A FRESH LOOK INTO THE BIOMETRIC AUTHENTICATION
Perspective from Shannon's Secrecy System and a Special Wiretap Channel

Yanling Chen[1] and A. J. Han Vinck[2]
[1]*Department of Telematics, Norwegian University of Science and Technology, O.S.Bragstads plass 2A, Trondheim, Norway*
[2]*Institute for Experimental Mathematics, University of Duisburg-Essen, Ellernstr. 29, Essen, Germany*

Keywords: Biometrics, Authentication, Error-correcting codes, Information theory.

Abstract: In this paper, we first look at the biometric authentication scheme as an extension of Shannon's secrecy system with an error-prone key, and derive the necessary condition for the perfect secrecy. Furthermore, we show that the Juels-Wattenberg scheme is optimal by fulfilling such a condition once the biometric key and its error pattern satisfy certain statistical distributions; otherwise, it is possible to improve its performance by coding on basis of the biometric. We further confirm this proposition by reformulating the Juels-Wattenberg scheme with a smart encoder to a specific model of wiretap channel with side information, where the side information is the enrolled biometric template and assumed to be known at the encoder. The idea of the smart encoder is inspired by the fact that the authorities are collecting the biometric information from people since years, and this knowledge could in turn be used to better design the biometric systems for people's good. From an information theoretic perspective, we explore the secrecy capacity of this specific wiretap channel and demonstrate that the knowledge of the enrolled biometric template at the smart encoder does provide an advantage so as to enhance the performance of the biometric authentication scheme.

1 INTRODUCTION

Traditional user authentication systems are mostly based on something one knows but one may forget (e.g.: a password), or something one has but one may lose (e.g.: a passport), whilst the new biometric systems are based on people's biometric characteristics which represent who one really is. Thus the new systems overcome the disadvantages of the traditional ones and provide an attractive solution to do user authentication in a fast, easy and convenient manner.

The biometric characteristics can be physiological (such as fingerprint, DNA, iris or face), or behavioral (such as typing rhythm, gait and voice). All of them are (or rather should be) unique, not duplicable or transferable. So it is important that no biometric image or template is stored. What is stored is so-called "encrypted biometric template". To do this, one can bind a random key to a biometric such that neither the random key nor the biometric can be retrieved from the stored data. We note that in practice, the biometric representations of a person vary dramatically depending on the acquisition method, acquisition environment and user's interaction with the acquisition device. Thus it arises the challenge how to tolerate the fuzziness of biometric readings and ensure the exactitude at the same time so as to fulfill the system requirements.

Among the emerging biometric authentication systems, fuzzy commitment (Juels and Wattenberg, 1999) and fuzzy vault (Juels and Sudan, 2006) are two of the most representative. In particular, according to a thorough study presented in (Cavoukian and Stoianov, 2009), the fuzzy commitment scheme, whose main spirit is to employ error correcting codes (J.MacWilliams and Sloane, 1977) to tackle the fuzziness problem of biometric templates, is conceptually the simplest, but also one of the best for the biometrics where the proper alignment of images is possible. Whilst when the biometric data is unordered or with arbitrary dimensionality (such as fingerprint minutiae), it is more suitable to apply the fuzzy vault scheme. As observed by (Y. Dodis and Smith, 2008), the fuzzy commitment scheme and the fuzzy vault scheme are essentially secure sketches in Hamming metric space and set difference metric space, respectively. In (Y. Dodis and Smith, 2008), the authors also introduced a scheme based on the constant-weight code and permutations. Balakirsky et al. in (V. B. Balakirsky and Vinck, 2009) reviewed this permutation block coding scheme and pointed out that most of the permutations could be "bad" in the manner that a ran-

Chen Y. and J. Han Vinck A..
A FRESH LOOK INTO THE BIOMETRIC AUTHENTICATION - Perspective from Shannon's Secrecy System and a Special Wiretap Channel.
DOI: 10.5220/0003519001680177
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2011), pages 168-177
ISBN: 978-989-8425-71-3

dom choice of permutations may result in a poor performance in secrecy. They further demonstrated that proper assignment of the permutations on the basis of the biometric and the employed constant-weight code could significantly reduces the probability of a successful attack.

In this paper, we take a fresh look into biometric authentication from an information-theoretic perspective. As a general model, we reconsider the biometric authentication as an extension of Shannon's secrecy system. Similarly to the results for Shannon's secrecy system, we derive a necessary condition for obtaining perfect secrecy, which do not depend on any specific metric spaces. More specifically, we put our focus on the fuzzy commitment scheme, i.e., Juels-Wattenberg scheme as referred in the rest of the paper, as the Hamming distance is perhaps the most natural metric to consider. We show that the Juels-Wattenberg scheme can be optimal in transmission/storage efficiency under some idealized settings.

Going one step further, we notice that since years the authorities have been collecting the biometric information from people. For instance, in most countries to apply for a visa, a digital photograph needs to be submitted; and when one enters the border of a country, she/he might be required to have her/his fingerprint scanned. So we could assume that there is a smart encoder which learns the enrolled biometric templates, and in turn may use this knowledge to improve the performance of the current biometric authentication scheme. Inspired by this observation and previous work, we investigate the Juels-Wattenberg scheme with a smart encoder which learns the enrolled biometric template. By remodeling it to a specific model of wiretap channel, we establish insights into limitations and possible improvement on the current biometric system.

