A POLYNOMIAL TIME PROCEDURE CONVERTING ERROR CORRECTING CODES TO SEMANTICALLY SECURE WIRETAP CODES

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Eric Kubischta

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Electrical and Computer Engineering

March 2018

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

A POLYNOMIAL TIME PROCEDURE CONVERTING ERROR CORRECTING CODES TO SEMANTICALLY SECURE WIRETAP CODES

By

Eric Kubischta

The supervisory committee certifies that this thesis complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Sanjay Karmakar

Chair

Dr. Indranil SenGupta

Dr. Ivan Lima

Approved:

6 April 2018	Dr. Benjamin Braaten
Date	Department Chair

ABSTRACT

We furnish a procedure based on universal hash families that can convert an error correcting code of rate R to a semantically secure wiretap code of rate $R - \xi$ where ξ is some parameter derived from the eavesdropper's channel. This conversion is shown to be polynomial time efficient with block length and is applicable to any *discrete time* channel.

To prove the induced wiretap code is semantically secure, we have upgraded recent leakage bounds by maximizing over all message distributions. The semantic leakage is shown to be exponentially decreasing with block length.

As an explicit application, we construct a concrete, polynomial time efficient, semantically secure wiretap code that can achieve the secrecy capacity of the AWGN wiretap channel. Moreover, this wiretap coding scheme has both probability of error and semantic leakage exponentially diminishing with block length.

ACKNOWLEDGEMENTS

I would like to thank my research partner Parker Pavlicek and adviser Dr. Sanjay Karmakar; their acuity was especially beneficial. I would also like to thank Mihir Bellare, Stefano Tessaro, Alexander Vardy, and Himanshu Tyagi for much of this thesis is inspired by their insightful works.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS iv

LIST OF FIGURES vii

1. INTRODUCTION 1

2. PRELIMINARIES 6

  2.1. Probability Theory 6

  2.2. Information Theory 9

  2.3. Modeling Communication 11

  2.4. Communication Theory 12

  2.5. Wiretap Channels and Secrecy Capacity 16

  2.6. Concepts from Computer Science 20

3. PRELUDE 24

  3.1. Comparison of Security Metrics 24

  3.2. Main Contributions 28

  3.3. Prior Work 29

4. A WIRETAP CODING SCHEME 30

  4.1. Transmission Procedure 30

  4.2. Implementation 32

5. SECURITY OF OUR WIRETAP SCHEME 39

  5.1. Max Information and Typical Sets 39

  5.2. Leakage bounds 41

6. ACHIEVABLE RATES OF OUR WIRETAP SCHEME 51

  6.1. Achievable Rates 51

  6.2. Removing the Public Seed Assumption 54

7. APPLICATION 58

7.1. Achieving Semantic security on AWGN wiretap channels 58

8. DISCUSSION 64

8.1. Looking Forward 64

8.2. Conclusion 65

REFERENCES 66

LIST OF FIGURES

Figure Page

4.1. Transmission Scheme. 30

7.1. AWGN Channel model. 59

1. INTRODUCTION

The first rigorous treatment of secure communication was put forth in 1949 by Claude Shannon in his paper "Communication theory of secrecy systems" [14] (declassified version). Shannon considered a scenario where a transmitter attempted to impart a message to a receiver while keeping the message *completely* hidden from an adversary. Shannon allowed the transmitter to encrypt the message first into some ciphertext (also called a codeword in the sequel) and that the adversary received an error-free copy of this ciphertext. In reality, environments engender errors so that this is a worst case scenario of sorts.

To measure if a message was completely hidden from the adversary, Shannon put forth the idea of *perfect secrecy*. He said *perfect secrecy* was obtained if the *mutual information* (to be defined in Chapter 2) between the message and the ciphertext was exactly 0. In other words, he said perfect secrecy was only obtained if the message and ciphertext were statistically independent. Shannon established that a "secret key" must be used that only the transmitter and intended receiver know otherwise secure communication in this sense would not be possible. Moreover, Shannon proved that this secret key must be *at least* as long as the message itself and that each key may *only* be used for one message.

This disappointing result implied that perfect secrecy was impossible for all intents and purposes. This split the field of cryptography into two major directions. In one direction (dubbed computational based security) the adversary is assumed to have bounded computational resources. In the other direction (dubbed physical layer security) the adversary is assumed to receive a copy of the ciphertext with errors induced by the environment.

Computational security took a strong foothold with the introduction of public key cryptography in the 1970's. It has remained the primary provider of security since then due to its ease of implementation, plethora of schemes, and well suited (so far) assumptions about resource boundedness. Indeed most computational security schemes are based on the assumption that some decision problems are *easy* to verify but *hard* to solve. For example, security of the widely used RSA scheme is based on the assumption that the factorization of the product of two large primes is a hard problem.

That being said, computational security is *unproven* security. **P** is a computational complexity class that basically contains all decision problems that can be *solved* efficiently fast. **NP** is a computational complexity class that basically contains all decision problems that can be *verified* efficiently fast. The **P** vs. **NP** problem asks if **P** = **NP**. It is one of the most renowned *open* problems in computer science and mathematics. Intuitively, if **P** = **NP** then every decision problem that can be verified efficiently can also be solved efficiently. This has the potential to break most computational security systems in use today for they rely on the often tacit assumption that some problems are in **NP** but not in **P**.

Nonetheless many cryptographers believe that **P** ≠ **NP**. In that case, computational security would be (nearly) immune to an adversary with a classical computer bounded sufficiently in time. However, **P** and **NP** are complexity classes applicable only to *classical* decision problems; they are not suited for *quantum* decision problems. For example, RSA is based on the assumption that factoring large numbers is an **NP** hard problem, however, in the 1990's Peter Shor [15] constructed a factoring algorithm that can factor large numbers efficiently fast on quantum computers. A (somewhat) comparable complexity class to **P** for quantum computers is **BQP**, the class containing all quantum decision problems that can be solved efficiently fast (with small bounded probability of error). It turns out that $\mathbf{P} \subset \mathbf{BQP}$ but it is also conjectured that **BQP** contains decision problems strictly in **NP** (assuming **P** ≠ **NP**); that is, **BQP** contains some problems that cannot be solved efficiently with a classical computer but can be solved efficiently with a quantum computer. The previous has sparked a program called "post quantum cryptography" directed toward advancing computational based security so that it is secure against an adversary with access to a quantum computer. However, even with post quantum cryptography, an adversary with sufficiently large computational resources could break computational based security in theory.

At this point, we have exclusively focused on computational based security, let us retreat back to the other direction stemming from Shannon's disappointing result: physical layer security. Physical layer security is based on the assumption that communication is not error free. Aaron Wyner in the 1970's created *the wiretap channel* [20], a way of modeling physical layer security. In this paper he proved that it was possible to simultaneously send a message with low probability of error to an intended receiver while keeping the message hidden from an adversary as long as the length of the codeword (ciphertext) was sufficiently large. He also proved that there was an optimal operating point in this case which he dubbed the *secrecy capacity* of the wiretap channel; this was the highest rate of information that could be sent reliably yet securely in his model. Wyner's measure of security is now called the weak metric and is similar to the perfect security metric of Shannon except it is asymptotic, assumes the message is uniformly distributed, and divides by the length of the codeword effectively making the metric a *rate*. Wyner's method of proof was by existence (rather than by construction) to show that such a coding scheme achieving said optimal operating point is possible.

Since Wyner, physical layer security (also called information theoretic security in the sequel) has flourished theoretically. Many new models of wiretap channels have come forth and subsequently been characterized with respect to their secrecy capacity. The main assumption in every one of the wiretap channel models is that the intended receiver be at some *physical layer advantage* over the adversary with respect to the transmitter. An intuitive way of thinking about this is that the adversary's received ciphertext is "more noisy" (or equivalently, contains more errors on average) than the intended receiver's received ciphertext/codeword. When this is not the case, information theoretic security does not allow us to transmit any information securely.

The assumption about an adversary's physical layer is a *strong* assumption. In reality, nothing prevents an adversary from obtaining a physical layer advantage effectively negating any security. This has been the main reason why computational based security is considerably more popular. Despite this, when the adversary *is* at a disadvantage to the intended receiver, this form of security *is provable*. This is in sharp contrast to the state of current affairs in computational security where we have the **P** vs. **NP** problem. Moreover, with the advent of quantum computers on the horizon, many researchers are looking for security solutions that transcend computational based methods. Indeed, information theoretic security (with a physical layer advantage) satisfies this quest.

Even with the proper motivation to use physical layer security in some instances, the field is not entirely ready for realistic applications. There are two considerable hurdles that need to be properly addressed before physical layer security can be considered *realistic*. As in the case of Wyner, many coding schemes to date for wiretap models are not concrete; that is, they are proved by existence rather than by construction. In reality, we need algorithmic coding schemes that can be implemented on computers. Proofs by existence are very powerful tools theoretically, but almost useless in practice. More in this line, coding schemes for wiretap channels that are concrete are rarely efficient. Efficiency is especially important in the wiretap domain since security is based on asymptotic codeword length. The second considerable hurdle for wiretap channels is in a completely different line and can be somewhat subtle: security metrics for wiretap channels *must coincide* with reality. Wyner used the "weak metric" as a way to measure how hidden a message was from the adversary. This was shown to be an insufficient measure of secrecy in reality and it was replaced by another security metric called the "strong metric." This has been the de facto metric for wiretap channels for some time but was criticized by cryptographers [2] only within the last decade. They put forth even another metric of security dubbed "semantic security" that is the exact asymptotic equivalent to Shannon's perfect secrecy. They advocated its use as we do so here: coding schemes can only be considered *properly realistic* when they are secure on wiretap channels under the semantic security metric (or its equivalent form as we shall later see). Thus in summary, the two main hurdles of information theoretic security can be overcome by finding concrete and efficient coding schemes for wiretap channels that are provably secure under the semantic security metric.

To this end, we have the motivation for this thesis. We present a new coding scheme for arbitrary wiretap channels based on the work of [1], [17], and [18]. Here the coding scheme is a concatenation of a *preprocessing scheme* with an *error correction code* (ECC) for the point to point channel between the transmitter and intended receiver. We prove that our preprocessing scheme can be implemented using an algorithm that has quadratic time complexity; thus, our preprocessor is concrete and efficient. We show our coding scheme provides semantic security so long as the secure rate of information is less than $R - \xi$ where R is the information rate of the ECC and ξ is some parameter of the adversary's channel. With this, we can equivalently say that given an ECC of rate R for a point to point channel between the transmitter and intended receiver, our coding scheme *efficiently converts* this ECC into a semantically secure *wiretap coding scheme* of rate $R-\xi$.

When the ECC is also given concretely and efficiently, then our entire coding scheme from front to end is also concrete and efficient and thus we overcome both hurdles of information theoretic security we previously mentioned. Furthermore, in some sense our coding scheme also *converts* the problem of finding a concrete and efficient *wiretap* scheme into a problem of finding a concrete and efficient *error correction coding* scheme. The quest to find the latter is already an extremely active field of study.

Lastly, as a point of emphasis, we show that our coding scheme can even be used to achieve the optimal operating point as put forth by Wyner called the secrecy capacity. In more detail, on the AWGN wiretap channel commonly used to model satellite and deep space transmissions, we use our preprocessing scheme with a concrete and efficient ECC for the point to point main AWGN channel and show that our entire coding scheme is concrete, quadratic time efficient, semantically secure and that both the semantic "leakage" and probability of error of this scheme go exponentially fast to 0 with respect to the codewords length.

This thesis will be organized as follows. In Chapter 2, we present background information necessary to understand the mathematically rich language of our setting. In Chapter 3, we demonstrate the necessity of using the semantic security metric and expound explicitly on what we have contributed to this thesis along with previous works toward this end. In Chapter 4, we explicitly define our coding scheme and prove that our preprocessor is both concrete and efficient. In Chapter 5, we prove that our coding scheme is semantically secure as described above. In Chapter 6, we give a characterization of which information rates are achievable using our coding scheme. In Chapter 7, we provide an application to the AWGN wiretap channel as mentioned. In the final chapter, we provide a concluding "program" for wiretap schemes looking forward.

2. PRELIMINARIES

In this introductory chapter, we will provide the reader with some background information necessary to understand the language in our setting. We will mainly be concerned with definitions and primary results from the fields of probability theory, information theory, and communication theory but will also cover some concepts from computer science.

First let us emphasize some notation and conventions. We shall denote n-dimensional vectors by a^n where a_i denotes the ith component; i.e., $a^n = (a_1, \ldots, a_n)$. We shall denote the indicator function (sometimes called characteristic function) by $\mathbb{1}_{\mathcal{A}}(x)$ or $\mathbb{1}\,(x \in \mathcal{A})$. We will take all logarithms in this paper to be base 2 unless we write ln, for which we mean the natural logarithm of base e.

2.1. Probability Theory

We will assume knowledge of basic measure theoretic probability in this thesis, but will review some concepts to elucidate our notation.

Let $(\Omega, \mathscr{B}, \mathbb{P})$ be some probability space. If X is some random variable taking values in the measurable space $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$ then basic results from measure theoretic probability theory prove that $(\mathcal{X}, \mathscr{B}_{\mathcal{X}}, \nu_X)$ is also a probability space such that

$$\nu_X(B) = (\mathbb{P} \circ X^{-1})(B) = \mathbb{P}(X = B) \qquad \forall B \in \mathscr{B}_{\mathcal{X}}.$$

The measure ν_X is called the ***distribution*** of X and the surjective image $\mathcal{X}$ is called the ***alphabet*** of X. We will only consider random variables in real coordinate space $(\mathbb{R}^n)$ in this thesis so that it follows that $\mathcal{X} \subset \mathbb{R}^n$. When the size of the alphabet $|\mathcal{X}|$ is countable X is called a ***discrete*** random variable; otherwise when the size of the alphabet $|\mathcal{X}|$ is uncountable, X is called a ***continuous*** random variable.

If μ is some other measure on $\mathcal{X}$ such that ν_X is absolutely continuous with respect to μ then the Radon-Nikodym Theorem implies that there exists some function $f : \mathcal{X} \to [0, \infty)$ (unique up to some μ-null set) such that

$$\nu_X(B) = \int\limits_B f d\mu \qquad \forall B \in \mathscr{B}_{\mathcal{X}},$$

where this is the Lebesgue–Stieltjes integral. This function f is often called the Radon-Nikodym derivative and denoted by $\frac{d\nu_X}{d\mu}$. In this work however we shall adopt a different convention standard in this literature. We shall denote such a function instead by ω_X and refer to this as the ***probability density*** of X (with respect to μ). Furthermore, we will *always* assume that X and μ are chosen so that the distribution ν_X is absolutely continuous with respect to μ (and thus the probability density ω_X always exists).

Remark. For an element $x \in \mathcal{X}$, the probability density maps $x \mapsto \omega_X(x)$ in our current notation; this will later become cumbersome. Thus we will drop the subscript X (as needed) in our notation and identify a probability density by its argument. For example, $\omega(y)$ will correspond to a probability density ω_Y for a random variable Y whereas $\omega(z)$ will correspond to a probability density ω_Z for a random variable Z.

Let X be a random variable on measurable space $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$. We will be mainly concerned with two measures μ on $\mathcal{X}$ in this thesis. When X is a continuous random variable we shall take μ to be the *Lebesgue measure*. In this case the distribution of X is given by:

$$\nu_X(B) = \mathbb{P}[X \in B] = \int_B \omega(x)dx \qquad \forall B \in \mathscr{B}_{\mathcal{X}}.$$

This is of course the familiar Lebesgue integral. On the other hand, when X is a discrete random variable, we shall take μ to be the *counting measure*. In this case the distribution of X is given by:

$$\nu_X(B) = \mathbb{P}[X \in B] = \sum_{x \in B} \omega(x) \qquad \forall B \in \mathscr{B}_{\mathcal{X}}.$$

We know from probability theory that probability densities can be represented by *probability mass functions* in the discrete case. To emphasize this, we will occasionally write $\omega(x)$ as $P_X(x)$ (the standard notation) so that the distribution of X can be given by:

$$\nu_X(B) = \mathbb{P}[X \in B] = \sum_{x \in B} P_X(x) \qquad \forall B \in \mathscr{B}_{\mathcal{X}}.$$

With this framework and notation in mind we will consider joint and conditional probability densities is the usual manner. Sticking with our notation, $\omega(x, y)$ corresponds to a joint probability density $\omega_{XY}(x, y)$ for the product random variable $X \times Y$. Furthermore, $\omega(x|y)$ will correspond to the conditional density $\omega_{X|Y}(x|y) = \omega(X = x|Y = y)$. Conditional probability densities can also be represented by the joint density over the conditioned density as $\frac{\omega(x,y)}{\omega(y)} = \omega(x|y)$. Another property we shall use is called the ***marginal density property*** and is given by:

$$\omega(x) = \int_{\mathcal{Y}} \omega(x, y)\mu(dy).$$

As usual we will say that random variables X and Y are ***independent*** if $\omega(x, y) = \omega(x)\omega(y)$ for every $x \in \mathcal{X}, y \in \mathcal{Y}$; we will write in this case $X \perp Y$. Consider the random variables M on $\mathcal{M}$, X on $\mathcal{X}$, and Y on $\mathcal{Y}$. If the product density factors as

$$\omega(y, x, m) = \omega(y|x)\omega(x|m)\omega(m) \quad \forall m \in \mathcal{M}, x \in \mathcal{X}, y \in \mathcal{Y},$$

then we say M, X, Y form a ***Markov chain*** (in that order); we will denote this by $M \to X \to Y$.

For a random variable X on $\mathcal{X}$ with μ a measure for $\mathcal{X}$, we define the ***expected value*** of X by:

$$\mathbb{E}[X] = \int_{\mathcal{X}} x\, \omega(x)\mu(dx),$$

where again the integral here is the Lebesgue–Stieltjes integral.

Lastly with respect to probability theory, we will need two extremely important tools that are crucial to the proofs.

Lemma 1 (Jensen's Inequality). Let X be a random variable and Ψ some function.

- If Ψ is *convex* then:

$$\Psi(\mathbb{E}[X]) \leq \mathbb{E}\left[\Psi(X)\right].$$

- If Ψ is *concave* then:

$$\Psi(\mathbb{E}[X]) \geq \mathbb{E}\left[\Psi(X)\right].$$

A corollary to Jensen's inequality is the following lemma. It is extremely useful in information theory.

