

$$\operatorname{var}\left(\tilde{\alpha}_n\right) \leq \frac{1}{N_{\text{trials}}}\left\{\alpha_n - \alpha_n^2\right\} = \operatorname{var}\left(\hat{\alpha}_n\right). \qquad (4.25)$$

□

Now we turn to prove that $\operatorname{var}\left(\tilde{\alpha}_n\right)$ converges to zero in probability when n is large enough.

From (4.18) and (4.24), $\operatorname{var}\left(\tilde{\alpha}_n\right)$ can be expressed as follows:

$$\begin{aligned}
\operatorname{var}\left(\tilde{\alpha}_n\right) &\leq \frac{1}{N_{\text{trials}}}\left[\sum_{o^n \in \mathcal{H}_1} P^n_{Y|X}\left(o^n \mid x^n\right) \exp\left(-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right) - \alpha_n^2\right] \\
&= \frac{1}{N_{\text{trials}}}\left[\exp\left(-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right) \sum_{o^n \in \mathcal{H}_1} P^n_{Y|X}\left(o^n \mid x^n\right) - \alpha_n^2\right] \\
&= \frac{1}{N_{\text{trials}}}\left[\exp\left(-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right) \alpha_n - \alpha_n^2\right] \\
&= \frac{1}{N_{\text{trials}}}\alpha_n\left[\exp\left(-nD\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right) - \alpha_n\right].
\end{aligned} \qquad (4.26)$$

To make it simple, let $D^*_s = D\left(P^*_s \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)$. We have already established the asymptotic behavior of α_n from (3.37), then:

$$\frac{1}{n}\log \alpha_n \overset{n\to\infty}{\to} -D^*_s. \qquad (4.27)$$

In addition from the Chernoff bound (3.61) we establish that

$$\frac{1}{n}\log \alpha_n \leq -D^*_s. \qquad (4.28)$$

Thus, for every $\epsilon > 0$, there exists n_ϵ such that for all $n > n_\epsilon$, we have:

$$-D_s^* - \epsilon \leq \frac{1}{n} \log \alpha_n \leq -D_s^*.$$

Equivalently, for all $n > n_\epsilon$, we have:

$$\exp\left[-n\left(D_s^* + \epsilon\right)\right] \leq \alpha_n \leq \exp\left[-nD_s^*\right]. \tag{4.29}$$

Or

$$0 \leq \exp\left(-nD_s^*\right) - \alpha_n \leq \exp\left(-nD_s^*\right) - \exp\left[-n\left(D_s^* + \epsilon\right)\right]. \tag{4.30}$$

Choose ϵ small enough such that the LHS and the RHS of (4.30) both go to zero when n is sufficiently large. It follows that $\exp\left(-nD_s^*\right) - \alpha_n$ goes to 0 as $n \to \infty$. Combining this fact and (4.26), we have $\operatorname{var}\left(\tilde{\alpha}_n\right) \to 0$ as $n \to \infty$. Using the same arguments we also have $\operatorname{var}\left(\tilde{\beta}_n\right) \to 0$. It is worth noting that using the conditional limit theorem (cf. Theorem 11.6.2, [19]) one can directly compute (4.22) for $n \to \infty$. More precisely, conditioned to $\hat{P}_{o^n|x^n} \in E_{x^n,H_1}$, applying the conditional limit theorem we have

$$\hat{P}_{o^n|x^n} \overset{n\to\infty}{\to} P_s^* \qquad \text{in probability} \tag{4.31}$$

By the continuity of relative entropy, we have

$$\begin{gathered} D(\hat{P}_{o^n|x^n} \parallel P_s^* \mid \hat{P}_{x^n}) \overset{n\to\infty}{\to} 0 \\ D\left(\hat{P}_{o^n|x^n} \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) \overset{n\to\infty}{\to} D\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) \end{gathered} \tag{4.32}$$

Plugging (4.32) into (4.22) we obtain:

$$\begin{aligned} & \mathbb{E}_{P_{Y|X}^n}\left[\phi\left(O^n \mid x^n; \mathcal{H}_1\right) \times \left(\tfrac{P_{Y|X}^n(O^n|x^n)}{Q_{Y|X}^n(O^n|x^n)}\right)\right] \\ \approx & \sum_{o^n \in \mathcal{H}_1} P_{Y|X}^n\left(o^n \mid x^n\right) \exp\left\{n\left[-D\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right]\right\}, \\ = & \alpha_n^2 \end{aligned}$$

and the result $\operatorname{var}(\tilde{\alpha}_n) = \alpha_n^2 - \alpha_n^2 = 0$ is followed immediately.

The numerical results using importance sampling to estimate probability of false alarm and probability of non-detection will be given in the next section.

4.2 Practical performance analysis

4.2.1 Experimental setup

Without loss of generality, we use in our analysis a generalized Gaussian distribution to model the physical device, i.e. the association of a printer with a scanner, used by the legitimate source $T_m(o \mid x)$ and by the counterfeiter $T_c(o \mid \hat{x})$. The probability density function of generalized Gaussian distribution is of following form

$$p(o \mid x) = \frac{b}{2a\Gamma(1/b)} e^{-(|o-m(x)|/a)^b}, \tag{4.33}$$

where $m(x)$ is the mean and the parameter a can be computed from the variance $\sigma^2 = \mathrm{var}[O]$:

$$a = \sqrt{\sigma\Gamma(1/b)/\Gamma(3/b)}. \tag{4.34}$$

The parameter b is used to control the sparsity of the the distribution, for example when $b = 1$ the distribution is Laplacian, $b = 2$ the distribution is Gaussian, and $b \to +\infty$ the distribution is uniform. The resulting distribution is first discretised then truncated to provide values within $[0, \ldots, 255]$ to model a scanning process. Each channel is parametrized in this case by four parameters, two per each type of dots, $m_0 = m(0)$ and σ_0 for black dots and $m_1 = m(1)$ and σ_1 for white dots. It is noted that other print and scan models that take into account the gamma transfer function or additive noise with input dependent variance can be found in [44], but the general methodology of this dissertation is not dependent on the model and can still be applied.

Figure 4.1 illustrates the different effects of the generalized Gaussian distributions on the main and the opponent channels of same mean and variance and $b = 1$ (Laplacian distribution), $b = 2$ (Gaussian distribution) and $b = 6$, i.e. close to a uniform distribution.

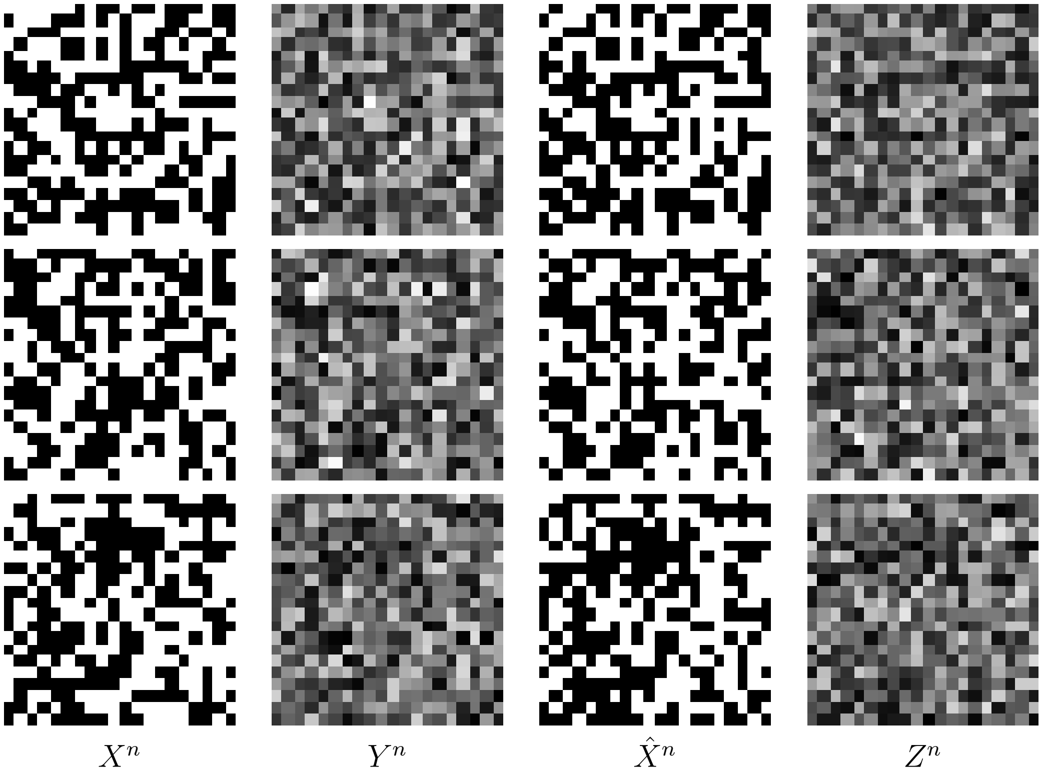


Figure 4.1: Generalized Gaussian distribution for $b = 1$ (first row), $b = 2$ (second row) and $b = 6$ (third row). Main and opponent channels are identical with $m_0 = 50$, $m_1 = 150$, $\sigma_0 = 40$, $\sigma_1 = 40$.

4.2.2 Comparison between the Gaussian approximation, the asymptotic expression and the MC simulations

In order to assess the accuracy of the computations of α and β we can use either the Gaussian approximation given by (3.23) and (3.24), or the asymptotic expression given by (3.41) and (3.42) or the classical MC simulation given by (4.4) and (4.5) or the MC simulations using importance sampling given by (4.12) and (4.13).

Figure 4.2 presents the curves of α and β with respect to the threshold λ of the test (3.29). It illustrates the gap between the estimation of α and β using the Gaussian approximation and the asymptotic expression. The classical MC simulations confirm the fact that the asymptotic expressions based on Sanov's theorem are tight.

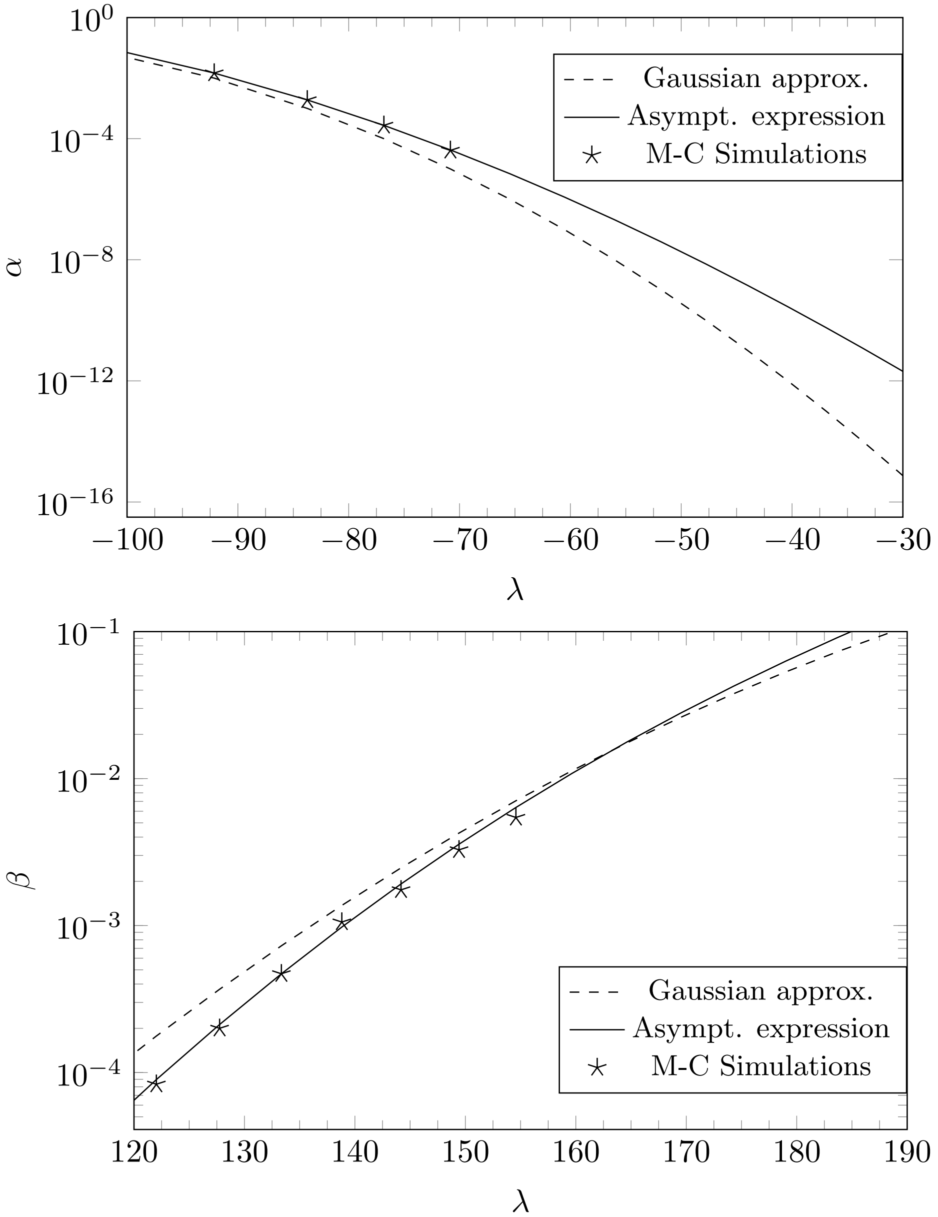
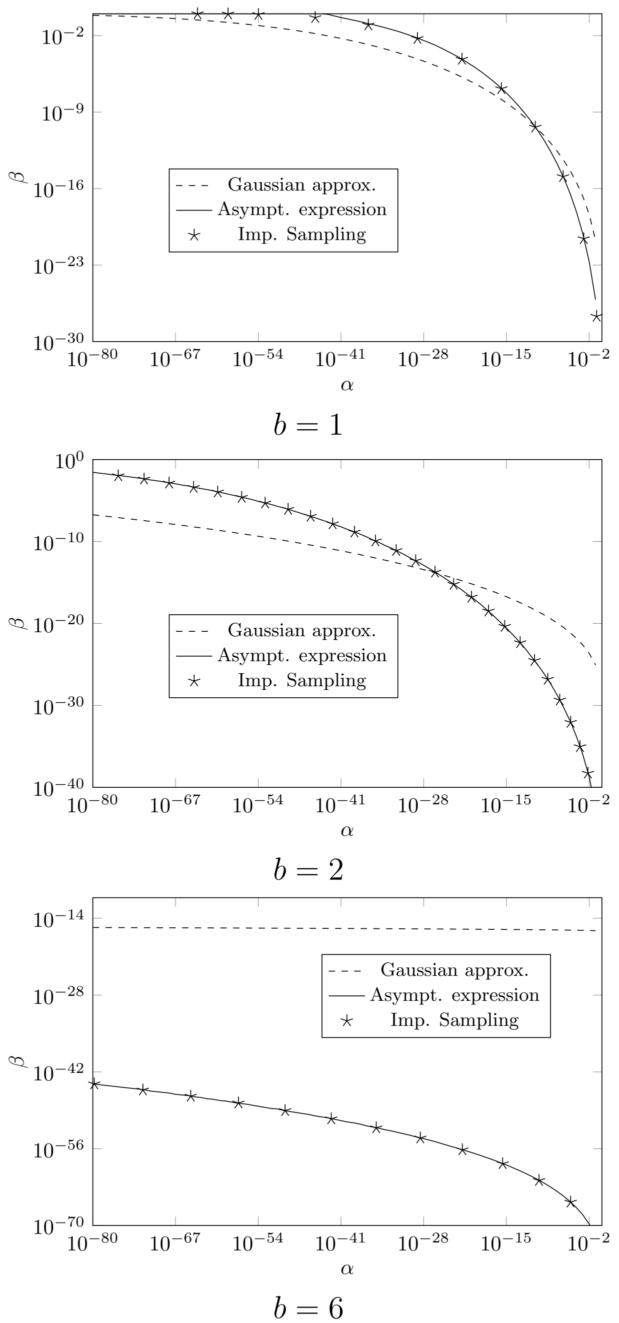


Figure 4.2: Comparison between the Gaussian approximation, the asymptotic expression and Monte-Carlo simulations (10^6 trials) for the second strategy, $N = 2000$, $\sigma = 50$.

Figure 4.3 shows the ROC curves for generalized Gaussian distributions and $b = \{1, 2, 6\}$. It illustrates the gap between the estimation of α and β using the Gaussian approximation and the asymptotic expression or the MC simulations using importance sampling. The MC simulations using importance sampling again confirm the fact that the derived Sanov bounds are tight, and the difference between the results obtained with the Gaussian approximation are very important especially for close to uniform channels. We can also notice that for the same channel power, the authentication performances are better for $b = 6$ then for $b = 2$ and $b = 1$. More interestingly, we can see that the MC simulations using importance sampling can give results for very small probabilities of false alarm and non-detection (α is up to 10^{-80} and β is up to 10^{-40}) while the classical MC simulations give the results for much larger value of α and β (α is up to 10^{-16} and β is up to 10^{-4}).



$b = 1$

$b = 2$

$b = 6$

Figure 4.3: Comparison between the Gaussian approximation, the asymptotic expression and Monte-Carlo simulations for $b = 1$ $b = 2$ and $b = 6$. Main and opponent channels are identical, $m_0 = 50$, $m_1 = 150$, $\sigma_0 = 40$, $\sigma_1 = 40$.

4.2.3 Optimization of the print and scan channel

In this subsection, by considering the print and scan models as the Generalized Gaussian distribution 4.33, we are able to maximize the authentication performances for two different security scenarios. The first one considers the opponent as passive and assume that his print-and-scan channel is the same as the legitimate channel. The second scenario devises a minimax game where an active opponent tries to maximize the probability of non-detection by choosing appropriate parameters on his channel. This authentication problem can be seen as a game where the main goal of the receiver, for a given false alarm probability α, is to find a channel that minimizes the probability of miss detection β.

Practically the channel can be chosen by using a given quality of paper, an ink of appropriate density or by adopting an given resolution. For example if the legitimate source wants to decrease the noise variance, he can choose to use oversampling to replicate the dots, on the contrary if the legitimate source wants to increase the noise variance, he can use a paper of lesser quality. It is important to recall that because the opponent will have to print a binary version of its observation, and because a printing device at this very high resolution can only print binary images, the opponent will in any case have to print with decoding errors after estimation $\hat{X}$.

We analyze two scenarios described as follows

- The legitimate source and the opponent have identical printing devices (by devices we mean printer, ink, paper, scanner), practically this means that they use exactly the same printing setup. In this case the legitimate source will try to look for the channel $\mathcal{C}$ such that for a given α, the legitimate party will have a probability of miss detection β^* such that:

$$\beta^* = \min_{\mathcal{C}} \beta(\alpha). \tag{4.35}$$

In this case, the opponent is defined to be passive.

- The opponent can modify its printing channel $\mathcal{C}_o$ (here we assume that he can change the variance of its noise), practically it means that he can modify one or several parameters of the printing setup. The opponent then tries to maximize the probability of false detection by choosing the adequate printing channel, and the legitimate sources will adopt the printing channel $\mathcal{C}_l$ which will minimize the probability of false detection. We end up with what is called a min-max game in game theory, where the optimal β^* is the solution of:

$$\beta^* = \min_{\mathcal{C}_l} \max_{\mathcal{C}_o} \beta(\alpha). \tag{4.36}$$

  In this case the opponent is active since he tries to adapt his strategy in order to degrade the authentication performance.

Because the expressions of $\beta(\alpha)$ is not simple and has to be computed using the asymptotic expressions (3.37) and (3.42), we cannot solve this problem analytically and we have to use numerical calculus instead.

For the Generalized Gaussian model, we assume that the means $m\left(0\right)$ and $m\left(1\right)$ and the modes $M\left(0\right)$ and $M\left(1\right)$ are respectively constant for all the players in the different channels (which implies that the scanning process has the same calibration for the two types of images). We assume also that variances of black and whites dots are equal at each channel and denote them σ_m^2 and σ_o^2 for main and opponent respectively.

Passive opponent

Here the opponent undergoes a channel identical to the main channel, the only parameter of the optimization problem (4.35) is consequently σ_m. Figure 4.4 presents the evolution of β w.r.t. σ_m for $\alpha = 10^{-6}$ with respectively $m_0 = 50$, $m_1 = 150$ for the Gaussian channel distribution.

For each channel configuration, we can find an optimal configuration, this configuration offers a smaller probability of error for $b = 6$ than for $b = 2$ or $b = 1$.

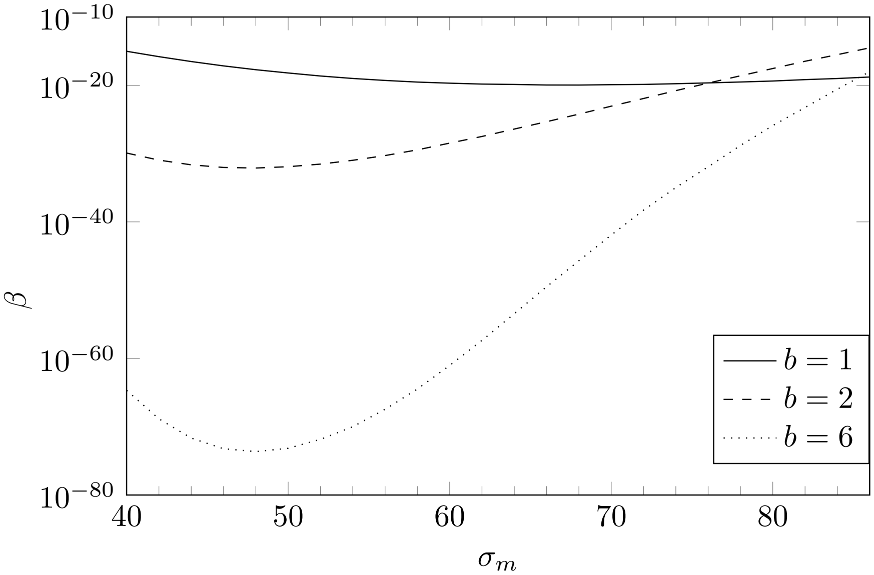
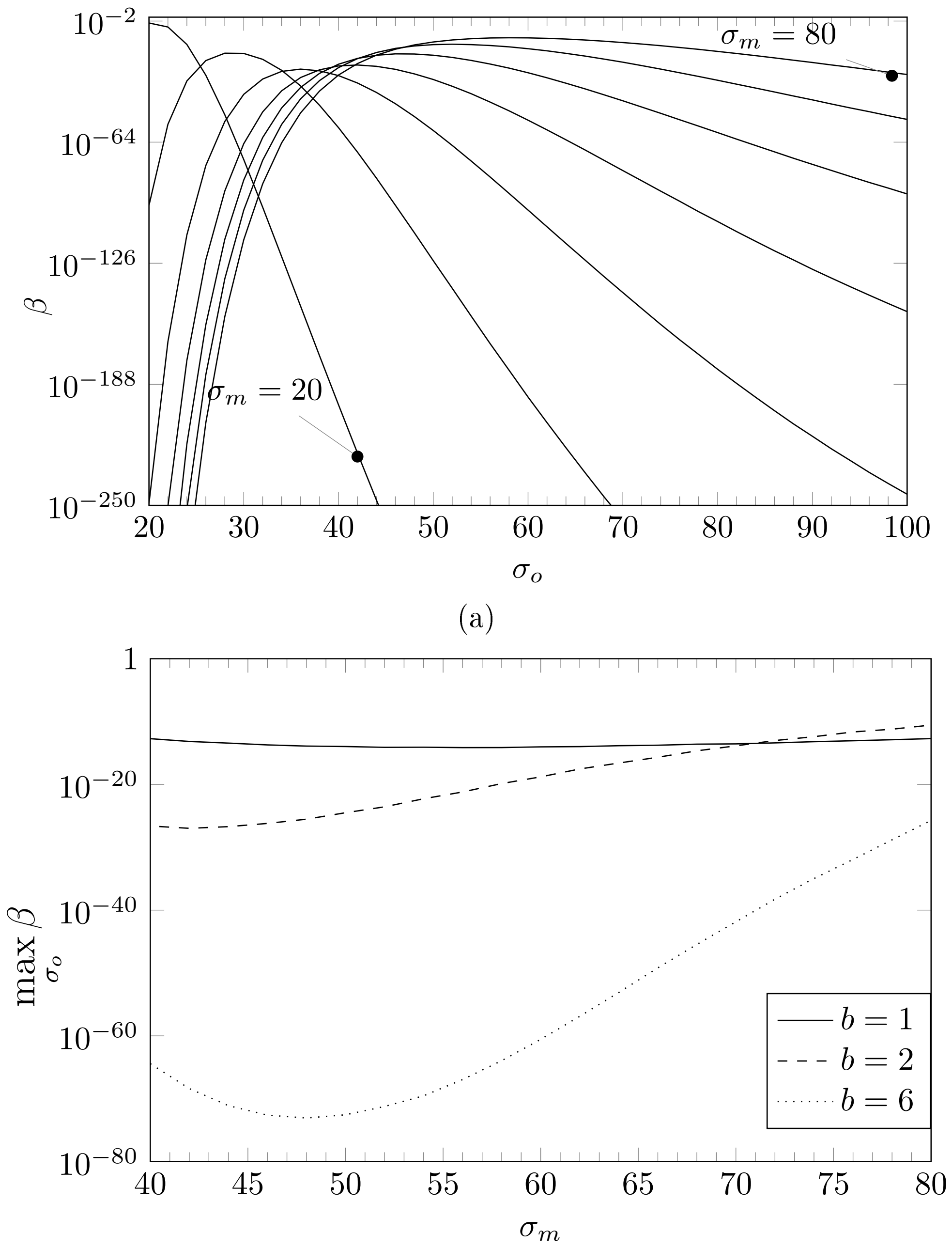


Figure 4.4: Evolution of the probability of non detection w.r.t the standard deviation of the channel ($\alpha = 10^{-6}$) for the Generalized Gaussian distribution.