There are two kinds of errors that biometric systems do: *false rejection* occurs when a legitimate user is rejected and *false acceptance* occurs when an imposter is accepted as a legitimate user. So the performance of the system is often illustrated by the *false rejection rate* (FRR) and *false accept rate* (FAR). The less are both rates, the better is the system performance. In the reformulated systems present in this paper, we use the terminologies *average probability of error* at the legitimate user and the *information leakage rate* to the eavesdropper to evaluate the accuracy and privacy performance. The former concept is by definition the FRR; whilst the latter, as its name suggests, characterizes the amount of information leak to a third party. If the best an attacker can do is to try to obtain the biometric template/key from the database of the "encrypted biometric templates", then due to Fano's inequality it can be shown that the FAR is upper bounded by the information leakage rate. One can refer to the Appendix for a detailed proof of this.

In this paper, we use $\mathbf{b}$ to denote the master biometric template, which is mostly generated from multiple biometric samples from the user at the enrollment phase; $\mathbf{b}'$ denotes the biometric template obtained at the time of authentication; while $\mathbf{e}$ represents the difference of the biometric readings of the same user at two different phases. For simplicity, the analysis of this paper is based on the following assumptions.

- information is represented and transmitted in bits.
- biometric characteristics contain enough randomness which can be extracted to guarantee the system performance in terms of accuracy and secrecy.
- variation in the biometric readings $\mathbf{e}$ is independent of the master biometric template $\mathbf{b}$.

Throughout this paper, between two binary sequences, the bitwise addition is carried out modulo 2. Besides, when the dimension of a sequence is clear from the context or to be defined, we denote the sequences in boldface letters for simplicity. A similar convention applies to random variables, which are denoted by upper-case letters. For the readers' convenience, we also provide a list of notations in Appendix.

The rest of the paper is organized as follows: in Section 2, we briefly review the Juels-Wattenberg scheme. In Section 3, we look into the biometric authentication scheme from the perspective of an extension of Shannon's secrecy system. In Section 4, we reformulate the Juels-Wattenberg scheme with a smart encoder to a specific wiretap channel with side information. We demonstrate how the knowledge of the enrolled biometrics can be employed to improve the performance of the Juels-Wattenberg scheme through both theoretical results and numerical examples. Finally we conclude in Section 5.

2 JUELS-WATTENBERG SCHEME

The Juels-Wattenberg scheme (Juels and Wattenberg, 1999) is described as follows:

At enrollment,

- choose a random vector $\mathbf{s}$ and accordingly construct a codeword $\mathbf{c}$ by a prespecified error correcting code.

- calculate and store Hash($\mathbf{s}$) and $\mathbf{r} = \mathbf{c} + \mathbf{b}$. Here Hash$(\cdot)$ is a cryptographic hash function.

At the authentication phase,

- input $\mathbf{b}'$ and calculate $\mathbf{c}' = \mathbf{b}' + \mathbf{r} = \mathbf{b}' + \mathbf{b} + \mathbf{c} = \mathbf{c} + \mathbf{e}$. Here $\mathbf{e} = \mathbf{b}' + \mathbf{b}$.
- decode $\hat{\mathbf{c}}$ and recover $\hat{\mathbf{s}}$. Correct decoding delivers $\hat{\mathbf{s}} = \mathbf{s}$.
- compare Hash($\hat{\mathbf{s}}$) with Hash($\mathbf{s}$). Accept if they are equal and reject otherwise.

3 PERSPECTIVE OF SHANNON'S SECRECY SYSTEM

In this section, we first briefly review Shannon's secrecy system (Shannon, 1949). Then we take the biometric authentication scheme as an extension of Shannon's secrecy system, addressing its performance limit in terms of efficiency at perfect secrecy.

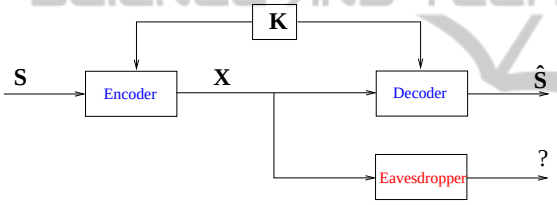

Figure 1: Shannon's secrecy system.

The model of Shannon's secrecy system (Shannon, 1949) is illustrated in Fig. 1. A sender (Alice) wants to communicate a message $\mathbf{s}$ to a receiver (Bob) over a public channel in the presence of an eavesdropper (Eve) who observes the channel output. Alice and Bob share a key $\mathbf{k}$, which is unknown to Eve. So to each message $\mathbf{s}$ and key $\mathbf{k}$, the encoder assigns a ciphertext $\mathbf{x}$, and to each ciphertext $\mathbf{x}$ and $\mathbf{k}$, the decoder assigns $\hat{\mathbf{s}}$ as a decoded message corresponding to $\mathbf{s}$.

Suppose that a $(2^{nR}, n)$ secrecy code is used to encode $\mathbf{s}$ into $\mathbf{x}$, where 2^{nR} is the number of the messages $\mathbf{s}$ and n is the length of the sequence $\mathbf{x}$. Clearly R is the *transmission rate* from the sender to the receiver. Further we denote the *information leakage rate* associated with the $(2^{nR}, n)$ secrecy code to be

$$R_l^{(n)} = \frac{1}{n} I(\mathbf{S}; \mathbf{X}); \tag{1}$$

and the *average probability of error* to be

$$P_e^{(n)} = \Pr\{\mathbf{S} \neq \hat{\mathbf{S}}\}. \tag{2}$$

Here $I(\cdot)$ is the *mutual information* function, which measures the amount of information shared by two variables. For its definition, one can refer to (Cover and Thomas, 2005).