Lemma 2 (Log-Sum Inequality). Let $\{a_i\}$ and $\{b_i\}$ be finite sequences of non-negative real numbers. Then

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\left(\sum_j a_j\right)}{\left(\sum_j b_j\right)}.$$

2.2. Information Theory

Information theory was nearly single handedly created by Claude Shannon in his pioneering paper "A mathematical theory of communication" [13]. In his paper, Shannon made rigorous the idea of *information*. In particular, Shannon viewed *information* in some sense as the reduction of uncertainty in a random variable. In this section we shall briefly review information theory and provide several important tools that will be used constantly. In particular, we shall consider a measure theoretic introduction to information theory; for reference, consider [8].

Consider the product space $\mathcal{X} \times \mathcal{Y}$ and let μ be a measure on this product space. Let the product density be given by $\omega(x, y)$ with respect to μ and the *independent* density be given by $\omega(x)\omega(y)$ with respect to μ.

Definition. We define the ***mutual information*** between X and Y by

$$I(X \wedge Y) = \int_{\mathcal{X}\times\mathcal{Y}} \omega(x, y) \log \frac{\omega(x, y)}{\omega(x)\omega(y)} \mu(d(x, y)).$$

Remark. Sometimes the mutual information between random variables X and Y is written $I(X; Y)$, but we shall follow the convention of [6] and [18] and write $I(X \wedge Y)$.

Intuitively, mutual information describes how much "information" X contains about Y (or vice versa as we shall see in the next lemma). In another sense, it also describes *how independent* X and Y are from each other. Indeed, if $X \perp Y$, then we easily see that $I(X \wedge Y) = 0$.

Suppose we want to know how much "information" some random variable X_1 contains about X_2: we can simply use mutual information in this case. However, if we are given another random variable X_3, then we need another tool, because X_3 could possibly contain some information about either X_1 or X_2; the following definition describes such a tool.

Definition. Let (X_1, X_2, X_3) be a joint random variable on the space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ where μ is some measure on this space. We define the ***conditional mutual information*** between X_1 and X_2 given X_3 by

$$I(X_1 \wedge X_2 | X_3) = \int_{\mathcal{X}} \omega(x_1, x_2, x_3) \log \left(\frac{\omega(x_1, x_2|x_3)}{\omega(x_1|x_3)\omega(x_2|x_3)} \right) \mu(dx_1 dx_2 dx_3).$$

With these two tools in hand, let us see some of the most useful properties they admit.

Lemma 3. [8, Theorem 1.6.3] *Let X, Y, Z, X_i, Y_i be random variables (for $i = 1, \ldots, n$). The following are properties of mutual information.*

1. $I(X \wedge Y) = I(Y \wedge X)$ (*symmetry*).

2. $I(X \wedge Y) \geq 0$ with equality iff $X \perp Y$ (*non-negativity*).

3. If $(X, Y) \perp Z$ then $I(X \wedge Y|Z) = I(X \wedge Y)$.

4. $I(X \wedge (Y, Z)) = I(X \wedge Z) + I(X \wedge Y|Z)$.

5. More generally than (4), we have

$$I(X \wedge Y_1, \ldots, Y_n) = \sum_{i=1}^{n} I(X \wedge Y_i | Y_{i-1}, Y_{i-2} \ldots, Y_1).$$

6. If $(X_1, Y_1), \ldots, (X_n, Y_n)$ are mutually independent then

$$I((X_1, \ldots, X_n) \wedge (Y_1, \ldots, Y_n)) = \sum_{i=1}^{n} I(X_i \wedge Y_i).$$

Remark. Property 4 and 5 above are often called the "chain rule of mutual information."

2.3. Modeling Communication

Communication of "content" from point A to point B is inherently random: electric signals are influenced by radiation and heat, digital packets can collide, and wireless signals self interfere not to mention interfere with other wireless signals. These are just a few of the many types of ***noise*** that can corrupt our content in reality. To model these scenarios mathematically, we use the concept of a channel which relies on probability theory.

2.3.1. Channels

Definition. Let $T : \mathcal{X} \to \mathcal{Y}$ be some stochastic map, X a random variable on $\mathcal{X}$, and μ some measure on $\mathcal{Y}$. When $Y = T \circ X$ is a random variable on $\mathcal{Y}$ we define the following.

- The ***transition density*** of T is the conditional density $\omega(y|x)$.

- A ***channel*** is given by the tuple $(\mathcal{X}, \omega(y|x), \mathcal{Y})$. We will often abuse notation and write $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ while referring to the map T as the channel itself.

Remark. Note that the concatenation of channels form Markov chains.

The transition density probabilistically tells us how the channel is mapping $\mathcal{X}$ to $\mathcal{Y}$. In essence, this transition density is modeling the noise present to reality we mentioned earlier. Indeed we see that given that some data point $x \in \mathcal{X}$ was *sent* across the channel, the probability that Y is in some subset $\mathcal{U} \subset \mathcal{Y}$ is given by

$$\int_{\mathcal{U}} \omega(y|x)\mu(dy).$$

2.3.2. Restricted Channels

For the rest of this paper, we will be considering *subnormalized* channels; i.e., channels with transition densities such that

$$\int_{\mathcal{Y}} \omega(y|x)\mu(dy) \leq 1.$$

This is a technical condition that allows us to define the following.

Definition. Given a channel $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ and subset $\mathcal{T} \subset \mathcal{X} \times \mathcal{Y}$ we define the following.

- A ***restricted transition density*** for T by

$$\omega_{\mathcal{T}}(y|x) = \begin{cases} \omega(y|x), & (x,y) \in \mathcal{T} \\ 0, & \text{Otherwise} \end{cases}.$$

- A ***restricted channel*** given by $T_{\mathcal{T}} = (\mathcal{X}, \omega_{\mathcal{T}}(y|x), \mathcal{Y})$.

With this new definition, given that $x \in \mathcal{X}$ was sent across the restricted channel $T_{\mathcal{T}}$, the probability that Y is in some subset $\mathcal{U} \subset \mathcal{Y}$ is given by

$$\int\limits_{\mathcal{U}} \omega_{\mathcal{T}}(y|x)\mu(dy) = \int\limits_{\mathcal{U}} \omega(y|x) \mathbb{1}_{\mathcal{T}}\left(x,y\right)\mu(dy).$$

Definition. Given a channel $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$.

- If $|\mathcal{X}|$ and $|\mathcal{Y}|$ are both uncountable, we call the channel T *continuous.*

- If $|\mathcal{X}|$ and $|\mathcal{Y}|$ are both countable, we call the channel T *discrete.*

2.4. Communication Theory

Let T be a channel given by $(\mathcal{X}, \omega(y|x), \mathcal{Y})$. In communication theory, we are concerned with sending a message from some finite index set $\mathcal{M}'$ across T: the $\underline{\text{T}}$ransmission channel. It was shown by Shannon in his pioneering work [13] that using a channel sequentially many times by means of some *code* will evoke redundancy, allowing *successful* communication. In this thesis, we will always refer to the number of channel uses as the *block length* (of the code) and denote it by n. We will consider all n channel uses simultaneously by way of a new *induced channel* T^n given by $(\mathcal{X}^n, \omega(y^n|x^n), \mathcal{Y}^n)$ where $\omega(y^n|x^n)$ is the obvious induced conditional probability density. We sometimes call T^n the n-letter extension of T.

We will now spend the rest of this section rigorously defining these concepts and first results. As a note, we are only considering *discrete-time* channels in this work; an extension to continuous-time channels is an interesting future line of work.

2.4.1. Codes

Definition. Given some finite message set $\mathcal{M}'$ and channel T consider the induced n-letter extension channel $T^n = (\mathcal{X}^n, \omega(y^n|x^n), \mathcal{Y}^n)$ and define the following.

- An (n-length) ***encoder*** is an injective function, $e_n : \mathcal{M}' \to \mathcal{X}^n$.
- The ***codebook*** $\mathscr{C}_n$ for encoder e_n is the (*bijective*) image of e_n, $\mathscr{C}_n = e_n(\mathcal{M}')$. We refer to elements of the codebook as ***codewords***.
- A (n-length) ***decoder*** is a function, $d_n : \mathcal{Y}^n \to \mathcal{M}'$.
- A ***code*** $\mathcal{C}_n$ for T of length n is a pair of encoding and decoding functions: $\mathcal{C}_n = (e_n, d_n)$. The ***rate*** of the n-length code $\mathcal{C}_n$ is given by $R_{\mathcal{C}_n} = \frac{\log|\mathcal{M}'|}{n}$.
- A ***coding scheme*** $\mathcal{C}$ for T is a family of codes $\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$. The ***rate*** of the coding scheme $\mathcal{C}$ is given by $R_{\mathcal{C}} = \lim_{n\to\infty} R_{\mathcal{C}_n}$ (when this exists).

The previous definition is merely here to make our language precise. We point out that (as usual) we allow the size of $\mathcal{M}'$ to change with n so that the rate of the coding scheme $R_{\mathcal{C}}$ makes sense.

2.4.2. Error Correction Codes

A code for channel T in the previous section is a very general object. It consists of an encoder and decoder but no restrictions are placed on said functions. For *successful* communication over T, we need the decoder to have a low probability of making an error. Now we make these concepts precise.

Definition. We define the ***maximum probability of error*** for code $\mathcal{C}_n = (e_n, d_n)$ used for T by

$$p_{e,n} = \max_{M'\in\mathcal{M}'} \mathbb{P}[(d_n \circ T^n \circ e_n)(M') \neq M'].$$

- If $p_{e,n}$ is sufficiently small (in the eyes of the system designer) we call $\mathcal{C}_n$ an ***error correcting code*** (ECC) of length n for T.
- If $\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$ is a family of ECC's then we call $\mathcal{C}$ an ***error correcting coding scheme*** (ECC scheme) for T.

- If $\mathcal{C}$ satisfies $\lim_{n\to\infty} p_{e,n} = 0$, then we call the ECC scheme $\mathcal{C}$ ***reliable***. If $\mathcal{C}$ satisfies $\log(p_{e,n}) \leq -an^b$ for some $a, b > 0$ then we call the ECC scheme $\mathcal{C}$ ***exceptionally reliable***.
- If an ECC scheme $\mathcal{C}$ of rate $R_{\mathcal{C}}$ is reliable, then we say $R_{\mathcal{C}}$ is an ***achievable rate*** on T. Subsequently, we define the ***operational capacity*** of a channel T as the supremum of all achievable rates.

Remark. It was noted in [2] that "good" error correcting codes in practice should satisfy reliability exponentially fast; they called such ECC's "strongly reliable". Due to the plethora of definitions containing the wording "strong" we have instead called such ECC's here "exceptionally reliable." As a note, all exceptionally reliable ECCs are also reliable.

2.4.3. Power Constraints

For continuous channels, we almost always require the codebook $\mathscr{C}_n$ to be contained in some ball (with respect to some measure μ on $\mathcal{X}^n$) whose radius depends only on n. The reason behind this is because probability of error is inversely correlated to the distance between codewords: the farther apart codewords are in $\mathcal{X}^n$, the lower the probability of error. Indeed, if the codebook is not required to be contained in some ball, then we can always place codewords sufficiently far apart in $\mathcal{X}^n$ to obtain a negligibly small probability of error. However, in reality, placing codewords significantly far apart in space is a heavy cost in terms of *power* at the transmitter. Motivated by this, we have the following definition.

Definition. We say a code $\mathcal{C}$ for channel $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ satisfies the ***average power constraint*** P if for every $n \in \mathbb{N}$ it follows that

$$\frac{1}{n}||x^n||^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P \qquad \forall x^n \in \mathscr{C}_n.$$

Remark.

1. *More generally, we could consider a cost function that is not $\frac{1}{n}||\cdot||^2$, but since we have assumed the input is a subset of real coordinate space this definition will be sufficient for our needs.*

2. When the input to the channel is a random variable X (such as in Shannon's channel coding theorem next), we will change the above property to $\mathbb{E}[X^2] \leq P$. Indeed by the law of large numbers the above will converge to this latter property.

2.4.4. Shannon's Channel Coding Theorem

In [13], Shannon proved two major results: one on the fundamental limits of data compression and one on the fundamental limits of communication. We will only be concerned with the second major result and will refer to it, as per the standard, as Shannon's channel coding theorem. In particular, we will consider a generalization of Shannon's result.

Definition. Let $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ be a channel and $T^n = (\mathcal{X}^n, \omega(y^n|x^n), \mathcal{Y}^n)$ be the induced n-letter extension channel. If the density for T^n factors as $\omega(y^n|x^n) = \prod_{i=1}^n \omega(y_i|x_i)$, then we say T is a ***memoryless channel***.

This property is dubbed memoryless because when the densities split as such the output at time i depends *only* on the input at time i and no other time. In other words, this characterization implies the channel has no memory, it only knows what is happening in the moment. We many times consider memoryless channels to greatly simplify the mathematical analysis, but also because they are not completely impractical models.

Definition. Suppose we are given a memoryless channel $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$. We define the ***information channel capacity*** of T by

$$C_T = \max_X I(X \wedge Y),$$

where the maximization is over all random variables $X \in \mathcal{X}$. If T is continuous, we maximize over all $X \in \mathcal{X}$ such that $\mathbb{E}[X^2] \leq P$.

The following is Shannon's channel coding theorem generalized to all memoryless channels. We note that this theorem can be stated analogously for non-memoryless channels, but in this thesis we do not need such a powerful theorem and have chosen to stick with this result for simplicity.

Fact 1 (Channel Coding Theorem). (cf. [8]) Suppose T is a memoryless channel. Then the operational channel capacity is exactly equivalent to the information channel capacity. Precisely, if $R < C_T$ then there exists a reliable ECC scheme $\mathcal{C}$ with rate R. Conversely, if $\mathcal{C}$ is a reliable ECC scheme of rate $R_{\mathcal{C}}$, then $R_{\mathcal{C}} \leq C_T$.

Due to this major result, we will drop the words *operational* and *information* from channel capacity and simply refer to these terms universally as ***channel capacity***. Furthermore, we will be considering wiretap channels in a moment where there is also a notion of "capacity"; to this end, we will oftentimes refer to the capacity in this section as the ***point-to-point channel capacity*** when clarity is required.

2.5. Wiretap Channels and Secrecy Capacity

A wiretap channel is the natural extension of the previous section on reliable communication theory to reliable *and secure* communication theory. In this setup, we will have a transmitter, an intended receiver, and a passive eavesdropper[1]. It will be the goal of the transmitter to successfully transmit a message to the intended receiver while keeping the message hidden from the eavesdropper. Let us make this more precise.

Definition. Let $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ be a channel modeling the communication between a transmitter and intended receiver. Let $A = (\mathcal{X}, \omega(z|x), \mathcal{Z})$ be a channel modeling the unintended communication between a transmitter and a passive eavesdropper. We call the pair $W = (T, A) = (\mathcal{X}, \omega(y|x), \omega(z|x), \mathcal{Y}, \mathcal{Z})$ the ***wiretap*** channel.

Remark. Wyner [20] coined the term wiretap channel when he was considering the situation when an eavesdropper received a noisy version of a message strictly after the intended receiver (this property of a channel is called physically degraded). Later, [6] generalized Wyner's result to include all discrete memoryless channels: they dubbed their model a broadcast channel with confidential messages. Csiszar and Körner's naming is much more appealing but unfortunately has not stuck. The term wiretap channel is nearly ubiquitous in literature even in wireless channel scenarios where

[1] We will refer to the *adversary* of Chapter 1 in the sequel as an *eavesdropper* to emphasize she can only listen but not interfere.

this naming makes little sense. For this reason, we shall stick with convention and call even these general scenarios wiretap channels.

Note that we have chosen the letters T, A, and W so as to denote the $\underline{\text{T}}$ransmission channel, $\underline{\text{A}}$dversary's channel, and $\underline{\text{W}}$iretap channel. We also note that the n letter induced wiretap channel is given by

$$W^n = (T^n, A^n) = (\mathcal{X}^n, \omega(y^n|x^n), \omega(z^n|x^n), \mathcal{Y}^n, \mathcal{Z}^n).$$

Sometimes we will refer to the n-letter extension itself as the wiretap channel.

Definition. Let $\mathcal{W} = \{\mathcal{W}_n\}_{n\in\mathbb{N}}$ be a coding scheme for channels $T = (\mathcal{X}, \omega(y|x), \mathcal{Y})$ and $A = (\mathcal{X}, \omega(z|x), \mathcal{Z})$ using message set $\mathcal{M}$. We define the following.

- A ***xs security metric*** is an association that maps the pair $(\mathcal{W}_n, A^n)$ to a non-negative real number $\mathbb{L}_n^{\text{xs}}$ for every $n \in \mathbb{N}$. We refer to $\mathbb{L}_n^{\text{xs}}$ as the ***xs-leakage***.
- If $\lim_{n\to\infty} \mathbb{L}_n^{\text{xs}} = 0$ then we say the scheme $\mathcal{W}$ is ***xs-secure*** relative to channel A. If $\log(\mathbb{L}_n^{\text{xs}}) \leq -an^b$ for some $a, b > 0$ then we say $\mathcal{W}$ is ***exceptionally xs-secure*** relative to channel A.
- We call $\mathcal{W}$ a (xs) ***wiretap coding scheme*** for wiretap channel $W = (T, A)$ if it satisfies the following two conditions:
  - *(Reliability):* $\mathcal{W}$ is a reliable ECC scheme for T.
  - *(Security):* $\mathcal{W}$ is secure relative to A.

  If these two conditions are satisfied exceptionally, then we say that $\mathcal{W}$ is an ***exceptional (xs) wiretap coding scheme*** for wiretap channel W.
- If R is the rate of an xs wiretap coding scheme, then we say R is an ***xs achievable secrecy rate.*** We call the supremum of all xs achievable secrecy rates the ***xs secrecy capacity***.

Remark. Again [2] noted that security should be ascertained exponentially fast with block length. They called such a condition "strongly xs secure." Noting again the plethora of definitions in this context whose name includes the wording "strong" we have instead opted here again for "exceptionally xs secure."