Active opponent

In this scenario, the opponent can tune his variance σ_o^2 to confuse the receiver with the higher β. Figure 4.5.a shows the evolutions of β w.r.t σ_o for different σ_m when a Generalized Gaussian channel is assumed. We can see that in each case it's in the opponent interest to optimize his channel.

Figures 4.5.b shows the evolution of the best opponent strategy $\max_{\sigma_o} \beta$ w.r.t σ_m. By comparing it with Figure 4.4, we can see that the opponent's probability of non detection can be multiplied by one or several orders of magnitude for the Generalized Gaussian distribution ($\times 10^6$ for $b = 1$, $\times 10^5$ for $b = 2$) but stays the same when the distribution is close to uniform ($b = 6$).



(a)

(b)

Figure 4.5: Evolution of opponent strategy β for the Generalized Gaussian distribution for $b = 2$ (a), and the best opponent strategy $\max(\beta)$ w.r.t the standard deviation of the channel ($\alpha = 10^{-6}$) for the Generalized Gaussian distribution (b) and the for the Lognormal distribution (c).

Discussion

When facing a passive opponent, it is not surprising to notice that in each case β is important whenever σ_m is very small, i.e. when the print and scan noise is negligible hence the estimation of the original code by the opponent is easy; or very large; i.e. when the print and scan noise is so important that the original and forgery become equally noisy. The legitimate source will consequently avoid channel generating noise of very small or very large variance.

The active scenario offers a saddle point satisfying (4.36). This means that even if the adversary owns ideally perfect print and scan devices ($\sigma_o \to 0$, $o^n = \hat{x}^n$), it is not to his advantage to use it since the authentication is still efficient due to the decoding errors he will create by generating the binary code $\hat{X}^n$.

Another general remark to notice is that the optimal parameters of the active scenario are very close to the ones of the passive scenario, which means that the adversary has little room to maneuver when choosing his best attack (see Figures 4.4 and 4.5 (b)) and nearly no room when the noise is close to uniform ($b = 6$).

It is also important to notice that for distributions of same variance, dense distributions yields to better authentication performance than sparse distributions for both scenarios (see Figures 4.4 and 4.5 (b)). This is due to the fact that a distribution close to uniform tends to create a bigger overlap between the two decision regions than a sparse distribution that will generate codes mainly lying in the original one.

4.3 Conclusions

In this chapter, we have presented some practical results for the authentication system without using channel coding. We have numerically estimated the false alarm and non-detection probabilities by using Gaussian approximation, asymptotic expressions and MC simulations. By comparing the results of using MC simulations and the theoretical approximation presented in the previous chapter, we can conclude that the results using asymptotic expressions based Sanov's theorem are more accurate than the ones based on the Gaussian approximation. More interestingly, our analysis shows that the MC simulations using the importance sampling are able to estimate very small values of the two types of error probabilities.

We have also showed that it is possible to optimize the authentication performance when we know the model of the print and scan channel. The results have revealed that for the Generalized distributions the game can be tractable, and that it is in the interest of the legitimate source to adopt a channel which is close to the uniform distribution. It should be noted that this optimization is possible due to not only the knowledge of the print and scan model but also the consideration of the accurate computations of the false alarm and non-detection probabilities. In the next chapter, we also investigate the behavior of these two types of error probabilities but in the flavor of channel coding.

Chapitre 5

Authentication performance using deterministic codes

5.1 Introduction

In the previous chapter, we have considered the setup in which a secret message is mapped into a binary GC without being encoded before, and then the GC is printed on a package of a product. We have expressed the Neyman-Pearson test to check whether the observed sequence comes from the legitimate source or from the opponent, given that the legitimate receiver has the knowledge of the secret.

In the current chapter, we study the setup where the secret message is encoded with a deterministic channel encoder before being mapped to a GC which is then printed on packages of products. Different from the approach presented in the previous chapter, which performs authentication by primarily testing the likelihood ratio between the main and opponent channels distributions as a basic discriminating measure, this chapter is mainly based on channel coding theorems to analyze the authentication problem.

It is well-known that for a rate R less than the capacity of a given channel, there exists a code with a small decoding error probability P_e (cf. (2.23)). This argument motivates firstly the use of coding regarding to the main channel in order to enhance the correctness property, which means reducing the probability of false alarm to an arbitrary small value. On the other hand, it is worth reminding that the receiver does not know whether the observed message comes from the legitimate source or from the opponent. He then simply uses one metric for its decoding rule which will, more naturally, be matched to the distribution law of the main channel. In consequence, when the observed sequence comes from the counterfeiter, the decoding rule will be mismatched with respect to the opponent channel. One version of the converse coding theorem for mismatched decoder (cf. Theorem 1, [51]), states that for codes with mismatched decoders, a rate R greater than the corresponding mismatched capacity

C_{LM} [1] implies that the average over the ensemble of codebook of correct decoding probability is asymptotically vanishing. Then a rate R code greater than C_{LM} exits with very small probability of non detection.

This second argument motivates the use of channel encoder with an appropriate rate in order to insure the desired security in the context of authentication, namely reducing to an arbitrary small value the probability of accepting a fake. Unfortunately the problem of the existence of one code verifying simultaneously the results of the two aforementioned theorems, i.e. achieving small type I and type II error probabilities at the same time remains unsolved.

Hence, for the existence purpose we have relaxed the condition $R > C_{LM}$ to rates greater than the true capacity of the opponent channel, this way the strong converse of the coding theorem can be applied. Under these conditions we establish the existence of codes achieving simultaneously the desired authentication performance, i.e. arbitrary small type I and type II error probabilities, which is not possible in uncoded case because of the unavoidable tradeoff imposed by hypothesis testing.

In this chapter, we first formulate the problem of our authentication model by using deterministic codes. We then show the existence of an encoder-decoder for which the probability of non detection is small for a given negligible probability of false alarm. We next discuss how to compute the lower bound on mismatched capacity, C_{LM}, which plays an important role for the development of this chapter. We finally conclude the chapter by proposing a practical coding scheme using parallel concatenated codes with turbo decoding.

5.2 Setup

The formulation of the problem is depicted in Figure 5.1. A secret message m chosen uniformly at random from the message set $\mathcal{M} = \left\{1, ..., 2^k\right\}$ and shared with the legitimate receiver is encoded into a codeword $\mathbf{x}^{(m)}$ belonging to the set of possible codewords or codebook $\mathcal{C} = \left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(|\mathcal{M}|)}\right\} \subseteq \mathcal{X}^n$. The encoding function is said to be deterministic in the sense that only one codeword is associated to a given message m. Because of the presence of this encoder, the noisy sequence at the output of the main channel may be decoded with very small amount of errors with an appropriate choice of the encoder and decoder parameters. This is true for the legitimate receiver as an evidence, but it is also unfortunately true for the counterfeiter who observes the same output statistic. Indeed, the gray level versions of the GC observed by the legitimate receiver and the one observed by the opponent have the same distribution law. Recalling the Kerckhoffs's principle (everything about the system is known except the secret message) the counterfeiter could be then able to decode Y^n and to extract the secret message with a negligible rate of errors. He then could possibly imitate the printed code

[1] In the random coding sense the lower bound of the mismatched capacity C_{LM} is indeed the mismatched capacity C_M.

of the original packages with high degree of similarities such that the receiver cannot detect a number of forged products. In order to avoid this information leakage about the secret message we then need to provide an additional level of security by encrypting codeword $\mathbf{x}^{(m)}$ with a secret key $\mathbb{K}$ known secretly by the legitimate receiver. The GC is then the graphical mapping of the encrypted codeword $\mathbf{x}_{\mathbb{K}}^{(m)}$. In all of this work we make the hypothesis that this key insures perfect secrecy, specifically that the published version Y^n does not convey any information about either the key or the secret message:

$$I(M\,;Y^n) \;=\; 0, \tag{5.1}$$

$$I(\mathbb{K}\,;Y^n) \;=\; 0. \tag{5.2}$$

If we consider that the message is related to the characteristics of the product or to some identity number, one may define quantities in (5.1) and (5.2) as “identity leakage” (which is similar to “privacy leakage” in biometry) and “secrecy leakage” respectively. Note that we can choose an encryption method such as the one-time pad [59] for which the main and the opponent channels will both corrupt the original code X^n sample-wise. After this small digression, we continue this chapter by describing the set up in order to study how the coded version of the secret message can be decoded by a receiver who is the unique owner of the key to decrypt $\mathbf{x}_{\mathbb{K}}^{(m)}$.

Once encrypted, the GC is then printed and scanned to be processed by the legitimate receiver and by the potential opponent. These physical operations are modeled by the fact that the encrypted word $\mathbf{x}_{\mathbb{K}}^{(m)}$ is published through the main channel $V\left(y \mid x\right)$ and the corresponding scanned digital code is $\mathbf{y}$. The distribution of the main channel $V\left(y \mid x\right)$ is thus equivalent to one print and scan process, and $V\left(y \mid x\right) = T_m\left(y \mid x\right)$ as presented in subsection 3.1.2. It should be noted that $\mathbf{x}_{\mathbb{K}}^{(m)}$ also takes values in $\mathcal{X}^n$. The opponent observes $\mathbf{y}$, without being able to decode it, but processes it to create his own GC which will be printed on his forged product or document, hoping that the corresponding scanned sequence $\mathbf{z}$ will be accepted by the legitimate receiver. The opponent channel is denoted by $W\left(z \mid x\right)$ and is a physically degraded channel of $V\left(y \mid x\right)$. More specifically,

$$W\left(z \mid x\right) = \sum_{y\in\mathcal{Y}} UT\left(z \mid y\right) V\left(y \mid x\right), \tag{5.3}$$

where $UT\left(z|y\right)$ is another DMC with finite input alphabet $\mathcal{Y}$, finite output alphabet $\mathcal{Z}$.

The receiver may then get one of the two possible sequences $\mathbf{y}$ and $\mathbf{z}$. Using his knowledge of the key, he applies a decoding function $\psi_{V,\mathbb{K}}(\cdot)$ and get either $\hat{m} = \psi_{V,\mathbb{K}}(\mathbf{y}, \mathcal{C})$ or $\tilde{m} = \psi_{V,\mathbb{K}}(\mathbf{z}, \mathcal{C})$ respectively. He checks then the correspondence between the output of his decoder and the secret message in his database to infer a decision about the authenticity of the tested product. It is in the interest of the authentication designer to find a scheme in such a way that the probability that the decoded message $\hat{m} \neq m$ is close to zero and that the probability that the decoded message $\tilde{m} = m$ is as small as possible.

Authentication performances are evaluated in terms of probability of false alarm α_m and probability of non-detection β_m, for a given message m. The probability of false alarm is the probability that a legitimate message is rejected by the receiver. More precisely,

$$\alpha_m = \Pr\{\hat{m} \neq m\}. \tag{5.4}$$

We denote by α the average probability of false alarm over the message set and recall that all messages are equally probable, then

$$\alpha = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\alpha_m. \tag{5.5}$$

It turns out that α_m is exactly the error probability on the main channel when sending the message m and α is the average block error probability on the main channel of a particular code (cf. (2.26) and (5.9)).

The probability of non detection is the probability that a forged message is accepted by the receiver. More precisely,

$$\beta_m = \Pr\{\tilde{m} = m\}. \tag{5.6}$$

Likewise we denote by β the average probability of non-detection and

$$\beta = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\beta_m. \tag{5.7}$$

Similarly, β is exactly the correct decoding probability on the opponent channel (cf. 5.10). This probability is also recognized as the success probability of the opponent P_S [2]. In the rest of this dissertation, the probability of non-detection and success probability are used interchangeably. The probability of non detection is equivalent to $1-Pe_{opp}$, where Pe_{opp} is the error probability on the opponent channel and will be defined in the next section.

In the sequel, we show that there exists a code with rate between the capacity of the main and the opponent channel such that the probability of false alarm α_m and probability of non-detection β_m can be made arbitrarily small simultaneously.

Because we are mainly concerned with coding improvement on authentication, we will simplify the studied model and subsequent notations by extracting a paradigm (see Figure 5.2) that resumes the principal elements involved in our development. We then

[2] P_S is the probability that the opponent's forgery code is accepted as authentic.

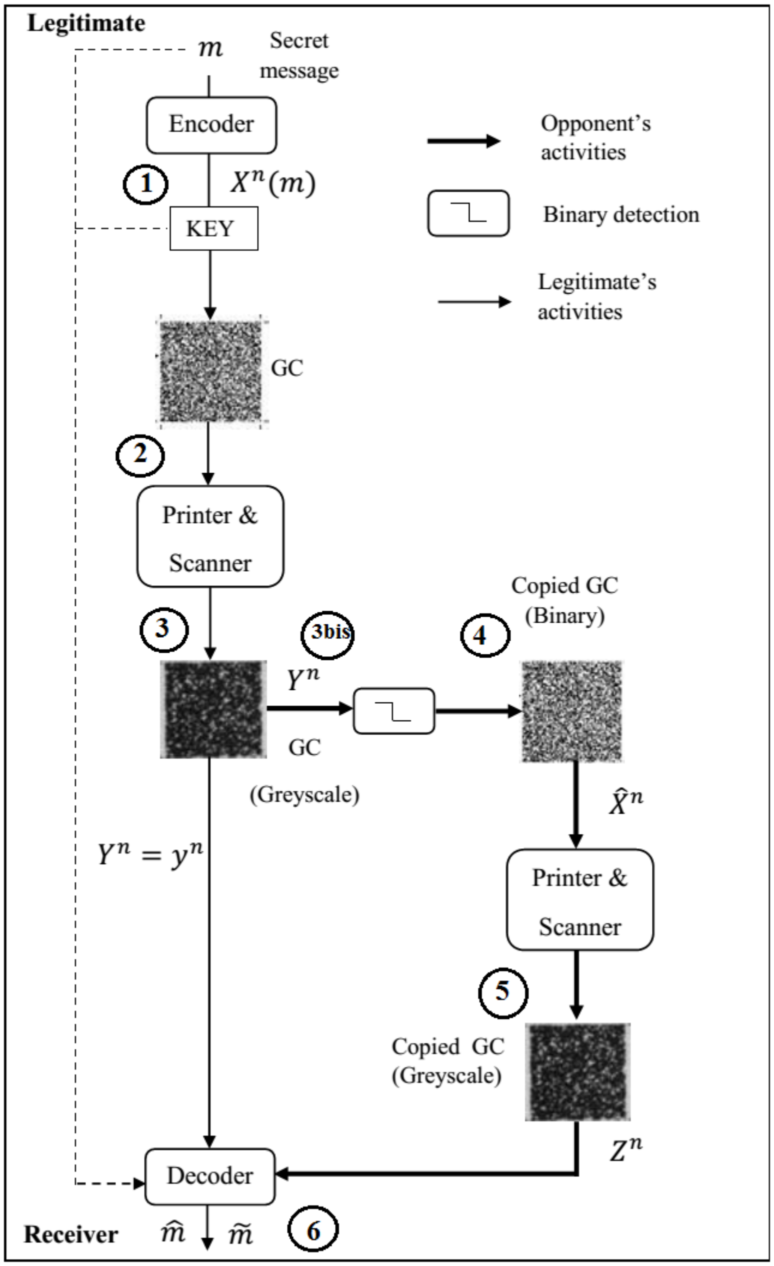


Figure 5.1: Channel coding based authentication model.

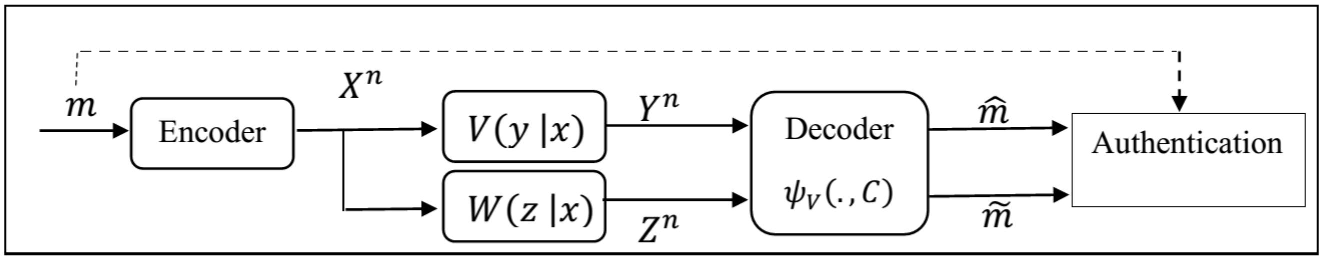


Figure 5.2: Channel model.

tackle the problem by considering that the succession of encrypting and decrypting functions forms an identity function and does not affect the noise of the memoryless channel. In the simplified model the subscript related to the key will be omitted, and the involved variables are the message m, its coded sequence $x^{(m)}$, the main and the opponent channels and the decoder.

5.3 Channel coding and authentication performance

As mentioned in the previous section, the probabilities of false alarm and non-detection are two typical measures of performances for authentication, and we demonstrate in this section how channel coding can enhance these quantities. More specifically, we will establish the existence of codes, without constructing them explicitly, which make the probability of false alarm α_m and the probability of non-detection β_m arbitrarily small simultaneously.

Recalling that the receiver has no idea whether the observed sequence $\boldsymbol{o}$ comes from the legitimate or from the opponent, and that his decoding rule matches with the main channel $\{\mathcal{X}, V\left(y \mid x\right), \mathcal{Y}\}$ but mismatches with the opponent channel $\{\mathcal{X}, W\left(z \mid x\right), \mathcal{Z}\}$, we define the decoding function for a given $(\mathcal{M}_n, n)$ code $\mathcal{C}_n$ as follows:

$$\begin{aligned} &\psi_V : \mathcal{O}^n \to \mathcal{M}_n \\ &\psi_V\left(\boldsymbol{o}, \mathcal{C}_n\right) = \arg\max_{j\in\mathcal{M}_n} \prod_{i=1}^{n} V\left(o_i \mid x_i^{(j)}\right). \end{aligned} \tag{5.8}$$

Consequently, the probability of error on the main and the opponent channels are as follow:

- The block error probability on the main channel is:

$$Pe_V^{(n)}\left(V, \mathcal{C}_n\right) \;\;=\;\; \Pr\left\{\hat{M} \neq M \mid \mathcal{C}_n\right\} \tag{5.9}$$

- The block error probability on the opponent channel is:

$$Pe_V^{(n)}\left(W, \mathcal{C}_n\right) = \Pr\left\{\tilde{M} \neq M \mid \mathcal{C}_n\right\} \tag{5.10}$$

In the following, we recall the achievability part of the channel coding theorem which plays a crucial role in this chapter. This theorem concerns the behavior of (5.9), and thus authentication performance in the main channel.

Theorem 5.1. *(Channel coding theorem: Achievability part) For a DMC $\{\mathcal{X}, V(y \mid x), \mathcal{Y}\}$, all rates below capacity $C(V)$ are achievable. Specifically, for every rate $R < C(V)$, there exists a sequence of $\left(e^{nR}, n\right)$ codes $\mathcal{C}_n$ such that $Pe_V^{(n)}(V, \mathcal{C}_n) \rightarrow 0$.*

We now study the converse of the channel coding theorem for mismatched decoding in order to upper bound the authentication performance in the opponent channel (5.10). However, as mentioned in subsection 2.3.2, the strong converse for a general channel remains unsolved. Balakirsky [8] showed a weak converse theorem of mismatched decoding in a DMC with binary inputs. More precisely, for any code such that $R > C_{LM}(W, V)$[3], there exists a positive δ such that $Pe_V^{(n)}(W, \mathcal{C}_n) > \delta$. It should be emphasized that this weak converse theorem is true for any code. Since theorem 5.1 shows the existence of codes achieving arbitrary small error probability, one can then conclude that there exists a code $\mathcal{C}_n$ whose rate $C_{LM}(W, V) \leq R \leq C(V)$ such that $Pe_V^{(n)}(V, \mathcal{C}_n) \rightarrow 0$ and $Pe_V^{(n)}(W, \mathcal{C}_n) > \delta$. In the authentication context, it then can be stated that there exists a code such that the probability of false alarm α is negligible and the probability of non-detection β is bounded away from zero. Unfortunately this is not meaningful for authentication purpose where we want both error probabilities to be as small as possible.

However, if we shrink the region of rates R, i.e. $C(W) \leq R \leq C(V)$ instead of $C_{LM}(W, V) \leq R \leq C(V)$ we are able to achieve the negligible probability of false alarm and non-detection simultaneously. In order to get this result, we need to use the strong converse of the channel coding theorem of Wolfowitz (1957).

Theorem 5.2. *(Strong converse theorem) [28] For an arbitrary DMC $\{\mathcal{X}, W(z \mid x), \mathcal{Z}\}$ of capacity $C(W)$ and any $\left(e^{nR}, n\right)$ code $\mathcal{C}_n$ with $R > C(W)$,*

$$Pe_B(\mathcal{C}_n) \geq 1 - \frac{4A}{n(R - C(W))^2} - \exp\left[\frac{-n(R - C(W))}{2}\right], \tag{5.11}$$

where A is a finite positive constant depending on the channel but not on n or e^{nR} and $Pe_B(\mathcal{C}_n)$ is the block error probability with respect to the code $\mathcal{C}_n$ (c.f (2.26)).

Remark 5.3. It is worth noting that the right-hand side of (5.11) is independent of the decoding rule. It means that the strong converse theorem is still true for mismatched decoding. More precisely, for an arbitrary DMC $\{\mathcal{X}, W(z \mid x), \mathcal{Z}\}$ of capacity $C(W)$ and any $\left(e^{nR}, n\right)$ code $\mathcal{C}_n$ with $R > C(W)$,

$$Pe_V^{(n)}(W, \mathcal{C}_n) \geq 1 - \frac{4A}{n(R - C(W))^2} - \exp\left[\frac{-n(R - C(W))}{2}\right], \tag{5.12}$$

[3] The lower bound C_{LM} on mismatched capacity coincides in this case with C_{LM}.

where $Pe_V^{(n)}(W, \mathcal{C}_n)$ is the probability of error decoding when the decoding metric $d(x,z) = \log V(z \mid x)$ is used. The decoding regions in the proof of [28] are specified for a mismatched decoding rule as follows:

$$\mathcal{Z}_m = \left\{ \mathbf{z} : V(\mathbf{z} \mid \mathbf{x}_m) \geq V(\mathbf{z} \mid \mathbf{x}_{m'}) \text{ for all } m' \neq m \right\}, \tag{5.13}$$

where $m \in \{1, ..., \mid \mathcal{M}_n \mid\}$.