The communication system shown in Fig. 1 is said to have *perfect secrecy* if

- $P_e^{(n)} = 0$, i.e., the message is decoded correctly, and
- $R_l^{(n)} = 0$, i.e., the *information leakage* $I(\mathbf{S}; \mathbf{X}) = 0$, the ciphertext reveals no information about the message.

By Shannon's perfect secrecy theorem (Shannon, 1949), the necessary condition for perfect secrecy is $H(\mathbf{K}) \geq H(\mathbf{S})$. Here $H(\cdot)$ is the *entropy* function (Cover and Thomas, 2005), which measures the uncertainty associated with a random variable.

3.1 An Extension of Shannon's Secrecy System

Now let us take a fresh look at the biometric authentication scheme from the point of view of Shannon's secrecy system. Alice and Bob can be considered as the same user at the enrollment phase and the authentication phase. The user uses his/her biometric property as a key. In Shannon's secrecy system model, the key used at the encoder and the decoder is the same. However, in the biometrics authentication scheme, the biometric key varies slightly in each reading: $\mathbf{b}$ at enrollment whilst $\mathbf{b}'$ at the authentication phase. Therefore, we could formulate the biometric authentication scheme as an extension of Shannon's secrecy system, as shown in Fig. 2. Similarly to Shannon's perfect secrecy theorem, we have the following theorem for the Shannon's secrecy system with an error-prone key.

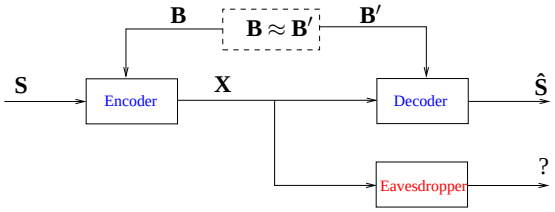

Figure 2: An extension of Shannon's secrecy system.

Theorem 3.1. *Consider the extension of Shannon's secrecy system as shown in Fig. 2. The necessary condition for perfect secrecy is* $H(\mathbf{S}) \leq I(\mathbf{B}; \mathbf{B}')$.

Proof. The proof uses definitions of entropy, mutual information as well as their properties such as chain rule, data-processing inequality and so on (Cover and Thomas, 2005). We consider

$$\begin{aligned} H(\mathbf{S}) &\overset{(a)}{=} H(\mathbf{S}|\mathbf{X}) \\ &= H(\mathbf{S},\mathbf{B}'|\mathbf{X}) - H(\mathbf{B}'|\mathbf{S},\mathbf{X}) \\ &\overset{(b)}{=} H(\mathbf{B}'|\mathbf{X}) - H(\mathbf{B}'|\mathbf{S},\mathbf{X}) \\ &\overset{(c)}{\leq} H(\mathbf{B}'|\mathbf{X}) - H(\mathbf{B}'|\mathbf{S},\mathbf{B}) \\ &\overset{(d)}{\leq} H(\mathbf{B}'|\mathbf{X}) - H(\mathbf{B}'|\mathbf{B}) \\ &\leq H(\mathbf{B}') - H(\mathbf{B}'|\mathbf{B}) \\ &= I(\mathbf{B};\mathbf{B}'), \end{aligned}$$

where (a) follows by definitions of entropy and mutual information: $H(\mathbf{S}) = H(\mathbf{S}|\mathbf{X}) + I(\mathbf{S};\mathbf{X})$; and by the secrecy constraint $I(\mathbf{S};\mathbf{X}) = 0$; (b) follows that by chain rule of entropy: $H(\mathbf{S},\mathbf{B}'|\mathbf{X}) = H(\mathbf{B}'|\mathbf{X}) + H(\mathbf{S}|\mathbf{X},\mathbf{B}')$; further by data-processing inequality: $H(\mathbf{S}|\mathbf{X},\mathbf{B}') \leq H(\mathbf{S}|\hat{\mathbf{S}})$, since $\mathbf{S} \to (\mathbf{X},\mathbf{B}') \to \hat{\mathbf{S}}$ forms a Markov chain; and by definition of entropy and the communication constraint $\Pr\{\mathbf{S} \neq \hat{\mathbf{S}}\} = 0$, we have $H(\mathbf{S}|\hat{\mathbf{S}}) = 0$; (c) follows that $H(\mathbf{B}'|\mathbf{S},\mathbf{X}) \geq H(\mathbf{B}'|\mathbf{S},\mathbf{X},\mathbf{B}) = H(\mathbf{B}'|\mathbf{S},\mathbf{B})$ since $\mathbf{B}' \to (\mathbf{S},\mathbf{B}) \to \mathbf{X}$ forms a Markov chain; and (d) follows the fact that given $\mathbf{B}$, $\mathbf{B}'$ is independent of $\mathbf{S}$. □

As one can see in next subsection, the equality in Theorem 3.1 can be achieved under some idealized settings. In fact, the achievability is largely depending on the statistical behavior of the error pattern $\mathbf{e}$.

We also note that the results we have obtained in this subsection are general, do not depend on any particular metric space. However, from next subsection on, our discussion will be mainly in Hamming metric space. Special focus will be on the Juels-Wattenberg scheme and its variants.

3.2 Juels-Wattenberg Scheme can be Optimal

Let k be the length of sequence $\mathbf{s}$; n be the length of sequences $\mathbf{b},\mathbf{b}'$ (thus $\mathbf{e}$) and $\mathbf{x}$. Recall that $\mathbf{e} = \mathbf{b} + \mathbf{b}'$ represents the difference of the two readings of the same biometric property at two different phases. It is reasonable to assume that there exists t, such that $hw(\mathbf{e}) \leq t$ always holds, where $hw(\cdot)$ is the *Hamming weight* function, representing the number of the non-zero bits in a sequence. Under this assumption, we can use an error correcting code to tackle the fuzziness problem of the biometric key as the Juels-Wattenberg scheme does.