When Wyner originally considered the wiretap channel he used a security metric defined via the leakage $\mathbb{L}_n = \frac{1}{n}I(M \wedge Z^n)$ (for M uniformly random). Just as in the channel coding theorem, Wyner showed that the operational secrecy capacity was exactly equal to the information secrecy capacity for all memoryless, *physically degraded* channels; meaning the eavesdropper was located physically after the intended receiver. Csiszar and Körner [6] generalized Wyner's result exactly for the case of all discrete memoryless channels getting the exact same expression. Again this was upgraded (cf. [3]) to all memoryless channels and the expression changed simply from a maximum to a supremum (since it includes continuous channels). We give this general result next and avoid the terminology operational/information secrecy capacity since they are equivalent.

Fact 2. [3] The secrecy capacity for arbitrary memoryless channels with power constraint P using security metric induced by $\mathbb{L}_n = \frac{1}{n}I(M \wedge Z^n)$ (for M uniform) is given by

$$C_s = \sup_{VX \in \mathcal{D}} \left(I(V \wedge Y) - I(V \wedge Z) \right)$$

where $\mathcal{D} = \{VX \mid V \to X \to YZ \text{ and } \mathbb{E}[X^2] \leq P\}$.

Remark. For a discrete channel with no power constraint we can choose $P = +\infty$.

Here V is an auxiliary random variable that is often called the channel prefix. It was the insight of Csiszar and Körner [6] who saw that this was needed before the channel input to *properly* characterize secrecy capacity.

Remark. It is interesting that our pseudo-message M' of Chapter 4 turns out to be an optimal channel prefix for the AWGN wiretap channel. Generalizing this result to other wiretap channels is a current line of future research.

2.5.1. Secrecy capacity under different metrics

We will also be interested how the secrecy capacity changes under different security metrics. First let us see a way to *order* security metrics.

Definition. Fix a scheme $\mathcal{W}$ over an eavesdropper's channel A. Suppose xs_1 and xs_2 are two security metrics with respective leakage $\mathbb{L}_n^{\mathrm{xs}_1}$ and $\mathbb{L}_n^{\mathrm{xs}_2}$. We say security metric xs_1 is ***stronger***

than security metric xs_2 if

$$\lim_{n\to\infty} \mathbb{L}_n^{\mathrm{xs}_1} = 0 \implies \lim_{n\to\infty} \mathbb{L}_n^{\mathrm{xs}_2} = 0.$$

Moreover we say security metric xs_1 is ***equivalent*** to security metric xs_2 if

$$\lim_{n\to\infty} \mathbb{L}_n^{\mathrm{xs}_1} = 0 \iff \lim_{n\to\infty} \mathbb{L}_n^{\mathrm{xs}_2} = 0.$$

Now consider the following lemma.

Fact 3. Fix a wiretap channel and denote the secrecy capacity under security metric xs by $C_s\big|_{\mathrm{xs}}$. Suppose xs_1 is a stronger security metric than xs_2. It follows that:

1. $$C_s\big|_{\mathrm{xs}_1} \leq C_s\big|_{\mathrm{xs}_2}.$$

2. If $C_s\big|_{\mathrm{xs}_2}$ is an xs_1-achievable secrecy rate then

$$C_s\big|_{\mathrm{xs}_1} = C_s\big|_{\mathrm{xs}_2}.$$

The first part of this lemma says that the secrecy capacity under a stronger metric can never go up (although it can stay the same). This coincides exactly with intuition since by strengthening our measure of security, we are being more restrictive.

The second part of this lemma also says that if we can achieve the weaker secrecy capacity using a stronger metric, then in fact, the secrecy capacity under the stronger metric is equivalent to the weaker secrecy capacity; that is, we don't need to prove a converse for a secrecy capacity result: we get the result for free with an achievability proof.

Remark. A characterization of the secrecy capacity like Fact 2 under different metrics has been done (cf. [3]). However, in the next section we will define semantic security and no characterization for this metric has been done.

2.6. Concepts from Computer Science

The last set of background we need in this thesis may seem at times disparate to the previous background, but its utility will be clear in time.

2.6.1. Guessing Probability

Definition. Let M be a discrete random variable with alphabet $\mathcal{M}$. We define the ***guessing probability*** of M by

$$\mathbf{GP}(M) = \max_{m \in \mathcal{M}} \mathbb{P}\left[M = m\right].$$

Moreover, given another random variable R on $\mathcal{R}$, we define the ***average guessing probability*** of M given R by

$$\mathbf{GP}(M|R) = \int_{\mathcal{R}} \omega(r) \max_{m \in \mathcal{M}} \omega(m|r)\mu(dr),$$

where μ is some metric on $\mathcal{R}$.

Remark. Here we interpret $\max_{m \in \mathcal{M}} \omega(m|r)$ *as a function of* r*: say* $f : \mathcal{R} \to [0, \infty)$*. Therefore the average guessing probability of* M *given* R *becomes* $\int_{\mathcal{R}} \omega(r) f(r) \mu(dr) = \mathbb{E}[f(R)]$*.*

Intuitively, the guessing probability of M is the probability of guessing the outcome of the random variable M correctly when using the best strategy, which is to guess the outcome of M with the highest *a priori* probability.

If we were given the outcome of another random variable, say $R = r$, our best strategy would change to picking the outcome of M with the highest *conditional* a priori[2] probability: $\max_m \omega(m|r)$. However, perhaps $R = r$ is a rare event (or conversely an extremely likely event), then this doesn't quite capture the randomness of R. What is better is to take the expected value of the previous conditional guessing probability over all possible outcomes of R; this way, we average out all outcomes of R and get the average chance of guessing the outcome of M correctly when the random variable R is also given. As a simple consequence, if $M \perp R$ we see that $\mathbf{GP}(M|R) = \mathbf{GP}(M)$ as one would expect.

[2] With respect to the random variable M.

2.6.2. Universal Hash Families

Over the years, universal hash families (introduced in [5]) have found utility in many fields of computer science. In particular, we will study their effectiveness in providing security.

Definition. Let $\mathcal{M} = \{0,1\}^k$ be a set of binary strings of length k and $\mathcal{M}'$ and $\mathcal{S}$ finite sets such that S is a random variable on $\mathcal{S}$. Consider now a finite family of functions indexed by $\mathcal{S}$:

$$\mathcal{F} = \{f_s : \mathcal{M}' \to \mathcal{M} \,|\, s \in \mathcal{S}\}.$$

(i) $\mathcal{F}$ is called a ***universal hash family*** (UHF) if for every $m_1' \neq m_2' \in \mathcal{M}'$ we have:

$$\mathbb{P}_{s\in\mathcal{S}}\left[f_s(m_1') = f_s(m_2')\right] \leq \frac{1}{2^k}.$$

(ii) $\mathcal{F}$ is called ***uniform*** if for every $m' \in \mathcal{M}'$ and for every $m \in \mathcal{M}$ we have:

$$\mathbb{P}_{s\in\mathcal{S}}\left[f_s(m') = m\right] = \frac{1}{2^k}.$$

(iii) We call a UHF $\mathcal{F}$ ***hash value independent*** (HV independent) if for every $m_1' \neq m_2' \in \mathcal{M}'$ and $m \in \mathcal{M}$ we have:

$$\mathbb{P}_{s\in\mathcal{S}}\left[f_s(m_1') = f_s(m_2') \,\middle|\, f_s(m_1') = m\right] \leq \frac{1}{2^k}.$$

(iv) $\mathcal{F}$ is called b***-regular*** if for every $s \in \mathcal{S}$ and for every $m \in \mathcal{M}$ we have:

$$|\{m' \in \mathcal{M}' \,|\, f_s(m') = m\}| = 2^b.$$

(v) $\mathcal{F}$ is called ***invertible*** if for each $s \in \mathcal{S}$ there exists some stochastic mapping $\phi_s : \mathcal{M} \to \mathcal{M}'$ such that for all $m \in \mathcal{M}$, $f_s(\phi_s(m)) = m$. If the stochastic map ϕ_s maps $m \in \mathcal{M}$ according to a uniform distribution to its image $\phi_s(m) \subset \mathcal{M}'$ for every $s \in \mathcal{S}$ and $m \in \mathcal{M}$ then we call $\mathcal{F}$ ***evenly invertible***.

(vi) Lastly, we call $\mathcal{F}$ a ***semantic security inducing universal hash family*** (SSI-UHF) if it is: (1) universal, (2) uniform, (3) HV independent, (4) b-regular, and (5) evenly invertible.

Remark. When $S \sim \text{unif}(\mathcal{S})$ (as will be done exclusively in this paper), the aforementioned probabilities can simply be rewrote as counting probabilities from combinatorics. For example consider the uniform property above. Fix $m \in \mathcal{M}$ and $m' \in \mathcal{M}'$ and let $\mathcal{U} = \{s \in \mathcal{S} \mid f_s(m') = m\}$. Then $\mathbb{P}_{s\in\mathcal{S}}[f_s(m') = m] = |\mathcal{U}|/|\mathcal{S}|$.

Many of the definitions here coincide with those found in computer science literature. Indeed, the conditions of being a *universal hash family* and *uniform* are found in most textbooks on hash families. The condition of being *b-regular* and *invertible* can be found in [1] and [18]. That being said, we have invented some terminology. We have dubbed hash families that are universal, uniform, hash value independent, b-regular, and evenly invertible as *semantic security inducing universal hash families* to emphasize hash families with these five properties as the proper ones for inducing semantic security (see Chapter 3) on a wiretap channel as we shall later see. We have also invented the terminology *hash value independent*. Intuitively, HV independence means the probability of a collision does not increase by *too much* even if we know where two values collide to. Note in particular that if a HV independent UHF is also uniform then for every $m'_1 \neq m'_2 \in \mathcal{M}'$ and $m \in \mathcal{M}$ we have

$$\mathbb{P}_{s\in\mathcal{S}}\left[f_s(m'_1) = f_s(m'_2) = m\right] \leq \frac{1}{2^k 2^k}.$$

.

2.6.3. Efficiency

In the sequel, codes, wiretap codes, UHFs, SSI-UHFs, and pre/post processors (as described in the next chapter) will all be described by ***algorithms*** (when referring to their implementation). In particular these *algorithms* will all inherently be functions of the block length n. We will be concerned with the complexity of their implementation with respect to n as follows.

Definition. Let f, g be algorithms. If there exists positive constants c and n_0 such that

$$f(n) \leq cg(n) \quad \forall n > n_0,$$

then we write $f = \mathcal{O}(g)$ and say f is "big-O" of g.

Remark. We will always suppress the constants c and n_0 in this thesis.

What this definition means is that asymptotically, the growth rate of f is approximately *at most* that of g.

Definition. Let f be an algorithm. Ordered from slowest growing to fastest growing, we say the time complexity of f (with respect to n) is:

- ***linear*** if $f = \mathcal{O}(n)$,
- ***polynomial***[3] if there exists a constant $c > 1$ such that $f = \mathcal{O}(n^c)$, and
- ***exponential*** if there exists a constant $c > 1$ such that $f = \mathcal{O}(c^n)$.

Remark. All linear algorithms are also polynomial.

Polynomial time complexity is often described as "fast" in the computer science world. With this we have the gold standard definition for efficiency.

Definition. Let f be an algorithm. We say that f (or the process described using f) is ***efficient*** if f has polynomial time complexity.

[3] If $c = 2$ here, we sometimes will be explicit and say the time complexity of f is ***quadratic***.

3. PRELUDE

Now that we have the proper language, in this chapter we provide *specific mathematical* motivation for our main result and discuss previous works leading to this point.

3.1. Comparison of Security Metrics

Guaranteeing wiretap security is only as good as the metric being used. If the metric does not coincide sufficiently with reality, then all results, however mathematically sound, will not be readily applicable in practice. With this intutition in mind, it is therefore of interest to examine the metrics used in the wiretap community and see how they fare when put under the microscope. This is exactly what was done in [2]. In that work, the authors have argued traditional means of measuring security in the information theory community fall short in practice. In particular, the original metric of Wyner and the so called *strong* security metric introduced by [12] do not properly gauge an eavesdropper's advantage in a practical setting. The authors thus introduce several new metrics to close this gap including one that looks identical to the strong metric but is maximized over all message distributions and one that is analogous to the gold-standard metric from cryptography: semantic security. Let us review these wiretap security metrics.

Definition. Fix a wiretap coding scheme $\mathcal{W}$ for an eavesdropper's channel $A = (\mathcal{X}, \omega(y|x), \mathcal{Z})$ using message set $\mathcal{M}$.

1. The ***weak security metric*** (weak) is the original metric of Wyner [20] and the leakage is given by
$$\mathbb{L}_n^{\text{weak}} = \frac{1}{n} I(M \wedge Z^n), \quad M \sim \text{unif}(\mathcal{M}).$$

2. The ***mutual-information security metric for random messages*** (mis-r) has leakage given by
$$\mathbb{L}_n^{\text{mis-r}} = I(M \wedge Z^n), \quad M \sim \text{unif}(\mathcal{M}).$$

3. The ***mutual-information security metric*** (mis) has leakage given by

$$\mathbb{L}_n^{\text{mis}} = \max_{P_M} I(M \wedge Z^n).$$

4. The ***semantic security metric*** (ss) has leakage given by

$$\mathbb{L}_n^{\text{ss}} = \sup_{f,M} \Big(\mathbf{GP}(f(M)|Z^n) - \mathbf{GP}(f(M)) \Big),$$

   where f is any function $\mathcal{M} \to \{0,1\}^*$ (finite).

Remark. The mutual information security metric for random messages (mis-r) was originally called the strong security metric. However, as we shall see shortly, this is an inappropriate name since the metric is not quite as "strong" as researchers once thought. We have thus chosen to stick with the name introduced in [2] and ignore the naming strong.

3.1.1. Weak Metric

When Wyner originally considered the weak metric in his pioneering work, he was thinking about the leakage $\mathbb{L}_n^{\text{weak}}$ as a *rate*; that is, Wyner said a wiretap scheme was secure if the rate of mutual information the eavesdropper obtained about the message went to 0 asymptotically with block length n. This seems reasonable, but can be shown to have severe problems. As a simple pathological example, suppose the mutual information between the message and the eavesdroppers output $I(M \wedge Z^n)$ (for M uniform) grows as $\log(n)$ with block length n. Then as $n \to \infty$, this mutual information term grows unbounded while the weak-leakage $\mathbb{L}_n^{\text{weak}}$ goes to 0. Thus, if we used the weak security metric to measure security in this situation, we would say our scheme is secure yet the eavesdropper is receiving an *infinite* amount of information as measured by mutual information. This is clearly unsettling, but admittedly this example is a bit ad hoc, however, even non-pathological examples can be constructed that show the weak metric is quite unappealing in a practical sense (cf. [4]).

3.1.2. MIS-R Metric

The pitfalls of the weak metric led the information theoretic community to define a new metric: the mutual information security metric with random messages [12]. The motivation behind this metric is that instead of the *rate* asymptotically going to 0, the *total* amount of mutual information between the uniformly distributed message and eavesdropper's output should asymptotically go to 0. Indeed our pathological example mentioned for weak security no longer works. This metric seems to be exactly what we want as information theorists. This metric even mimics *perfect* security as originally suggested by [14] except there the total amount of mutual information is *exactly* 0 for a finite blocklength. However, there is one serious problem.

3.1.3. MIS Metric

The *mis-r* metric has recently been questioned by the likes of cryptographers [2]. In particular, the message of the mis-r metric is assumed to be uniformly distributed, however, messages in real life are rarely this structured. It was originally argued by information theorists that since a message is always assumed to be compressed before being transmitted that the assumption of a uniform distribution was correct. However, as pointed out in [2], compression in real life is a deterministic function that cannot possibly change entropy. Indeed by this motivation, [2] defined the mutual-information security metric exactly as the *mis-r* security metric *except* maximized over all message distributions. This metric has the information theorists intuition that for block length sufficiently large, the total amount of information the eavesdropper receives about the message is negligible *for any kind of message.*

3.1.4. Semantic Security Metric

At this point, an information theorist could be satisfied since there does not seem like much more one could ask for in an asymptotic security metric, however, as cryptographers, [2] put forth another metric that is the information theorists analog of the gold standard in cryptography: the semantic security metric. Let us break apart the intuition behind how $\mathbb{L}_n^{\text{ss}}$ is defined.

Temporarily fix f as the identity function and M as a random variable with some arbitrary message distribution. In this case the semantic leakage becomes $\mathbf{GP}(M|Z^n) - \mathbf{GP}(M)$ and we are basically asking: "how much will the eavesdropper's odds of correctly guessing the realization of M increase when given Z^n"? Ideally, we do not want the eavesdropper's odds to increase at all, but we would be satisfied if this difference went to 0 asymptotically with block length n. Now relax the distribution on M and optimize over all message distributions. At this point, we are asking for how much the eavesdropper's odds increase no matter how unstructured our message is. Finally, we need to relax f and consider all possible functions: why? Well, as argued in [2], we should not be only content that the eavesdropper's odds of guessing the entire message barely increase, rather, we should also guarantee that the eavesdropper's odds of guessing *any* portion of the message barely increase. Indeed, the first bit of a real life message could possibly contain very important information yet, if we don't optimize over f, we would be allowing the case where the eavesdropper could guess the first bit just not the entire message. Hence, optimizing over f allows us to guarantee that the eavesdropper's odds of guessing any partial part or any manipulation of the message do not increase when given Z^n.

The intuition behind semantic security in this regime is admittedly much more sound than any of the other metrics mentioned. Asymptotically, semantic security guarantees that given Z^n, an eavesdropper has no better chance of guessing any conceivable manipulation of M then by the obvious strategy of picking the most probable realization. This is extremely satisfying and pragmatic not to mention there does not seem like any other metric we could define that could give us more asymptotically. An immediate question that arises is then: "how much stronger is semantic security than mutual-information security?". The surprising and celebrated result of [2] is that asymptotically, semantic security is **equivalent** to mutual information security.

3.1.5. Metric Ordering

We collect this result along with the other orderings below.

Fact 4. Fix a wiretap coding scheme $\mathcal{W}$ for a wiretap channel $W = (T, A)$.