We can get a tighter bound for $Pe_V^{(n)}(W, \mathcal{C}_n)$ by using Chernoff bound rather than the Chebyshev inequality. Then $Pe_V^{(n)}(W, \mathcal{C}_n)$ is lower bounded as follows:

$$Pe_V^{(n)}(W, \mathcal{C}_n) \geq 1 - 2\exp[-n\gamma(R)], \tag{5.14}$$

where

$$\gamma(R) = \min\left\{ \frac{R - C(W)}{2}, \max_{s>0}\left[s\left(C(W) + \frac{R - C(W)}{2} \right) - \log g(s) \right] \right\} > 0, \tag{5.15}$$

$$g(s) = \sum_z W(z \mid x) \exp\left[s \log \frac{W(z \mid x)}{\sum_{x'} P_X(x') W(z \mid x')} \right]. \tag{5.16}$$

As mentioned above, the success probability P_S of the opponent is equal to $1 - Pe_V^{(n)}(W, \mathcal{C}_n)$. Hence, P_S is upper bounded as follows

$$P_S \leq 2\exp[-n\gamma(R)]. \tag{5.17}$$

The following proposition shows that we can find an authentication code whose rate is between the capacity of the main channel and the opponent channel, and such that we can achieve jointly negligible probabilities of false alarm and non-detection.

Proposition 5.4. *Given a system of channel $\{\mathcal{X}, V(y \mid x), W(z \mid x), \mathcal{Y}, \mathcal{Z}\}$, a rate $R \in [C(W), C(V)]$, for all $\epsilon > 0$, there exists a sequence $(\mathcal{M}_n, n)$ code $\mathcal{C}_n$ a such that*

$$\alpha_m(\mathcal{C}_n) \to 0$$

$$\beta_m(\mathcal{C}_n) \to 0,$$

for all $m \in \mathcal{M}_n$.

Proof. To prove this proposition, we employ the achievability part of channel coding theorem and its strong converse. As $R < C(V)$, then by theorem 5.1, we know that there exists a sequence of $\left(e^{nR}, n\right)$ codes $\mathcal{C}_n$ such that $Pe_V^{(n)}(V, \mathcal{C}_n) \to 0$. Since $Pe_V^{(n)}(V, \mathcal{C}_n) = \alpha(\mathcal{C}_n)$ the probability of false alarm, we can say that for any $\epsilon > 0$, there exists n_ϵ such that for all $n > n_\epsilon$, the $\left(e^{nR}, n\right)$ code $\mathcal{C}_n$ makes $\alpha(\mathcal{C}_n) < \epsilon$. Recall that

$$\alpha\left(\mathcal{C}_n\right)=\frac{1}{\left|\mathcal{M}_n\right|}\sum_{m=1}^{\left|\mathcal{M}_n\right|}\alpha_m\left(\mathcal{C}_n\right)<\epsilon. \tag{5.18}$$

Therefore it follows that at least half of the messages m and their associated codewords $x^n(m)$ must have a probability of false alarm α_m less than 2ϵ, otherwise (5.18) would be violated. Choosing the codebook $\mathcal{C}_n^*$ which consists of these codewords $x^n(m)$, we then have $\alpha_m\left(\mathcal{C}_n^*\right)<2\epsilon$. It should be noted that code $\mathcal{C}_n^*$ have $\left|\mathcal{M}_n^*\right|$ codewords and

$$\begin{aligned}\left|\mathcal{M}_n^*\right| &= {}^{\left|\mathcal{M}_n\right|}/_{2},\\ R^* &= \log\tfrac{\left|\mathcal{M}_n\right|/2}{n}=R-\log\tfrac{2}{n}.\end{aligned} \tag{5.19}$$

On the other hand, as $R>C(W)$, by the strong converse theorem 5.3 (c.f 5.14) we have:

$$Pe_V^{(n)}(W)\left(\mathcal{C}_n^*\right)\geq 1-2\exp\left[-n\gamma(R)\right],$$

where $\gamma(R)$ is defined in (5.15).

In addition, we have

$$Pe_V^{(n)}\left(W,\mathcal{C}_n^*\right)=1-\beta\left(\mathcal{C}_n^*\right). \tag{5.20}$$

Hence we obtain

$$\beta\left(\mathcal{C}_n^*\right)\leq 2\exp\left[-n\gamma(R)\right]. \tag{5.21}$$

Moreover, $2\exp\left[-n\gamma(R)\right]\rightarrow 0$ as n gets larger. In other words, for every $\epsilon^{'}>0$, there exists $n_{\epsilon^{'}}>0$ such that for all $n>n_{\epsilon^{'}}$ we have $\beta\left(\mathcal{C}_n^*\right)\leq\epsilon^{'}$. Recall that

$$\beta\left(\mathcal{C}_n^*\right)=\frac{1}{\left|\mathcal{M}_n^*\right|}\sum_{m=1}^{\left|\mathcal{M}_n^*\right|}\beta_m\left(\mathcal{C}_n^*\right). \tag{5.22}$$

It follows that at least half of the messages m and their associated codewords $x^n(m)$ must have a probability of false alarm β_m less than $2\epsilon^{'}$, otherwise (5.22) would be violated. Choosing the codebook $\mathcal{C}_n^{**}$ which consists of these codewords $x^n(m)$, we then have $\beta_m\left(\mathcal{C}_n^{**}\right)<2\epsilon^{'}$. The code $\mathcal{C}_n^{**}$ have $\left|\mathcal{M}_n^{**}\right|$ codewords and

$$\begin{aligned}\left|\mathcal{M}_n^{**}\right| &= {}^{\left|\mathcal{M}_n\right|}/_{4},\\ R^* &= \log\tfrac{\left|\mathcal{M}_n\right|/4}{n}=R-\log\tfrac{4}{n}.\end{aligned} \tag{5.23}$$

It should be noted that R^* approaches R as n gets larger. Thus we can draw the conclusion that there exists a sequence of $\left(e^{nR},n\right)$ codes $\mathcal{C}_n$ such that

$$\alpha_m\left(\mathcal{C}_n\right)\rightarrow 0,$$

$$\beta_m\left(\mathcal{C}_n\right) \rightarrow 0.$$

□

We have proved that when $C\left(W\right) \leq R \leq C\left(V\right)$, it is entirely possible to find codes so that we can achieve arbitrary small probabilities of false alarm and non-detection at the same time, which is extremely meaningful for authentication. For the rates $C_{LM}\left(W, V\right) \leq R \leq C\left(V\right)$, it is definitely possible to find a code $\mathcal{C}_n$ that makes the probability of false alarm arbitrarily small but how about the probability of non-detection?

As mentioned above, if we employ the weak converse theorem of mismatched decoding proved by Balakirsky [8] and the achievability part of channel coding theorem, we can find a code making the probability of false alarm arbitrarily small but the probability of non-detection is said to be bounded away from a positive number whose behavior is not specified and thus may be innocuous. Merhav et al. [51] proposed a converse to the mismatched decoding theorem in the sense of random coding. This means that the average probability of decoding error over the ensemble of codes approaches one with increasing n. Therefore it is possible to find a code $\mathcal{C}_n^{'}$ that makes the probability of non-detection negligible.

However, from code $\mathcal{C}_n$, making the probability of false alarm arbitrary small and from code $\mathcal{C}_n^{'}$, making the probability of non-detection arbitrary small, it would also be desirable to extract a common code verifying the two error probabilities simultaneously small. Unfortunately, this is still an unsolved problem and we will deal with it in the future work.

In the sequel, we recall the converse of the channel coding theorem with mismatched decoding stated in a random coding sense [51]. In order to understand the random coding technique we first remind the definition of a random code ensemble [28].

A random code ensemble of length n and rate R with an input distribution P_X consists of all codebooks $\mathcal{C}_n = \left\{\mathbf{x}_1, ..., \mathbf{x}_{|\mathcal{M}|}\right\}$ with $|\mathcal{M}| = e^{nR}$ codewords of length n in which each symbol of each codeword is chosen independently and randomly according to the distribution P_X. Thus in this ensemble codes, the probability of a particular code $\mathcal{C}_n$ is

$$\Pr\left(\mathcal{C}_n\right) = \prod_{m=1}^{|\mathcal{M}|} P_X^n\left(\mathbf{x}_m\right). \tag{5.24}$$

Each particular code $\mathcal{C}_n$ has its own probability of error decoding with respect to a given decoding rule. Random coding is a method used to prove that the decoding error probability over the ensemble of codes is small. Therefore, there are at least one code whose error decoding probability is as small as the average one.

Theorem 5.5. *Given a channel $\{\mathcal{X}, W(z \mid x), \mathcal{Z}\}$ and a mismatched decoding metric $d_V(x,z) = \log V(z \mid x)$. Assume that there exists a channel f that satisfies the constraints of* (2.31) *with a strict inequality in the last constraint. Then for any random code ensemble of length n and rate R with the input distribution P_X, if rate $R > C_{LM}(W,V)$ then*

$$\begin{aligned}&\overline{Pe}_B(W,V) \geq 1 - \exp\left(-e^{n(R - C_{opp}(W))}\right) - \\ &-\sum_{b \in \mathcal{Z}} \left[e^{-nD(P_Z(b)+\epsilon \| P_Z(b))} + e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}\right] - \exp\left[-nD\left(P_{XZ}^{*} \parallel P_X W\right)\right].\end{aligned} \tag{5.25}$$

where

$$P_Z(z) = \sum_{x \in \mathcal{X}} P_X(x) W(z \mid x),$$

$$P_{XZ}^{*} = \underset{\hat{P}_{XZ} \in E_m}{\arg\min} D\left(\hat{P}_{XZ} \parallel P_X W\right),$$

and E_m is defined in (5.71).

Proof. See appendix 5.7.3. □

Remark 5.6. Using similar arguments to those of Proposition 5.4, one can establish the existence of a sequence of codes $\mathcal{C}_n^{'}$ such that

$$\begin{aligned}&Pe_V^{(n)}\left(W, \mathcal{C}_n^{'}\right) \geq 1 - \exp\left(-e^{n(R - C_{opp}(W))}\right) - \\ &-\sum_{b \in \mathcal{Z}} \left[e^{-nD(P_Z(b)+\epsilon \| P_Z(b))} + e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}\right] - \exp\left[-nD\left(P_{XZ}^{*} \parallel P_X W\right)\right].\end{aligned} \tag{5.26}$$

Discussion In Figure 5.3, we plot the success probability of the opponent without channel coding and the upper bound on the success probability (cf. (5.17) and (5.12)) when using channel coding, with respect to the standard deviation of the counterfeit channel for a given small probability of false alarm $\log \alpha = -70$.

From the figure we can see that when the standard deviation of the counterfeit/opponent device model T_c (3.6) is such that $\sigma_o > 25$, channel coding gives better authentication performance than without channel coding. This can be explained due to Proposition 5.4. When σ_o is small so is the gap between the two capacities $C(W)$ and $C(V)$. This fact makes the decoding error probability on the main channel to converge to 1 slowly, i.e. the probability of non detection is still rather important for moderate n. However, the probability of success without channel coding is still very small as explained in the discussion 4.2.3. Then it might be better to not use channel coding in this case.

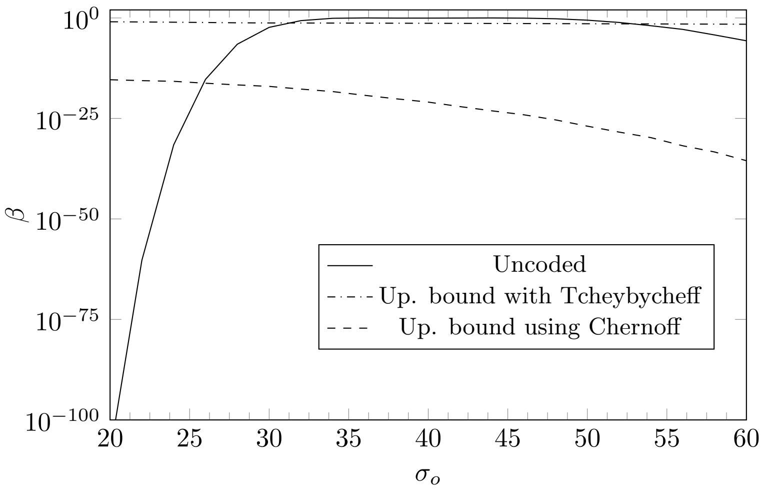


Figure 5.3: $\log \alpha = -70$,$\sigma_{main} = 50$.

In the next section , we present how to compute the lower bound of mismatched capacity C_{LM} which facilitates plotting some curves in practical coding scheme.

5.4 Computation of lower bound mismatched capacity C_{LM}

In this section, we discuss the computation of the lower bound C_{LM} on mismatched capacity. We apply an equivalent strategy as the Arimoto-Blahut algorithm, i.e. alternating maximization procedures as it can be observed from equations (5.37). Recalling theorem 2.24, we know indeed that the mismatched capacity is lower bounded by:

$$C_{LM} = \max_{P_X(x)} I_{LM}\left(P_X\left(x\right)\right), \tag{5.27}$$

where

$$I_{LM}\left(P_X\left(x\right)\right) = \min_{f \in \mathcal{F}(P_X, W, V)} I_f\left(X; Z\right). \tag{5.28}$$

and the minimum is over all $F\left(P_X, W, V\right)$ defined in (2.31).

First we focus on minimizing $I_f\left(X;Z\right)$ on the region $f\left(x,z\right) \in \mathcal{F}\left(P_X, W, V\right)$. This problem can be transformed to a non-linear optimization problem with equalities and inequalities constraints as follows

$$\begin{array}{ll} \text{minimize} & F(f) = \sum\limits_{x \in \mathcal{X}} \sum\limits_{z \in \mathcal{Z}} f(x,z) \log \frac{f(x,z)}{P_X(x) P_Z(z)}, \\ \text{subject to} & h_z(f) = 0, \text{ for all } z \in \mathcal{Z} \\ & h_x(f) = 0, \text{ for all } x \in \mathcal{X} \\ & h(f) = 0, \\ & g(f) \leq 0, \\ & f \in \mathcal{F}(P_X, W, V) \subseteq [0,1]^{|\mathcal{X}| \times |\mathcal{Z}|}. \end{array} \tag{5.29}$$

where

$$\begin{aligned} h_z(f) &= \sum_{x \in \mathcal{X}} f(x,z) - P_Z(z) \\ h_x(f) &= \sum_{z \in \mathcal{Z}} f(x,z) - P_X(x) \\ h(f) &= \sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} f(x,z) - 1 \\ g(f) &= -\sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} f(x,z) \log V(z \mid x) - D. \end{aligned} \tag{5.30}$$

The Lagrangian associated with the problem (5.29) is

$$L(f, \lambda, \boldsymbol{\nu}) = F(f) + \lambda g(f) + \sum_{x \in \mathcal{X}} \nu_x h_x(f) + \sum_{z \in \mathcal{Z}} \nu_z h_z(f) + \nu h(f), \tag{5.31}$$

where $\boldsymbol{\nu} = (\nu_x, \nu_z, \nu) \in \mathbb{R}^3$, $\lambda \geq 0$.

The Lagrange dual function is

$$\varphi(\lambda, \boldsymbol{\nu}) = \min_{f \in \mathcal{F}(P_X, W, V)} L(f, \lambda, \boldsymbol{\nu}). \tag{5.32}$$

It is easy to check that (5.29) is a convex optimization problem i.e. $F(f)$ and $g(f)$ are convex and $h_x(f)$, $h_z(f)$, $h(f)$ are affine. Moreover, it satisfies the Slater's conditions (cf. appendix 5.7.6). Thus (5.29) holds the strong duality. In other words, $I_{LM}(P_X(x))$ can be alternatively written as follows

$$I_{LM}(P_X(x)) = \max_{\lambda \geq 0, \nu_x \in \mathbb{R}} \varphi(\lambda, \boldsymbol{\nu}). \tag{5.33}$$

Proposition 5.7. *The Lagrangian $L(f, \lambda, \boldsymbol{\nu})$ achieves the minimum when*

$$f(x,z) = \frac{P_X(x) P_Z(z) \exp(\lambda \log V(z \mid x) - \nu_x)}{\sum\limits_{x' \in \mathcal{X}} P_X(x') \exp(\lambda \log V(z \mid x') - \nu_{x'})}, \tag{5.34}$$

and the minimum of $L\left(f,\lambda,\boldsymbol{\nu}\right)$ *is*

$$\varphi\left(\lambda,\boldsymbol{\nu}\right)=\sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}}P_X\left(x\right)W\left(z\mid x\right)\log\frac{\exp\left(\lambda\log V\left(z\mid x\right)-\nu_x\right)}{\sum\limits_{x'\in\mathcal{X}}P_X\left(x'\right)\exp\left(\lambda\log V\left(z\mid x'\right)-\nu_{x'}\right)}. \quad (5.35)$$

Proof. see appendix 5.7.1. □

From Proposition 5.7, we have

$$I_{LM}\left(P_X\left(x\right)\right)=\max_{\lambda\geq 0,\nu_x\in\mathbb{R}}\sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}}P_X\left(x\right)W\left(z\mid x\right)\log\frac{\exp\left(\lambda\log V\left(z\mid x\right)-\nu_x\right)}{\sum\limits_{x'\in\mathcal{X}}P_X\left(x'\right)\exp\left(\lambda\log V\left(z\mid x'\right)-\nu_{x'}\right)}. \quad (5.36)$$

From (5.27) and (5.36) we have

$$C_{LM}=\max_{P_X(x)}\max_{\lambda\geq 0,\nu_x\in\mathbb{R}}\sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}}P_X\left(x\right)W\left(z\mid x\right)\log\frac{\exp\left(\lambda\log V\left(z\mid x\right)-\nu_x\right)}{\sum\limits_{x'\in\mathcal{X}}P_X\left(x'\right)\exp\left(\lambda\log V\left(z\mid x'\right)-\nu_{x'}\right)}. \quad (5.37)$$

Now we will use the alternating optimization algorithm (see chapter 10, [74]) to compute C_{LM} as follows:

- Maximizing the objective $\varphi\left(\lambda,\boldsymbol{\nu}\right)$ over $P_X\left(x\right)\in\mathcal{P}\left(\mathcal{X}\right)$ for a given $\left(\lambda,\nu_x\right)$, we can compute the maximizing P_X (see 5.7.2) using

$$P_X\left(x\right)=\frac{\sum\limits_{z}W\left(z\mid x\right)\log h\left(x,z\right)}{\sum\limits_{z}W\left(z\mid x\right)h\left(x,z\right)}. \quad (5.38)$$

- Maximizing the objective $\varphi\left(\lambda,\boldsymbol{\nu}\right)$ over $\left(\lambda,\nu_x\right)$ with $\lambda\geq 0$ for a fixed P_X whose form is in (5.38). In order to do so, we use the augmented Lagrange method to solve the constraints optimization (see [38]).

Figure 5.4 plots the lower bound on the mismatched capacity of the opponent channel. The main channel is governed by a Gaussian distributions whose means are $\mu_{m0}=50$ or $\mu_{m1}=150$ for black and white dots respectively and the same standard deviation $\sigma_m=\sigma_{m0}=\sigma_{m1}$ varying from 10 to 90, $\in[10,90]$. The opponent channel is a degraded version of the main channel (5.3) and practically the resulting model is a mixture of Gaussian with the same means as the main channel and standard deviations set to $\sigma_o=30$ for both black and white components.

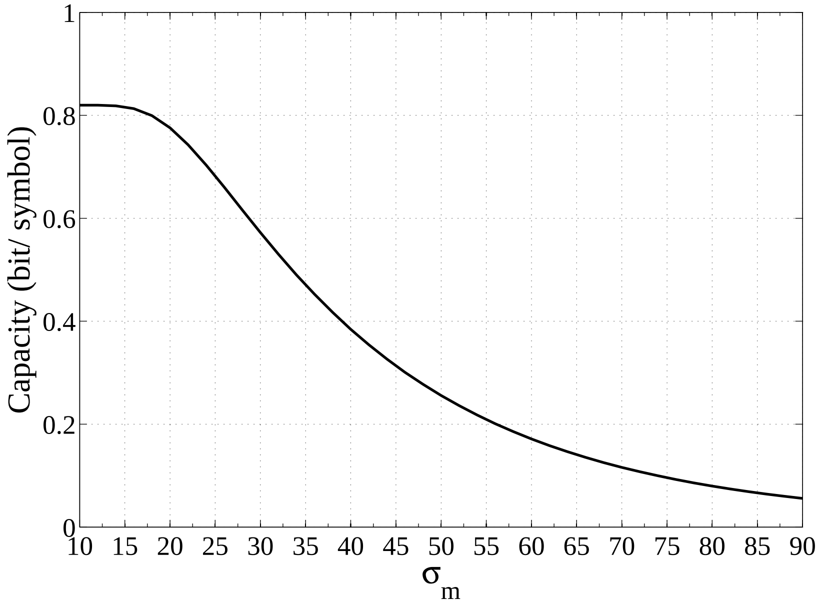


Figure 5.4: Lower bound on mismatched capacity.

5.5 Practical coding using Parallel Concatenated Codes

In the previous section we establish the existence of codes which can yield arbitrarily small authentication errors as long as the transmission rate is between the two channel capacities. However it is not specified how to practically construct or design such codes.

In the research field on channel coding it was suggested in the two last decades new coding schemes or enhanced decoding algorithms for older codes, approaching successfully Shannon limit rather for rates not greater than 1 bit/channel use. Among these codes the most popular ones are concatenated codes, low density parity check codes or polar codes. We suggest here the use of concatenated codes to possibly improve the authentication performance. This choice is a way to analyze the use of deterministic codes in authentication based on printed GC, and does not intend to be restrictive about the nature of the encoder.

We study more specifically a concatenation of several codes separated by interleavers referred as Turbo codes by their inventors Berrou, Glavieux and Thitimajshima in [12]. Due to the presence of the interleaver and with a good design of constituent codes, concatenated codes achieve surprising performance near Shannon limit. Concatenated codes are used in many applications such as 3G mobile communications and deep space communications. We will focus here on parallel concatenation of convolutional codes which is the structure proposed in [12].

5.5.1 Preliminaries

A. The Parallel Concatenated Convolutional Code (PCCC)

A PCCC is shown in Figure 5.5. The proposed code consists of two binary rate $1/2$ constituent convolutional encoders separated by an interleaver. These two convolutional encoders are two recursive systematic encoders (RSC), as illustrated on Figure 5.6. The input sequence is then sent through the channel along with the coded sequences. The input information sequence $\mathbf{u} = (u_1, u_2, ..., u_K)$ enters the first encoder and then is permuted by the interleaver before entering the second encoder. The first RSC encoder outputs the systematic sequence $\mathbf{x}^s = (x_1^s, x_2^s, ..., x_K^s) = \mathbf{u}$, and the parity check sequence $\mathbf{x}_1^p = \left(x_{1,1}^p, x_{1,2}^p, ..., x_{1,K}^p\right)$ while the second RSC encoder discards its systematic sequence and only outputs the parity check sequence $\mathbf{x}_2^p = \left(x_{2,1}^p, x_{2,2}^p, ..., x_{2,K}^p\right)$ because its systematic output is just a scrambled version of the first one. It can be summarized that when the input sequence $\mathbf{u} = (u_1, u_2, ..., u_K)$ is fed into the encoder, the output sequence is $\mathbf{x} = (\mathbf{x}^s, \mathbf{x}_1^p, \mathbf{x}_2^p)$. It means that for each input bit u_t, the outputs are $x_t = \left(x_t^s, x_{1,t}^p, x_{2,t}^p\right)$, where $t \in [1...K]$.