Suppose that there exists an (n,k,d) linear code C of length n, dimension k and minimum distance d, where $d = 2t+1$. For any k-bit secret $\mathbf{s}$, first we encode it into an n-bit sequence $\mathbf{c}$, where $\mathbf{c} \in C$. Then take $\mathbf{x} = \mathbf{c} + \mathbf{b}$.

To recover the secret $\mathbf{s}$ from the ciphertext $\mathbf{x}$, we simply add $\mathbf{b}'$. Thus we have $\mathbf{c}' = \mathbf{x} + \mathbf{b}' = \mathbf{c} + \mathbf{e}$. Due to the fact that $hw(\mathbf{e}) \leq t$ and $\mathbf{c} \in C$, $\mathbf{s}$ will be correctly decoded. Thus we have $P_e^{(n)} = 0$.

If the key $\mathbf{B}$ is uniformly distributed over C, it can be readily checked that $\mathbf{X}$ is also uniformly distributed over C and $\Pr\{\mathbf{s}\} = \Pr\{\mathbf{s}|\mathbf{x}\}$. Thus $\mathbf{S}$ and $\mathbf{X}$ are independent and we have $H(\mathbf{S}|\mathbf{X}) = H(\mathbf{S}) = k$, i.e., $R_l^{(n)} = 0$.

Furthermore, if the error sequence $\mathbf{E}$ is uniformly distributed among the sequences, which are of length n, have Hamming weight $\leq t$ and are in total $M = \sum_{i=0}^{t} \binom{n}{i}$ of them, then it is easy to check that $\mathbf{B}' = \mathbf{B} + \mathbf{E}$ has a uniform distributions among $M \cdot 2^k$ different n-bit sequences. Straightforwardly we have

$$\begin{aligned} I(\mathbf{B};\mathbf{B}') &= H(\mathbf{B}') - H(\mathbf{B}'|\mathbf{B}) \\ &= H(\mathbf{B}') - H(\mathbf{E}) \\ &= \log\{M \cdot 2^k\} - \log M \\ &= k = H(\mathbf{S}). \end{aligned}$$

Therefore, in order to tolerate t errors in the key $\mathbf{b}$, the amount of information can be carried in n bits is bounded by the largest dimension k of a t-error correcting linear code of length n, which turns out to be a coding problem. If we characterize the transmission/storage efficiency R of the above scheme by the information rate k/n of the linear code C, then according to the Singleton bound (J.MacWilliams and Sloane, 1977), we easily derive an upper bound $R \leq 1 - \frac{d-1}{n} = 1 - \frac{2t}{n}$. Besides, if we allow n to grow since we have assumed that the biometric under consideration has enough randomness, then due to the Gilbert-Varshamov bound (J.MacWilliams and Sloane, 1977), the rate $R \geq 1 - h(\frac{d}{n})$ is achievable for $d/n < 1/2$. Here $h(\cdot)$ is the binary entropy function.

Recall that in the Juels-Wattenberg scheme, the biometric vector $\mathbf{b}$ is assumed to be uniformly distributed among vectors of a given length n. In that case, the published vector $\mathbf{c} + \mathbf{b}$ is also uniformly distributed among the n-bit vectors and thus yields no information on the secret $\mathbf{s}$ or the biometric template $\mathbf{b}$. In the extension of Shannon's secrecy system discussed above, we see that the Juels-Wattenberg scheme is still optimal once $\mathbf{b}$ is uniformly distributed over a linear $(n,k,2t+1)$ code C and $\mathbf{e}$ is uniformly distributed over n-bits sequences of Hamming weight $\leq t$.

However, in reality, the statistical distributions of the extracted biometrics and its error pattern can be far different (one can refer to (A. Pankanti and Jain, 2002) for a comprehensive survey on the probability of the fingerprint configuration). The equality in Theorem 3.1 can be hard to achieve or not achievable at

all. So open problems are challenging up to the real biometric readings and the real error patterns.

3.3 Improvement of Juels-Wattenberg Scheme

In last subsection, we show that once $\mathbf{b},\mathbf{e}$ satisfy certain statistical distributions, the Juels-Wattenberg scheme provides with optimal performance by employing an error correcting code. However, we wonder if this still holds otherwise. We further notice that in the Juels-Wattenberg scheme, the choice of the codeword $\mathbf{c}$ (not necessarily from a linear code) is independent of the enrolled biometric reading $\mathbf{b}$. In the following, we present an example and show that by exploring their dependency, i.e., taking codeword $\mathbf{c}$ on the basis of $\mathbf{b}$, it is possible to improve the secrecy performance of the Juels-Wattenberg scheme.