- Semantic Security is equivalent to mutual-information security:

$$\lim_{n\to\infty} \mathbb{L}_n^{\text{ss}} = 0 \iff \lim_{n\to\infty} \mathbb{L}_n^{\text{mis}} = 0.$$

- Mutual information security is *strictly stronger* than mutual information security for random messages:

$$\lim_{n\to\infty} \mathbb{L}_n^{\text{mis}} = 0 \overset{\not\Leftarrow}{\Longrightarrow} \lim_{n\to\infty} \mathbb{L}_n^{\text{mis-r}} = 0.$$

- Mutual information security for random messages is *strictly stronger* than weak security:

$$\lim_{n\to\infty} \mathbb{L}_n^{\text{mis-r}} = 0 \overset{\not\Longleftarrow}{\Longrightarrow} \lim_{n\to\infty} \mathbb{L}_n^{\text{weak}} = 0.$$

Remark. This result was originally proved for discrete random variables in [2] but was identically replicated for continuous random variables by [10], thus we make no distinction between the two cases.

The semantic security metric correctly captures the notion of *asymptotic* perfect secrecy, however, it is somewhat arduous to work with. Since it is equivalent to mutual information security, which is considerably easier to work with we can consider mis while getting semantic security for free. With this we have a conceivable way to achieve what we set out to do in the introduction.

3.2. Main Contributions

With the preceding intuition regarding security metrics in mind, the main contributions of this thesis are to partially rectify the two main problems of physical layer security as mentioned in Chapter 1. In particular, in this thesis we do the following.

- We construct a wiretap coding scheme that consists of a preprocessing scheme with a reliable ECC scheme. We show that the preprocessing scheme can be described algorithmically guaranteeing that it is concrete and realizable in practice. Furthermore, we prove that our preprocessing scheme is efficient in block length (quadratic time).

- We construct a bound on the *mis* leakage when using our wiretap scheme by upgrading a direct approach found in [18]. In more detail, [18] provided a leakage bound when the message M was uniformly distributed (that is, they provided a bound on *mis-r* leakage); we remove this restriction and provide an analogous bound for any distribution on the message M.

- We show that our wiretap coding scheme can achieve a positive achievable rate under the *mis* security metric for arbitrary channels so that asymptotically our wiretap scheme is semantically secure. Our wiretap coding scheme is modular and acts like a converter that takes as input an error correcting code and outputs a semantically secure wiretap code under certain ECC rate and eavesdropper channel conditions.

- We show that our coding scheme can achieve the secrecy capacity of the AWGN wiretap channel with semantic security in a concrete and efficient way, thus completely solving the two main problems of physical layer security at least for the AWGN case.

3.3. Prior Work

With regard to concrete and efficient schemes under the *mis-r* metric, [17] was able to achieve the *mis-r* secrecy capacity on AWGN channels and [18] later extended this result to both AWGN and *certain* discrete memoryless channels. Indeed [18] provided a method of proof to bound their *mis-r* leakage of which our bounds in Chapter 5 are inspired by.

Semantic security for the wiretap channel was first introduced in [2]. In that work the authors provided a wiretap scheme based on UHFs that achieved positive rates for certain discrete memoryless channels but their wiretap coding scheme could not achieve the secrecy capacity. An extension of that work by the same authors was presented in [1]. There the authors constructed a polynomial time efficient wiretap scheme that *could* achieve the secrecy capacity of *certain* discrete memoryless channels. This class of channels was fairly restrictive, but this result was generalized to even more discrete memoryless channels in [16]. All of the preceding schemes were concrete and established in polynomial time but the proofs restrict them to discrete alphabets (and certain error correcting codes).

With respect to the AWGN wiretap channel, [11] only recently provided a concrete and efficient wiretap coding scheme that could achieve the secrecy capacity. We were only recently made aware of this result and it is noted that even though we obtain the same result in Chapter 7, our wiretap scheme is based on *completely* different principles than the former. It is also noted that an attempt at a wiretap scheme that could achieve the semantic secrecy capacity for AWGN channels in an analogous way to the one we present in Chapter 4 was redacted due to an error (according to the authors). We are unaware of the exact error therein but our approach is relatively different.

4. A WIRETAP CODING SCHEME

In this chapter we will furnish a wiretap coding scheme $\mathcal{W}$ for an arbitrary wiretap channel $W = (T, A)$ which is based on a wiretap scheme put forth in [1], [17], and [18]. We will first define each step of this scheme and show that it is reliable. Then we will give a particular implementation and show that this implementation is efficient with respect to the block length n.

4.1. Transmission Procedure

Over an arbitrary wiretap channel $W = (T, A)$, our transmission procedure involves combining a SSI-UHF with a reliable ECC already in use over the main point to point channel. This modular wiretap scheme is precisely the scheme put forth in [1, 17, 18] except there, the UHF was only required to be b-regular and evenly invertible. Since here we require a SSI-UHF, we are also demanding that our UHF be *hash value independent* and *uniform*. The necessity of these two extra properties will be fleshed out in the next chapter (particularly in the Leftover Hash Lemma).

Consider Figure 4.1; this describes our transmission procedure.

$$\begin{tikzcd}
 & & & Y^n \arrow[r, "d_n"] & \widehat{M'} \arrow[r, "\Gamma^{-1}"] & \widehat{M} \\
(M, S) \arrow[r, "\Gamma"] & M' \arrow[r, "e_n"] & X^n \arrow[ur, "T^n"] \arrow[dr, "A^n"'] & & & \\
 & & & Z^n & &
\end{tikzcd}$$

Figure 4.1. Transmission Scheme.

4.1.1. Preprocessing Layer

Consider the finite sets $\mathcal{M} = \{0,1\}^k$ and $\mathcal{M}' = \{0,1\}^l$ with $l > k$. We shall refer to $M \in \mathcal{M}$ as the ***actual message*** and $M' \in \mathcal{M}'$ as the ***pseudo-message*** because M represents the information the transmitter *actually* wishes to impart to the intended receiver securely, whereas M'

is some random variation of the actual message necessary for security. We will *not* assume which distribution the message M takes so that our transmission procedure can lead to semantic security.

Over a fixed arbitrary finite set $\mathcal{S}$, the transmitter will first draw a seed $S \sim \text{unif}(\mathcal{S})$ to be used for the remainder of transmission. We assume the seed is independent of the message M and that the realized seed is *publicly* available to all parties. All communication must take place over the wiretap channel; however, we will show later in this thesis that the transmitter can send the seed before the transmission of an actual message with no asymptotic rate or security loss.

The transmitter now chooses a SSI-UHF $\{f_s : \mathcal{M}' \to \mathcal{M} \,|\, s \in \mathcal{S}\}$. Suppose each function f_s in the SSI-UHF has its inverse given by ϕ_s. The transmitter uses this inverse to "inversely hash" an actual message to a pseudo message using a channel $\Gamma : \mathcal{M} \times \mathcal{S} \to \mathcal{M}'$ given by $\Gamma(m, s) = \phi_s(m)$, which we call the *pre-processing layer*. In particular, since this UHF is evenly invertible, if some message $m \in \mathcal{M}$ and seed $s \in \mathcal{S}$ are realized, the pseudo-message M' is chosen according to $\text{unif}(\phi_s(m))$. Since the SSI-UHF is b-regular, each image $\phi_s(m)$ has 2^b elements. Thus, Γ has transition density $\omega(m'|m, s) = 2^{-b}$ for any choice of $m \in \mathcal{M}$, $s \in \mathcal{S}$, and $m' \in \phi_s(m)$.

4.1.2. Coding Layer

The transmitter chooses some reliable ECC scheme $\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$ that satisfies the power constraint for the channel if there is one. We will assume (as per standard) that each party has full knowledge of $\mathcal{C}$. Thus, for a given blocklength n, each party knows $\mathscr{C}_n$ is the codebook and we have inherently induced new channels: $T^n : \mathscr{C}_n \to \mathcal{Y}^n$ and $A^n : \mathscr{C}_n \to \mathcal{Z}^n$. We will henceforth be considering *these* as the main and eavesdropper's channels for the remainder of the thesis. At this point the transmitter channel encodes the pseudo-message M' using e_n, this will be a random variable $X^n = e_n(M')$ over $\mathscr{C}_n$. Next the transmitter sends X^n over the wiretap channel $W = (T, A)$; that is, the channel input X^n is sent across T^n but also across A^n inherently.

4.1.3. Decoding

Let us first focus on the intended receivers channel. The intended receiver will receive a (potentially) noisy version of the channel input: $Y^n = T^n(e_n(M'))$. The goal of the intended receiver is to correctly guess M' from Y^n. This is accomplished using the estimate $\widehat{M'} = d_n(Y^n)$. Since we have assumed $\mathcal{C}$ to be reliable, each $\mathcal{C}_n$ is an ECC. Thus, the probability of error $p_{e,n}$

is considerably low. In particular, this means there is a high probability that $\widehat{M'} = M'$ where asymptotically with block length n, this equality happens almost surely.

Next, the intended receiver shall *post-process* $\widehat{M'}$ to an estimate of the actual message $\widehat{M}$ using the hash function corresponding to the public seed S. That is, the post processing channel $\Gamma^{-1} : \mathcal{M}' \times \mathcal{S} \to \mathcal{M}$ is given by $f_s(m')$. Since our UHF is invertible, if $\widehat{M'} = M'$ then the UHF is guaranteed to map $\widehat{M'}$ to M (the original message). In this sense, the pre-post processing layers do not subtract anything from our reliability. In more detail, if $\mathcal{C}$ is reliable to begin with then our entire wiretap scheme will also satisfy reliability. Furthermore, if $\mathcal{C}$ is exceptionally reliable, then our wiretap scheme is exceptionally reliable as well.

4.1.4. Eavesdropper's Channel layer

Once the eavesdropper receives her channel output $Z^n = A^n(e_n(M'))$ she will attempt to decode it in a similar fashion to that of the intended receiver; however, we will not assume *how* she decodes her output since that could affect our measure of security. As a side note, in contrast to computational based security methods, we also do *not* assume the boundedness of resources at the eavesdropper.

4.1.5. Discussion

As in [18], we call the preceding scheme *modular* since the preprocessing layer is not intrusive to the main channel in any way. That is, our preprocessing layer could be added to any already existing communication system without changing any core components of the original system.

4.2. Implementation

Our wiretap scheme consists of two components. First we have a pre/post processing scheme based on a SSI-UHF. Second we have a reliable ECC scheme $\mathcal{C}$ for the main point to point channel T. The first question we might ask is: "does such a wiretap scheme even exist?" Well Shannon's channel coding theorem can be extended to almost *any* point to point channel T. This means that a reliable ECC scheme $\mathcal{C}$ will always exist for any channel so long as the rate of the ECC scheme satisfies $R_{\mathcal{C}} \leq C_T$, the point to point main channel capacity. Thus, we only need to be concerned if such a pre/post processing scheme exists; in particular, if a SSI-UHF exists.

In this section we construct a SSI-UHF to be used in the pre/post processing layers for our wiretap coding scheme. We start with the UHF construction used in the implementation of the scheme in [1, 17, 18] upon which our wiretap scheme is based. We show that this construction is very close to what we need but it is not quite a SSI-UHF. However, this construction gives way to a similar construction that *is* a SSI-UHF, which we prove. Lastly, we prove that this construction is implementable in polynomial time with respect to block length n.

4.2.1. A Close Construction

Denote the all-0 bit string of length l by 0^l. Let $\mathcal{M} = \{0,1\}^k$, $\mathcal{M}' = \{0,1\}^l$, and $\mathcal{S} = \{0,1\}^l \setminus 0^l$. Interpret the l-bit strings of $\mathcal{M}'$ and $\mathcal{S}$ as elements in the finite field $GF(2^l)$ and the k-bit strings of $\mathcal{M}$ as elements in the finite field $GF(2^k)$. We will denote the multiplicative operation of the field $GF(2^l)$ by $\odot$. With this, we can define a family of functions (cf. [1, 18, 17]) by

$$\mathcal{F}^* = \{f_s : \mathcal{M}' \to \mathcal{M} \,|\, s \in \mathcal{S}\},$$

where $f_s(m') = \mathfrak{sel}_k(s \odot m')$ and $\mathfrak{sel}_k(\cdot)$ is a function that simply selects the k most significant bits. This family is easily seen to be universal, $(l-k)$-regular, and properly invertible. The inverse is given by $\phi_{s,R}(m) = s^{-1} \odot (m||R)$ where R is some *uniform* random variable over $\{0,1\}^{l-k}$, s^{-1} is the inverse element of s in $GF(2^l)$, and $(\cdot||\cdot)$ is the *concatenation* function. Note, that since R is uniformly random, so is $\phi_{s,R}(m)$ over the preimage $f_s^{-1}(m)$, thus this UHF is evenly invertible. Moreover it is mentioned in [1, 18, 17] that this family is polynomial time efficient.

Due to the plethora of satisfying properties this family has, an obvious first attempt to instantiate the pre/post processor of *our* coding scheme is certainly this family. Indeed, this family already satisfies three of the five properties of SSI-UHFs and the pre and post processors induced by this scheme are polynomially time computable with block length n. We simply need to show this family is both uniform and HV independent.

Unfortunately, $\mathcal{F}^*$ does **not** satisfy the uniformity property. Suppose $m' = 0^l$, then $f_s(0^l) = \mathfrak{sel}_k(s \odot 0^l) = \mathfrak{sel}_k(0^l) = 0^k$ for every $s \in \mathcal{S}$ and so $\mathbb{P}[f_s(m') = m] = 1 \not\lesssim 2^{-k}$. If we try to remove the problem point 0^l from $\mathcal{M}'$, then we ought to remove 0^k from $\mathcal{M}$ because no combination of (s, m') will map to 0^k and we would not satisfy regularity.

4.2.2. Fixing the previous construction

In order to “fix” the previous construction, consider the following family of functions

$$\mathcal{F} = \{f_{s,t} : \mathcal{M}' \to \mathcal{M} \mid s \in \{0,1\}^l \setminus 0^l,\ t \in \{0,1\}^l\}$$

where $f_{s,t}(m') = \mathfrak{sel}_k\left((s \odot m') \oplus t\right)$. Here, $\oplus$ and $\odot$ denote addition and multiplication respectively in $GF(2^l)$ and $\mathfrak{sel}_k(\cdot)$ selects the k most significant bits just as in the previous family. As a remark, we note that $\oplus$ here is equivalent simply to modulo-2/bitwise/xor addition. Moreover, we note that the entire seed in this case is given by $\mathcal{S} = \{0,1\}^l \setminus 0^l \times \{0,1\}^l$ where $|\mathcal{S}| = (2^l - 1)2^l$.

Theorem 1. The family of functions

$$\mathcal{F} = \{f_{s,t} : \mathcal{M}' = \{0,1\}^l \to \mathcal{M} = \{0,1\}^k \,|\, s \in \{0,1\}^l \setminus 0^l, t \in \{0,1\}^l\}$$

given by $f_{s,t}(m') = \mathfrak{sel}_k\left((s \odot m') \oplus t\right)$ with inverses given by $\phi_{s,t,R}(m) = s^{-1} \odot ((m||R) \oplus t)$ (with $R \sim \text{unif}(\{0,1\}^{l-k})$) is a SSI-UHF.

Proof. We will show $\mathcal{F}$ is universal, uniform, HV independent, $(l-k)$-regular, and evenly invertible.

- *Universality:* Fix $m_1' \neq m_2' \in \mathcal{M}'$. We wish to count how many (s,t) satisfy:

  $$\mathfrak{sel}_k\left(\left(s \odot m_1'\right) \oplus t\right) = \mathfrak{sel}_k\left(\left(s \odot m_2'\right) \oplus t\right).$$

  Since $\oplus$ is given by bitwise addition, we can distribute $\mathfrak{sel}_k(\cdot)$ and reduce the equation to: $\mathfrak{sel}_k(s \odot m_1') \oplus_k \mathfrak{sel}_k(t) = \mathfrak{sel}_k(s \odot m_2') \oplus_k \mathfrak{sel}_k(t)$ where $\oplus_k$ is addition over $GF(2^k)$. This reduces even further to $\mathfrak{sel}_k(s \odot m_1') = \mathfrak{sel}_k(s \odot m_2')$, however, this is an equation that does not involve t so that indeed, any choice of t satisfies the original equation.

  This equation can be rewritten as

  $$0^k = \mathfrak{sel}_k(s \odot m_1') \oplus_k \mathfrak{sel}_k(s \odot m_2') = \mathfrak{sel}_k((s \odot m_1') \oplus (s \odot m_2')) = \mathfrak{sel}_k(s \odot m''),$$

where we have defined $m'' = m'_1 \oplus m'_2$. Now since $m'_1 \neq m'_2$ then $m'' = m'_1 \oplus m'_2 \neq 0^l$. Moreover by assumption $s \neq 0^l$ so that for each choice of s, the multiplication $s \odot m''$ is a unique element in $\{0,1\}^l \setminus 0^l$. Since there are $2^{l-k} - 1$ elements in $\{0,1\}^l \setminus 0^l$ that have the first k bits set to 0, then there are $2^{l-k} - 1$ choices of s that satisfy $0^k = \mathfrak{sel}_k(s \odot m'')$.

In summary, we have 2^l choices for t and $2^{-k}(2^l - 2^k)$ choices for s, thus we have $2^{-k}2^l(2^l - 2^k)$ choices for (s,t) that satisfy $\mathfrak{sel}_k((s \odot m'_1) \oplus t) = \mathfrak{sel}_k((s \odot m'_2) \oplus t)$. However, $2^{-k}2^l(2^l - 2^k) \leq 2^{-k}2^l(2^l - 1)$ since $k \geq 1$ so that (noting $|\mathcal{S}| = 2^l(2^l - 1)$) we have proved that $\mathcal{F}$ is a universal hash family.

- *Uniformity:* Fix $m' \in \mathcal{M}'$ and $m \in \mathcal{M}$. We wish to count how many (s,t) satisfy:

$$\mathfrak{sel}_k\left((s \odot m') \oplus t\right) = m.$$

We can distribute $\mathfrak{sel}_k(\cdot)$ and view this as the equation $\mathfrak{sel}_k(t) = m \oplus_k \mathfrak{sel}_k(s \odot m')$. For each choice of s the first k bits of t are fixed and the last $l-k$ bits are free; thus there are 2^{l-k} choices for t. Since there are no restrictions at all on s, we can choose any of the $2^l - 1$ l-length bits strings (excluding 0^l) for s.