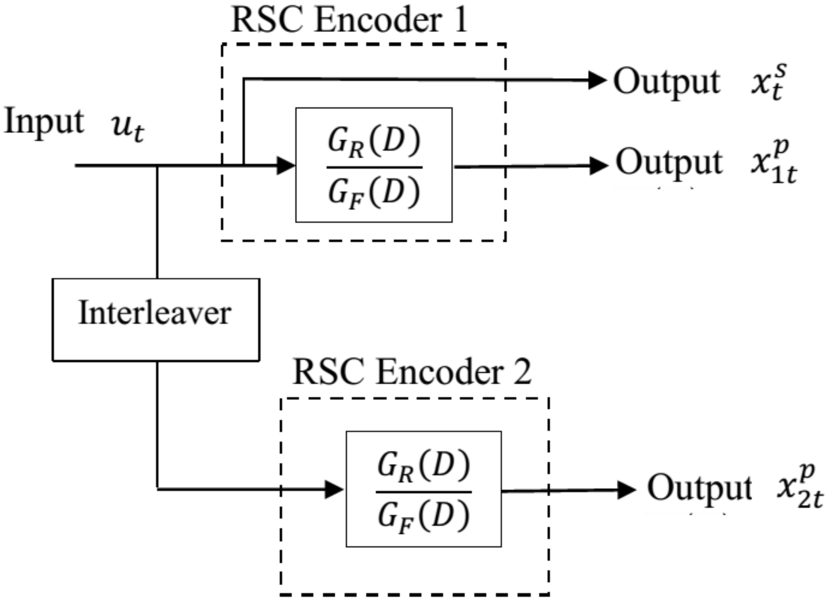


Figure 5.5: Turbo Encoder.

The RSC encoder The RSC encoder is practically implemented with ν registers set in a pipeline structure with a feedback loop that enables to use one or many of the registers output. It is represented by two generator vectors $G_R = [g_{R0}, g_{R1}, .., g_{R\nu}]$ and $G_F = [g_{F0}, g_{F1}, .., g_{F\nu}]$ where element g_{Ri} , $1 \leq i \leq \nu$ indicates if the output of register i is used again in the input, element g_{Fi}, $1 \leq i \leq \nu$ indicates if this register output is

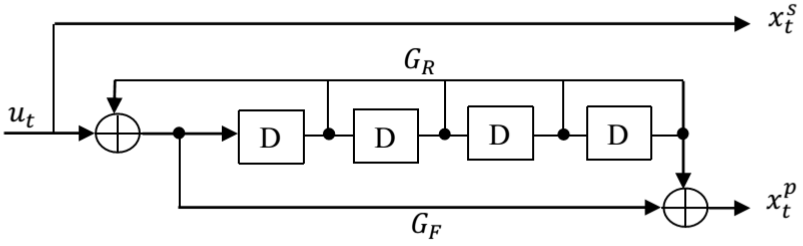


Figure 5.6: The recursive systematic encoder (RSC).

involved in the parity check output bit, and finally elements g_{R0} and g_{F0} refer to the input of all these registers and their implications in the recursive structure. Introducing a dummy variable D representing time delay, the generator vectors may be expressed as polynomials in D. For example in Figure 5.6, we have $G_R(D) = 1 + D + D^2 + D^3 + D^4$ and $G_F = 1 + D^4$, where the sign + counts for modulo 2 adder. The RSC encoder in Figure 5.6 is denoted as RSC$\left(1, \frac{G_F(D)}{G_R(D)}\right)$ where 1 represents the systematic output.

The Interleaver An interleaver is used between the two constituent convolutional encoders to provide randomness to the input sequence. The most common used interleaver is pseudo-random interleaver which rearranges bits of the input sequence in a random manner. More importantly, the interleaver is used to increase the weights of the output sequences.

B. The Turbo Decoder

The ensemble performance of concatenated codes interconnected with interleavers has been investigated widely (see [11], [9] and [58]) shedding light on their surprising performance. The main result is stated as upper bounds of the word and bit error probabilities by averaging over all random interleavers of a given length, and using maximum likelihood decoding. However, the problem of implementing an optimal decoder arises because of the complexity generated by the inherent concatenation structure. It was suggested using sub-optimal decoder where each constituent code is decoded separately in a cascaded structure. In other words the output of one decoder feeds the input of the next one with a soft (real) quantity called “innovation” about bit u_t for all t. It was therefore important to design decoders with soft outputs. This idea was addressed earlier for serially concatenated codes [46]. Each decoder takes advantage from the preceding one and is always able to compute real metrics related to its decoding rule.

In order to limit the number of decoders in the cascaded structure, the authors in [12] suggested to use only two constituent codes and then to feed-back the first decoder with the output of the second decoder with a recursive loop. This is why the code is referred to as a Turbo code. The decoding is then an iterative procedure where the “innovation” the constituent decoders have to exchange via interleavers and deinterleavers is an "extrinsic information". The need of extrinsic information is essential to favor the convergence of the iterative decoding. An efficient extrinsic information on bit u_t must be independent of the extrinsic information on any u_l, $l \neq t$. Furthermore the extrinsic information must be independent of any direct observation coming from previous decoder or channel output concerning [55]. The presence of very large pseudo-random interleaver enables these properties to be relevant in practice. How to compute the extrinsic information depends on the criterion of the decoders. The most commonly used criterion guiding the decision is the maximum of the marginal *a posteriori*:

$$\Lambda(u_t) = \frac{P(u_t = 0 \mid o^n)}{P(u_t = 1 \mid o^n)} \underset{\hat{u}_t=0}{\overset{\hat{u}_t=1}{\lessgtr}} 1. \tag{5.39}$$

In their original paper, Berrou and al. [12] proposed a modified version of the BCJR algorithm[4] [7] in order to compute the extrinsic information along with the marginal *a posteriori*, the latter being useful to make a decision at the final iteration. The Turbo decoder is depicted in Figure 5.7.

In our authentication model the secret message m is then the binary sequence $\mathbf{u} = (u_1, u_2, ..., u_K)$ at the input of a PCCC, and the coded is sequence $\mathbf{x} = (\mathbf{x}^s, \mathbf{x}_1^p, \mathbf{x}_2^p)$ of length $n = 3K$. One can puncture this sequence by deleting one parity check bit alternatively from each constituent code such that $n = 2K$. We propose then two possible coding rates, $R = 1/3$ or $R = 1/2$. The decoding metric is adapted to the main channel so that the extrinsic information for bit u_t computed at decoder stage a and using innovations from decoder b (where $a, b \in \{1, 2\}$) is:

$$\mathrm{Ext}^{(\mathrm{a})}(u_t = u) \quad = \quad \frac{1}{\Phi} \sum_{\mathbf{x}: x_t^s = u} \prod_{\substack{i=1 \\ i \neq t}}^{K} V(y_i^s \mid x_i^s) \mathrm{Ext}^{(\mathrm{b})}(u_i) \prod_{i=K+1}^{n} V(y_{a,i}^p \mid x_{a,i}^p),$$

where Φ is a normalizing factor. It turns out that the extrinsic information takes place as an novelty updating the *a priori* information $\pi(u_t)$. We will see in the next section how this extrinsic information can help us to study practically the convergence of the iterative decoder and thus guide our choice about the parameters of specific channel models.

5.5.2 Channel optimization for authentication purpose

Starting ideas w.r.t. the channel capacity:

As stated above, a code with rate $C(W) \leq R \leq C(V)$ exists such that one can achieve arbitrary small false alarm and small non detection probabilities. The larger is the difference between these capacities the more flexible is the choice of R. On the other hand, one may choose the channel's parameters such that the corresponding capacities

[4] The BCJR algorithm is in turn a version of the Forward-Bakward algorithm designed to estimate the state of a hidden Markov source

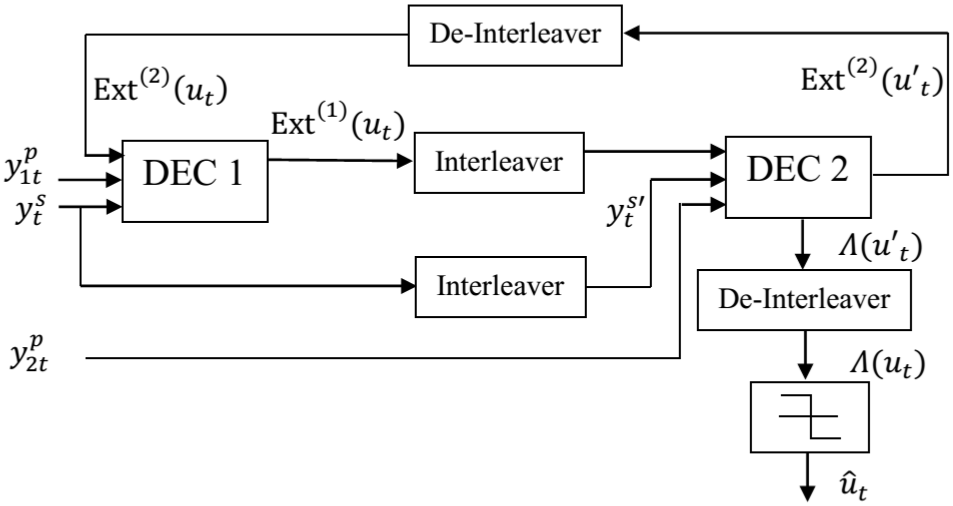


Figure 5.7: Turbo Decoder.

surround a given fixed rate. Naturally, the opponent will try to push the capacity of his channel to be maximal, hopping that its value will exceed the coding rate. If this happens the receiver cannot detect any forgery. In Figure 5.9 we report the capacities of the main and the opponent channels with the following setup:

- we consider the truncated Gaussian channel for the main and a mixture of two Gaussian for the opponent channel.
- we let the main channel's variance σ_m^2 vary and consider two values of the variance of the opponent print and scan devices. The two channels exhibit the same gray level spread around black and around white.

In Figure 5.8 we present the capacity of the main channel w.r.t its standard deviation. Is represented also the lower bound on the mismatched capacity of the opponent when the decoding metric is the distribution of the main channel. The standard deviation of the print and scan devices of the opponent is null, i.e. the print and scan is considered as perfect in this case, or a moderate black and white spread. Figure 5.9 compares the lower bound on mismatched capacity of the opponent channel and the true capacity, for the same set of parameters.

From these figures we notice that the best attack the opponent may launch when using channel coding is to set a perfect printer, thus reducing the gap between the two capacities. We remind that for uncoded case, this setting was not on his advantage where he rather has to imitate the printer of the main channel to maximize his chance of cheating.

Another point we can notice for this channel model, is that the lower bound on mismatch capacity is very close to the true one. In consequence the lower limiting rate for the existence of a code achieving very small type I and type II errors can be set to $C(W)$.

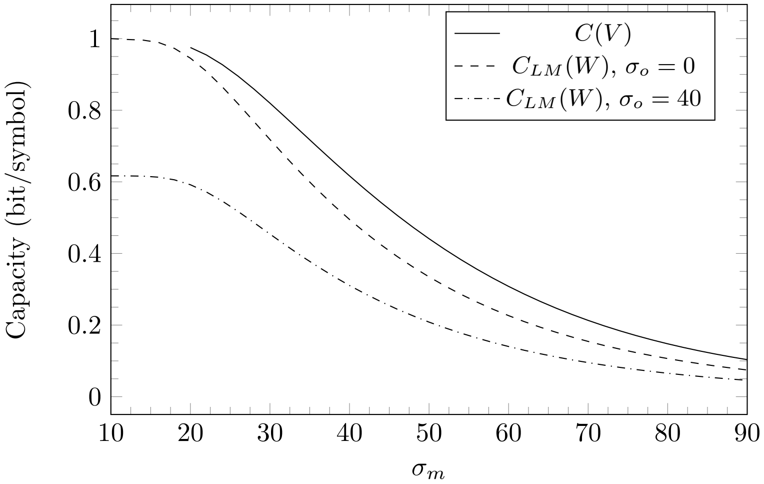


Figure 5.8: Capacity of the main $C(V)$ and the lower bound on mismatched capacity of the opponent channel $C_{LM}(W)$.

Finally, the interval $[C(W), C(V)]$ including a 1/3 coding rate, is obtained when the main channel's standard deviation takes values within [50, 58] where the lower limit corresponds to the most significant attack of the opponent. For rate 1/2 this interval reduces to [38, 46].

Numerical optimization using EXtrinsic Information Transfer chart (EXIT):

We analyze the EXIT chart tool to study practically the behavior of the iterative decoder for different parameters of the channel. We aim then at selecting the parameters that allow us the best authentication performances.

The EXIT chart tool informs us essentially about the possible convergence (or non-convergence) of the iterative decoder. Using this tool, one can also estimate the minimal number of iterations needed to achieve this convergence. In our authentication setup, it is desired that the decoder converges when the observed sequence comes from the main channel. On the opposite side, one wishes that it fails to converge when the observed sequence comes from the opponent.

The EXIT chart tool was first introduced by S. Ten Brink [66]-[67]. It is a function relating a specific measure of the extrinsic information at the input and the one obtained at the output of each constituent decoder. Since the extrinsic information is a probability distribution, the author measures the mutual information between this distribution and the *a priori* of the source. In consequence, this chart will describe how a decoder behaves when small or higher amount of reliable information feeds its input. Since in the iterative decoding of PCCC, constituent decoders feed each other with extrinsic information, the EXIT chart tool will help to visualize the decoding trajectory at each step of the iterative process. The author insists however that this tool is not a rigorous proof of stability and convergence of turbo decoders. His study is also based

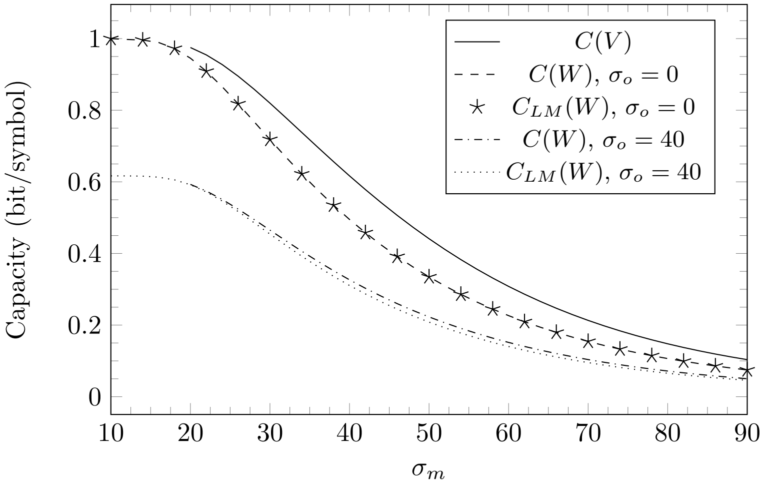


Figure 5.9: Capacity of the main and the opponent channel $C(V)$, $C(W)$ and the lower bound on mismatched capacity of the opponent channel $C_{LM}(W)$.

on two assumptions namely:

- the extrinsic information on bit u_t is not influenced (independence property) by channel observation of the systematic output at time t neither by the innovation about this bit coming from the previous decoder.
- the probability density function of the $L-$value given $u_t = u$ (defined as $\log \frac{\text{Ext}(u_t=1)}{\text{Ext}(u_t=0)}$) approaches the Gaussian distribution with the increasing number of iterations.

The first assumption is important in our analysis and is practically satisfied with large interleaver. The second one may be omitted when the binary input DMC channel is symmetric and the sequence of $L-$values L_t $1 \leq t \leq K$ is ergodic [29]. Recalling the definition of the mutual information we have (in bits/symbol):

$$I(L;U) = \sum_{u=0,1} \frac{1}{2} \int_{-\infty}^{+\infty} p(\ell \mid u) \log_2 \frac{2p(\ell \mid u)}{p(\ell \mid 0) + p(\ell \mid 1)} d\ell. \tag{5.40}$$

For a symmetric and memoryless binary input channel modeling the print and scan devices, the distribution of the $L \mid u$ is also symmetric:

$$p(\ell \mid 0) = p(-\ell \mid 1). \tag{5.41}$$

In consequence one can write the following simplification in (5.40):

$$\int_{-\infty}^{+\infty} p(\ell \mid 1) \log_2 \frac{2p(\ell \mid 1)}{p(\ell \mid 0) + p(\ell \mid 1)} d\ell = \int_{-\infty}^{+\infty} p(\ell \mid 0) \log_2 \frac{2p(\ell \mid 0)}{p(\ell \mid 0) + p(\ell \mid 1)} d\ell (5.42)$$

The symmetry property insures also consistency of the $L-$value [56], i.e.:

$$p(\ell \mid 1) \;\; = \;\; \exp(\ell)\, p(-\ell \mid 1). \qquad (5.43)$$

Plugging this into (5.40) the mutual information becomes:

$$I(L;U) \;\; = \;\; 1 - \underset{p(\ell|1)}{\mathbb{E}} \left[\log_2\left(1 + \exp(-\ell)\right)\right].$$

One can estimate $I(L;U)$ over the ergodic K-sequence of computed $\ell_t = \log \frac{\mathrm{Ext}(u_t=1)}{\mathrm{Ext}(u_t=0)}$ values for a given constituent decoder by taking the sample mean:

$$I(L;U) \;\; \approx \;\; 1 - \frac{1}{K}\sum_{t=1}^{K} \left[\log_2\left(1 + \exp(-(2u_t-1)\ell_t)\right)\right],$$

where $(2u_t - 1)$ is a correction factor to account in the K-sequence of ℓ_t for distributions of $L \mid u = 0$. Now let $I_{in} = I(L_{in}; U)$ and $I_{out} = I(L_{out}; U)$, be the mutual information computed at the input and at the output of a given decoder respectively. Precisely, L_{in} is the log-ratio of the two probabilities forming the new *a priori* distribution, and L_{out} is the log-ratio of the two extrinsic information ($Ext(1)$ and $Ext(0)$) computed by the decoder. Because of the feed-back loop, it turns out that at the second decoding stage I_{out} becomes I_{in}. The EXIT chart draws then simultaneously on the same graph the two non explicit functions:

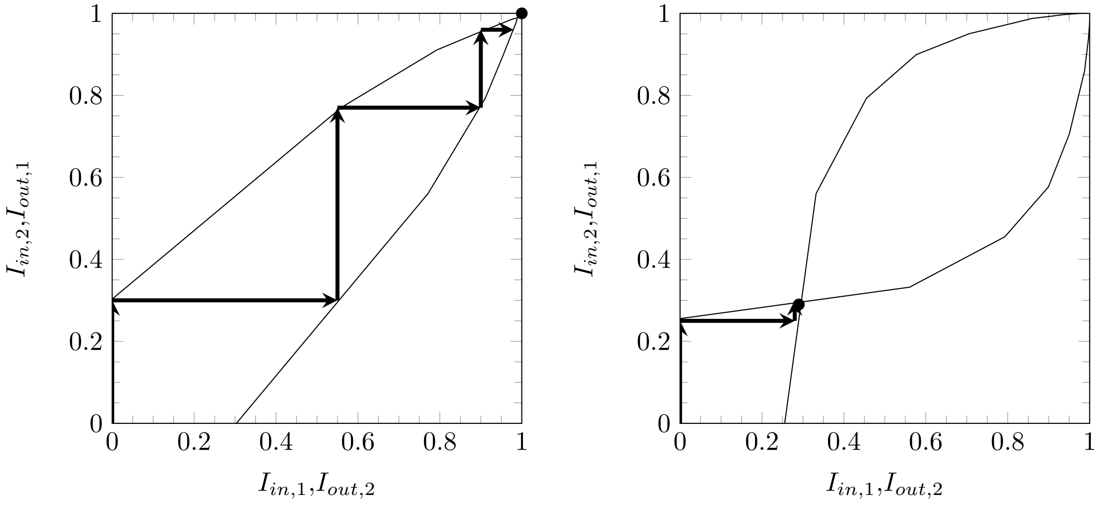
$$I_{out,1} \;\; = \;\; f_1(I_{in,1}),$$

$$I_{out,2} \;\; = \;\; f_2^{-1}(I_{in,2}).$$

The x-axis of the chart reports the mutual information $I_{in,1}$ at the input of the first decoder, and the y-axis reports the mutual information at its output $I_{out,1}$. The chart $I_{out,1} = f_1(I_{in,1})$ draws then the response of the first decoder for each amount of information feeding it. The second chart $I_{out,2} = f_2^{-1}(I_{in,2})$ draws actually the response of the second decoder when $I_{in,2} = I_{out,1}$ obtained from the first decoder and reported on the y-axis feeds it. Its output $I_{out,2}$ becomes the new $I_{in,1}$and is reported on the x-axis.

For practical results we consider again the truncated Gaussian channel for the main and a mixture of two Gaussian distributions for the opponent channel. The parameters of these distributions are set as above and we focus on the worse case attack, i.e. when the opponent print and scan devices are modeled by impulses around μ_0 and μ_1. The length of the code is set to $n = 2000$.

Let us here insist on the fact that a perfect print and scan model at the opponent end doesn't mean an opponent channel without noise. We recall indeed that the opponent processes first an observed gray level version of a genuine GC in order to estimate a binary one and prints it on the package of his counterfeit product. The opponent channel in this case is a binary-input binary-output channel with crossover probabilities



(a) main

(b) opponent, mismatch

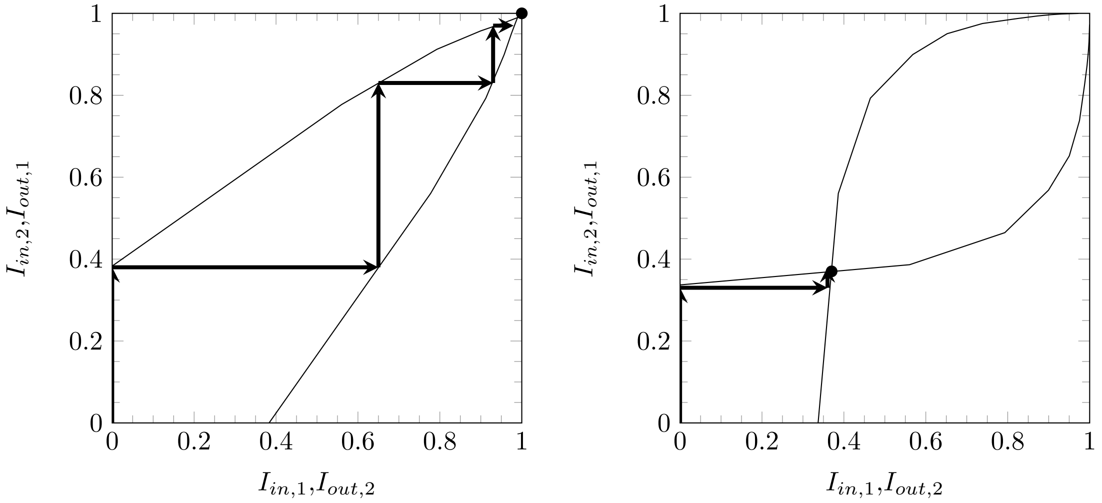
Figure 5.10: R=1/3 $\sigma_m = 48$, $\sigma_o = 0$.

depending on how noisy is the main channel. In other words the equivocation function on the main channel $H(X \mid Y)$ acts as a lower bound on these crossover probabilities p_e in regard to Fano's inequality: $h(p_e) \geq H(X \mid Y)$ where $h(p_e)$ is the binary entropy function.

Results are given in Figures 5.10 a and b. As stated in the previous section, for coding rate 1/3 we choose a main channel with a standard deviation of about 50. The EXIT chart on the main channel shows a trajectory that insures the convergence of the iterative decoder (the arrows drawing a zigzag path into the EXIT chart). The turbo decoder is then able to converge within 4 iterations. Whereas in Figure 5.10.b, drawing the EXIT for the opponent channel, the trajectory stops and is confined in a small region where the two decoder characteristics intersect. The decoder fails to converge. Figure 5.11 extends the results for code rate 1/2. Similarly to the previous setting, the lower capacity is obtained for a standard deviation of about 40. The trajectory of the iterative decoder in the main channel shows its convergence, while in the opponent channel it stops at the intersection of the two charts.

5.5.3 Authentication performances

Arbitrary small block error probability is generally hard to achieve practically. We aim however at this phase of our work, to give first results in order to evaluate numerically authentication performances in this setup. To do this, we will employ the bit error probability at the output of the iterative decoder given that the marginal *a posteriori* criterion described above optimizes the decision rule over each symbol or bit. In future research we will explore other coding strategy and/or decoding rule in order to fill the gap with the theoretical results stated above.