Example 3.2. *Consider the Juels-Wattenberg scheme. We let $\mathcal{C},\mathcal{B}$ be sets of the codewords $\mathbf{c}$ and master biometric templates $\mathbf{b}$, respectively. Suppose $\mathcal{C}=\{\mathbf{c_1},\mathbf{c_2},\mathbf{c_3}\}$ and $\mathcal{B}=\{\mathbf{b_1},\mathbf{b_2},\mathbf{b_3},\mathbf{b_4},\mathbf{b_5},\mathbf{b_6}\}$. In particular, we specify $\mathbf{c}$ and $\mathbf{b}$ by the following matrices:*

$$\begin{bmatrix}\mathbf{c_1}\\ \mathbf{c_2}\\ \mathbf{c_3}\end{bmatrix}=\begin{bmatrix}00110011\\ 01010101\\ 10101010\end{bmatrix},$$

$$\begin{bmatrix}\mathbf{b_1}\\ \mathbf{b_2}\\ \mathbf{b_3}\\ \mathbf{b_4}\\ \mathbf{b_5}\\ \mathbf{b_6}\end{bmatrix}=\begin{bmatrix}00001111\\ 00110011\\ 01010101\\ 10101010\\ 11001100\\ 11110000\end{bmatrix}.$$

For simplicity we assume that $\mathbf{b}$ is uniformly distributed over $\mathcal{B}$.

In the original Juels-Wattenberg scheme, $\mathbf{c}$ is chosen randomly independent of $\mathbf{b}$. Thus $\mathbf{c}$ is uniformly distributed over $\mathcal{C}$. It is easy to check that $\mathbf{c}+\mathbf{b}$ has a non-uniform distribution over 8 different sequences. In fact, $\mathbf{c}+\mathbf{b}$ is 00111100 or 11000011 either with probability $1/18$; 01011010 or 10100101 with either probability $1/9$; and 00000000, 01100110, 10011001 or 11111111 each with probability $1/6$.

In particular, we notice that if $\mathbf{c}+\mathbf{b}$ turns out to be either 00111100 or 11000011, then it corresponds to one possibility $\mathbf{c}=\mathbf{c_1}$. Besides, with the observation $\mathbf{c}+\mathbf{b}$, one can easily calculate that the probability of correct guess of $\mathbf{c}$ (thus $\mathbf{b}$) is

$$2*\frac{1}{18}+2*\frac{1}{9}*\frac{1}{2}+4*\frac{1}{6}*\frac{1}{3}=0.4444;$$

the information leakage from $\mathbf{c}+\mathbf{b}$ on $\mathbf{c}$ is

$$\begin{aligned}I(\mathbf{C};\mathbf{C}+\mathbf{B})=&H(\mathbf{C})+H(\mathbf{C}+\mathbf{B})-H(\mathbf{C},\mathbf{B})\\ \overset{(a)}{=}&H(\mathbf{C}+\mathbf{B})-H(\mathbf{B})\\ =&2*\frac{1}{18}*\log 18+2*\frac{1}{9}*\log 9\\ &+4*\frac{1}{6}\log 6-\log 6\\ =&0.3061;\end{aligned}$$

and the information leakage from $\mathbf{c}+\mathbf{b}$ on $\mathbf{b}$ is

$$\begin{aligned}I(\mathbf{B};\mathbf{C}+\mathbf{B})=&H(\mathbf{B})+H(\mathbf{C}+\mathbf{B})-H(\mathbf{C},\mathbf{B})\\ \overset{(a)}{=}&H(\mathbf{C}+\mathbf{B})-H(\mathbf{C})\\ =&2*\frac{1}{18}*\log 18+2*\frac{1}{9}*\log 9\\ &+4*\frac{1}{6}\log 6-\log 3\\ =&1.3061;\end{aligned}$$

where (a) is due to the fact that the choice of $\mathbf{c}$ is independent of $\mathbf{b}$. The logarithm is to the base 2.

Now we slightly modify the Juels-Wattenberg scheme by introducing a bit dependency to the choice of $\mathbf{c}$ on the basis of $\mathbf{b}$. When $\mathbf{b}=\mathbf{b_1}$ or $\mathbf{b_6}$, we do not take $\mathbf{c_1}$ as a candidate of $\mathbf{c}$ so as to avoid the output 00111100 or 11000011 of $\mathbf{c}+\mathbf{b}$. That is, if $\mathbf{b}\neq\mathbf{b_1}$ or $\mathbf{b_6}$, we choose $\mathbf{c}$ randomly from set $\mathcal{C}$; and if $\mathbf{b}=\mathbf{b_1}$ or $\mathbf{b_6}$, we choose $\mathbf{c}$ randomly from set $\mathcal{C}\setminus\{\mathbf{c_1}\}$. Then it is easy to check that $\mathbf{c}+\mathbf{b}$ is uniformly distributed among 6 different sequences, 01011010, 10100101, 00000000, 01100110, 10011001 and 11111111, each with probability $1/6$. Thus $\mathbf{C}+\mathbf{B}$ has the same entropy as $\mathbf{B}$. In this case, with the observation $\mathbf{c}+\mathbf{b}$, one can calculate and see that the probability of correct guess of $\mathbf{c}$ (thus $\mathbf{b}$) is reduced to

$$2*\frac{1}{6}*\frac{1}{2}+4*\frac{1}{6}*\frac{1}{3}=0.3889;$$

the information leakage on $\mathbf{c}$ is reduced to

$$\begin{aligned}I(\mathbf{C};\mathbf{C}+\mathbf{B})=&H(\mathbf{C})+H(\mathbf{C}+\mathbf{B})-H(\mathbf{C},\mathbf{B})\\ \overset{(b)}{=}&H(\mathbf{C})-H(\mathbf{C}|\mathbf{B})\\ =&\frac{2}{9}*\log\frac{9}{2}+2*\frac{7}{18}*\log\frac{18}{7}\\ &-\frac{1}{6}(2*\log 2+4*\log 3)\\ =&0.1520;\end{aligned}$$

and the information leakage on $\mathbf{b}$ is reduced to

$$\begin{aligned}I(\mathbf{B};\mathbf{C}+\mathbf{B})=&H(\mathbf{B})+H(\mathbf{C}+\mathbf{B})-H(\mathbf{C},\mathbf{B})\\ \overset{(b)}{=}&H(\mathbf{C}+\mathbf{B})-H(\mathbf{C}|\mathbf{B})\\ =&\log 6-\frac{1}{6}(2*\log 2+4*\log 3)\\ =&1.1950;\end{aligned}$$

where (b) is due to the fact that the choice of $\mathbf{c}$ is dependent on $\mathbf{b}$, and $H(\mathbf{B}) = H(\mathbf{C}+\mathbf{B})$.