In aggregate there are $2^{l-k}(2^l - 1)$ choices of (s,t) that satisfy $\mathfrak{sel}_k((s \odot m') \oplus t) = m$. Noting again that $|\mathcal{S}| = 2^l(2^l - 1)$ we have proven that our family $\mathcal{F}$ is uniform.

- *HV Independence:* Fix some $m'_1 \neq m'_2 \in \mathcal{M}'$ and $m \in \mathcal{M}$. We wish to count how many (s,t) satisfy

$$m = \mathfrak{sel}_k((s \odot m'_1) \oplus t) = \mathfrak{sel}_k((s \odot m'_2) \oplus t).$$

Focus first on $m = \mathfrak{sel}_k((s \odot m'_1) \oplus t)$. It follows that $\mathfrak{sel}_k((s \odot m'_1) \oplus t) = \mathfrak{sel}_k(s \odot m'_1) \oplus_k \mathfrak{sel}_k(t)$ so that $\mathfrak{sel}_k(t) = m \oplus_k \mathfrak{sel}_k(s \odot m'_1)$. This means that for each choice of s, the first k bits of t must be chosen as $m \oplus_k \mathfrak{sel}_k(s \odot m'_1)$ in order to satisfy the above equation however the last $l-k$ bits of t are free so that for every choice of s we have 2^{l-k} choices of t.

Now focus on $m = \mathfrak{sel}_k((s \odot m'_2) \oplus t) = \mathfrak{sel}_k(s \odot m'_2) \oplus_k \mathfrak{sel}_k(t)$. From the above paragraph it immediately follows that $m = \mathfrak{sel}_k(s \odot m'_2) \oplus_k m \oplus_k \mathfrak{sel}_k(s \odot m'_1)$. This implies $\mathfrak{sel}_k(s \odot m'_1) =$

$\mathfrak{sel}_k(s \odot m_2')$. However, we have already seen from the universality of $\mathcal{F}$ that the number of s that satisfy this is upper bounded by $2^{-k}(2^l - 1)$.

In aggregate, there are never more than $2^{-k}2^{-k}2^l(2^l - 1)$ choices of (s, t) that satisfy the equation: $m = \mathfrak{sel}_k((s \odot m_1') \oplus t) = \mathfrak{sel}_k((s \odot m_2') \oplus t)$. Since the number of seeds is given by $|\mathcal{S}| = 2^l(2^l - 1)$, we have shown our family $\mathcal{F}$ is indeed hash independent.

- *Regularity:* Fix some $m \in \mathcal{M}$, $s \in \{0,1\}^l \setminus 0^l$, and $t \in \{0,1\}^l$. We wish to count how many m' satisfy:

  $$\mathfrak{sel}_k\left(\left(s \odot m'\right) \oplus t\right) = m.$$

  As usual break up this equation to $\mathfrak{sel}_k(s \odot m') = m \oplus_k \mathfrak{sel}_k(t)$. Since we are working in $GF(2^l)$ and $s \neq 0^l$, for each choice of $m' \in \{0,1\}$ the product $s \odot m'$ will be a *unique* element in $\{0,1\}^l$. But the first k bits of this product are fixed at $m \oplus_k \mathfrak{sel}_k(t)$ due to the previous equation while the last $l - k$ bits are free. Hence there will be 2^{l-k} choices of m' that satisfy the original equation.

  Therefore, $\mathcal{F}$ is $(l - k)$-regular.

- *Invertibility:* Let $m \in \mathcal{M}$, $s \in \{0,1\}^l \setminus 0^l$, and $t \in \{0,1\}^l$. Then,

  $$\begin{aligned} f_{s,t}(\phi_{s,t,R}(m)) &= \mathfrak{sel}_k\left(s \odot \left(s^{-1} \odot ((m||R) \oplus t)\right) \oplus t\right) \\ &= \mathfrak{sel}_k\left((m||R) \oplus t \oplus t\right) \\ &= \mathfrak{sel}_k\left(m||R\right) \\ &= m, \end{aligned}$$

  where the penultimate equality follows from the fact $GF(2^l)$ has characteristic 2.

- *Even Invertibility:* Suppose we are given a $m \in \mathcal{M}$, $s \in \{0,1\}^l \setminus 0^l$, and $t \in \{0,1\}^l$. Since $R \sim \text{unif}(\{0,1\}^{l-k})$ the pseudo-message M' will also be chosen uniformly over $\phi_{s,t,R}(m)$.

■

With this theorem we have found a concrete (algorithmic) implementation of a SSI-UHF. This means our wiretap coding scheme of the previous section is well defined.

With this, let us see how the time complexity of our pre/post processing scheme fares for this implementation.

Proposition 1. Noting that l and k are functions of the coding block length n, we have:

1. Given $m \in \mathcal{M}$, $s \in \{0,1\}^l \setminus 0^l$, $t \in \{0,1\}^l$, and $r \in \{0,1\}^{l-k}$, the inverse $\phi_{s,t,r}(m) = s^{-1} \odot ((m||r) \oplus t)$ can be computed in quadratic-time with respect to n.

2. Given $s \in \mathcal{S}$ and $m' \in \mathcal{M}'$, the function $f_{s,t}(m') = \mathfrak{sel}_k((m' \odot s) \oplus t)$ can be computed in quadratic time with respect to n.

Proof.

1. Concatenation here has time complexity $\mathcal{O}(k+(l-k))$ and thus is linear with n: $\mathcal{O}(k+(l-k)) = \mathcal{O}(l) = \mathcal{O}(nR_{\mathcal{C}_n}) = \mathcal{O}(n)$. Addition in $GF(2^l)$ operates as bitwise addition (or XOR) and thus the time complexity is also linear with n: $\mathcal{O}(l) = \mathcal{O}(nR_{\mathcal{C}_n}) = \mathcal{O}(n)$. Therefore, the operation $(m||r) \oplus t$ has time complexity $\mathcal{O}(n+n) = \mathcal{O}(n)$; i.e. it is linear.

   Now inversion and multiplication in $GF(2^l)$ is known to be computed in quadratic time in l (cf. [7, Chapter 2]). Thus computing s^{-1} is $\mathcal{O}(n^2)$ and computing the multiplication $s^{-1} \odot ((m||r) \oplus t)$ is $\mathcal{O}(n^2)$. Computing the entire inverse $\phi_{s,r}(m)$ is therefore on the order of $\mathcal{O}(n + 2n^2) = \mathcal{O}(n^2)$.

   In aggregate, this entire first step by the transmitter can be computed in polynomial time in n; in particular, the entire preprocessing scheme is quadratic.

2. Using the same arguments as above, the operation $m \odot s$ can be implemented in quadratic time and addition can be implemented in linear time. Clearly, $\mathfrak{sel}_k(\cdot)$ can be implemented in $\mathcal{O}(k) = \mathcal{O}(n)$: linear time with n. Thus, the entire post-processing scheme also can be implemented in quadratic time in n.

■

Thus, in conclusion of this section, we have constructed a concrete and efficient SSI-UHF so that the pre and post processors induced by this SSI-UHF are polynomially time computable with block length n. Indeed there is not much more one could ask for in a family of hash functions but we do stress that the construction given here is by no means unique. When designing a system, any such SSI-UHF will do for the pre/post processing schemes of our wiretap scheme although it does behoove one to find a SSI-UHF that is concrete and *optimally* efficient[1].

[1] Indeed the main reason our SSI-UHF is efficient is based on the efficiency of multiplication and inversion in $GF(2^l)$. Constructing a SSI-UHF using another more efficient method is of considerable interest.

5. SECURITY OF OUR WIRETAP SCHEME

We have already seen that the scheme we constructed in Chapter 4 satisfies the reliability property of a wiretap scheme (and does so exceptionally when $\mathcal{C}$ is chosen exceptionally). Now we need to show that the scheme satisfies the security property as well. In this section we will do just that by constructing leakage bounds for the *mis* metric. It will turn out that under certain conditions our leakage bounds asymptotically go to 0 implying that our scheme is a *mis* wiretap scheme and hence a semantic wiretap scheme. In particular, under further restrictions, our wiretap scheme is shown to be exceptional.

It is noted that leakage bounds for arbitrary wiretap channels using evenly invertible, b-regular UHFs are already given in [18]; however, the leakage there assumes M follows a uniform distribution and hence will only lead to *mis-r* security at best. We therefore need to generalize the *leftover hash lemma* (channel version) in [18] to overcome this obstacle. What becomes obvious upon proof is that considering UHF's that are only evenly invertible and b-regular is not quite restrictive enough to lead to semantic security; this explains why in our wiretap coding scheme of Chapter 4 we chose our UHF to also be *uniform* and *hash value independent.* Now before we begin, we introduce some definitions and notation to facilitate our claim.

5.1. Max Information and Typical Sets

The leakage bounds we present in the next subsection and those found in [18] depend on a curious metric called max-information. Intuitively, max-information measures the *most* amount of "information" that could be sent across a given channel A^n using a specific code; however, "information" in this context may not be equivalent to *mutual information.*

Definition. Let μ be some measure on $\mathscr{Z}^n$. Then with respect to some n length ECC with codebook $\mathscr{C}_n$ used over wiretap channel $W = (T, A)$, we define ***max-information*** (over A^n) by

$$\boldsymbol{\mathcal{I}}_n = \log\left(\int_{\mathscr{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n|x^n)\mu(dz^n)\right).$$

In general, $\omega(z^n|x^n)$ finds the *relative likelihood* that an output z^n occurs given that $x^n \in \mathscr{C}_n$ occurred. Therefore $\max\limits_{x^n \in \mathscr{C}_n} \omega(z^n|x^n)$ measures the *highest likelihood* a particular eavesdropper output z^n could have over all codewords. Integrating with respect to z^n converts this likelihood into something that resembles a conditional "*probability*"; however, it is not a true probability. Taking the logarithm normalizes this between 0 and $\log|\mathscr{C}_n|$.

Max-information is concerned with the space of events $\mathscr{C}_n \times \mathscr{Z}^n$; the space that contains the 2-tuple realizations of codewords and eavesdropper outputs respectively. Working with this entire space is both difficult and unnecessary since we are really only concerned with those events that happen with sufficiently high probability; that is, the entire space clearly contains realizations that are extremely unlikely to happen. Therefore, we will end up restricting the entire space onto subspaces that contain the most likely realizations.

Definition. For $\epsilon \geq 0$, we call a subset $\mathcal{T} \subset \mathscr{C}_n \times \mathscr{Z}^n$ a $(1-\epsilon)$-*typical set* if

$$\mathbb{P}\left[(X^n, Z^n) \in \mathcal{T} \,|\, X^n = x^n\right] \geq 1 - \epsilon, \qquad \forall x^n \in \mathscr{C}_n.$$

Furthermore, we will denote the set of all $(1-\epsilon)$-typical sets by $\mathscr{T}_\epsilon$.

Typical sets intuitively contain almost all that there is to know about our space up to some ϵ. Using $(1-\epsilon)$-typical sets, we can define another max-information over this reduced space that will be crucial to our proofs later on.

Definition. Given $\epsilon \geq 0$ define ***ϵ-smooth max-information*** by

$$\boldsymbol{\mathcal{I}}_n^\epsilon = \inf_{\mathcal{T} \in \mathscr{T}_\epsilon} \boldsymbol{\mathcal{I}}_n \bigg|_{\mathcal{T}},$$

where max-information evaluated on $\mathcal{T}$ is given by

$$\boldsymbol{\mathcal{I}}_n \bigg|_{\mathcal{T}} = \log \left(\int\limits_{\mathscr{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n|x^n) \mu(dz^n) \right).$$

That is, given some threshold ϵ, we find the smallest value that max-information could possibly be when defined on the *subnormalized* channels corresponding to those sets that contain *enough* probability with respect to our threshold. Later, we will bound the leakage between the transmitter and eavesdropper as an increasing function of this metric; thus, defining ϵ-smooth max-information using the infinum provides the tightest bound we should expect when ϵ is our threshold.

In this paper we will only be concerned with ϵ as a function of n and will mainly be concerned with the cases for which $\epsilon \to 0$ as $n \to \infty$.

5.2. Leakage bounds

In this section, we will present one of the main original works of this thesis: we will bound the *mis* leakage for the scheme we provided in Chapter 4. But first, let us see that assuming the eavesdropper has the seed only hurts our measure of security.

Lemma 4.

$$\max_{P_M} I(M \wedge Z^n) \leq \max_{P_M} I(M \wedge Z^n, S).$$

Proof. Let M have some distribution P_M, then:

$$\begin{aligned}
I(M \wedge Z^n) &= \int\limits_{\mathcal{Z}^n} \sum_{m \in \mathcal{M}} \omega(z^n, m) \log \left(\frac{\omega(z^n, m)}{P_M(m)\omega(z^n)} \right) \mu(dz^n) \\
&= \int\limits_{\mathcal{Z}^n} \sum_{m \in \mathcal{M}} \left(\sum_{s \in \mathcal{S}} \omega(z^n, m, s) \right) \log \left(\frac{\sum\limits_{s' \in \mathcal{S}} \omega(z^n, m, s')}{\sum\limits_{s'' \in \mathcal{S}} P_M(m)\omega(z^n, s'')} \right) \mu(dz^n) \\
&\leq \int\limits_{\mathcal{Z}^n} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \omega(z^n, m, s) \log \left(\frac{\omega(z^n, m, s)}{P_M(m)\omega(z^n, s)} \right) \mu(dz^n) \\
&= I(M \wedge Z^n, S).
\end{aligned}$$

The second equality follows from the marginal property of densities whereas the inequality follows from the log-sum inequality. Maximizing over all probability distributions P_M implies the claim. ■

With this lemma, we see that if we can bound $\max_{P_M} I(M \wedge Z^n, S)$ then we can also bound $\max_{P_M} I(M \wedge Z^n)$. What this means is that giving the eavesdropper the seed (as we have assumed in Chapter 4) can never *decrease* the leakage. Fortunately, even with this assumption that the eavesdropper knows the seed perfectly, we can still bound the former term. With this, we are now ready for the following theorem; it is our main contribution.

Theorem 2 (Leftover Hash Lemma). Using the transmission procedure $\mathcal{W}$ outlined in Chapter 4, for any wiretap channel $W = (T, A)$ and $\epsilon \geq 0$ it follows that the mis-leakage of $\mathcal{W}$ over W is bounded as:

$$\mathbb{L}_n^{\text{mis}} = \max_{P_M} I(M \wedge Z^n) \leq \frac{1}{\ln 2} 2^{-b+\boldsymbol{\mathcal{I}}_n^{\epsilon}} + \epsilon k.$$

Proof. First note that by Lemma 4, we have $\max_{P_M} I(M \wedge Z^n) \leq \max_{P_M} I(M \wedge Z^n, S)$ so that it is sufficient to bound $\max_{P_M} I(M \wedge Z^n, S)$.

We will split the proof into two parts: $\epsilon > 0$ and $\epsilon = 0$; let us start with the $\epsilon = 0$ case. Here 1-typical sets $\mathcal{T}$ are equal to the entire space $\mathscr{C}_n \times \mathscr{Z}^n$ less a set of measure 0, so that $\boldsymbol{\mathcal{I}}_n^0 = \boldsymbol{\mathcal{I}}_n$. To show our claim is valid, it is therefore sufficient in the case $\epsilon = 0$ to show:

$$\max_{P_M} I(M \wedge Z^n, S) \leq \frac{1}{\ln 2} 2^{-b+\boldsymbol{\mathcal{I}}_n}.$$

To begin, suppose M has some arbitrary distribution. Since $\mathcal{S}$ and $\mathcal{M}$ are finite by assumption we will use the counting measure on their space. Thus, using the definition of conditional mutual information we have

$$I(M \wedge Z^n | S) = \int_{\mathscr{Z}^n} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \omega(m, z^n, s) \log \left(\frac{\omega(m, z^n | s)}{\omega(m|s)\omega(z^n|s)} \right) \mu(dz^n),$$

where μ is some measure on $\mathscr{Z}^n$.

From the *chain rule of mutual information*, since $M \perp S$ by assumption, we have $I(M \wedge Z^n, S) = I(M \wedge Z^n | S)$. It then follows that

$$\begin{aligned} I(M \wedge Z^n, S) &= I(M \wedge Z^n | S) \\ &= \int_{\mathscr{Z}^n} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \omega(m, z^n, s) \log \left(\frac{\omega(m, z^n | s)}{\omega(m|s)\omega(z^n|s)} \right) \mu(dz^n) \\ &= \int_{\mathscr{Z}^n} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \omega(m, z^n, s) \log \left(\frac{\omega(z^n | m, s)}{\omega(z^n|s)} \right) \mu(dz^n). \end{aligned}$$

Let us now expand each conditional density of the numerator and denominator of the logarithm in the previous term. Starting with the numerator we have:

$$\begin{aligned}\omega(z^n|m,s) &= \sum_{m'\in f_s^{-1}(m)} \omega(z^n|m')\omega(m'|m,s)\\ &= 2^{-b} \sum_{m'\in f_s^{-1}(m)} \omega(z^n|m')\\ &= 2^{-b} \sum_{m'\in \mathcal{M}'} \omega(z^n|m')\mathbb{1}\left(f_s(m') = m\right).\end{aligned}$$

The first equality follows from the fact that we can take M' as an intermediate node and sum over all possible realizations of M'; by assumption, since we are given m and s, then M' can only be found in $\phi_s(m)$ where ϕ_s is the even-inverse of f_s. The second equality follows since, once given m and s, the density of M' is uniform on a set with 2^b elements.

The expansion of the conditional density in the denominator is given as:

$$\begin{aligned}\omega(z^n|s) &= \frac{\omega(z^n,s)}{P_S(s)}\\ &\overset{\mathit{1}}{=} \sum_{m\in\mathcal{M}} \frac{\omega(z^n,m,s)P_M(m)}{P_S(s)P_M(m)}\\ &\overset{\mathit{2}}{=} \sum_{m\in\mathcal{M}} \omega(z^n|m,s)P_M(m)\\ &= 2^{-b} \sum_{m'\in\mathcal{M}'} \omega(z^n|m') \sum_{m\in\mathcal{M}} P_M(m)\mathbb{1}(f_s(m') = m)\\ &\overset{\mathit{3}}{=} 2^{-b} \sum_{m'\in\mathcal{M}'} \omega(z^n|m')P_M(f_s(m')).\end{aligned}$$

Justification.