(a) main (b) opponent, mismatch

Figure 5.11: R=1/2 $\sigma_m = 38$, $\sigma_o = 0$.

We design now the PCCC and channel parameters to the optimized settings presented in the previous subsection and corresponding to the worst case attack launched by the opponent. Precisely, we model his print and scan devices as impulses around gray levels μ_0 and μ_1 for black and white respectively. The standard deviation of the main channel is set to 50 when rate 1/3 code is used and to 38 for a 1/2 code rate. The length of the output codeword is set to $n = 2000$ and we compare the authentication performances to the uncoded configuration with the same GC length proposed in Chap 3. The constituent decoders' metrics are fixed to the transition distribution of the main channel, and according to the trajectories of the EXIT, we take the minimal number of iterations to favor the convergence in the main channel.

In authentication with channel coding, the receiver infers a decision about authenticity by comparing the decoded message to the one in his database. As we are dealing with bit error probability and binary secret message sequence it turns out that this test may be carried on by comparing the Hamming distance $d_H(\mathbf{m}, \hat{\mathbf{m}})$ between the decoded binary message and the one in the database to a given threshold λ. Obviously when this threshold is set to 0, the block error probability becomes the measure used to evaluate the authentication performances.

Dealing with counting the number of errors (the aforementioned Hamming distance) makes it easier to tune one of the two types of authentication error probabilities to a predefined value and minimizes the other regarding to the Neyman-Pearson optimal test. Type I and II error probabilities can be upper bounded using the union bound involving the transfer function of the constituent convolutional codes, namely the IR-WEF (Input-Redundancy Weight Enumerating Function) [11]-[25]. The probability of false alarm may be bounded as follows:

$$\alpha \leq \sum_{w=\lambda}^{K} \sum_{d=w}^{n-w} A_{w,d} \Pr \left\{ \text{error event with Hamming distance } d \right\}. \tag{5.44}$$

For small probability of false alarm the threshold is set far from the mean of the distribution of error events with Hamming distance $d_H(\mathbf{m}, \hat{\mathbf{m}})$ obtained in the main channel. For large K, coefficients $A_{w,d}$ approaches a binomial [23].

The probability of non detection may be bounded by:

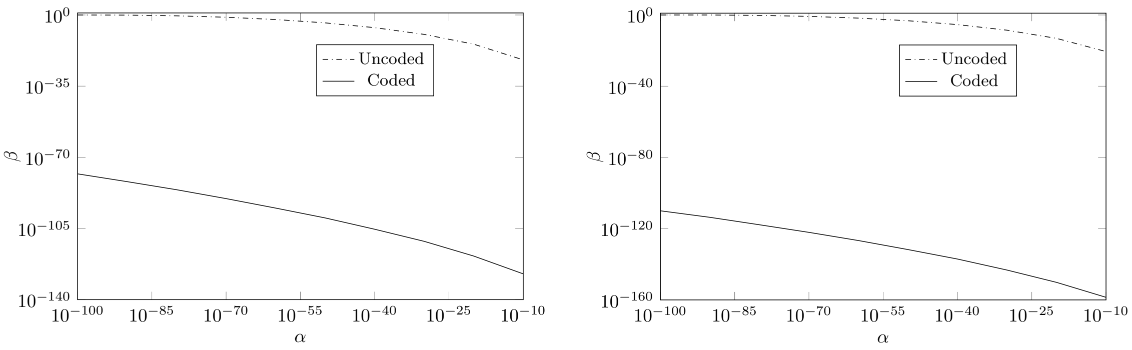
$$\beta \leq \sum_{w=0}^{\lambda} \sum_{d=d_{min}}^{n-w} A_{w,d} \Pr \left\{ \text{error event with Hamming distance } d \right\}. \tag{5.45}$$

The probability of an error event with very small value of the Hamming distance $d_H(\mathbf{m}, \tilde{\mathbf{m}})$ at the output of the opponent channel with a mismatched decoder is negligible because $R > C(W)$. The rate of bit errors in each message approaches 0.5 with high probability, so that predominant coefficients $A_{w,d}$ in the sum (5.45) are near binomial [23].

Type I and II error probabilities can then be derived directly from $Pe_{b,V}(V)$ and $Pe_{b,V}(W)$ (2.28), where we add letters V and W to distinguish the corresponding channel and decoding metric. We then apply the Sanov theorem on binomial distributions with parameters $Pe_{b,V}(V)$ and K for the false alarm and $Pe_{b,V}(W)$ and K for the probability of non detection. Then α and β can be upper bounded as follows:

$$\begin{aligned} \alpha &\doteq \exp\left[-KD\left(\tfrac{\lambda}{K} \parallel Pe_{b,V}(V)\right)\right], \\ \beta &\doteq \exp\left[-KD\left(\tfrac{\lambda}{K} \parallel Pe_{b,V}(W)\right)\right], \end{aligned} \tag{5.46}$$

where it is easy to check that both P^* in (3.38) and 3.40 are equal $\frac{\lambda}{K}$. The results are given in Figure 5.12. It is shown that a setup including a channel encoder gives better authentication performance with a good choice of the coding rate.



$R = 1/3$ $R = 1/2$

Figure 5.12: Comparison between coded and uncoded after worst case attack and optimization of the main channel. Uncoded: $\sigma_m = 42$, $\sigma_o = 40$ (corresponding to the min max game optimal parameters). Coded: $R = 1/3$:$\sigma_m = 50$, $\sigma_o = 0$, ; for $R = 1/2$: $\sigma_m = 38$, $\sigma_o = 0$. These values correspond to the larger gap between C_m and C_o when the opponent launch the worst case attack.

5.6 Conclusions

In this chapter, we have presented a framework for analyzing the performance of authentication using channel coding. Our analysis has shown that it is possible to enhance authentication performance by choosing a code whose rate is between the capacity of the opponent channel and that of the main channel. Such a code is capable of making the probability of false alarm and probability of non-detection arbitrarily small at the same time which is an ideal achievement in authentication context. We have also compared the success probability of the opponent without channel code and the upper bound of success probability with channel coding. These curves support the possibility of using channel coding to improve authentication performance. We have also discussed the achievable rate regions of such codes. More precisely, we have showed that for rates between the mismatched capacity of the opponent and the capacity of the main channel, it is possible to find a code achieving small probability of false alarm and a code achieving small probability of non-detection. However, showing that these two codes are identical or extracting a common code from such codes is left for future research. The lower bound of mismatched capacity C_{LM} of a DMC can be computed using alternating optimization and the computation of C_{LM} plays a significant role to design practical codes.

We have finally proposed a practical coding scheme using parallel concatenated codes with turbo decoding. We have analyzed the EXIT chart tool which facilitates the choice of channels parameters in order to optimize authentication performance when the Turbo codes is employed.

5.7 Appendix

5.7.1 Proof of Proposition 5.7

$$\begin{aligned} L\left(f,\lambda,\boldsymbol{\nu}\right) = &\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\log\tfrac{f(x,z)}{P_X(x)P_Z(z)} + \\ &-\lambda\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\log V\left(z\mid x\right) - \lambda D + \sum_{x\in\mathcal{X}}\nu_x\left[\sum_{z\in\mathcal{Z}} f\left(x,z\right) - P_X\left(x\right)\right] + \\ &+\sum_{z\in\mathcal{Z}}\nu_z\left[\sum_{x\in\mathcal{X}} f\left(x,z\right) - P_Z\left(z\right)\right] + \nu\left[\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right) - 1\right]. \end{aligned} \tag{5.47}$$

Differentiating $L\left(f,\lambda,\boldsymbol{\nu}\right)$ with respect to $f\left(x,z\right)$and setting to 0, we have

$$\log\tfrac{f(x,z)}{P_X(x)P_Z(z)} + 1 + \nu_x + \nu_z + \nu - \lambda\log V\left(z\mid x\right) \ = 0, \tag{5.48}$$

for all $x\in\mathcal{X}$, $z\in\mathcal{Z}$.

Solving this set of equations we get

$$f\left(x,z\right) = P_X\left(x\right)P_Z\left(z\right)\exp\left(\lambda\log V\left(z\mid x\right) - \nu_x - \nu_z\right)c. \tag{5.49}$$

Using the assumption that $\sum\limits_{x\in\mathcal{X}} f\left(x,z\right) = P_Z\left(z\right)$, we have

$$P_Z\left(z\right) = \sum_{x\in\mathcal{X}} P_X\left(x\right)P_Z\left(z\right)\exp\left(\lambda\log V\left(z\mid x\right) - \nu_x - \nu_z\right)c.$$

Thus

$$c = \frac{e^{\nu_z}}{\sum\limits_{x\in\mathcal{X}} P_X\left(x\right)\exp\left(\lambda\log V\left(z\mid x\right) - \nu_x\right)}. \tag{5.50}$$

Hence $f\left(x,z\right)$ is of form

$$f\left(x,z\right) = \frac{P_X\left(x\right)P_Z\left(z\right)\exp\left(\lambda\log V\left(z\mid x\right) - \nu_x\right)}{\sum\limits_{x'\in\mathcal{X}} P_X\left(x'\right)\exp\left(\lambda\log V\left(z\mid x'\right) - \nu_{x'}\right)}. \tag{5.51}$$

Plugging $f\left(x,z\right)$ into (5.47) we have the minimum value of $L\left(f,\lambda,\boldsymbol{\nu}\right)$ as follows

$$\begin{aligned}
L_{\min} &= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\Big[\ \log\tfrac{f(x,z)}{P_X(x)P_Z(z)} + \nu_x + \nu_z + \nu - \lambda\log V\left(z\mid x\right)\ \Big] + \\
&\quad -\lambda D - \sum_{x\in\mathcal{X}}\nu_x P_X\left(x\right) - \sum_{z\in\mathcal{Z}}\nu_z P_Z\left(z\right) - \nu \\
&= -\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right) + \lambda\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} P\left(x\right)W\left(z\mid x\right)\log V\left(z\mid x\right) + \\
&\quad -\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\nu_x - \sum_{z\in\mathcal{Z}}\sum_{x\in\mathcal{X}} f\left(x,z\right)\nu_z - \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\nu \\
&= -\sum_{x,z} f\left(x,z\right)\left(1+\nu_x+\nu_z+\nu\right) + \lambda\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} P_{XZ}\left(x,z\right)\log V\left(z\mid x\right) \\
&= -\sum_{x,z} f\left(x,z\right)\left[\lambda\log V\left(z\mid x\right) - \log\tfrac{f(x,z)}{P_X(x)P_Z(z)}\right] + \lambda\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} f\left(x,z\right)\log V\left(z\mid x\right) \\
&= \sum_{x,z} f\left(x,z\right)\log\tfrac{f(x,z)}{P_X(x)P_Z(z)} \\
&= \sum_{x,z} P_X\left(x\right)W\left(z\mid x\right)\log\frac{\exp\left(\lambda\log V\left(z\mid x\right)-\nu_x\right)}{\sum\limits_{x^{'}\in\mathcal{X}} P_X\left(x'\right)\exp\left(\lambda\log V\left(z\mid x'\right)-\nu_{x'}\right)}.
\end{aligned} \tag{5.52}$$

5.7.2 Maximizing $\varphi\left(\lambda,\boldsymbol{\nu}\right)$ over $P\left(x\right)$ for a given $\left(\lambda,\nu_x\right)$

Taking the derivative of $\varphi\left(\lambda,\boldsymbol{\nu}\right)$ w.r.t. $P_X\left(x\right)$ we have

$$\begin{aligned}
\tfrac{\partial\varphi(\lambda,\boldsymbol{\nu})}{\partial P_X(x)} &= \sum_z W\left(z\mid x\right)\log\exp\left(\lambda\log V\left(z\mid x\right)-\nu_x\right) - \\
&\quad -\sum_z W\left(z\mid x\right)\log\sum_{x^{'}\in\mathcal{X}} P_X\left(x^{'}\right)\exp\left(\lambda\log V\left(z\mid x^{'}\right)-\nu_{x'}\right) - \\
&\quad -P_X\left(x\right)\sum_z W\left(z\mid x\right)\frac{\exp(\lambda\log V(z|x)-\nu_x)}{\sum\limits_{x^{'}\in\mathcal{X}} P_X\left(x^{'}\right)\exp\left(\lambda\log V\left(z|x^{'}\right)-\nu_{x^{'}}\right)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_z W(z \mid x) \log \frac{\exp(\lambda \log V(z|x) - \nu_x)}{\sum\limits_{x' \in \mathcal{X}} P_X\left(x'\right) \exp\left(\lambda \log V\left(z|x'\right) - \nu_{x'}\right)} - \\
&\quad - P_X(x) \sum_z W(z \mid x) \frac{\exp(\lambda \log V(z|x) - \nu_x)}{\sum\limits_{x' \in \mathcal{X}} P_X\left(x'\right) \exp\left(\lambda \log V\left(z|x'\right) - \nu_{x'}\right)} \\
&= \sum_z W(z \mid x) \log h(x,z) - P_X(x) \sum_z W(z \mid x) h(x,z),
\end{aligned}
$$

where

$$
h(x,z) = \frac{\exp(\lambda \log V(z|x) - \nu_x)}{\sum\limits_{x' \in \mathcal{X}} P_X\left(x'\right) \exp\left(\lambda \log V\left(z|x'\right) - \nu_{x'}\right)}.
$$

Thus the maximizing $P_X(x)$ is of following form

$$
P_X(x) = \frac{\sum\limits_z W(z \mid x) \log h(x,z)}{\sum\limits_z W(z \mid x) h(x,z)}. \tag{5.53}
$$

5.7.3 Proof of the converse of the channel coding theorem with mismatched decoder 5.5

The proof of this converse theorem is given in [51]. In this subsection, we present some complements which were unproved in [51] . Take a random code ensemble of length n and rate R with the input distribution P_X, we will show that

$$
\begin{aligned}
&\overline{Pe}_B(W,V) \geq 1 - \exp\left(-e^{n(R - C_{opp}(W))}\right) - \\
&- \sum_{b \in \mathcal{Z}} \left[e^{-nD(P_Z(b)+\epsilon \| P_Z(b))} + e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}\right] - \exp\left[-nD\left(P^*_{XZ} \parallel P_X W\right)\right].
\end{aligned} \tag{5.54}
$$

Let $|\mathcal{M}_n| = e^{nR}$ codewords of codebook $\mathcal{C}_n$ be chosen randomly and independently codeword by codeword and symbol by symbol according to distribution $P_X = \{P(x), x \in \mathcal{X}\}$. Moreover, the empirical distribution of each codeword is constrained to be in the set $T^n_\epsilon(P)$,

$$
T^n_\epsilon(P) = \left\{x^n : \left|\frac{n(a \mid x^n)}{n} - P(a)\right| \leq \epsilon, \forall a \in \mathcal{X}\right\}. \tag{5.55}
$$

Given a channel $\{\mathcal{X}, W^n(\mathbf{z} \mid \mathbf{x}), \mathcal{Z}\}$ and a mismatched decoding metric $d_V(x,z) = \log V(z \mid x)$, consider a codebook of e^{nR} codewords where each codeword is generated at random according to the probability distribution $P(x)$ and independently of all other codewords. Furthermore, the empirical distribution of each codeword is constrained to be in the set $T^n_\epsilon(P)$,

$$T_{\epsilon}^{n}\left(P_X\right)=\left\{x^n:\left|\frac{n\left(x\mid x^n\right)}{n}-P_X\left(x\right)\right|\leq\epsilon,\forall x\in\mathcal{X}\right\}. \tag{5.56}$$

The output sequence z^n should be in the set

$$T_{\epsilon}^{n}\left(P_Z\right)=\left\{x^n:\left|\frac{n\left(b\mid z^n\right)}{n}-P_Z\left(b\right)\right|\leq\epsilon,\forall b\in\mathcal{Z}\right\}, \tag{5.57}$$

where $P_Z\left(z\right)=\sum_{x\in\mathcal{X}}P_X\left(x\right)W\left(z\mid x\right)$ (cf. (2.31)).

The average over the ensemble of code is

$$\begin{aligned}Pe_V^{(n)}\left(W\right) &= \sum_{m=1}^{|\mathcal{M}_n|}Pe_V^{(n)}\left(W\mid m\right)\Pr\left(m\right)\\ &= \sum_{i=1}^{|\mathcal{M}_n|}Pe_V^{(n)}\left(W\mid m\right)\frac{1}{|\mathcal{M}_n|}=Pe_V^{(n)}\left(W\mid m\right),\end{aligned}$$

where the last inequality is due to the symmetry of the random code construction.

Consider the threshold decoder which find all $m^{'}$ such that $\left(x^n\left(m^{'}\right),z^n\right)\in T_{\delta}^{-}$, where $\delta>0$,

$$T_{\delta}^{-}=\left\{\left(\boldsymbol{x},\boldsymbol{z}\right):\log V\left(\boldsymbol{z}\mid\boldsymbol{x}\right)\leq n\left(-D+\delta\right)\right\}. \tag{5.58}$$

The error decoding probability when sending message m

$$\begin{aligned}Pe_V^{(n)}\left(W\mid m\right) &= 1-Pr\left\{V\left(Z^n\mid X^n\left(m^{'}\right)\right)<V\left(Z^n\mid X^n\left(m\right)\right)\quad\text{for all}\,m^{'}\neq m\mid m\right\}\\ &\geq 1-Pr\left\{V\left(Z^n\mid X^n\left(m^{'}\right)\right)<V\left(Z^n\mid X^n\left(m\right)\right)\right.\\ &\qquad \left.\text{for all}\,m^{'}\neq m,\,\left(X^n\left(m\right),Z^n\right)\in T_{\delta}^{-},Z^n\in T_{\epsilon}\left(P_Z\right)\mid m\right\}\\ &\qquad -\Pr\left\{Z^n\in T_{\epsilon}^{c}\left(P_Z\right)\mid m\right\}-\Pr\left\{\left(X^n\left(m\right),Z^n\right)\in\left[T_{\delta}^{-}\right]^{c}\mid m\right\}.\end{aligned} \tag{5.59}$$

As in [51], for ϵ small enough , we have the upper bound for the first term of 5.59 :

$$\begin{gathered}Pr\left\{V\left(Z^n\mid X^n\left(m^{'}\right)\right)<V\left(Z^n\mid X^n\left(m\right)\right)\right.\\ \left.\text{for all}\,m^{'}\neq m,\,\left(X^n\left(m\right),Z^n\right)\in T_{\delta}^{-},Z^n\in T_{\epsilon}\left(q\right)\mid m\right\}\\ \leq\exp\left(-e^{n\left(R-I_{LM}\left(X;Z\right)\right)}\right)\\ \leq\exp\left(-e^{n\left(R-C_{LM}\left(W,V\right)\right)}\right).\end{gathered}$$

The last two terms of (5.59), $\Pr\left\{Z^n\in T_{\epsilon}^{c}\left(P_Z\right)\mid m\right\}$ and $\Pr\left\{\left(X^n\left(m\right),Z^n\right)\in\left[T_{\delta}^{-}\right]^{c}\mid m\right\}$, will vanish as $n\to\infty$ by the weak law of large numbers.

Firstly, we will find the upper bound for the term $\Pr\left\{Z^n \in T_\epsilon^c\left(P_Z\right) \mid m\right\}.$

It is demonstrated in the subsection 5.7.4 that for every $b \in \mathcal{Z}$ such that $P_Z\left(b\right) + \epsilon < 1$ and $P_Z\left(b\right) - \epsilon > 0$,

$$Pr\left\{\mid \frac{n\left(b \mid Z^n\right)}{n} - P_Z\left(b\right) \mid > \epsilon \mid m\right\} \leq e^{-nD\left(P_Z(b)+\epsilon \| P_Z(b)\right)} + e^{-nD\left(P_Z(b)-\epsilon \| P_Z(b)\right)}, \quad (5.60)$$

where

$$D\left(P_Z\left(b\right) + \epsilon \parallel P_Z\left(b\right)\right) = \left[q\left(b\right) + \epsilon\right] \log \frac{q\left(b\right) + \epsilon}{q\left(b\right)} + \left[1 - q\left(b\right) - \epsilon\right] \log \frac{1 - q\left(b\right) - \epsilon}{1 - q\left(b\right)},$$

$$D\left(P_Z\left(b\right) - \epsilon \parallel P_Z\left(b\right)\right) = \left[q\left(b\right) - \epsilon\right] \log \frac{q\left(b\right) - \epsilon}{q\left(b\right)} + \left[1 - q\left(b\right) + \epsilon\right] \log \frac{1 - q\left(b\right) + \epsilon}{1 - q\left(b\right)}.$$

Thus,

$$\begin{aligned} \Pr\left\{Z^n \in T_\epsilon^c\left(P_Z\right) \mid m\right\} &= \Pr\left\{\bigcup_{b \in \mathcal{Z}} \left[\mid \frac{n\left(b \mid Z^n\right)}{n} - q\left(b\right) \mid > \epsilon\right] \mid m\right\} \\ &\leq \sum_{b \in \mathcal{Z}} \Pr\left\{\mid \frac{n\left(b \mid Z^n\right)}{n} - P_Z\left(b\right) \mid > \epsilon \mid m\right\} \qquad (5.61) \\ &\leq \sum_{b \in \mathcal{Z}} \left[e^{-nD\left(P_Z(b)+\epsilon \| P_Z(b)\right)} + e^{-nD\left(P_Z(b)-\epsilon \| P_Z(b)\right)}\right]. \end{aligned}$$

Now we will find the upper bound for the last term $\Pr\left\{\left(X^n\left(m\right), Z^n\right) \in \left[T_\delta^-\right]^c \mid m\right\}.$

Using Sanov's theorem, we prove the following result in the subsection 5.7.5.

For n large enough we have

$$\Pr\left\{\left(X^n\left(m\right), Z^n\right) \in \left[T_\delta^-\right]^c \mid m\right\} \leq \exp\left[-nD\left(P_{XZ}^* \parallel P_X W\right)\right], \qquad (5.62)$$

where

$$P_{XZ}^* = \underset{\hat{P}_{XZ} \in E_m}{\arg\min} D\left(\hat{P}_{XZ} \parallel P_X W\right), \qquad (5.63)$$

and E_m is defined in (5.71).