This example demonstrates how one can enhance the secrecy performance of the Juels-Wattenberg scheme by choosing codewords based on knowledge of the biometric template. Clearly, the Juels-Wattenberg scheme does not always provide optimal solutions. To achieve a good performance accuracy and secrecy, the appropriate error correcting code, in particular the choice of the codeword, should be chosen largely based on the biometric template and its error pattern.

One may also notice that the example looks very similar to the one in (V. B. Balakirsky and Vinck, 2009). However, the underneath coding methods are different, where in our example we use the coding from Juels-Wattenberg scheme while a permutation block coding is employed in (V. B. Balakirsky and Vinck, 2009).

In next section, we build an information theoretic framework for the Juels-Wattenberg scheme with a smart encoder which learns the biometric templates at enrollment, further confirm its advantage in obtaining better trade-offs between the accuracy and secrecy performance.

4 PERSPECTIVE OF THE WIRETAP CHANNEL

The concept of the wiretap channel was first introduced by Wyner in (Wyner, 1975). Its goal is to achieve not only *efficient*, *reliable* but also *secure* communication between the sender and legitimate receiver. Here being secure is against an eavesdropper, who is assumed to know the deployed encoding and decoding scheme and observe a degraded version of the output at the legitimate receiver. Similarities can be easily recognized between the model of the wiretap channel, and a biometric authentication system with an attacker who is assumed to have access to the encrypted biometrics stored in the database. So an information theoretic approach can be taken by looking into the biometric authentication system from a perspective of the wiretap channel. Such an instance is first given in (Cohen and Zémor, 2004). Other trials can be found in (V. B. Balakirsky and Vinck, 2009) and (Vinck and Balakirsky, 2010), etc.

In (Cohen and Zémor, 2004), the authors reconsider the Juels-Wattenberg scheme as a Wyner's wiretap channel. The reformulation is shown in Fig. 3, where $\mathbf{S},\mathbf{X}$ serve similar roles as $\mathbf{s},\mathbf{c}$ in the Juels-Wattenberg scheme, respectively, representing the secret chosen randomly and the corresponding codeword; $\mathbf{Y} = \mathbf{X}+\mathbf{E}$ at the legitimate receiver, is the sequence recovered by the legal user at the authentication phase; $\mathbf{Z} = \mathbf{X}+\mathbf{B}$ is the encrypted biometric template stored in database and assumed to be accessible to an attacker. $\mathbf{E}$, as the variation of two biometric readings, is assumed to be *less noisy* than $\mathbf{B}$ and independent of $\mathbf{B}$. That is, the channel from $\mathbf{X}$ to $\mathbf{Y}$, to the legitimate receiver, is *less noisy* than the one from $\mathbf{X}$ to $\mathbf{Z}$, to the eavesdropper.

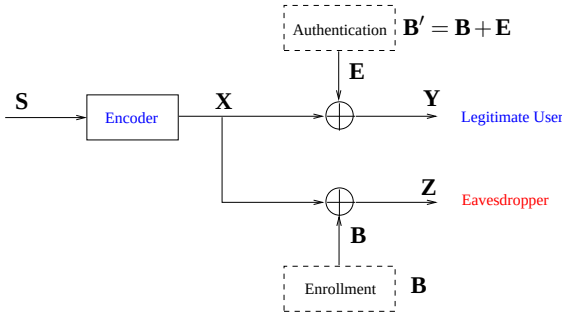

Figure 3: Reformulation of Juels-Wattenberg scheme as a wiretap channel (Cohen and Zémor, 2004).

For the model shown in Fig. 3, we recall the *information leakage rate* associated with a $(2^{nR}, n)$ secrecy code

$$R_l^{(n)} = \frac{1}{n} I(\mathbf{S};\mathbf{Z});$$

and the *average probability of error* at the legitimate receiver

$$P_e^{(n)} = \Pr\{\mathbf{S} \neq \hat{\mathbf{S}}\}.$$

We note that in most communication systems based on a statistical channel model, it is often impossible to achieve a positive transmission rate at the absolute perfect secrecy (i.e. $R_l^{(n)} = P_e^{(n)} = 0$). For instance, the zero-error capacity (maximum transmission rate at $P_e^{(n)} = 0$) of a binary symmetric channel is zero. So it is necessary to consider a weaker concept: asymptotic perfect secrecy.

A communication system is said to have *asymptotic perfect secrecy*, if for any arbitrary small $\varepsilon, \varepsilon' > 0$, there exists a secrecy code $(2^{nR}, n)$ such that

- $P_e^{(n)} < \varepsilon$, i.e., decoding error occurs only with arbitrarily small probability;
- $R_l^{(n)} < \varepsilon'$, i.e., the information leakage can be made arbitrarily small.

We say a secrecy rate R^* *achievable* if for any arbitrary small $\varepsilon, \varepsilon', \varepsilon'' > 0$, there exists a secrecy code $(2^{nR}, n)$ such that

$$R > R^* - \varepsilon'', \quad P_e^{(n)} < \varepsilon, \quad R_l^{(n)} < \varepsilon'. \qquad (3)$$

The maximum secrecy rate is called *asymptotic secrecy capacity* or for short *secrecy capacity*.