1. *Marginal density property.*

2. *By assumption, $M \perp S$.*

3. *When s is fixed, f_s is a well defined function. Thus, inside the sum over $\mathcal{M}'$, $f_s(m')$ can map to only a single $m \in \mathcal{M}$. Therefore, the indicator is 1 only for a single value of m; namely, when $m = f_s(m')$.*

We now continue expanding the leakage using these two conditional densities.

$$
\begin{aligned}
&I(M \wedge Z^n, S) \\
&\overset{\mathit{4}}{=} \int_{\mathcal{Z}^n} \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \omega(m, z^n, s) \log\left(\frac{2^{-b}\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=m)}{2^{-b}\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n) \\
&\overset{\mathit{5}}{=} \int_{\mathcal{Z}^n} \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \omega(z^n|m, s)P_M(m)P_S(s) \log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=m)}{\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n) \\
&\overset{\mathit{6}}{=} \frac{1}{|\mathcal{S}|}\int_{\mathcal{Z}^n} \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \omega(z^n|m, s)P_M(m) \log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=m)}{\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n).
\end{aligned}
$$

Justification.

4. *We will break with our convention slightly. Here we have written $\omega(z^n|u')$ as shorthand for $\omega_{Z^n|M'}(z^n|u')$; analogously for $\omega(z^n|u'')$. We will stick with this new convention for the remainder of the proof; i.e. $\omega(\cdot|u^*)$ and $\omega(u^*)$ will be shorthand for densities with respect to M'.*

5. *By assumption, $M \perp S$.*

6. *By assumption, $S \sim \text{unif}(\mathcal{S})$.*

At this point we can expand the conditional density $\omega(z^n|m, s)$ as before and continue:

$$
\begin{aligned}
&\overset{\boldsymbol{I}}{=} \frac{2^{-b}}{|\mathcal{S}|}\int_{\mathcal{Z}^n} \sum_{\substack{m\in\mathcal{M}\\ s\in\mathcal{S}\\ m'\in\mathcal{M}'}} \omega(z^n|m')P_M(m)\mathbb{1}(f_s(m')=m) \log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=m)}{\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n) \\
&\overset{\mathit{7}}{=} \frac{2^{-b}}{|\mathcal{S}|}\int_{\mathcal{Z}^n} \sum_{\substack{m\in\mathcal{M}\\ s\in\mathcal{S}\\ m'\in\mathcal{M}'}} \omega(z^n|m')P_M(m)\mathbb{1}(f_s(m')=m) \log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=f_s(m'))}{\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n) \\
&\overset{\mathit{8}}{=} \frac{2^{-b}}{|\mathcal{S}|}\int_{\mathcal{Z}^n} \sum_{\substack{s\in\mathcal{S}\\ m'\in\mathcal{M}'}} \omega(z^n|m')P_M(f_s(m')) \log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega(z^n|u')\mathbb{1}(f_s(u')=f_s(m'))}{\sum\limits_{u''\in\mathcal{M}'}\omega(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n)
\end{aligned}
$$

$$\overset{9}{=} 2^{-b} \int\limits_{\mathcal{Z}^n} \left[\frac{1}{|\mathcal{S}|} \sum_{\substack{s \in \mathcal{S} \\ m' \in \mathcal{M}'}} \omega(z^n|m') P_M(f_s(m')) \log \left(\sum_{u' \in \mathcal{M}'} \omega(z^n|u') \mathbb{1}(f_s(u') = f_s(m')) \right) \right] + \cdots$$

$$+ \left[-\frac{1}{|\mathcal{S}|} \sum_{\substack{s' \in \mathcal{S} \\ m'' \in \mathcal{M}'}} \omega(z^n|m'') P_M(f_{s'}(m'')) \log \left(\sum_{u'' \in \mathcal{M}'} \omega(z^n|u'') P_M(f_{s'}(u'')) \right) \right] \mu(dz^n).$$

Justification.

7. *The entire summand is 0 unless $m = f_s(m')$, so we can replace the m in the indicator function of the log as such as long as we stick with the convention that $0 \log 0 = 0$ as the limit suggests.*

8. *As before, the indicator will filter all but a single m; namely, when $m = f_s(m')$.*

9. *We can break up the logarithm into a subtraction where we change indices of the summation so as not to become confused.*

We will now consider each of the square brackets separately, starting with the first. The first square bracket can be written (after multiplying by the unit $2^{-k}2^k$) as

$$2^{-k} \sum_{m' \in \mathcal{M}'} \omega(z^n|m') \sum_{s \in \mathcal{S}} \left(\frac{2^k}{|\mathcal{S}|} P_M(f_s(m')) \right) \log \left(\sum_{u' \in \mathcal{M}'} \omega(z^n|u') \mathbb{1}(f_s(u') = f_s(m') \right).$$

Since our preprocessor is a SSI-UHF it is a *uniform* hash family. Thus for any $m' \in \mathcal{M}'$ we have:

$$\sum_{s \in \mathcal{S}} \left(\frac{2^k}{|\mathcal{S}|} P_M(f_s(m')) \right) = \sum_{m \in \mathcal{M}} P_M(m) \frac{2^k}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{1} \left(f_s(m') = m \right) = 1;$$

thus we can aptly apply Jensen's inequality and move the preceding term inside of the logarithm at the expense of an inequality to get:

$$2^{-k} \sum_{m' \in \mathcal{M}'} \omega(z^n|m') \log \left(2^k \sum_{u' \in \mathcal{M}'} \omega(z^n|u') \sum_{s \in \mathcal{S}} \frac{P_M(f_s(m'))}{|\mathcal{S}|} \mathbb{1}(f_s(u') = f_s(m')) \right).$$

If $u' = m'$, it is clear that the indicator will always return 1 regardless of $s \in \mathcal{S}$ so that the argument of the logarithm becomes

$$\sum_{u' \in \mathcal{M}'} \omega(z^n|u') \mathbb{1}(u' = m') = \omega(z^n|m'),$$

where we have again used the fact that our preprocessor is a SSI-UHF and is hence uniform.

On the contrary, if $u' \neq m'$, the indicator will only return 1 some of the time, and a nice simplification of the expression is not obvious at this time.

Combining these cases together, the entire first square bracket is less than or equal to:

$$2^{-k} \sum_{m' \in \mathcal{M}'} \omega(z^n|m') \log \Bigg[\omega(z^n|m') + \dots \\ +2^k \sum_{u' \in \mathcal{M}'} \omega(z^n|u') \sum_{s \in \mathcal{S}} \frac{P_M(f_s(m'))}{|\mathcal{S}|} \mathbb{1}(f_s(u') = f_s(m')) \mathbb{1}(u' \neq m') \Bigg].$$

Let us now move onto the second square bracket. We can write this term as

$$- \frac{1}{|\mathcal{S}|} \sum_{s' \in \mathcal{S}} \left(\sum_{m'' \in \mathcal{M}'} \omega(z^n|m'') P_M(f_{s'}(m'')) \right) \log \left(\sum_{u'' \in \mathcal{M}'} \omega(z^n|u'') P_M(f_{s'}(u'')) \right)$$
$$\leq - \frac{1}{|\mathcal{S}|} \left(\sum_{s' \in \mathcal{S}} \sum_{m'' \in \mathcal{M}'} \omega(z^n|m'') P_M(f_{s'}(m'')) \right) \log \left(\frac{\sum_{s'' \in \mathcal{S}} \sum_{u'' \in \mathcal{M}'} \omega(z^n|u'') P_M(f_{s''}(u''))}{\sum_{s''' \in \mathcal{S}} 1} \right),$$

where the inequality follows from the log-sum inequality. Now again using the fact that our preprocessor is a SSI-UHF and hence uniform, the entire second square bracket becomes:

$$-2^{-k} \sum_{m'' \in \mathcal{M}'} \omega(z^n|m'') \log \left(2^{-k} \sum_{u'' \in \mathcal{M}'} \omega(z^n|u'') \right).$$

We are now at a point where each square bracket is properly simplified and we would like to recombine them with the main string of inequalities. But before we get started, denote the following sum by $\psi_{z^n,m'}$:

$$\frac{2^k 2^k}{\sum_{u'' \in \mathcal{M}'} \omega(z^n|u'')} \sum_{u' \in \mathcal{M}'} \omega(z^n|u') \sum_{m \in \mathcal{M}} P_M(m) \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{1}(m = f_s(m')) \mathbb{1}(f_s(m') = f_s(u')) \mathbb{1}(u' \neq m').$$

Then it follows,

$$\begin{aligned}
I(M \wedge Z^n, S) &\leq 2^{-b-k} \int\limits_{\mathcal{Z}^n} \sum_{m' \in \mathcal{M}'} \omega(z^n|m') \log \left(\frac{2^k \omega(z^n|m')}{\sum\limits_{u'' \in \mathcal{M}'} \omega(z^n|u'')} + \psi_{z^n,m'} \right) \mu(dz^n) \\
&\overset{10}{\leq} 2^{-b-k} \int\limits_{\mathcal{Z}^n} \sum_{m' \in \mathcal{M}'} \omega(z^n|m') \log \left(\frac{2^k \omega(z^n|m')}{\sum\limits_{u'' \in \mathcal{M}'} \omega(z^n|u'')} + 1 \right) \mu(dz^n) \\
&\overset{11}{\leq} \frac{2^{-b}}{\ln 2} \int\limits_{\mathcal{Z}^n} \frac{\sum_{m' \in \mathcal{M}'} \omega(z^n|m')^2}{\sum_{v \in \mathcal{M}'} \omega(z^n|v)} \mu(dz^n) \\
&\overset{12}{\leq} \frac{2^{-b}}{\ln 2} \int\limits_{\mathcal{Z}^n} \max_{m' \in \mathcal{M}'} \omega(z^n|m') \mu(dz^n) \\
&\overset{13}{=} \frac{2^{-b}}{\ln 2} \int\limits_{\mathcal{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n|x^n) \mu(dz^n) \\
&= \frac{1}{\ln 2} 2^{-b + \boldsymbol{\mathcal{I}}_n}.
\end{aligned}$$

Justification.

10. *Note that since our preprocessor is a SSI-UHF, it is hash value independent so that* $\psi_{z^n,m'} \leq 1$.
11. *Convert* log *to the natural logarithm so that we can use the inequality* $\ln(x+1) \leq x$ *for all* $x > 0$.
12. *The weighted sum, when choosing each weight to be equal to each element of data, is always smaller than the max of the data.*
13. *Although the code* $\mathcal{C}_n$ *(and hence codebook* $\mathscr{C}_n$*) may be initially generated at random, it is fixed and assumed to be known to all parties at the time of transmission. Thus there is a non-random bijective mapping* $\mathcal{M}' \to \mathscr{C}_n$ *and if* $m' \mapsto x^n$*, it follows that* $\omega(z^n|m') = \omega(z^n|x^n)$.

With this, we have constructed an upper bound to $I(M \wedge Z^n, S)$ for an arbitrary message distribution P_M. However, since the bound did not depend on the specific choice of P_M, the bound also holds for $\max_{P_M} I(M \wedge Z^n, S)$. Therefore, we have concluded the $\epsilon = 0$ case.

Let us move onto the $\epsilon > 0$ case. Fix some $\epsilon > 0$ and consider some $(1-\epsilon)$ typical set $\mathcal{T} \subset \mathscr{C}_n \times \mathcal{Z}^n$. To this typical set, we define an associative set $\mathcal{T}_* \subset \mathcal{M}' \times \mathcal{Z}^n$. In more detail, this set is needed for technical reasons and is defined in an obvious way as

$$\mathcal{T}_* = \{(m', z^n) \,|\, (e_n(m'), z^n) \in \mathcal{T}\},$$

where e_n is the encoding function of the fixed code $\mathcal{C}_n$ (with codebook $\mathscr{C}_n$).

Now consider step I. in the previous string of inequalities written as:

$$\frac{2^{-b}}{|\mathcal{S}|} \int\limits_{\mathcal{Z}^n} \sum_{\substack{m \in \mathcal{M} \\ s \in \mathcal{S}}} P_M(m) \left[\sum_{m' \in \mathcal{M}'} \omega(z^n|m') \mathbb{1}(f_s(m') = m) \log \left(\frac{\sum\limits_{u' \in \mathcal{M}'} \omega(z^n|u') \mathbb{1}(f_s(u') = m)}{\sum\limits_{u'' \in \mathcal{M}'} \omega(z^n|u'') P_M(f_s(u''))} \right) \right] \mu(dz^n).$$

Inside of the square bracket, z^n and m can be considered *fixed*, and thus, each of the 3 sums over $\mathcal{M}'$ can be considered as a sum over two other sets:

$$\mathcal{M}'_1 = \{m' \in \mathcal{M}' : (m', z^n) \in \mathcal{T}_*\} \text{ and}$$

$$\mathcal{M}'_2 = \{m' \in \mathcal{M}' : (m', z^n) \in \mathcal{T}_*^{\complement}\},$$

where $\mathcal{T}_*^{\complement}$ denotes the complement of $\mathcal{T}_*$ in $\mathcal{M}' \times \mathcal{Z}^n$.

With this, we can then apply the log-sum inequality to the above term to yield the following at the expense of an inequality:

$$\begin{aligned} &\frac{2^{-b}}{|\mathcal{S}|} \int\limits_{\mathcal{Z}^n} \sum_{\substack{m \in \mathcal{M} \\ s \in \mathcal{S}}} P_M(m) \cdots \\ &\quad \cdots \left[\left(\sum_{m' \in \mathcal{M}'_1} \omega(z^n|m') \mathbb{1}(f_s(m') = m) \right) \log \left(\frac{\sum\limits_{u' \in \mathcal{M}'_1} \omega(z^n|u') \mathbb{1}(f_s(u') = m)}{\sum\limits_{u'' \in \mathcal{M}'_1} \omega(z^n|u'') P_M(f_s(u''))} \right) + \cdots \right. \\ &\quad \left. \cdots + \left(\sum_{m' \in \mathcal{M}'_2} \omega(z^n|m') \mathbb{1}(f_s(m') = m) \right) \log \left(\frac{\sum\limits_{u' \in \mathcal{M}'_2} \omega(z^n|u') \mathbb{1}(f_s(u') = m)}{\sum\limits_{u'' \in \mathcal{M}'_2} \omega(z^n|u'') P_M(f_s(u''))} \right) \right] \mu(dz^n). \end{aligned}$$

Now define $\mathcal{Q}_{\mathcal{T}_*}$ by:

$$\frac{2^{-b}}{|\mathcal{S}|} \int\limits_{\mathcal{Z}^n} \sum_{\substack{m \in \mathcal{M} \\ s \in \mathcal{S}}} P_M(m) \sum_{m' \in \mathcal{M}'} \omega_{\mathcal{T}_*}(z^n|m') \mathbb{1}(f_s(m') = m) \log \left(\frac{\sum\limits_{u' \in \mathcal{M}'} \omega_{\mathcal{T}_*}(z^n|u') \mathbb{1}(f_s(u') = m)}{\sum\limits_{u'' \in \mathcal{M}'} \omega_{\mathcal{T}_*}(z^n|u'') P_M(f_s(u''))} \right) \mu(dz^n),$$

so that the leakage becomes

$$I(M \wedge Z^n, S) \leq \mathcal{Q}_{\mathcal{T}_*} + \mathcal{Q}_{\mathcal{T}_*^{\complement}}.$$

When considering just $\mathcal{Q}_{\mathcal{T}_*}$ we can continue where we left off on line I. of the previous proof ($\epsilon = 0$ case). In fact, it is not hard to see that almost nothing changes and we end up with:

$$\begin{aligned}\mathcal{Q}_{\mathcal{T}_*} &\leq \frac{2^{-b}}{\ln 2}\int_{\mathcal{Z}^n} \max_{m'\in\mathcal{M}'} \omega_{\mathcal{T}_*}(z^n|m')\mu(dz^n)\\ &= \frac{2^{-b}}{\ln 2}\int_{\mathcal{Z}^n} \max_{x^n\in\mathscr{C}_n} \omega_{\mathcal{T}}(z^n|x^n)\mu(dz^n).\end{aligned}$$

Now let's focus on $\mathcal{Q}_{\mathcal{T}_*^{\complement}}$. It follows that:

$$\begin{aligned}
&\mathcal{Q}_{\mathcal{T}_*^{\complement}}\\
&\overset{14}{\leq} \frac{2^{-b}}{|\mathcal{S}|}\int_{\mathcal{Z}^n}\sum_{s\in\mathcal{S}}\left(\sum_{m'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|m')P_M(f_s(m'))\right)\log\left(\frac{\sum_{u'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u')}{\sum\limits_{u''\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u'')P_M(f_s(u''))}\right)\mu(dz^n)\\
&\overset{15}{\leq} \frac{2^{-b}}{|\mathcal{S}|}\int_{\mathcal{Z}^n}\left(\sum_{s\in\mathcal{S}}\sum_{m'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|m')P_M(f_s(m'))\right)\log\left(\frac{|\mathcal{S}|\sum_{u'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u')}{\sum\limits_{s'\in\mathcal{S}}\sum\limits_{u''\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u'')P_M(f_{s'}(u''))}\right)\mu(dz^n)\\
&\overset{16}{=} 2^{-(b+k)}\int_{\mathcal{Z}^n}\sum_{m'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|m')\log\left(\frac{2^k\sum_{u'\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u')}{\sum\limits_{u''\in\mathcal{M}'}\omega_{\mathcal{T}_*^{\complement}}(z^n|u'')}\right)\mu(dz^n)\\
&= k2^{-(b+k)}\sum_{m'\in\mathcal{M}'}\int_{\mathcal{Z}^n}\omega(z^n|m')\mathbb{1}\Big((m',z^n)\in\mathcal{T}_*^{\complement}\Big)\mu(dz^n)\\
&\overset{17}{=} k2^{-(b+k)}\sum_{x^n\in\mathscr{C}_n}\int_{\mathcal{Z}^n}\omega(z^n|x^n)\mathbb{1}\Big((x^n,z^n)\in\mathcal{T}^{\complement}\Big)\mu(dz^n)\\
&= k2^{-(b+k)}\sum_{x^n\in\mathscr{C}_n}\mathbb{P}\left[(X^n,Z^n)\in\mathcal{T}^{\complement}\,|\,X^n=x^n\right]\\
&\overset{18}{\leq} k\epsilon
\end{aligned}$$

Justification.