In conclusion, for n sufficiently large we have,

$$1 - Pe_{dV}^{(n)}\left(W\right) \leq \exp\left(-e^{n\left(R - C_{opp}(W)\right)}\right) + \sum_{b \in \mathcal{Z}} \left[e^{-nD\left(P_Z(b)+\epsilon \| P_Z(b)\right)} + e^{-nD\left(P_Z(b)-\epsilon \| P_Z(b)\right)}\right] +$$

$$+ \exp\left[-nD\left(P_{XZ}^* \parallel P_X W\right)\right]. \qquad (5.64)$$

5.7.4 Proof of (5.60)

We will show that for every $b \in \mathcal{Z}$, such that $P_Z(b) + \epsilon < 1$, $P_Z(b) - \epsilon > 0$

$$\Pr\left\{ \mid \frac{n(b \mid Z^n)}{n} - P_Z(b) \mid > \epsilon \mid m \right\} \leq e^{-nD(P_Z(b)+\epsilon \| P_Z(b))} + e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}.$$

Using Chernoff bound, for all $\nu > 0$,we have

$$\begin{aligned}
\Pr\left\{ \mid \frac{n(b \mid Z^n)}{n} - P_Z(b) \mid > \epsilon \mid m \right\} &\leq \Pr\left\{ \frac{n(b \mid Z^n)}{n} \geq \epsilon + P_Z(b) \mid m \right\} \\
&\leq \mathbb{E}\left[e^{\nu \frac{n(b|Z^n)}{n} | m} \right] e^{-\nu(\epsilon + P_Z(b))} \\
&= \left[\sum_{k=0}^{n} \Pr\left\{ n(b \mid Z^n) = k \mid m \right\} e^{\nu k/n} \right] e^{-\nu(\epsilon + P_Z(b))} \\
&= \left[\sum_{k=0}^{n} C_n^k P_Z(b)^k (1 - P_Z(b))^{n-k} e^{\nu k/n} \right] e^{-\nu(\epsilon + P_Z(b))} \\
&= \left[(1 - P_Z(b)) + P_Z(b) e^{\nu/n} \right]^n e^{-\nu(\epsilon + P_Z(b))}.
\end{aligned}$$

Therefore,

$$\Pr\left\{ \frac{n(b \mid Z^n)}{n} > \epsilon + P_Z(b) \mid m \right\} \leq \min_{\nu > 0} \left[(1 - P_Z(b)) + P_Z(b) e^{\nu/n} \right]^n e^{-\nu(\epsilon + P_Z(b))}. \tag{5.65}$$

The right hand side of 5.65 achieves the minimum value when

$$\begin{aligned}
\nu &= \infty && \text{if } P_Z(b) + \epsilon \geq 1, \\
e^{\nu/n} &= \frac{(1 - P_Z(b))(\epsilon + P_Z(b))}{P_Z(b)(1 - P_Z(b) - \epsilon)} && \text{if } P_Z(b) + \epsilon < 1.
\end{aligned} \tag{5.66}$$

Plugging $e^{\nu/n} = \dfrac{(1 - P_Z(b))(\epsilon + P_Z(b))}{P_Z(b)(1 - P_Z(b) - \epsilon)}$ into 5.65, we have

$$\begin{aligned}
\Pr\left\{ \frac{n(b \mid Z^n)}{n} > \epsilon + P_Z(b) \mid m \right\} &\leq \left[\left(\frac{P_Z(b)}{P_Z(b) + \epsilon} \right)^{P_Z(b)+\epsilon} \left(\frac{1 - P_Z(b)}{1 - P_Z(b) - \epsilon} \right)^{1 - P_Z(b) - \epsilon} \right]^n \\
&= e^{-nD(P_Z(b)+\epsilon \| P_Z(b))}.
\end{aligned} \tag{5.67}$$

Similarly, for $P_Z(b) - \epsilon > 0$, we have

$$\Pr\left\{\frac{n(b \mid Z^n)}{n} < P_Z(b) - \epsilon \mid m\right\} \leq e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}. \tag{5.68}$$

Therefore,

$$\Pr\left\{\mid \frac{n(b \mid Z^n)}{n} - P_Z(b) \mid > \epsilon \mid m\right\} \leq e^{-nD(P_Z(b)+\epsilon \| P_Z(b))} + e^{-nD(P_Z(b)-\epsilon \| P_Z(b))}. \tag{5.69}$$

5.7.5 Proof of (5.62)

$$\begin{aligned}
&\Pr\left\{(X^n(m), Z^n) \in \left[T_\delta^-\right]^c \mid m\right\} \\
&= \Pr\left\{\log V(Z^n \mid X^n(m)) > n(-D+\delta) \mid m\right\} \\
&= \Pr\left\{\frac{1}{n}\sum_{j=1}^{n} \log V(Z_j \mid X_j(m)) > (-D+\delta) \mid m\right\} \\
&= \Pr\left\{\frac{1}{n}\sum_{a\in\mathcal{X}, b\in\mathcal{Z}} n(a,b \mid X^n(m), Z^n) \log V(b \mid a) > -D+\delta \mid m\right\} \\
&\leq \Pr\left\{\sum_{a\in\mathcal{X}, b\in\mathcal{Z}} \hat{P}_{X^n(m)Z^n}(a,b) \log V(b \mid a) \geq -D+\delta \mid m\right\}.
\end{aligned} \tag{5.70}$$

Denote

$$E_m = \left\{\hat{P}_{X^n(m)Z^n} : \sum_{a\in\mathcal{X}, b\in\mathcal{Z}} \hat{P}_{X^n(m)Z^n}(a,b) \log V(b \mid a) \geq -D+\delta\right\}. \tag{5.71}$$

Applying Sanov's theorem 2.15 and the Chernoff bound (3.61) we have

$$\begin{aligned}
\Pr\left\{(X^n(m), Z^n) \in \left[T_\delta^-\right]^c \mid m\right\} &= P_{XZ}^n\left(\hat{P}_{X^n(m)Z^n} \in E_m\right) \\
&\leq \exp\left[-nD\left(P_{XZ}^* \| P_X W\right)\right],
\end{aligned} \tag{5.72}$$

where

$$P_{XZ}^* = \underset{\hat{P}_{XZ}\in E_m}{\arg\min} D\left(\hat{P}_{XZ} \| P_X W\right). \tag{5.73}$$

5.7.6 The Lagrange dual problem

For the sake of completeness, we briefly present some basic knowledge about optimization which we used in this dissertation. For more general discussion, we refer readers to the book of Stephen Boyd [14].

Primal problem P

$$\begin{array}{ll} \text{minimize} & f\left(x\right) \\ & \\ \text{subject to} & \begin{array}{c} g\left(x\right)\leq 0 \\ h\left(x\right)=0 \\ x\in D \end{array} \end{array} , \tag{5.74}$$

where $f:\mathbb{R}^m\rightarrow\mathbb{R}$ and $g:\mathbb{R}^m\rightarrow\mathbb{R}$ and $h:\mathbb{R}^m\rightarrow\mathbb{R}$.

The problem 5.74 is call the *primal problem.*

We define the *Lagrangian* $F:\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}\rightarrow\mathbb{R}$ associated with the problem 5.74 as follows:

$$F\left(x,u,v\right)=f\left(x\right)+ug\left(x\right)+vh\left(x\right). \tag{5.75}$$

Lagrange dual problem D

$$\begin{array}{ll} \text{maximize} & \theta\left(u,v\right) \\ & \\ \text{subject to} & u\geq 0 \end{array} , \tag{5.76}$$

where

$$\theta\left(u,v\right)=\inf_{x\in D}F\left(x,u,v\right), \tag{5.77}$$

is the Lagrange dual function.

Note that when the Lagrangian is unbounded in D then the Lagrange dual function $\theta\left(u,v\right)=-\infty$.

The problem 5.76 is called the *Lagrange dual problem* associated with the original problem 5.74, which is called the *primal problem.*

Let p^* be the optimal value of the problem 5.74. For any $u\geq 0$and any v, we have

$$\theta\left(u,v\right)\leq p^*. \tag{5.78}$$

Denote $d^*$the optimal value of the Lagrange dual problem 5.76 , then we always have $d^*\leq p^*$ which is called weak duality. If the equality $d^*=p^*$holds, we call that strong duality holds. The strong duality in general does not hold. The following theorem says that under some condition, we have the strong duality. Without loss of generality we can assume that the primal problem 5.74 has a finite optimal value i.e. $p^*>-\infty$, because if $p^*=-\infty$, then by weak duality we have $d^*=-\infty$.

Theorem 5.8. *(Strong Duality Theorem.) Let D be a nonempty convex set in $\mathbb{R}^m$. Let $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$ be convex and $h : \mathbb{R}^m \to \mathbb{R}$ be affine. Suppose that the Slater's condition is satisfied. It means that there exists an $\widetilde{x} \in D$ such that $g\left(\widetilde{x}\right) < 0$ and $h\left(\widetilde{x}\right) = 0$ and $0 \in \operatorname{int} h\left(D\right)$, where $h(D) = \{h\left(x\right) : x \in D\}$. Then*

$$\inf\left\{f\left(x\right) : x \in D, g\left(x\right) \leq 0, h\left(x\right) = 0\right\} = \sup\left\{\theta\left(u, v\right) : u \geq 0\right\}. \tag{5.79}$$

Moreover, if $p^* > -\infty$ then $\sup\left\{\theta\left(u, v\right) : u \geq 0\right\}$ is achieved at $\left(u^*, v^*\right)$ with $u^* \geq 0$. If $\inf\left\{f\left(x\right) : x \in D, g\left(x\right) \leq 0, h\left(x\right) = 0\right\}$ is achieved at x^* then $u^* g\left(x^*\right) = 0$.

Chapter 6

Conclusions and perspectives

6.1 Conclusions

In this dissertation we have studied the problem of authentication using graphical codes from an information-theoretic security point of view. Particularly, the core of this dissertation focuses on minimizing both the probability of false alarm and the probability of non-detection.

In the chapter 3 and 4, we have developed and analyzed the framework of authentication without channel coding. Tackling the problem using hypothesis testing, we have invoked the optimal Neyman-Pearson test to perform authentication and two types of error probability have been computed by using either Gaussian approximation or arguments relying on Sanov's theorem. It should be mentioned that in this analysis the knowledge about the opponent channel is required, and this can be obtained by estimating the parameters of the opponent channel based on the observation of GCs coming from the opponent. We have also introduced two possible strategies for the receiver which are binary thresholding and gray level observation. Using asymptotic results of probabilities type I and II, we have showed that the gray level observation strategy gives a better authentication performance than binary thresholding.

The numerical estimation of the probabilities of false alarm and non-detection have been obtained by specific Monte-Carlo simulations and these results have confirmed that the error probabilities using asymptotic expressions based Sanov's theorem are accurate comparing to the Gaussian approximation. We have also proved that it is entirely possible to estimate very small values of the two types of error probabilities by employing MC simulation using the importance sampling. It has been presented in chapter 4 that we are also able to optimize the authentication performance provided that the knowledge about the model of the print and scan channel is available.

Chapter 5 analyzed the authentication problem by the use of channel coding. In this setting, a secret message is encoded into a codeword by a deterministic encoder before mapping to the GC with the presence of the secret key which is only known by the receiver. As the receiver does not know whether the observed sequence comes from the legitimate source or from the opponent, we propose to use only one decoding rule matching with the main channel and mismatching with the opponent channel, and consequently to use the theory of mismatch decoding for authentication purposes. We have demonstrated that the enhancement of authentication performance is possible by choosing a code with rate between the mismatch capacity of the opponent channel and that of the main channel. In particular, by choosing such a code, we can achieve simultaneously as small as possible probabilities of false alarm and non-detection. However, we have only established the existence of such codes but not specified how to construct them. We have also discussed the extension of achievable rates of such codes. More precisely, if the rate is between the mismatched capacity of the opponent and the capacity of the main channel it is then possible to show the existence of a code making the probability of false alarm arbitrarily small and the existence of a code making the probability of non detection negligible. Unfortunately whether these two codes are identical or not is still an open problem. We have also presented curves plotting the success probability in the uncoded scheme and the upper bound of success probability in the case using channel coding. These results again confirm the theoretical advantages of using channel coding. The last part of chapter 5 proposed a practical scheme using parallel concatenated codes with turbo decoding. By employing a specific concatenation of several codes separated by interleavers referred as turbo codes, we have analyzed the channel optimization for authentication. The investigation of the EXIT chart has played an important role in choosing channels' parameters so that the best authentication is achieved.

6.2 Perspectives

The following points are some of the directions where our work could be extended.

In the chapter 3, we primarily employed the likelihood ratio test between the main and opponent channels distributions as a basic discriminating measure to perform authentication and the knowledge about the distribution of the opponent channel is consequently necessary. An interesting direction to consider is that how to do hypothesis testing without knowing the distribution of the opponent channel. Research on hypothesis testing in which the hypothesis H_1 is unknown has been studied extensively [75], [27], [24]. In [75], O. Zeitouni et al. discussed when the generalized likelihood ratio test is optimal. The behavior of two types of errors when they are very small is also studied [24].

In this dissertation we concentrated on the i.i.d. sequences. It is interesting to investigate what we can do when the data is just independent but not identically distributed. One promising solution is that we may use Gartner-Ellis's theorem [16] instead of Sanov to estimate asymptotically the two types of error when they are in large deviation. In

Gartner-Ellis, one might say that we are able to relax the condition on independence and perhaps identical distribution too, in a controlled way [2].

It is also essential to study if our work can be applicable for other applications such as Steganalysis [17] and Forensics [68]. In particular, one might use Sanov's theorem to asymptotically estimate the very small values of probabilities of false alarm and non-detection when doing hypothesis testing.

As noted in the chapter 5, the region of rate R is relaxed in the interval $\left[C\left(W\right), C\left(V\right)\right]$ so that we can establish the existence of codes such that the two types of error can be made arbitrarily small at the same time. Therefore, a question that still remains is whether one can show the existence of codes with rate in the larger interval, greater than the mismatched capacity of the opponent and less than the capacity of the main channel, so that the two types of error probability are negligible. One might probably establish these desired codes by deleting the worse codewords or concatenating the two codes which makes the false alarm small and which makes the probability of non-detection small on the opponent channel.

A very promising problem relating to using channel coding in authentication performance is to employ the stochastic coding strategies. More precisely, we use a randomized encoder in which the transmitted codeword is chosen randomly in a set of codewords representing for the same authentication message. This randomness plays a crucial role in confusing the opponent. Research on the wiretap codes has been studied extensively [20], [73], [70], [47]. Remarkably, in [47], Bloch and Laneman study the channel-resolvability-based constructions, which associate to each message a subcode that operates just above the resolvability of the eavesdropper's channel. It is showed that the channel resolvability enables to achieve strong secrecy. In this dissertation, we often circumvented the problem of designing codes achieving authentication performance. It might be possible to construct such codes thanks to many works related to code designs achieving strong secrecy such as polar codes [48], [6], [57], LDPC codes [69], [63], [64] etc.

Chapitre 7

Résumé en Français

7.1 Introduction

7.1.1 Présentation du sujet

Cette thèse apporte une approche théorique sur le problème pratique de l'authentification de produits ou documents par codes graphiques. Par "authentification", nous entendons ici la confirmation que le support sur lequel se trouve le code graphique est original, c'est à dire qu'il ne provient pas d'une copie, qu'il n'est pas faux. Ce type de méthodes peut être utilisé pour attester qu'un document d'identité, qu'un diplôme, qu'une boite de médicaments, ... ne sont pas contrefaits.

A l'instar des méthodes biométriques, le principe d'authentification étudié repose sur des caractéristiques uniques et non clonables. Si dans le domaine de la biométrie nous pouvons utiliser des caractéristiques issue d'empreintes digitales, elles proviennent ici de l'interaction entre la disposition aléatoire des fibres de papier et de l'encre d'impression. Le principe d'authentification par codes graphiques (CG) repose sur une technologie développée par l'entreprise "Advanced Track and Trace" participant au projet ANR "Estampille", il se décompose en plusieurs étapes illustrées sur la Figure 7.1 et détaillées dans la liste suivante :

1. Le code graphique (CG) est représenté par un code binaire ou chaque 0 sera lu par l'imprimante comme l'impression d'un point (dot en anglais) et chaque 1 comme un endroit laissé vierge. Ce code pourra être aléatoire (voir chapitres 3 et 4 et 1.a sur la figure) ou structuré (cf chapitre 5 et 1.b sur la figure). Il sera également généré à partir d'une clé secrète connue uniquement par l'imprimeur et le receveur cherchant à authentifier le code.

2. Le CG est imprimé sur un support papier à l'aide d'une imprimante professionnelle haute résolution de type "offset" pouvant fournir 2400 points par pouce. Dans ce contexte chaque point fera $10\mu m$ de diamètre (voir étape 2 sur la figure). A cette échelle, l'impression d'un point est un processus aléatoire. La forme et la position

finale du point dépendent de la disposition des fibres de papiers, de la précision de l'imprimante (qui grave les points sur une plaque métallique), mais aussi des particules qui composent l'encre. Une fois imprimé le code est donc dégradé de manière irréversible (il est théoriquement impossible d'estimer précisément le code original à partir du code imprimé), c'est cette priorité qui sera utilisée afin de vérifier l'authenticité d'un CG.

3. Si un contrefacteur souhaite copier le document ou l'emballage sur lequel se trouve le CG, il devra dans un premier temps le scanner (voir étape 3 sur la figure). Il s'agit également d'un processus aléatoire (ajout d'un bruit provenant du capteur utilisé lors de l'acquisition).

4. Dans un second temps il devra également le binariser puisqu'une imprimante n'est capable d'imprimer qu'une information binaire à haute résolution (voir étape 4 de la figure). Comme dans l'étape 1, cette information code la présence ou l'absence d'un point. De part le caractère aléatoire du processus d'impression acquisition, le code binarisé sera en pratique différent du code original.

5. Le contrefacteur génère son faux CG en imprimant le code binaire estimé (étape 5 de la figure).

6. Le receveur observe un CG imprimé et scanné et doit décider s'il provient d'un code original ou d'un code contrefait. Pour cela il utilise le code graphique original (généré à partir de la clé secrète) et le code observé pour construire un test d'hypothèse permettant d'accepter ou de rejeter le code observé. Dans le chapitre 5 nous envisagerons l'utilisation d'une version binaire du code observé (étapes 6a et 7a de la figure) ou l'utilisation de sa version en niveaux de gris (étapes 6b et 7b de la figure).

7.1.2 Problèmes étudiés

Cette thèse distingue deux scénarios : dans le premier le message binaire pseudo-aléatoire servant à l'authentification constitue directement le CG, dans le second ce message est codé par l'utilisation de systèmes de codage canal.

Lorsqu'il s'agira des CG non-codés, nous répondrons aux questions suivantes :

- sous l'hypothèse que les modèles statistiques des canaux impression/acquisition de l'imprimeur légitime et du contrefacteur sont connus, comment obtenir un test optimal permettant de garantir une probabilité de fausse alarme[1] donnée tout en minimisant la probabilité de non détection[2] ?

- comment calculer théoriquement cette probabilité de non détection, de surcroit lorsqu'elle est très faible ?

[1] c.a.d. la probabilité de juger un code original comme étant falsifié.

[2] c.a.d. la probabilité de juger un code falsifié comme étant original.

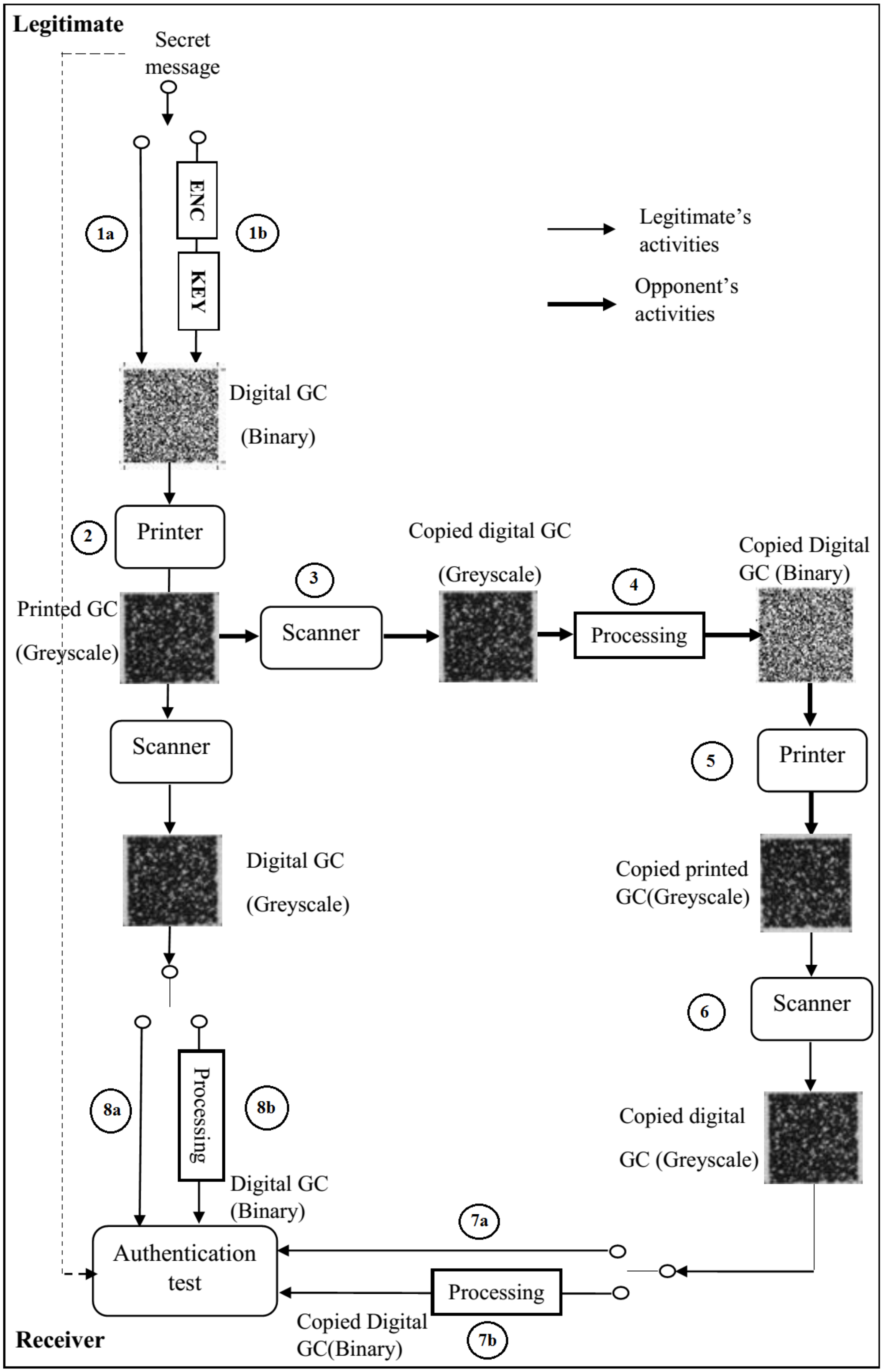


Figure 7.1: Principe de l'authentification basée sur des Codes Graphiques (CG).

- comment la calculer pratiquement ?
- quel est l'apport d'une observation en niveau de gris par rapport à une observation binaire ?
- quel est l'impact du canal d'impression ?

Lorsqu'il s'agira d'utiliser le codage canal déterministe, nous traiterons les problèmes suivants :

- quel modèle de codage canal est associé à notre problème d'authentification ?
- comment mesurer la probabilité de non détection pour une probabilité de fausse alarme très faible ?
- dans quels cas pouvons-nous montrer que l'utilisation de codes déterministes permettra d'améliorer les performances d'authentification ?
- comment pratiquement implémenter de tels codes, et quels performances obtient-on en pratique ?
- lors de l'utilisation de codes concaténés, comment les construire de manière à obtenir de bonnes performances en authentification ?

7.2 Fondements théoriques et pratiques

7.2.1 Fondements théoriques

Cette thèse utilise des éléments propres à la théorie de l'information et à la théorie du codage.

Le calcul précis des probabilités de fausse alarme et de non-détection repose sur la méthode des types et le théorème de Sanov.

La méthode des types [19] permet d'encadrer la probabilité $Q^n\left(\hat{P}_{X^n} = \hat{P}\right)$ qu'un type $\hat{P}$ (aussi appelé histogramme en traitement du signal) soit observé en fonction de la divergence $D\left(\hat{P} \parallel Q\right)$ entre le type et sa distribution sous-jacente Q (voir Lemme (2.13)).

Le théorème de Sanov est un résultat important lié à l'utilisation des types : il permet de calculer la limite, lorsque la taille de la séquence servant à construire le type tend vers l'infini, de la probabilité $Q^n\left(\hat{P}_{X^n} \in E\right)$, c'est à dire la probabilité qu'un type appartienne a une région convexe E de l'espace des distributions (voir Théorème (2.15)). Cette fois-ci c'est la borne inférieure de la divergence $(\inf_{P \in E} D(P \parallel Q))$ entre les distributions appartenant à la région et la distribution sous-jacente qui intervient.

Dans le cas où le CG est généré à partir d'un système de codage canal, le receveur de notre schéma d'authentification fait face à un problème de décodage inadapté ("mismatched decoding" en anglais). En effet le message qu'il cherche à décoder peut soit venir d'un CG original (décodage complet), soit d'un CG falsifié (décodage inadapté). Dans ce deuxième scénario, la quantité d'information transmissible sans erreurs dans ce canal peut être maximisée par la capacité inadaptée ("mismatched capacity" en anglais) étudiée par Lapidoth [42] et présentée dans le théorème (2.31).

Notre problème peut se rapprocher du problème d'authentification de messages proposé par Simmons [61] puis repris par Maurer [50], et qui consiste à authentifier un message crypté transmis sur un canal non bruité. Dans ce contexte, le schéma d'authentification doit faire un compromis entre deux types d'attaques : une attaque par substitution pour laquelle l'adversaire cherchera à inférer la clé secrète et substituer un message observé par le sien, et une attaque où l'adversaire cherchera à tromper le receveur et à prendre la place de l'émetteur en envoyant un message quelconque .