As a direct consequence of Wyner's result (Wyner, 1975), the (asymptotic) secrecy capacity: C_s^1, of the model in Fig. 3 is

$$C_s^1 = \max_{\Pr\{x\}} \{I(X;X+E) - I(X;X+B)\}. \quad (4)$$

4.1 A Special Wiretap Channel with Side Information

Motivated by the observation that the authorities collect the biometric information from people since years, we introduce a smart encoder into the biometric authentication scheme. The smart encoder learns the enrolled biometric templates, and in turn uses its knowledge to enhance the performance of scheme. To explore how much gain can be achieved, we can reformulate the Juels-Wattenberg scheme with a smart encoder to 1): a wiretap channel with two-sided information, one (i.e., $\mathbf{b}$) available at the encoder and the other (i.e., $\mathbf{b}'$) at the decoder of the legitimate receiver; 2) a wiretap channel with side information available at the encoder, where the side information $\mathbf{b}$, is at the same time, the noise in the eavesdropper channel, as shown in Fig. 4. The idea 2) is originally proposed in (Vinck and Balakirsky, 2010).

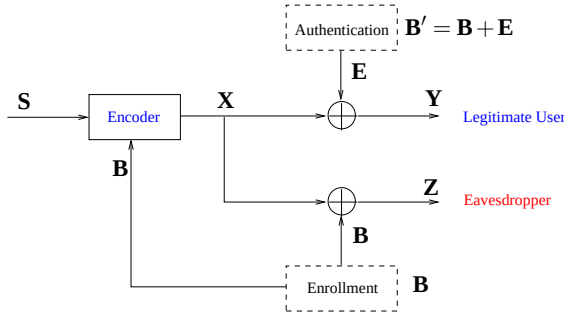

Figure 4: Reformulation of Juels-Wattenberg scheme with a smart encoder as a wiretap channel with side information.

Following the results from (Chen and Vinck, 2008) and (Liu and Chen, 2007), one can obtain some direct results on the secrecy rate (achievable transmission rate at asymptotic perfect secrecy) of the reformulated model. However, the single-letter characterization of the secrecy rate for both models involves an auxiliary parameter, and at present its calculation still remains an unsolved problem. To avoid this, we take a new approach and investigate the reformulation 2).

In this subsection, we consider the secure communication problem via the channel shown in Fig. 4. We note that both $\mathbf{S}$ and $\mathbf{B}$ are known at the encoder. We look at the extreme case, where $\mathbf{X}$ is chosen totally based on knowledge of $\mathbf{B}$. In particular, we take $\mathbf{X} = \mathbf{B}$. Then the legitimate receiver receives $\mathbf{Y} = \mathbf{B} + \mathbf{E}$; whilst the eavesdropper receives $\mathbf{Z} = \mathbf{X} + \mathbf{B}$. Since the information is transmitted in bits, $\mathbf{Z}$ results in a zero sequence. That is, the eavesdropper constantly receives a zero sequence no matter what $\mathbf{B}$ is. Clearly it does not help him to have a better guess of the information being transmitted. Therefore in this case, according to Shannon's coding theorem for noisy channels, $I(\mathbf{X};\mathbf{Y}) = I(\mathbf{B};\mathbf{B}+\mathbf{E})$ bits information can be reliably transmitted to the legitimate receiver while keeping it ignorant to the eavesdropper. If the length of the codeword is n, then we obtain a secrecy rate

$$R_s^2 = \frac{1}{n} I(\mathbf{B};\mathbf{B}+\mathbf{E}). \quad (5)$$

If we let C_s^2 be the secrecy capacity of the channel shown in Fig. 4, then we have $R_s^2 \le C_s^2$. Next we will show the converse $R_s^2 \ge C_s^2$ and thus prove the following theorem.

Theorem 4.1. $C_s^2 = R_s^2$.

Proof. The proof uses the definitions of entropy, mutual information, data-processing inequality and Fano's inequality, for which one can refer to (Cover and Thomas, 2005).

Consider a $(2^{nR}, n)$ secrecy code with a secrecy constraint: $R_l^{(n)} = \frac{1}{n} I(\mathbf{S};\mathbf{Z}) \le \varepsilon_0$ and a communication constraint $P_e^{(n)} = \Pr\{\mathbf{S} \ne \hat{\mathbf{S}}\} \le \varepsilon_1$.

$$\begin{aligned}
nC_s^2 &\le H(\mathbf{S}) \\
&\overset{(a)}{\le} H(\mathbf{S}|\mathbf{Z}) + n\varepsilon_0 \\
&= H(\mathbf{S},\mathbf{B}'|\mathbf{Z}) - H(\mathbf{B}'|\mathbf{S},\mathbf{Z}) + n\varepsilon_0 \\
&= H(\mathbf{B}'|\mathbf{Z}) + H(\mathbf{S}|\mathbf{B}',\mathbf{Z}) - H(\mathbf{B}'|\mathbf{S},\mathbf{Z}) + n\varepsilon_0 \\
&\overset{(b)}{\le} H(\mathbf{B}'|\mathbf{Z}) + H(\mathbf{S}|\mathbf{Y}) - H(\mathbf{B}'|\mathbf{S},\mathbf{Z}) + n\varepsilon_0 \\
&\overset{(c)}{\le} H(\mathbf{B}'|\mathbf{Z}) - H(\mathbf{B}'|\mathbf{S},\mathbf{Z},\mathbf{X}) + n\varepsilon_0 + n\varepsilon_1 \\
&\overset{(d)}{\le} H(\mathbf{B}'|\mathbf{Z}) - H(\mathbf{B}'|\mathbf{B}) + n\varepsilon_0 + n\varepsilon_1 \\
&\le H(\mathbf{B}') - H(\mathbf{B}'|\mathbf{B}) + n\varepsilon_0 + n\varepsilon_1 \\
&= I(\mathbf{B};\mathbf{B}') + n\varepsilon_0 + n\varepsilon_1,
\end{aligned}$$