14. In the numerator of the logarithm, we have used $\mathbb{1}(f_s(u') = m) \leq 1$ *for all* s, u', m.

15. Log-sum inequality.

16. *Our preprocessor is a SSI-UHF and hence it is uniform.*

17. *As argued in the previous block, there is a bijective mapping $\mathcal{M}' \to \mathscr{C}_n$.*

18. *We chose $\mathcal{T}$ to be a $(1-\epsilon)$ typical set and there are at most $2^{(b+k)}$ pseudo messages (and hence codewords) because our UHF is b-regular.*

Again, just as in the $\epsilon = 0$ case, we have provided an upper bound to $I(M \wedge Z^n, S)$ for an arbitrary message distribution P_M so that the upper bound also holds for $\max_{P_M} I(M \wedge Z^n, S)$. This concludes the $\epsilon > 0$ case.

Combining both cases, we have for any $\epsilon \geq 0$:

$$\max_{P_M} I(M \wedge Z^n, S) \leq \frac{2^{-b}}{\ln 2} \int\limits_{\mathcal{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n|x^n)\mu(dz^n) + \epsilon k.$$

Since this inequality was derived using an *arbitrary* $(1-\epsilon)$-typical set $\mathcal{T}$, we may as well optimize our choice of $\mathcal{T}$ while keeping ϵ fixed so as to obtain the *tightest* possible bound. With this and Lemma 4 we have proven the claim of the theorem. ■

We see that if $\mathcal{I}_n^\epsilon < b$ and $\epsilon = 0$ or $\epsilon \to 0$ as $n \to \infty$ then Theorem 2 implies our wiretap scheme of Chapter 4 is mis-secure and hence semantically secure. In particular, if $\epsilon = 0$ or ϵ goes to 0 exponentially fast with n, then our wiretap scheme is exceptional so long as our ECC $\mathcal{C}$ is exceptional. With this, we have proven that our scheme is a wiretap scheme that is semantically secure so long as we can prove that the max-information is bounded by b. We note that surprisingly, our bound is *exactly* the bound put forth by [18] (their leakage bound only used the *mis-r* metric).

Remark. Here ϵ is a facilitator of mathematical (and numerical) ease. We could pick $\epsilon = 0$ and attempt to calculate $\mathcal{I}_n$ outright, however this turns out to be significantly more arduous than calculating $\mathcal{I}_n^\epsilon$.

6. ACHIEVABLE RATES OF OUR WIRETAP SCHEME

In this chapter we will be concerned with characterizing the achievable rates of our wiretap scheme. We will show that the leakage bounds of the previous chapter immediately give a characterization of the achievable rates. Moreover, at the end of this section, we will remove the public seed assumption and actually transmit the seed over the wiretap channel.

6.1. Achievable Rates

Theorem 2 is very useful in its own right, however, we can manipulate it just a bit further to really flesh out some utility in the wiretap domain where we are most concerned with rates. The following corollary follows immediately.

Corollary 1. Using the transmission scheme of Chapter 4, let $k = nR_n$ and $l = nR_{\mathcal{C}_n}$ where R_n and $R_{\mathcal{C}_n}$ are the overall security rate and error correction coding rate respectively. It follows for any wiretap channel and $\epsilon \geq 0$ that the *mis* leakage is bound as:

$$\mathbb{L}_n^{\text{mis}} = \max_{P_M} I(M \wedge Z^n) \leq \frac{1}{\ln 2} 2^{-n\left(R_{\mathcal{C}_n} - R_n - \frac{\mathcal{I}_n^\epsilon}{n}\right)} + \epsilon n R_n.$$

This corollary basically says we can convert the language of Theorem 2 to that of rates. With this we come to a significant proposition.

Proposition 2. Suppose we are using the transmission scheme $\mathcal{W}$ of Chapter 4 with a sequence of ECC's $\mathcal{C} = \{\mathcal{C}_n\}$ each with rate $R_{\mathcal{C}_n}$ and asymptotic rate $R_{\mathcal{C}} \leq C_T$ (the point-to-point capacity of the main channel). Let R_s denote the asymptotic overall security rate. Then for any wiretap channel $W = (T, A)$ if ϵ is chosen so that $\epsilon n \to 0$ as $n \to \infty$ we have the following.

(1) We can achieve all secure asymptotic rates

$$R_s < \left(R_{\mathcal{C}} - \lim_{n\to\infty} \frac{\mathcal{I}_n^\epsilon}{n}\right)^+ \quad \text{or} \quad R_s = 0$$

with *semantic* security.

(2) If $\lim\limits_{n\to\infty} \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n} \leq \xi$ then we can achieve all secure rates

$$R_s < (R_{\mathcal{C}} - \xi)^+ \quad \text{or} \quad R_s = 0$$

with *semantic* security.

(3) If ϵ is *exponentially* diminishing to 0 with n, then for any secure rates as in (1) and (2), then the scheme $\mathcal{W}$ is exceptionally semantically secure. In particular if $\mathcal{C}$ is exceptionally reliable, then $\mathcal{W}$ is an exceptionally semantically secure *wiretap* scheme.

Proof.

1. Consider Corollary 1. We need $\mathbb{L}_n^{\text{mis}} \to 0$ as $n \to \infty$ to get *mis*-security (and thus semantic). We have as $n \to \infty$ that $R_n \to R_s$ and $R_{\mathcal{C}_n} \to R_{\mathcal{C}}$. Since R_s is bounded then $\lim_{n\to\infty} \epsilon n R_n = 0$ by assumption that $\epsilon n \to 0$ as $n \to \infty$. Now if $\lim_{n\to\infty}(R_{\mathcal{C}_n} - R_n - \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}) > 0$ then the first term in the sum on the right hand side of Corollary 1 will also go to 0. But this is equivalent to
$$R_s < R_{\mathcal{C}} - \lim_{n\to\infty} \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}.$$
If the right hand side is negative however, we will instead choose $R_s = 0$ since rates cannot be negative.

2. Consider Corollary 1 again. Since $\lim_{n\to\infty} \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n} \leq \xi$, we can bound the asymptotic leakage as
$$\begin{aligned}\lim_{n\to\infty} \mathbb{L}_n^{\text{mis}} &= \lim_{n\to\infty} \left(\frac{1}{\ln 2} 2^{-n(R_{\mathcal{C}_n} - R_n)} 2^{n \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}} + \epsilon n R_n \right) \\ &= \frac{1}{\ln 2} 2^{\lim\limits_{n\to\infty}(-n(R_{\mathcal{C}_n} - R_n))} 2^{\lim\limits_{n\to\infty}(n) \cdot \lim\limits_{n\to\infty} \frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}} + \lim_{n\to\infty} \epsilon n R_n \\ &\leq \frac{1}{\ln 2} 2^{\lim\limits_{n\to\infty}(-n(R_{\mathcal{C}_n} - R_n))} 2^{\lim\limits_{n\to\infty}(n)\cdot\xi} + \lim_{n\to\infty} \epsilon n R_n \\ &= \frac{1}{\ln 2} 2^{\lim\limits_{n\to\infty}(-n(R_{\mathcal{C}_n} - R_n - \xi))} + \lim_{n\to\infty} \epsilon n.\end{aligned}$$
At this point we can continue exactly as in (1).

3. Clearly the first of the two summands on the right hand side of the conclusion of Corollary 1 is exponentially decreasing when R_s satisfies the rates given in either (1) or (2) above. Thus, if

$\epsilon n R_n$ is exponentially decreasing with n, the mis leakage $\mathbb{L}_n^{\text{mis}}$ is exponentially decreasing to 0; i.e. $\mathcal{W}$ is mis-exceptionally secure. Since mis is equivalent to semantic security asymptotically we know the semantic leakage must be decreasing *at least* as fast as the mis-leakage. Thus, if $\epsilon n R_n$ is exponentially decreasing with n then $\mathcal{W}$ is exceptionally semantically secure as well.

Now as mentioned in (1), $\lim_{n\to\infty} R_n = R_s$ is bounded. Thus, if ϵ is decreasing to 0 with n then so is $\epsilon n R_n$.

■

Remark. The notation $(\cdot)^+$ *means we only consider that term if it is strictly positive.*

This proposition gives us an interesting consideration of what semantically secure rates are achievable using our wiretap scheme. Proposition 2.1 is the best consideration but admittedly arduous. It is of future interest to calculate the rate of max-information $\frac{\mathcal{I}_n^\epsilon}{n}$ for specific channels especially for finite n. This will give us a bound on the *finite leakage*, a very interesting line of future research.

Proposition 2.2 is a much easier consideration as instead of calculating the rate of max-information *exactly* we only need a proper estimate. This part of the proposition says that given a reliable ECC scheme $\mathcal{C}$ of rate $R_{\mathcal{C}}$ for use on channel T, we can use our procedure given in Chapter 4 to *convert* this ECC to a *semantically secure wiretap code* for $W = (T, A)$ of rate $R_{\mathcal{C}} - \xi$ where ξ is *only* dependent on the eavesdroppers channel A. Since we provided a polynomially time efficient implementation of our procedure in Chapter 4, this conversion is also done in polynomial time.

Now that we have a good characterization of achievable rates. Let us consider some special cases that will be relevant in the next chapter.

Corollary 2. As a special case of Proposition 2.2, if the upper bound $\xi = C_A$ and we pick a reliable ECC scheme $\mathcal{C}$ with rate C_T then we can achieve all secure rates

$$R_s < (C_T - C_A)^+ \quad \text{or} \quad R_s = 0$$

with *semantic* security.

This characterization is useful because for many wiretap channels, the secrecy capacity is given by $C_s = C_T - C_A$. In other words, Corollary 2 implies that on those channels, if we can prove that $\xi = C_A$ and we choose $R_{\mathcal{C}} = C_T$, then we can achieve the secrecy capacity.

6.2. Removing the Public Seed Assumption

We have just seen in the preceding section that our scheme can provide semantic security for certain achieveable rates (provided that we prove a bound on the max-information rate), however, we have assumed hitherto that the seed S was publicly available to all parties. This is in *strict* violation of assumptions on a wiretap channel; that is, *all* communication must take place over the wiretap channel. In this section, we remove this assumption and transmit the seed over the wiretap channel. We will show asymptotically that no rate, security, or reliablity is lost.

6.2.1. A First Attempt

As a first attempt to resolve this violation, suppose the seed is transmitted before beginning transmission of an actual message. This is a problem, however, because it leads to *information rate loss* as follows. Suppose the seed can be transmitted with a probability of error less than some $p'_{e,n}$ to the intended receiver in nc channel uses for some constant $c > 1$. Then the transmitter sends k message bits of information in another n channel uses. Overall, k bits of information were transferred in $n + nc = n(1 + c)$ channel uses, thus our overall secure information rate in this case is given asymptotically by

$$\lim_{n \to \infty} \frac{k}{n(1+c)} = \frac{1}{1+c} R_s < R_s,$$

where R_s is the previous secure achievable rate assuming the seed was public. In other words, the possible asymptotic rates now achievable when sending the seed before message transmission is *stricly* less than before. Therefore the rates achieved using Proposition 2 are *not* possible to achieve anymore.

6.2.2. A resolution

As a better attempt to resolve this problem, suppose we use the same seed to send η messages $M_1, M_2, \ldots, M_\eta$ using η *independent* instances of the wiretap channel. First we will pick a blocklength n and on the first instance of the wiretap channel, we will send the seed over in nc channel uses, where $c > 1$ is chosen so that the seed's probability of error at the intended receiver is less than or equal to $p'_{e,n}$. Pessimistically (from the point of view at the transmitter), we will assume that the eavesdropper always receives a perfect copy of the seed. Now on each of the η independent channel instances, we will send a corresponding message using the same scheme as outlined in chapter 4 except using the *same seed* for each instance as that was sent across the wiretap channel initially. Let $\mathbf{M} = (M_1, M_2, \ldots, M_\eta)$ be the η length message. Let $\mathbf{Z} = (Z^n(1), Z^n(2), \ldots, Z^n(\eta))$ where $Z^n(i)$ is the n-letter eavesdropper output corresponding to both message and channel instance i.

Now consider the rate of this scheme. In each of the η channel uses we are sending k bits of information. Morover, we will end up using the channel ηn times for the message and nc times for the seed. Overall, the asymptotic secure rate of this new procedure is given by

$$\lim_{n\to\infty} \frac{\eta k}{\eta n + cn} = \lim_{n\to\infty} \frac{k}{n(1 + c/\eta)} = \frac{R_s}{\lim_{n\to\infty}(1 + c/\eta)},$$

where R_s is again the previous asymptotic secure acheivable rate when the seed was public. Since c is a constant, the only way to avoid information rate loss asymptotically is if $\eta \to \infty$ as $n \to \infty$.

Now consider the reliability of this new procedure. If each message has probability of error (at the intended receiver) bounded by $p_{e,n}$, then $\mathbf{M}$ has probability of error (at the intended receiver) bounded by $\eta \cdot p_{e,n}$. Thus to avoid information rate loss *and* transmit reliably, we need $\eta \to \infty$ and $\eta \cdot p_{e,n} \to 0$ as $n \to \infty$.

Now lets consider the leakage.

Lemma 5. For some $i \in \{1, \ldots, \eta\}$ the following holds:

$$\max_{P_{\mathbf{M}}} I(\mathbf{M} \wedge \mathbf{Z}) \leq \eta \cdot \max_{P_M} I(M_i \wedge Z^n(i)|S).$$

Proof. Let $\mathbf{M}$ have an arbitrary distribution $P_{\mathbf{M}}$. We see that the proof of Lemma 4 still applies and we have:

$$I(\mathbf{M} \wedge \mathbf{Z}) \leq I(\mathbf{M} \wedge \mathbf{Z}, S).$$

Since $S \perp M_i$ for each i, then $S \perp \mathbf{M}$. Then by the chain rule of mutual information we also have:

$$I(\mathbf{M} \wedge \mathbf{Z}) \leq I(\mathbf{M} \wedge \mathbf{Z}|S).$$

Now $(M_1, Z^n(1)), \ldots, (M_\eta, Z^n(\eta))$ are mutually independent once we are given S, thus we can use Lemma 3 to get

$$I(\mathbf{M} \wedge \mathbf{Z}) \leq \sum_{i=1}^{\eta} I(M_i \wedge Z^n(i)|S).$$

Now we want to maximize $I(\mathbf{M} \wedge \mathbf{Z})$ over all probability distributions $P_{\mathbf{M}}$. However that is equivalent to maximizing over each choice of P_{M_i} individually. The above becomes:

$$\max_{P_{\mathbf{M}}} I(\mathbf{M} \wedge \mathbf{Z}) \leq \sum_{i=1}^{\eta} \max_{P_{M_i}} I(M_i \wedge Z^n(i)|S).$$

Here i represents an instance of the wiretap channel. Choose the channel instance j that corresponds to the most leakage $\max\limits_{P_{M_j}} I(M_j \wedge Z^n(j)|S)$ leaked to the eavesdropper. The above then becomes

$$\max_{P_{\mathbf{M}}} I(\mathbf{M} \wedge \mathbf{Z}) \leq \eta \max_{P_{M_j}} I(M_j \wedge Z^n(j)|S).$$

■

This lemma intuitively says that the message leakage of all η wiretap channel instances is no more than the number of channel instances multiplied by the leakage over the "worst case" wiretap channel (worst here is with respect to the transmitter). Combining this result with the proof of Theorem 2 and the result of Corollary 1 gives the following corollary.

Corollary 3. Let i be the wiretap channel instance where the transmitter leaks the most information to the eavesdroper. Let $R_{\mathcal{C}_n}$ be the rate of the ECC and R_n the secure rate of transmission for that wiretap channel instance. It follows that

$$\max_{P_{\mathbf{M}}} I(\mathbf{M} \wedge \mathbf{Z}) \leq \frac{\eta}{\ln 2} 2^{-n\left(R_{\mathcal{C}_n} - R_n - \frac{\mathcal{I}_n^\epsilon}{n}\right)} + \epsilon \eta n R_n.$$

With this, just as in the last section, we see that if $R_n < R_{\mathcal{C}_n} - \frac{\mathcal{I}_n^\epsilon}{n}$ for each n, then so long as η grows with n strictly slower than exponential, the first term will go to 0. Furthermore, η must be chosen slow enough so that $\epsilon \eta n \to 0$ as $n \to \infty$.

In summary, with regards to how η must grow with n we need $\eta \to \infty$ as $n \to \infty$ but it must do so sufficiently slow. It suffices to pick something along the lines of $\eta = \mathcal{O}(\log(n))$ (if ϵ and

$p_{e,n}$ are chosen sufficently fast). In other words, as long as we keep on adding new independent messages when increasing the block length, we can still achieve the same rate, reliablility, and security asymptotically as before when we assumed the seed to be public. This trick was extorted in [1] and was called therein *seed recycling*. Our result follows very similar to that of [18].

7. APPLICATION

In this chapter, we show that our wiretap scheme of Chapter 4 can achieve the secrecy capacity of the AWGN wiretap channel. With this we also demonstrate the utility of ϵ-smooth max information. Indeed we will use the prescription put forth by Corollary 2 and bound the max-information.

7.1. Achieving Semantic security on AWGN wiretap channels

We consider now the additive white Gaussian noise wiretap channel (AWGN). The AWGN channel is arguably the most popular continuous alphabet channel model due to its simplicity. In this model, the output signal is a layering of the input signal with additive white Gaussian noise, or rather noise with spectral power at all frequencies. We represent the input signal by the random variable X and the additive white Gaussian noise by U. As is usual in this case, we will assume the AWGN wiretap channel to be memoryless.

The output at the intended receiver is represented by Y over $\mathcal{Y}$ and the output at the eavesdropper is represented by Z over $\mathcal{Z}$. We will suppose $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ and that the channels T and A can be described by their outputs given respectively as

$$\begin{aligned} Y &= X + U_T \quad \text{and} \\ Z &= X + U_A. \end{aligned}$$

The random variables U_T and U_A are assumed mutually independent and sampled independent and identically (iid) according to $\mathcal{N}(0, \sigma_T^2)$ and $\mathcal{N}(0, \sigma_A^2)$ respectively. As is usual in the continuous regime, to avoid the case of infinite channel capacity we assume there is some average power constraint P on the input. A figure of this setup is given below in Figure 7.1.