L'utilisation du bruit comme moyen d'authentification a été étudiée par Lai et al. [41] en appliquant la théorie des canaux sur écoute à l'authentification [73]. En utilisant une stratégie de codage aléatoire, les auteurs montrent que dans ce cas-ci la probabilité de succès de l'adversaire est encadrée par l'entropie de la clé utilisée pour chiffrer le message à authentifier. Notons toutefois que l'authentification par codes graphiques ne peut pas directement utiliser la théorie des canaux sur écoute puisque l'adversaire n'a dans notre cas pas accès à une version plus dégradée que le receveur légitime.

7.2.2 Fondements pratiques

L'authentification de personnes par moyens biométriques montrent des similarités avec l'authentification de messages mais aussi des différences. Les systèmes biométriques diffèrent de l'authentification de messages car ils se décomposent en deux étapes : l'enrôlement (acquisition et protection de la caractéristique biométrique de référence) et l'authentification (comparaison entre la caractéristique test et la caractéristique enrôlée). Comme les systèmes d'authentification de message, elle doit faire face à un canal bruité (ici provenant de l'acquisition), mais aussi assurer que la caractéristique enrôlée reste privée et ne permette pas de retrouver la biométrie de référence. Il y a encore une fois un compromis à trouver entre robustesse et confidentialité des données biométriques.

Les extracteur flous ("fuzzy extractors" en anglais) permettent de satisfaire ces deux contraintes via l'extraction d'une clé et une donnée auxiliaire lors de l'enrôlement. La clé sera stockée sur le serveur et la donnée auxiliaire restera publique et aidera à faire face au bruit lors de l'étape d'authentification. Ce concept est étendu par Ignatenko et

Willems [35] qui le rapproche du principe de partage de clé introduit par Maurer[49]. Ici la caractéristique enrôlée et la caractéristique légitime à authentifier partagent une clé commune et le message échangé représente la donnée auxiliaire. Ce message doit être généré de façon à minimiser la fuite d'information sur la clé et sur la biométrie de référence.

L'authentification de contenus physiques a été étudiée par Beekhof et al. [10], elle repose ici sur l'utilisation d'empreintes numériques binaires (pouvant s'apparenter à des CG) et étudie le cas ou un adversaire à une connaissance partielle du code original. En utilisant un test d'hypothèse, les auteurs montrent qu'il est possible de trouver un seuil minimisant le maximum de la probabilité de non-détection et de la probabilité de fausse-alarme.

Enfin, le cadre pratique de cette thèse prend source dans les travaux de Picard et al. [53],[54] qui présentent le principe de base d'authentification par CG. Ces travaux proposent d'utiliser le taux d'erreur binaire entre le CG original et le CG soumis comme score permettant l'authentification. Comme nous le verrons dans la section (7.3), cette stratégie n'est pas optimale.

7.3 Authentification par tests d'hypothèses

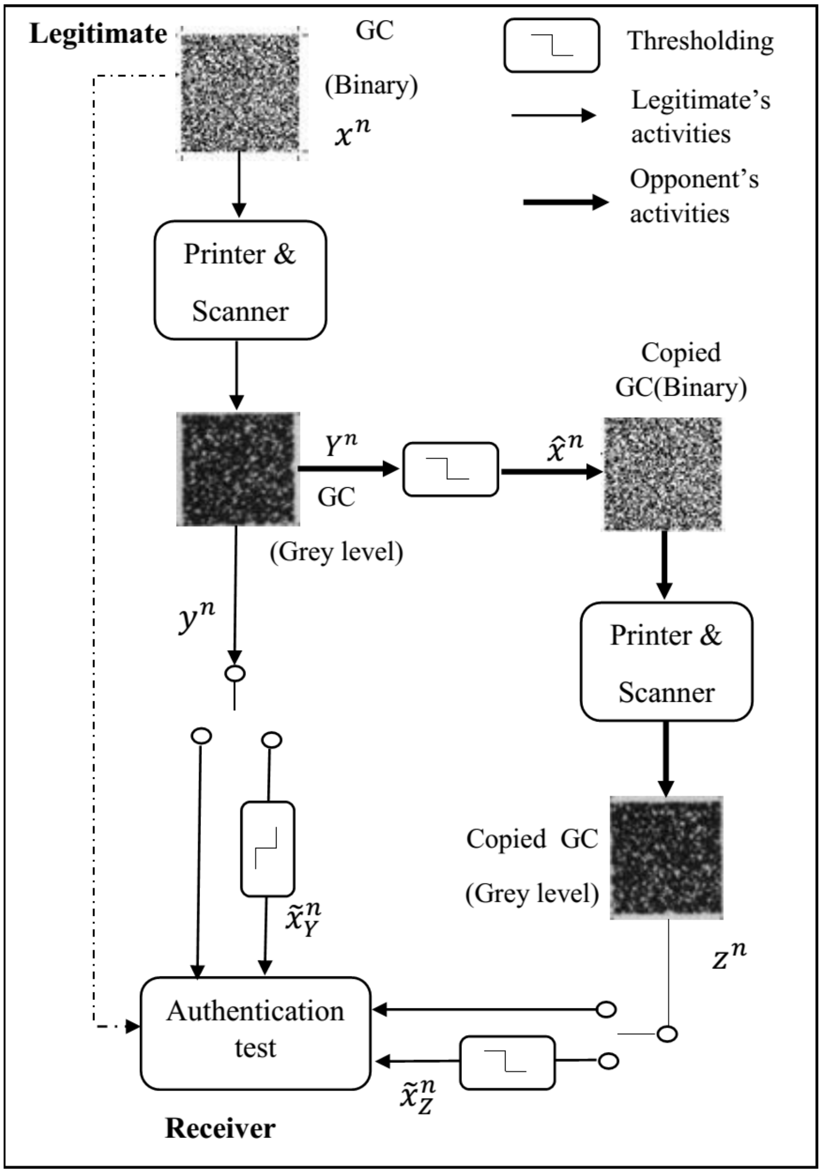


Figure 7.2: Authentification en utilisant des codes graphiques.

7.3.1 Principes du système d'authentification

Dans un premier temps, nous sommes partis de l'hypothèse que le receveur a en sa connaissance :

- d'une part le modèle statistique d'impression-acquisition des codes imprimés originaux $P_{Y^N|X^N}\left(v^N \middle| x^N\right)$ où X^N est le vecteur aléatoire représentant le code original et Y^N le vecteur aléatoire représentant le code original imprimé,

- d'autre part le modèle statistique d'impression-acquisition des codes contrefaits $P_{Z^N|X^N}\left(v^N\middle|\, x^N\right)$ où Z^N est le vecteur aléatoire représentant le code contrefait.

Ces deux hypothèses se traduisent pratiquement par le fait que si les codes sont imprimés dans des conditions constantes (même papier, même encre, même imprimante, ...) à la fois chez l'imprimeur légitime et le contrefacteur, il sera possible d'estimer précisément les modèles d'impression-acquisition. Nous supposons que ces deux modèles d'impression-acquisition sont i.i.d., et notre méthodologie peut s'appliquer pour différentes distributions comme par exemples des modèles Gaussien ou Lognormaux. Le modèle du contrefacteur pour une valeur de code x donnée est calculé à partir d'un mélange de deux distributions, l'une pour l'impression-acquisition d'un point noir, l'autre pour l'impression-acquisition d'une zone restée blanche et les paramètres de ce mélange sont déterminés à partir de l'erreur commise par le contrefacteur lors de la binarisation du code original.

Soit H_0 l'hypothèse traduisant le fait que l'observation du code reçu o^n est un code original et soit H_1 l'hypothèse traduisant le fait que l'observation du code reçu o^n est un code contrefait. Dans ces conditions, le receveur peut utiliser la stratégie de Neyman-Pearson qui consiste à calculer le rapport de vraisemblance :

$$L = \log\frac{P_{Z^n|X^n}\left(o^n\middle|\, x^n, H_1\right)}{P_{Y^n|X^n}\left(o^n\middle|\, x^n, H_0\right)}, \tag{7.1}$$

et à décider H_0 ou H_1 en comparant ce rapport à un seuil λ garantissant une probabilité de non détection minimale pour une probabilité de fausse alarme inférieure à un niveau α :

$$L \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{7.2}$$

Nous avons dans un premier temps comparé deux types d'observations, le premier suppose que le receveur binarise le code observé avant de calculer son test d'hypothèse, alors que le second type suppose que c'est l'image scannée en niveau de gris qui est directement utilisée comme observation o^n. Nous avons montré que la stratégie consistant à utiliser un code binaire n'est pas optimale dans le sens où pour une probabilité de fausse alarme donnée, la probabilité de non-détection d'un code contrefait est plus importante qu'avec l'utilisation d'un code scanné en niveau de gris. En utilisant des mesures informationnelles, nous pouvons montrer (voir appendice (3.6.1)) que la divergence entre les deux canaux est plus importante sans binarisation des observations.

Il est à noter cependant que d'un point de vue pratique la stratégie de binarisation peut comporter plusieurs avantages puisqu'elle ne nécessite pas la connaissance du canal d'impression du contrefacteur et qu'elle se traduit par un comptage du nombre d'erreurs entre le code observé et le code original.

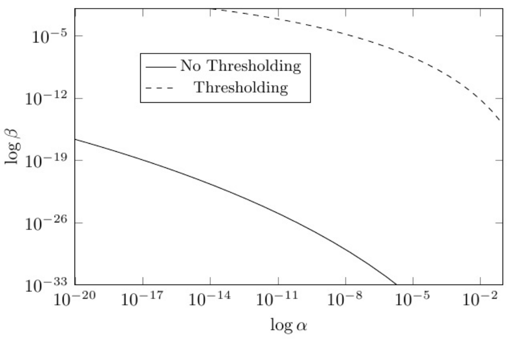


Figure 7.3: Comparaison des courbes ROC avec et sans binarisation. Ici l'utilisation directe de codes scannés en niveau de gris permet d'obtenir des performances en authentification bien supérieures. Modèle Gaussien, $N = 2.10^3$, $\sigma_b = \sigma_w = 52$.

7.3.2 Calcul précis des probabilités d'erreur

La probabilité de fausse alarme α, c'est à dire la probabilité de détecter un code original comme faux, s'exprime comme :

$$\alpha = \Pr\left(L \geq \lambda \mid H_0\right). \tag{7.3}$$

Classiquement, cette probabilité est calculée en invoquant le théorème central limite qui approxime la distribution de la variable aléatoire L par une distribution Gaussienne. Il s'avère cependant que pour une valeur très faible de cette probabilité, cette approximation n'est pas réaliste. Ce problème est identique pour la probabilité de non-détection β.

Nous avons ici utilisé le théorème de Sanov pour calculer une limite α à partir de la divergence entre une loi P_s^* et la loi du processus d'impression-acquisition $P_{Y|X}$, ainsi :

$$\alpha \overset{n\to\infty}{\to} \exp\left(-nD\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right), \tag{7.4}$$

avec P_s^* donnée par (voir aussi la Figure (7.4)) :

$$P_s^*\left(o \mid a\right) = \frac{P_{Y|X}^{1-s}\left(o \mid a\right) P_{Z|X}^{s}\left(o \mid a\right)}{\sum\limits_{o'} P_{Y|X}^{1-s}\left(o' \mid a\right) P_{Z|X}^{s}\left(o' \mid a\right)}, \tag{7.5}$$

s satisfaisant l'équation $D\left(P_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right) - D\left(P_s^* \parallel P_{Z|X} \mid \hat{P}_{x^n}\right) = \dfrac{\lambda}{n}$.

Une expression raffinée peut également être écrite en faisant intervenir la fonction génératrice des moments $g_L(s \mid H_0) = \mathbb{E}_{P_L(L|H_0)}\left[e^{sL}\right]$ et la fonction semi-invariante associée $\mu(s; H_0) = \ln g_L(s \mid H_0)$:

$$\alpha \simeq \frac{1}{\tilde{s}\sqrt{2\pi\mu_L^{''}\left(\tilde{s};\, x^n, H_0\right)}} \exp\left(-nD\left(\hat{P}_s^* \parallel P_{Y|X} \mid \hat{P}_{x^n}\right)\right). \tag{7.6}$$

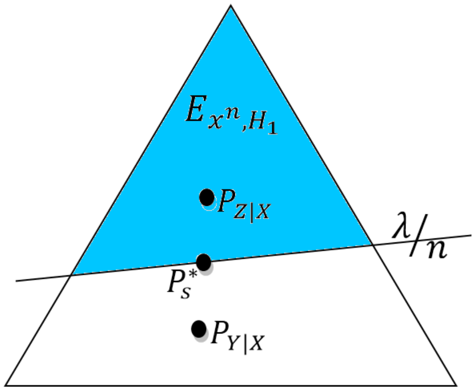


Figure 7.4: Représentation géométrique des distributions $P_{Y/X}$, $P_{Z/X}$ et P_s^*.

Il est également possible d'obtenir des expressions similaires pour la probabilité de non détection β.

7.3.3 Simulations par échantillonnage d'importance

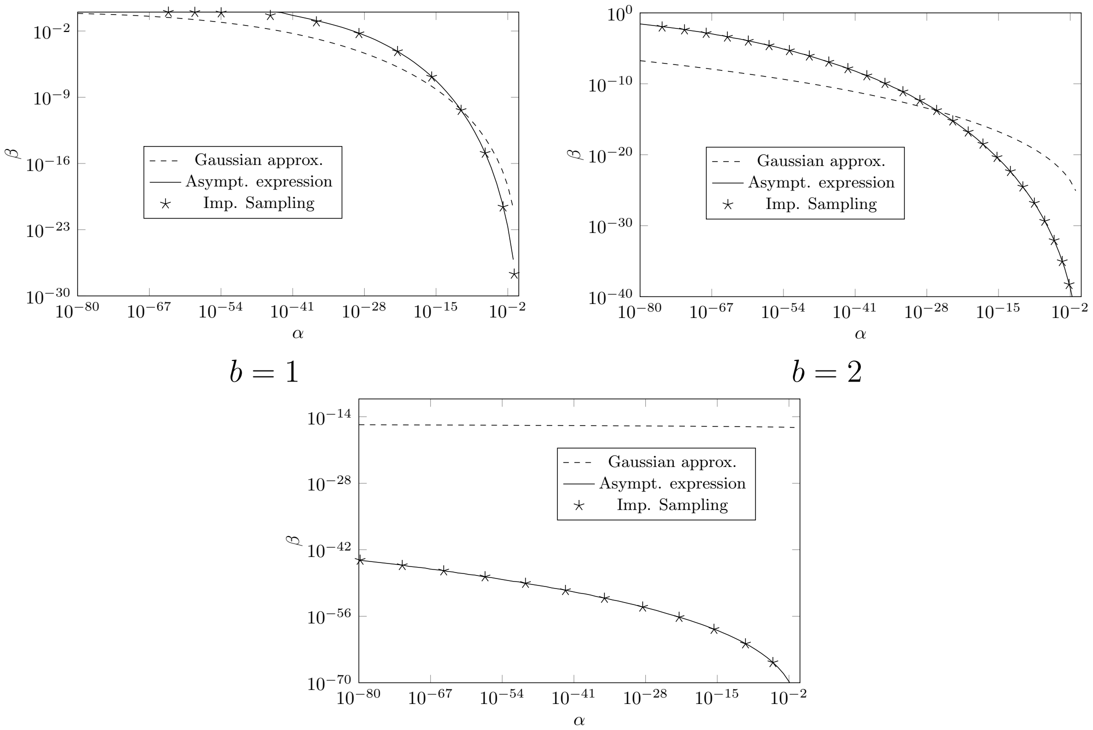
Afin d'analyser la précision des expressions (7.4) et (7.6), nous avons développé une méthode de Monte-Carlo basée sur l'échantillonnage d'importance. Pour ce faire, nous avons utilisé une loi de proposition égale à la densité P^* et nous avons montré que nous obtenions ainsi un estimateur non-biaisé dont la variance était inférieure à celle obtenue par l'estimateur de Monte-Carlo, tout en tendant asymptotiquement vers 0.

7.3.4 Résultats obtenus

La figure 7.5 présente une comparaison entre les courbes ROC obtenues via l'expression asymptotique et via l'approximation Gaussienne et ce pour différents paramètres de la distribution Gaussienne généralisée. Nous pouvons constater que dans certains cas, notamment pour des distributions approchant la loi uniforme, que cette différence est conséquente. La précision de l'expression asymptotique est également corroborée par des simulations de Monté-Carlo qui utilisent un échantillonnage d'importance.

La connaissance précise des probabilités d'erreurs nous a également permis d'optimiser les canaux d'impression acquisition. Nous avons analysé deux scénarios pratiques :

1. A partir d'un modèle d'impression-acquisition donné, nous cherchons dans un premier temps à trouver les paramètres du modèle qui permettront de minimiser la probabilité de non-détection β du système d'authentification. Cette optimisation revient en pratique à sélectionner le type d'imprimante, d'encre, et de papier qui permettront d'obtenir les meilleurs performances. Dans ce cas ci, nous faisons l'hypothèse que l'adversaire est passif et qu'il se contentera d'utiliser le même système d'impression-acquisition que l'imprimeur légitime.



$b = 1$

$b = 2$

$b = 6$

Figure 7.5: Comparaison entre l'approximation Gaussienne, l'expression asymptotique et les simulations de Monte-Carlo via échantillonage d'importance dans le cas de distributions Gaussienne généralisées $b = 1$, $b = 2$ and $b = 6$. Les canaux d'impression-acquisition pour l'imprimeur légitime et le contrefacteur sont identiques, $\mu_b = 50$, $\mu_w = 150$, $\sigma_b = 40$, $\sigma_w = 40$.

2. Le deuxième scénario correspond à un scénario de sécurité à proprement dit puisqu'ici nous prenons en compte un adversaire cherchant à modifier son modèle d'impression-acquisition afin de détériorer les performances du système d'authentification. L'objectif ici est d'envisager une attaque au pire des cas en cherchant le modèle d'impression-acquisition de l'imprimeur légitime qui permettra d'obtenir les meilleurs performances de détection une fois que l'adversaire aura sélectionné son modèle le plus néfaste. Dans ce scénario l'adversaire est actif puisqu'il est capable de modifier son canal d'impression-acquisition et nous partons du principe que le receveur connait le canal du contrefacteur.

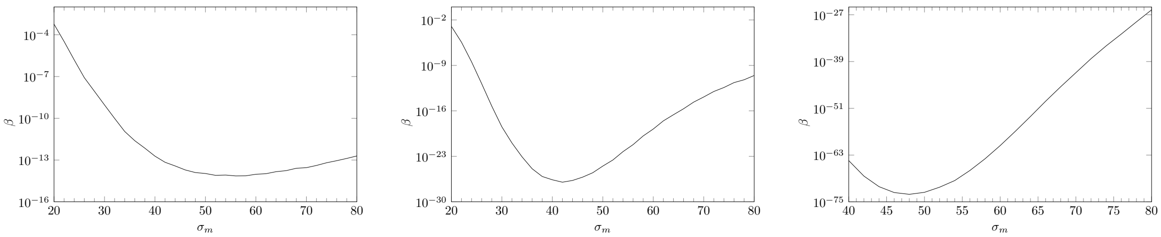
Le premier problème peut être formalisé par la recherche au sein d'une famille paramétrique donnée de canaux d'impression-acquisition $\mathcal{C}$, les paramètres du canal minimisant la probabilité de non détection β, nous cherchons donc la probabilité β^* telle que :

$$\beta^* = \min_{\mathcal{C}} \beta(\alpha). \tag{7.7}$$

Dans le second cas, l'optimisation consiste à résoudre un jeu min max pour deux familles de canaux, l'un appelé $\mathcal{C}_l$ pour l'imprimeur légitime, l'autre appelé $\mathcal{C}_o$ pour le contrefacteur. Dans le cas où le receveur connait le canal du contrefacteur, nous cherchons donc la probabilité β^* telle que :

$$\beta^* = \min_{\mathcal{C}_l} \max_{\mathcal{C}_o} \beta(\alpha). \tag{7.8}$$

La figure 7.6 présente un exemple de résultats obtenus dans le scénario qui considère un contrefacteur actif. Nous voyons que pour chacun de ces exemples (ce n'est cependant pas vrai dans tous les cas), la stratégie optimale pour l'imprimeur certifiée est d'éviter un procédé d'impression-acquisition peu bruité qui favoriserai une estimation facile du code original par le contrefacteur, mais d'éviter également un procédé trop dégradé pour lequel le bruit important empêcherait la distinction entre code originaux et codes contrefaits. Ces résultats montrent également l'intérêt d'utiliser un canal proche de la loi uniforme, c'est à dire paramètre b grand qui amène un β faible, par rapport à un canal proche d'une loi parcimonieuse, c'est à dire un$b$ faible qui amène un β grand.



$b = 1$, $\beta^* \approx 10^{-14}$ $b = 2$, $\beta^* \approx 10^{-27}$ $b = 6$, $\beta^* \approx 10^{-73}$

Figure 7.6: Evolution de la meilleure stratégie du contrefacteur $\max_{\sigma_o} \beta$ en fonction de l'écart type σ_m d'une distribution Gaussienne généralisée pour différents paramètres b de cette distribution. $\mu_b = 50$, $\mu_w = 150$, $\alpha = 10^{-6}$.

7.4 Authentification via l'utilisation de codes déterministes

Après avoir étudié un schéma d'authentification utilisant des CGs non structurés, nous cherchons maintenant à évaluer l'influence du codage canal sur notre système d'authentification en utilisant des codes déterministes. Cette démarche est motivée par le théorème du codage canal, qui montre qu'il est possible de construire un code qui sera décodé sans erreur après une étape d'impression-acquisition provenant de l'imprimeur légitime mais qui sera décodé avec erreurs après ré-impression par le contrefacteur. Dans notre contexte, il est cependant important de préciser que le receveur effectue un décodage inadapté car il ne connait pas l'origine du CG utilisé lors de l'authentification. Puisque nous faisons l'hypothèse que le receveur utilisera une stratégie de

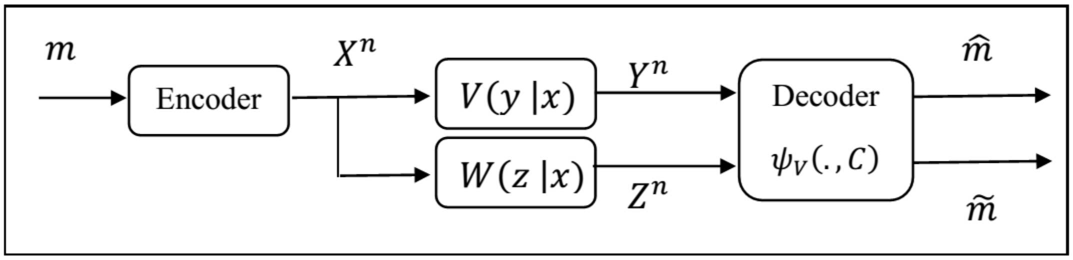


Figure 7.7: Décodeur inadapté et authentification.

codage-décodage liée au canal principal ($X^n \to Y^n$), la capacité inadaptée liée au canal du contrefacteur ($X^n \to Z^n$) sera différente de sa capacité réelle (voir figure 7.7).

7.4.1 Cadre théorique

Lorsqu'un codage canal est utilisé, la probabilité de fausse alarme α se traduit par la probabilité que le message $\hat{m}$ décodé soit différent du message m original ($\alpha = \Pr\{\hat{m} \neq m\}$) et la probabilité de non détection β s'écrit comme la probabilité qu'un message décodé $\tilde{m}$ provenant du contrefacteur soit égale au message original ($\beta = \Pr\{\tilde{m} = m\}$). Si le théorème du codage canal montre que la probabilité α peut être infiniment petite, nous chercherons à calculer la β à travers le calcul de la probabilité d'erreur dans le canal du contrefacteur $Pe_{opp} = 1 - \beta$.

Enfin, pour éviter que le contrefacteur soit à même de décoder le message original à partir du CG imprimé, nous faisons l'hypothèse que le message codé X^n (voir figure 7.8) est ensuite protégé par un système de cryptage tel que le "one-time pad" qui protège le code élément par élément.

La capacité inadaptée $C_M > C_{LM}$ (avec C_{LM} calculable) représente le taux d'information au dessus duquel la probabilité d'erreur dans le canal du contrefacteur devient non négligeable. C_{LM} se calcule en effectuant une optimisation non-linéaire avec contraintes (voir section 5.4). La construction de notre code qui est détaillée dans la section suivante s'appuie sur trois résultats théoriques :

1. Dans [8], l'auteur montre que pour un canal à entrée binaire, pour tout code tel que $R > C_{LM}$ la probabilité d'erreur liée au canal du contrefacteur est supérieure à une valeur positive et que dans ce cas là $C_{LM} = C_M$. Nous en déduisons donc que la borne C_{LM} est la borne de capacité limite au dessus de laquelle l'authentification est possible.

2. Dans [72], l' auteur propose la réciproque forte du théorème du codage canal, à savoir le fait que pour tout code tel que $R > C_{opp}$ et ce quelque-soit le décodeur, nous avons $Pe_{opp} > 1 - \delta$, δ pouvant s'écrire de façon explicite et tendant vers 0 avec $n \to \infty$. A partir de C_{opp} nous pouvons donc calculer précisément une borne supérieure de la probabilité d'erreur atteignable pour tout code.

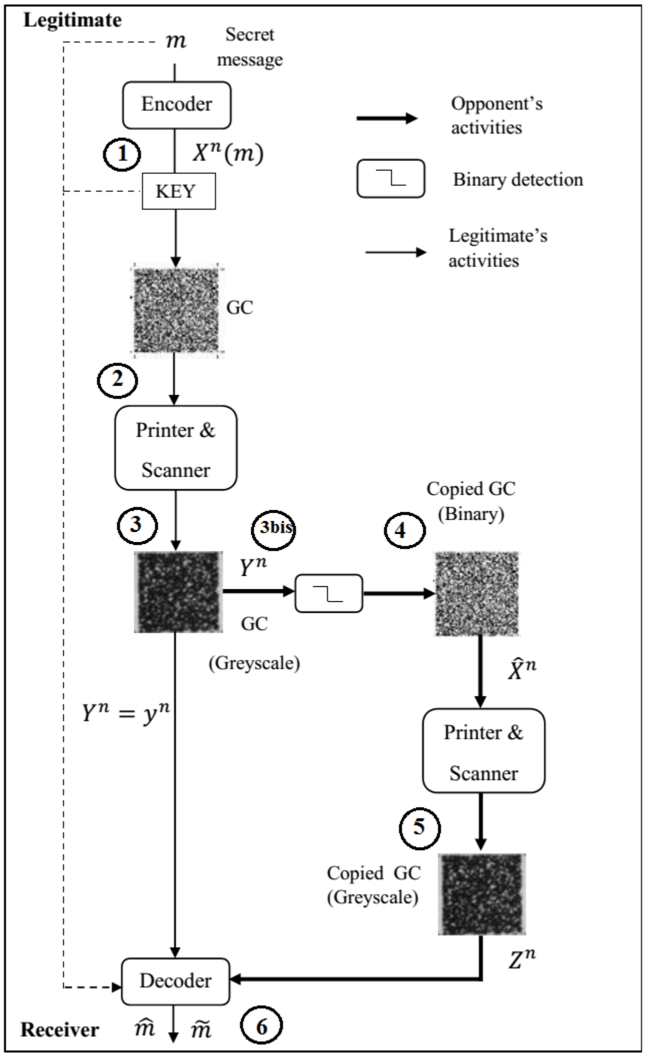


Figure 7.8: Modèle d'authentification basé sur le codage canal.

3. En s'inspirant de [51], qui souligne que pour $R > C_{LM}$ il existe un code tel que la probabilité moyennée sur tous les dictionnaires possibles $\bar{P}e_{opp}$ satisfait $\bar{P}e_{opp} > 1 - \delta$, et en montrant que δ peut s'écrire de façon explicite, nous sommes capables d'expliciter la borne supérieure de $\bar{P}e_{opp}$. Il convient toutefois de souligner que rien ne garantit (pour le moment) qu'un tel code soit également capable d'atteindre la capacité du canal principal pour obtenir une probabilité de fausse alarme nulle.

7.4.2 Cadre pratique

Nous avons sélectionné le codage Turbo [12] pour ses bonnes performances pratiques et la possibilité, via l'information extrinsèque associée, d'analyser ses performances de décodage sur le canal de l'adversaire. Le principe de ce codage est illustré sur la figure

7.9, il est constitué de deux codes convolutifs (ici des codes récursifs) de rendement 1/2 séparés d'un entrelaceur.

Le décodage turbo est un algorithme récursif (voir figure 7.10) où les deux décodeurs utilisent à tour de rôle l'information extrinséque provenant du décodage précédent. La fonction de transfert de l'information extrinsèque est utilisée pour analyser pratiquement les propriétés de convergence du décodeur Turbo. Nous sommes ainsi capable de régler les paramètres du code qui permettront de garantir une convergence du décodeur lorsque le CG arrive du canal principal mais aussi une divergence lorsque le CG arrive du canal provenant du contrefacteur.

La figure 7.11 montre une différence considérable en terme de performances d'authentification entre l'utilisation d'un CG non-codé et l'utilisation d'un CG codé par codage concaténé dans les cas les plus favorables pour l'adversaire.

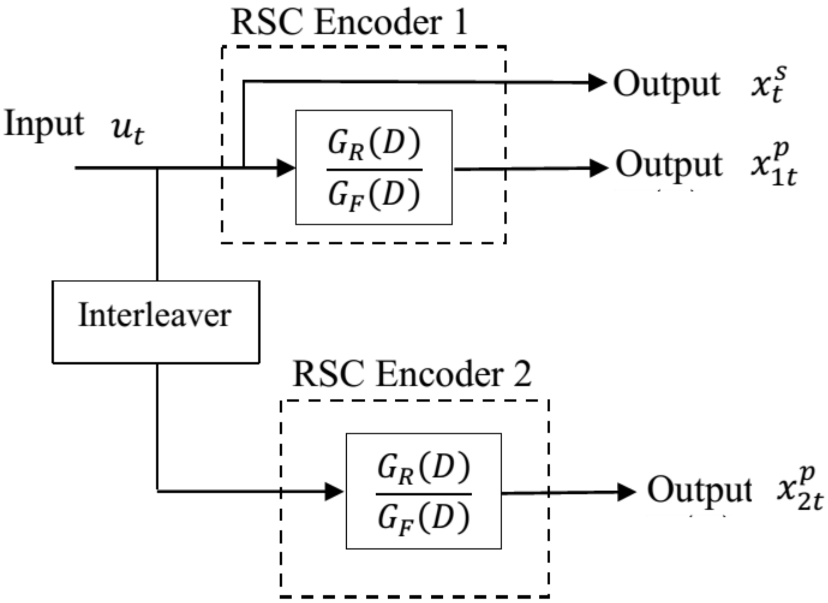


Figure 7.9: Principe du codage Turbo.

7.5 Conclusions et perspectives

Cette thèse a permis de donner un éclairage théorique au problème de l'authentification par codes graphiques. Nous avons envisagé deux scénarios, l'utilisation ou non d'un système de codage canal pour l'authentification. Lorsque le CG n'est pas codé, le test par rapport de vraisemblance s'avère être la solution optimale. Pour être convenablement utilisé, il faut cependant veiller à estimer précisément les probabilités de fausse alarme et de non détection. Ce problème n'est pas trivial lorsque ces probabilités sont faibles (i.e. $< 10^{-3}$) et nous avons utilisé la théorie des grandes déviations d'une part, et les

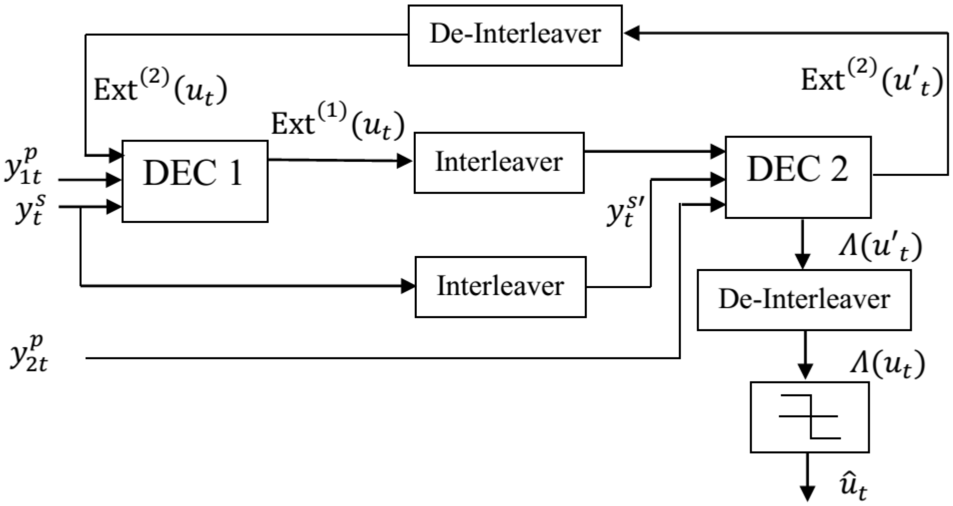
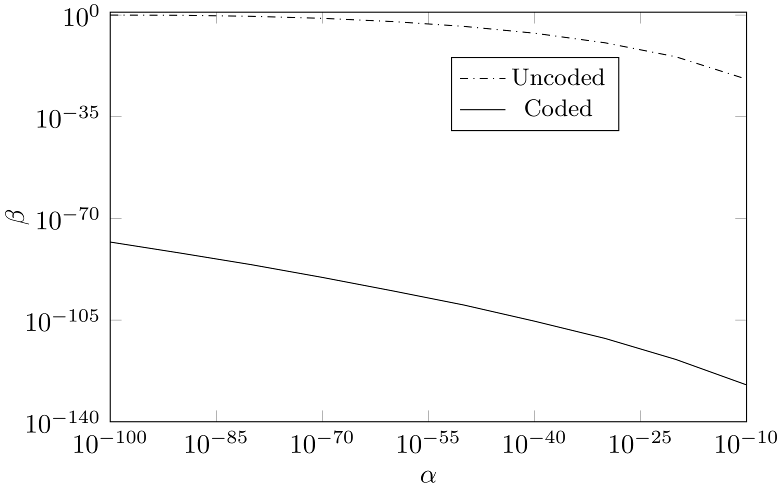


Figure 7.10: Principe du décodage Turbo.



$R = 1/3$

Figure 7.11: Comparaison entre des CG codés ou non-codés après l'attaque au pire des cas et l'optimisation du canal principal. Non-codé : $\sigma_m = 42$, $\sigma_o = 40$ (correspondant aux paramètres solutionnant le jeu min-max). Codé : $\sigma_m = 48$, $\sigma_o = 0$ (correspondant à la différence la plus importante entre C_m et C_o lorsque l'adversaire utilise l'attaque au pire des cas.

méthodes d'échantillonnage d'importance d'autre part pour calculer théoriquement et pratiquement ces probabilités. Cette étape est un pré-requis à l'optimisation du canal d'impression-acquisition, ou à l'analyse des performances en fonction de la parcimonie du canal.

Dans un second temps nous nous sommes placé dans le paradigme du codage canal et avons cherché à voir si les méthodes de codage déterministes pouvaient améliorer le système d'authentification étudié. En replaçant notre problème dans le cadre théorique du décodage inadapté, nous avons cherché d'une part à calculer la capacité inadaptée du canal provenant du contrefacteur, et de choisir un rendement de codage compris entre celle-ci et et la capacité du canal légitime. En utilisant des codes concaténés et leur décodage Turbo, nous avons proposé une implémentation qui permet via l'information extrinsèque, de régler les paramètres du canal de tel manière à assurer un décodage sans erreur dans le canal légitime, et à générer un nombre important d'erreurs dans le canal lié au contrefacteur.

Nos perspectives sont multiples. Lorsque l'on s'affranchit du codage canal, nous devrons chercher à voir si notre méthodologie peut être étendue à des signaux non-i.i.d et à des applications autres comme la stéganalyse et l'extraction de preuves numériques. Lorsque le codage canal est utilisé, nous chercherons à voir si d'autres codes (LDPC, *polar codes* ou stratégies de codage aléatoire) permettent d'obtenir des performances supérieurs aux codes concaténés aussi bien au niveau de la sécurité (minimisation de la fuite d'information) que des probabilités d'erreur.

Bibliography

[1] The deadly world of fake medicine. http://www.cnn.com/2012/07/17/health/living-well/falsified-medicine-bate/, 2012.

[2] Large deviation and sanov's theorem, 2013.

[3] Counterfeit consumer goods. http://en.wikipedia.org/wiki/Counterfeit_consumer_goods, 2014.

[4] The us sees more money lost to credit card fraud than the rest of the world combined. http://www.businessinsider.com/the-us-accounts-for-over-half-of-global-payment-card-fraud-sai-2014-3, 2014.

[5] Rudolf Ahlswede and Imre Csiszar. Common randomness in information theory and cryptography. part i: secret sharing. *IEEE Transactions on Information Theory*, 39(4), 1993.

[6] Mattias Andersson, Vishwambhar Rathi, Ragnar Thobaben, Jorg Kliewer, and Mikael Skoglund. Nested polar codes for wiretap and relay channels. *Communications Letters, IEEE*, 14(8):752–754, 2010.

[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *Information Theory, IEEE Transactions on*, 20(2):284–287, Mar 1974.

[8] Vladimir B Balakirsky. A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels. *IEEE Transactions on Information Theory*, 41(6):1889–1902, 1995.

[9] Gerard Battail. A conceptual framework for understanding turbo codes. *Selected Areas in Communications, IEEE Journal on*, 16(2):245–254, 1998.

[10] Fokko Beekhof, Sviatoslav Voloshynovskiy, and Farzad Farhadzadeh. Content authentication and identification under informed attacks. In *WIFS*, pages 133–138. Citeseer, 2012.

[11] Sergio Benedetto and Guido Montorsi. Unveiling turbo codes: Some results on parallel concatenated coding schemes. *Information Theory, IEEE Transactions on*, 42(2):409–428, 1996.

[12] Glavieux Alain Berrou, Claude and P.Thitimajshima. Near shannon limit error correcting coding and decoding: Turbo-codes. *Proceeding of ICC '93, Geneva, Switzerland*, pages 1064–1070, 1993.

[13] R.E. Blahut. *Principles and practice of information theory*, volume 1. Addison-Wesley, 1987.

[14] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[15] James DR Buchanan, Russell P Cowburn, Ana-Vanessa Jausovec, Dorothee Petit, Peter Seem, Gang Xiong, Del Atkinson, Kate Fenton, Dan A Allwood, and Matthew T Bryan. Fingerprinting documents and packaging. *Nature*, 436(28):475, 2005.

[16] Po-Ning Chen. Generalization of gartner-ellis theorem. *Information Theory, IEEE Transactions on*, 46(7):2752–2760, 2000.

[17] Rémi Cogranne, Cathel Zitzmann, Florent Retraint, Igor V Nikiforov, Philippe Cornu, and Lionel Fillatre. A local adaptive model of natural images for almost optimal detection of hidden data. *Signal Processing*, 100:169–185, 2014.

[18] Gérard Cohen and Gilles Zémor. Generalized coset schemes for the wire-tap channel: Application to biometrics. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, page 46. IEEE, 2004.

[19] Thomas M Cover and Joy A Thomas. Elements of information theory 2nd edition. 2006.

[20] Imre Csiszár. Almost independence and secrecy capacity. *Problemy Peredachi Informatsii*, 32(1):48–57, 1996.

[21] Imre Csiszár and Janos Korner. Graph decomposition: A new key to coding theorems. *Information Theory, IEEE Transactions on*, 27(1):5–12, 1981.

[22] Imre Csiszar and Janos Körner. Information theory: Coding theorems for discrete memoryless channels. *Budapest: Akadémiai Kiadó*, 2011.

[23] F. Pollara R.J. McEliece D. Divsalar, S. Dolinar. Transfer function bounds on the performance of turbo codesl. *TDA Progress Report*, pages 44–55, 1995.

[24] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 2. Springer, 1998.

[25] Dariush Divsalar and Fabrizio Pollara. Turbo codes for deep-space communications. *TDA progress report*, 42:120, 1995.

[26] Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *Advances in cryptology-Eurocrypt 2004*, pages 523–540. Springer, 2004.

[27] Meir Feder and Neri Merhav. Universal composite hypothesis testing: A competitive minimax approach. *Information Theory, IEEE Transactions on*, 48(6):1504–1517, 2002.

[28] Robert G Gallager. *Information theory and reliable communication*, volume 2. Springer, 1968.

[29] Joachim Hagenauer. The exit chart-introduction to extrinsic information transfer in iterative processing. In *Proc. 12th European Signal Processing Conference (EUSIPCO)*, pages 1541–1548. Citeseer, 2004.

[30] Tobias Haist and Hans J Tiziani. Optical detection of random features for high security applications. *Optics communications*, 147(1):173–179, 1998.

[31] John Michael Hammersley and David Christopher Handscomb. *Monte carlo methods*, volume 1. Springer, 1964.

[32] Anthony TS Ho, Xunzhan Zhu, and Yong Liang Guan. Image content authentication using pinned sine transform. *EURASIP Journal on Advances in Signal Processing*, 2004(14):2174–2184, 1900.

[33] Joseph Yu Ngai Hui. Fundamental issues of multiple accessing. 1983.

[34] Tanya Ignatenko. *Secret-key rates and privacy leakage in biometric systems*. PhD thesis, PhD thesis, Technical University of Eindhoven, 2009.

[35] Tanya Ignatenko and Frans MJ Willems. Biometric systems: Privacy and secrecy aspects. *Information Forensics and Security, IEEE Transactions on*, 4(4):956–973, 2009.

[36] Tanya Ignatenko and Frans MJ Willems. Privacy-leakage codes for biometric authentication systems. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1601–1605. IEEE, 2014.

[37] Anil K Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*. Springer, 1999.

[38] Nocedal Jorge and J Wright Stephen. Numerical optimization. *Springerverlang, USA*, 1999.

[39] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Proceedings of the 6th ACM conference on Computer and communications security*, pages 28–36. ACM, 1999.

[40] J Kiirner and Katalin Marton. General broadcast channels with degraded message sets. *IEEE Trans. Info. Theory*, 23:60–64, 1977.

[41] Lifeng Lai, Hesham El Gamal, and H Vincent Poor. Authentication over noisy channels. *Information Theory, IEEE Transactions on*, 55(2):906–916, 2009.

[42] Amos Lapidoth. Mismatched decoding and the multiple-access channel. *Information Theory, IEEE Transactions on*, 42(5):1439–1452, 1996.

[43] Qiming Li, Yagiz Sutcu, and Nasir Memon. Secure sketch for biometric templates. In *Advances in Cryptology–ASIACRYPT 2006*, pages 99–113. Springer, 2006.

[44] Ching-Yung Lin and Shih-Fu Chang. Distortion modeling and invariant extraction for digital image print-and-scan process. In *Int. Symp. Multimedia Information Processing*, 1999.

[45] Ching-Yung Lin and Shih-Fu Chang. A robust image authentication method distinguishing jpeg compression from malicious manipulation. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(2):153–168, 2001.

[46] J Lodge, P Hoeher, and J Hagenauer. The decoding of multidimensional codes using separable map filters. In *16th Biennial Symposium on Communications*, pages 343–346, 1992.

[47] N Laneman M Bloch. Strong secrecy from channel resolvability. *IEEE Trans. Info. Theory*, 2013.

[48] Hessam Mahdavifar and Alexander Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. *Information Theory, IEEE Transactions on*, 57(10):6428–6443, 2011.

[49] Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993.

[50] Ueli M Maurer. Authentication theory and hypothesis testing. *Information Theory, IEEE Transactions on*, 46(4):1350–1356, 2000.

[51] Neri Merhav, Gideon Kaplan, Amos Lapidoth, and S Shamai Shitz. On information rates for mismatched decoders. *Information Theory, IEEE Transactions on*, 40(6):1953–1967, 1994.

[52] Ravikanth Pappu, Ben Recht, Jason Taylor, and Neil Gershenfeld. Physical one-way functions. *Science*, 297(5589):2026–2030, 2002.

[53] J. Picard, C. Vielhauer, and N. Thorwirth. Towards fraud-proof id documents using multiple data hiding technologies and biometrics. *SPIE Proceedings–Electronic Imaging, Security and Watermarking of Multimedia Contents VI*, pages 123–234, 2004.

[54] J Picard and J Zhao. Improved techniques for detecting, analyzing, and using visible authentication patterns, july 28 2005. *WO Patent WO/2005/067,586.*

[55] O. Pothier. *Codes Composites Construits À Partir de Graphes Et de Leur Décodage Itératif.* ENST, E. École supérieure des télécommunications, 2000.

[56] Tom Richardson and Ruediger Urbanke. *Modern coding theory.* Cambridge University Press, 2008.

[57] Eren Sasoglu and Alexander Vardy. A new polar coding scheme for strong security on wiretap channels. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 1117–1121. IEEE, 2013.

[58] Igal Sason and Shlomo Shamai. Improved upper bounds on the ml decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum. *Information Theory, IEEE Transactions on*, 46(1):24–47, 2000.

[59] Claude E Shannon. Communication theory of secrecy systems*. *Bell system technical journal*, 28(4):656–715, 1949.

[60] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[61] Gustavus J Simmons. Authentication theory/coding theory. In *Advances in Cryptology*, pages 411–431. Springer, 1985.

[62] I Stiglitz. Coding for a class of unknown channels. *Information Theory, IEEE Transactions on*, 12(2):189–195, 1966.

[63] Arunkumar Subramanian, Ananda T Suresh, Safitha Raj, Andrew Thangaraj, Matthieu Bloch, and Steven McLaughlin. Strong and weak secrecy in wiretap channels. In *Turbo Codes and Iterative Information Processing (ISTC), 2010 6th International Symposium on*, pages 30–34. IEEE, 2010.

[64] Arunkumar Subramanian, Andrew Thangaraj, Matthieu Bloch, and Steven W McLaughlin. Strong secrecy on the binary erasure wiretap channel using large-girth ldpc codes. *Information Forensics and Security, IEEE Transactions on*, 6(3):585–594, 2011.

[65] G Edward Suh and Srinivas Devadas. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of the 44th annual Design Automation Conference*, pages 9–14. ACM, 2007.

[66] Stephan ten Brink. Iterative decoding trajectories of parallel concatenated codes. *ITG FACHBERICHT*, pages 75–80, 2000.

[67] Stephan Ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *Communications, IEEE Transactions on*, 49(10):1727–1737, 2001.

[68] Thanh Hai Thai, Florent Retraint, and Rémi Cogranne. Statistical detection of data hidden in least significant bits of clipped images. *Signal Processing*, 98:263–274, 2014.

[69] Andrew Thangaraj, Souvik Dihidar, A Robert Calderbank, Steven W McLaughlin, and J-M Merolla. Applications of ldpc codes to the wiretap channel. *Information Theory, IEEE Transactions on*, 53(8):2933–2945, 2007.

[70] Matthieu R Bloch Vincent YF Tan. Information spectrum approach to strong converse theorems for degraded wiretap channels. 2014.

[71] Frans MJ Willems and Tanya Ignatenko. Authentication based on secret-key generation. In *ISIT*, pages 1792–1796, 2012.

[72] Jacob Wolfowitz, Jacob Wolfowitz, Polen Statistiker, Jacob Wolfowitz, and Jacob Wolfowitz. *Coding theorems of information theory.* Number 31. Springer, 1961.

[73] Aaron D Wyner. The wire-tap channel. *Bell System Technical Journal, The*, 54(8):1355–1387, 1975.

[74] Raymond W Yeung. *A first course in information theory*, volume 1. Springer, 2002.

[75] Ofer Zeitouni, Jacob Ziv, and Neri Merhav. When is the generalized likelihood ratio test optimal? *Information Theory, IEEE Transactions on*, 38(5):1597–1602, 1992.