Fact 5. The capacity of an AWGN channel with average input power constraint P and additive noise variance σ^2 is given by

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2} \right).$$

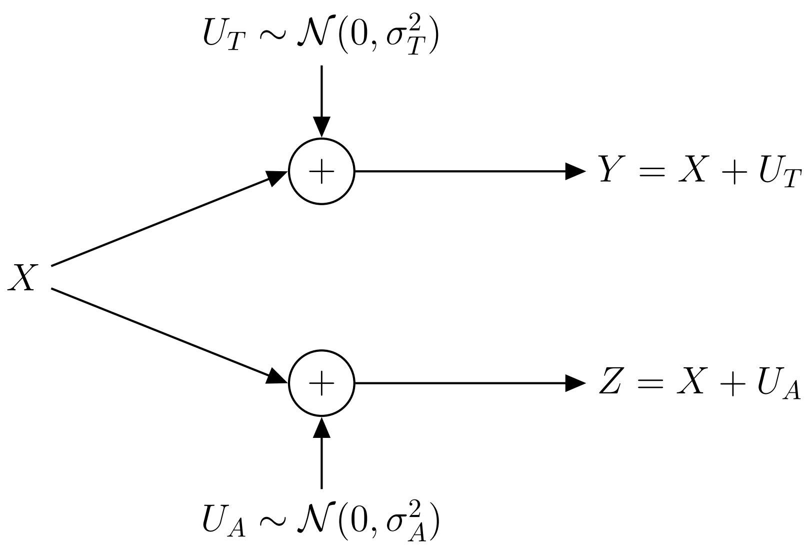
$U_T \sim \mathcal{N}(0, \sigma_T^2)$
$Y = X + U_T$
X
$Z = X + U_A$
$U_A \sim \mathcal{N}(0, \sigma_A^2)$

Figure 7.1. AWGN Channel model.

In particular, this means the capacity of the intended receivers point to point channel is given by $C_T = \frac{1}{2}\log(1 + \frac{P}{\sigma_T^2})$ and the capacity of the eavesdroppers point to point channel is given by $C_A = \frac{1}{2}\log(1 + \frac{P}{\sigma_A^2})$. Moreover, the secrecy capacity for AWGN wiretap channels is given by a simple subtraction of these terms.

Fact 6. [9] On an AWGN wiretap channel $W = (T, A)$, the (weak) secrecy capacity is given as:

$$C_s = \begin{cases} C_T - C_A, & \text{if } \sigma_T^2 < \sigma_A^2 \\ 0, & \text{Otherwise.} \end{cases}$$

In other words, so long as the noise variance of the main channel is strictly less than the noise variance of the eavesdroppers channel, secure communication is possible.

7.1.1. Max-info bounds for Gaussian

Our main goal of this section is to show that our scheme given in Chapter 4 can achieve the secrecy capacity of the AWGN wiretap channel under semantic security. Indeed we already have a prescription of how to do this considering the secrecy capacity in this case is written as the difference of point to point channel capacities. Namely, we can use Corollary 2 if we can bound the asymptotic rate of max-information by C_A. To this end, we devote this subsection.

We will be considering the measure μ for $\mathcal{Z}^n$ as the Lebesgue measure on $\mathbb{R}^n$ so that in particular, max-information is given by the standard Lebesgue integral over n-dimensional space:

$$\mathcal{I}_n^\epsilon = \log \int_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n|x^n)dz^n.$$

The next lemma can be found nearly exactly in [18] and we include this here simply for the sake of completeness and demonstrating the utility of ϵ smooth max information. Note that there, the authors used this lemma to show their UHF based scheme was *mis-r* secure whereas here, we are using this lemma to show our altered UHF scheme is semantically secure.

Lemma 6. Using any ECC $\mathcal{C}$, the max-information of an AWGN eavesdropper channel A is asymptotically bound as

$$\lim_{n\to\infty} \frac{\mathcal{I}_n^\epsilon}{n} \leq C_A.$$

Proof. Fix $\delta > 0$ small. Define a set $\mathcal{P}_{\text{out}} = \{z^n \in \mathbb{R}^n \,|\, ||z^n||^2 \leq n(P + \sigma_A^2)(1+\delta)\}$. Also for each $x^n \in \mathscr{C}_n$ define a set $\mathcal{P}_{\text{noise}}^{x^n} = \{z^n \,|\, ||z^n - x^n||^2 \geq n\sigma_A^2(1-\delta)\}$.

Now let $\mathcal{T}_{\text{out}}, \mathcal{T}_{\text{noise}} \subset \mathscr{C}_n \times \mathbb{R}^n$ be sets defined as $\mathcal{T}_{\text{out}} = \mathscr{C}_n \times \mathcal{P}_{\text{out}}$ and $\mathcal{T}_{\text{noise}} = \{(x^n, z^n) \,|\, z^n \in \mathcal{P}_{\text{noise}}^{x^n}$ for each $x^n \in \mathscr{C}_n\}$. Lastly define a set $\mathcal{T} = \mathcal{T}_{\text{out}} \cap \mathcal{T}_{\text{noise}}$.

It was shown in [18] that $\mathcal{T}$ is a $(1-\epsilon_n)$-typical set where $\epsilon_n = e^{-n\delta^2/8}$. Hence $\epsilon \to 0$ exponentially fast with n. With this we have the following.

$$\begin{aligned}
2^{\mathcal{I}_n^\epsilon} &\overset{\mathit{1}}{\leq} 2^{\left(\mathcal{I}_n\big|_{\mathcal{T}}\right)} \\
&= \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n|x^n)\mathbb{1}((x^n, z^n) \in \mathcal{T})dz^n \\
&\overset{\mathit{2}}{=} \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \left[\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{(z_i - x_i)^2}{2\sigma_A^2}\right)\right) \mathbb{1}((x^n, z^n) \in \mathcal{T})\right] dz^n \\
&= \frac{1}{(2\pi\sigma_A^2)^{\frac{n}{2}}} \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \left[\exp\left(-\frac{\|z^n - x^n\|^2}{2\sigma_A^2}\right) \mathbb{1}((x^n, z^n) \in \mathcal{T})\right] dz^n \\
&\overset{\mathit{3}}{\leq} \frac{\exp\left(-\frac{n}{2}(1-\delta)\right)}{(2\pi\sigma_A^2)^{\frac{n}{2}}} \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \mathbb{1}((x^n, z^n) \in \mathcal{T})dz^n \\
&\overset{\mathit{4}}{\leq} \frac{\exp\left(-\frac{n}{2}(1-\delta)\right)}{(2\pi\sigma_A^2)^{\frac{n}{2}}} \int\limits_{\mathcal{P}_{\text{out}}} dz^n
\end{aligned}$$

$$\begin{aligned}&\overset{5}{=} \frac{\exp\left(-\frac{n}{2}(1-\delta)\right)}{(2\pi\sigma_A^2)^{\frac{n}{2}}}\mathrm{Vol}(\mathcal{P}_{\mathrm{out}})\\&\overset{6}{=} \frac{\exp\left(-\frac{n}{2}(1-\delta)\right)}{(2\pi\sigma_A^2)^{\frac{n}{2}}}\frac{(\pi n(P+\sigma_A^2)(1+\delta))^{\frac{n}{2}}}{\Gamma(n/2+1)}.\end{aligned}$$

Justification.

1. *$\mathcal{T}$ is a $(1-\epsilon)$ typical set; however, it may not be the set corresponding to the "smallest" ϵ smooth max-information.*

2. *Each output is a normal random variable. Since we assume the channel is memoryless, we can split this density simply into a product.*

3. *We are working on $\mathcal{T}$ in the integral and thus $\mathcal{P}_{noise}$. Thus, $\|z^n - x^n\|^2 \geq n\sigma_A^2(1-\delta)$.*

4. *The indicator function returns either 0 or 1 in the area of interest $\mathcal{P}_{out}$ and 0 elsewhere. Thus, we can simply upper bound the indicator by 1 everywhere inside of $\mathcal{P}_{out}$.*

5. *The preceding integral indeed represented the volume of $\mathcal{P}_{out}$ where we note that $\mathcal{P}_{out}$ is a ball in real n space.*

6. *The volume of an n ball of radius r is given by*

$$\frac{\pi^{n/2}}{\Gamma(n/2+1)}r^n,$$

where here Γ is the gamma function (generalized factorial) from analysis.

Taking the logarithm of both sides of the preceding and dividing by n yields:

$$\begin{aligned}\frac{1}{n}\boldsymbol{\mathcal{I}}_n^\epsilon &\leq \frac{1}{n}\log\left(\frac{\exp\left(-(1-\delta)\right)}{2}\frac{(n(1+P/\sigma_A^2)(1+\delta))}{\Gamma(n/2+1)^{2/n}}\right)^{n/2}\\&= \frac{1}{2}\left(\log\left(1+\frac{P}{\sigma_A^2}\right)+\log\left((1+\delta)e^{\delta}\right)+\log\left(\frac{1}{2e}\cdot\frac{n}{\Gamma(n/2+1)^{2/n}}\right)\right)\\&= C_A + \frac{1}{2}\log\left((1+\delta)e^{\delta}\right)+\frac{1}{2}\log\left(\frac{1}{2e}\cdot\frac{n}{\Gamma(n/2+1)^{2/n}}\right).\end{aligned}$$

Fortunately, $\frac{n}{\Gamma(n/2+1)^{2/n}} \to 2e$ as $n \to \infty$. Moreover, our choice of δ is not restricted and can be made arbitrarily small. Thus, $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{\mathcal{I}}_n^\epsilon \leq C_A$. This completes the proof. ■

7.1.2. Implications

As stated before, using Corollary 2 we have:

Proposition 3. On an AWGN wiretap channel with power constraint P, if $\mathcal{C}$ is a reliable ECC scheme such that the rate of the scheme $R_{\mathcal{C}} = C_T$ then we can achieve the secrecy capacity using the scheme given in Chapter 4 under the semantic security metric.

This proposition combined with Fact 3 immediately give us something stronger.

Corollary 4. On the AWGN wiretap channel the semantic secrecy capacity is *equivalent* to the weak secrecy capacity.

With this we have officially shown that our wiretap coding scheme can achieve the secrecy capacity of AWGN wiretap channels so long as we can find a reliable ECC $\mathcal{C}$ that can achieve the main point to point channel capacity C_T[1]. Also, since ϵ is exponentially going to 0 with n in the proof of Lemma 6, if $\mathcal{C}$ is also chosen to be *exceptionally* reliable, then our entire wiretap coding scheme is *exceptional*.

Indeed an ECC scheme is given in [19] that is concrete, reliable, and has quadratic time complexity with respect to block length n in both encoding and decoding. Moreover, it has probability of error exponentially decreasing to 0 so that it is exceptionally reliable.

Thus using this ECC scheme with our efficient implementation given in Chapter 4 gives an end-to-end wiretap coding scheme for the AWGN wiretap channel that is concrete, efficient, exceptional (in both reliability and security), semantically secure, and can achieve the secrecy capacity.

Remark. Using the UHF $\mathcal{F}^*$ as given in Chapter 4, it is possible to use the results of [18] and this same ECC to obtain an end-to-end wiretap coding scheme that satisfies all of these properties besides semantic security (it will be only mis-r secure). However, as motivated in Chapter 3, mis-r security should not be used in practice.

A wiretap scheme for the AWGN wiretap channel achieving every single one of these properties already exists, however it was only introduced recently [11] and has come to the authors

[1] Indeed we know one exists by Shannon's channel coding theorem.

attention only upon the time of this writing. However, there the wiretap coding scheme is based on polar lattices and is not modular. In other words, the scheme must be considerably reworked so as to be applicable to other channels and perhaps it wont even be applicable to some channels. In contrast, our scheme *is* modular, the exact same pre/post processor used here can be used on any channel, one just needs to find a proper ECC scheme. As an example, [18, Lemma 5] proves that our exact same scheme (less the ECC) can also achieve the secrecy capacity on certain discrete memoryless channels. Moreover, Theorem 2 can be reworked to allow the concept of *side information* so as to prove our scheme works on fading channels modeling wireless communications. Indeed, in the journal version of this thesis, we prove that in that sense our scheme can achieve the secrecy capacity in the case of full CSIT fading channels. Thus, our *same scheme* achieves the semantic secrecy capacity in three disparate situations.

It is of course of considerable interest to see which other situations our scheme achieves the semantic secrecy capacity for; perhaps even in general situations. For instance, if we can show that $\lim\limits_{n \to \infty} \frac{\boldsymbol{\mathcal{I}}^{\epsilon}_{n}}{n} \leq I(V^* \wedge Z)$ (for some optimal prefix V^*) then our coding scheme implies that the weak secrecy capacity is equivalent to the semantic security metric on *all* discrete-time memoryless channels and even better, *our transmission procedure* can be used to achieve such a rate. This would indeed be a surprising result but initial investigation looks optimistic.

In closing remarks to this section, we note that our scheme/procedure is in some sense *universal* and reduces problems of the wiretap variety down to problems of the error correcting code variety. On any given wiretap channel, one needs only to find the bound ξ (dependent on the eavesdropper's channel only) and a "good ECC scheme" $\mathcal{C}$; i.e. a ECC scheme that is concrete, efficient, and reliable. However, finding good ECC schemes is already an active area of research and in fact is probably the biggest area of active research with respect to communication and information theory. Thus, our procedure basically merges the wiretap community *into* the ECC community. This is of interest in both a theoretic and practical sense.

8. DISCUSSION

8.1. Looking Forward

In Chapter 1, we argued that there were two main hurdles that information theoretic security needed to overcome. Here we demand even more from information theoretic coding schemes so as to bolster their realism. The following is such a program.

- ***Tangibility:*** Wiretap coding schemes should be concrete so as to be implemented in real life systems in an algorithmic way.
- ***Reliability:*** The probability of error should become negligible as block length becomes arbitrarily large: $p_{e,n} \to 0$ as $n \to \infty$.
- ***Security:*** The *semantic* leakage should become negligible as block length becomes arbitrarily large: $\mathbb{L}_n^{\mathrm{ss}} \to 0$ as $n \to \infty$.
- ***Efficiency:*** The wiretap scheme should have polynomial end-to-end time complexity with respect to the block length n.
- ***Finite Efficacy:***
  - *Reliability:* Wiretap coding schemes should be *exceptionally* reliable.
  - *Security:* Wiretap coding schemes should be *exceptionally* semantically secure.
- ***Rate Supremacy:*** The wiretap coding scheme should achieve the secrecy capacity.

This program seems demanding at first sight but it *ensures* coding schemes satisfy the rigorous demands of reality.

We *have* in this thesis constructed a wiretap coding scheme that satisfies every one of these requirements at least for the AWGN wiretap channel. It is of future interest to see how our scheme fares in regards to this program for other wiretap channels. Luckily, our scheme is modular so that this consideration does not require significant overhead.

8.2. Conclusion

Physical layer security does have its drawbacks to computational based security, but in instances where this form of security is possible and/or appropriate, *realistic* wiretap coding schemes ensure that this *unbreakable* form of security can actually be useful in practice. We have presented a method for taking error correcting codes and making them secure with respect to the best asymptotic metric: semantic security. Moreover, our conversion is a concrete process based on finite field arithmetic that is polynomial time efficiently implementable. Even further, for the case of the AWGN channel, our conversion takes *optimal error correcting codes*[1] and converts them into *optimal wiretap codes*[2].

[1] ECC's that can achieve the point to point AWGN capacity.

[2] Wiretap codes that can achieve the AWGN secrecy capacity.

REFERENCES

[1] Mihir Bellare and Stefano Tessaro. Polynomial-time, semantically-secure encryption achieving the secrecy capacity. *CoRR*, abs/1201.3160, 2012.

[2] Mihir Bellare, Stefano Tessaro, and Alexander Vardy. A cryptographic treatment of the wiretap channel. *CoRR*, abs/1201.2205, 2012.

[3] M. R. Bloch and J. N. Laneman. Strong secrecy from channel resolvability. *IEEE Transactions on Information Theory*, 59(12):8077–8098, Dec 2013.

[4] Matthieu Bloch and Joao Barros. *Physical-layer security: from information theory to security engineering.* Cambridge University Press, 2011.

[5] J.Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143 – 154, 1979.

[6] I. Csiszar and J. Korner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, May 1978.

[7] Darrel Hankerson, Alfred Menezes, and Scott Vanstone. *Guide to Elliptic Curve Cryptography.* Springer-Verlag., 2004.

[8] Shunsuke Ihara. *Information theory for continuous systems.* World Scientific, 1993.

[9] S. Leung-Yan-Cheong and M. Hellman. The gaussian wire-tap channel. *IEEE Transactions on Information Theory*, 24(4):451–456, Jul 1978.

[10] C. Ling, L. Luzzi, J. C. Belfiore, and D. Stehlé. Semantically secure lattice codes for the gaussian wiretap channel. *IEEE Transactions on Information Theory*, 60(10):6399–6416, Oct 2014.

[11] L. Liu, Y. Yan, and C. Ling. Achieving secrecy capacity of the gaussian wiretap channel with polar lattices. *IEEE Transactions on Information Theory*, 64(3):1647–1665, March 2018.

[12] Ueli M. Maurer. *The Strong Secret Key Rate of Discrete Random Triples*, pages 271–285. Springer US, Boston, MA, 1994.

[13] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, Oct 1948.

[14] C. E. Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715, Oct 1949.

[15] P. W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *eprint arXiv:quant-ph/9508027*, August 1995.

[16] I. Tal and A. Vardy. Channel upgrading for semantically-secure encryption on wiretap channels. In *2013 IEEE International Symposium on Information Theory*, pages 1561–1565, July 2013.

[17] H. Tyagi and A. Vardy. Explicit capacity-achieving coding scheme for the gaussian wiretap channel. In *2014 IEEE International Symposium on Information Theory*, pages 956–960, June 2014.

[18] H. Tyagi and A. Vardy. Universal hashing for information-theoretic security. *Proceedings of the IEEE*, 103(10):1781–1795, Oct 2015.

[19] S. Vatedka and N. Kashyap. A capacity-achieving coding scheme for the awgn channel with polynomial encoding and decoding complexity. In *2016 Twenty Second National Conference on Communication (NCC)*, pages 1–6, March 2016.

[20] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387, 1975.