where (a) follows by the definitions of entropy and mutual information: $H(\mathbf{S}) = H(\mathbf{S}|\mathbf{Z}) + I(\mathbf{S};\mathbf{Z})$; and by the secrecy constraint $I(\mathbf{S};\mathbf{Z}) \le n\varepsilon_0$; (b) follows by the data-processing inequality that $H(\mathbf{S}|\mathbf{B}',\mathbf{Z}) \le H(\mathbf{S}|\mathbf{Y})$ since $\mathbf{Y} = \mathbf{Z} + \mathbf{B}'$; (c) follows the fact that $H(\mathbf{B}'|\mathbf{S},\mathbf{Z}) \ge H(\mathbf{B}'|\mathbf{S},\mathbf{Z},\mathbf{X})$ and $H(\mathbf{S}|\mathbf{Y}) \le H(\mathbf{S}|\hat{\mathbf{S}}) \le nP_e^{(n)} + h(P_e^{(n)}) - 1 \le n\varepsilon_1$, where the last two inequalities are due to the Fano's inequality and the communication constraint $P_e^{(n)} \le \varepsilon_1$, respectively; (d) follows that $H(\mathbf{B}'|\mathbf{S},\mathbf{Z},\mathbf{X}) = H(\mathbf{B}'|\mathbf{S},\mathbf{X},\mathbf{B}) = H(\mathbf{B}'|\mathbf{B})$. □

As a conclusion, we have proven that the special wiretap channel with side information as shown in Fig. 4 has secrecy capacity $C_s^2 = \frac{1}{n} I(\mathbf{B};\mathbf{B}')$. Besides, if $\mathbf{B},\mathbf{E}$ and $\mathbf{B}'$ are sequences of i.i.d random variables B,E and B', respectively, where $B' = B+E$ (this corresponds a discrete memoryless channel with additive noises scenario), we have the following single-letter characterization of the secrecy capacity:

$$C_s^2 = \frac{1}{n} I(\mathbf{B};\mathbf{B}+\mathbf{E}) = I(B;B+E). \tag{6}$$

Remarks on the secrecy constraint: It is easy to see that the above proof also applies to a stricter secrecy constraint $\frac{1}{n} I(\mathbf{S},\mathbf{B};\mathbf{Z}) \leq \varepsilon_0$, since $I(\mathbf{S};\mathbf{Z}) \leq I(\mathbf{S},\mathbf{B};\mathbf{Z})$. Although the stricter constraint $\frac{1}{n} I(\mathbf{S},\mathbf{B};\mathbf{Z}) \leq \varepsilon_0$ is more suitable for biometric security concerns, we still use $\frac{1}{n} I(\mathbf{S};\mathbf{Z}) \leq \varepsilon_0$ since it is enough for the proof of the converse and consistent with the terminology used in the study on wiretap channels.

4.2 Advantage of Knowing the Noise in the Eavesdropper Channel

In this subsection, we will show that knowing the noise in the eavesdropper channel provides an advantage, in the manner that the secrecy capacity of communication model in Fig. 4 is no less than the one in Fig. 3. Intuitively this is true, since any secrecy rate achievable for the model in Fig. 3 is also achievable for the model in Fig. 4 (because the encoder can always ignore the side information). Theoretically we confirm this by the following theorem and further provide numerical comparisons in next subsection.

Theorem 4.2. *$C_s^1 \leq C_s^2$, i.e.,*

$$I(X;X+E) - I(X;X+B) \leq I(B;B+E). \tag{7}$$

Proof. First we consider the term $I(X;X+B)$. By chain rule of the mutual information (Cover and Thomas, 2005), we have

$$\begin{aligned} I(X;X+B) &= I(X,B;X+B) - I(B;X+B|X) \\ &= I(B;X+B) + H(X|B) - H(B|X). \end{aligned}$$

Similarly we have

$$I(X;X+E) \overset{(a)}{=} I(E;X+E) + H(X) - H(E);$$

$$I(B;B+E) \overset{(a)}{=} I(E;B+E) + H(B) - H(E),$$

where (a) follows that E is independent of X and B.

Applying the above equalities to (7), we derive the following inequality, which is equivalent to (7).

$$I(E;X+E) - I(B;X+B) \leq I(E;B+E).$$

So in order to prove the theorem, it is enough to show

$$I(B;X+B) \geq I(E;X+E) - I(E;B+E).$$

This inequality is valid, since

$$\begin{aligned} I(B;X+B) &\overset{(b)}{\geq} I(B;X+B|B+E) \\ &= H(B|B+E) - H(B|B+E,X+B) \\ &= H(E|B+E) - H(E|B+E,X+E) \\ &\overset{(c)}{\geq} H(E|B+E) - H(E|X+E) \\ &= I(E;X+E) - I(E;B+E), \end{aligned}$$

where (b) is due to Lemma 4.3; (c) follows the fact that $H(E|B+E,X+E) \leq H(E|X+E)$. □

Lemma 4.3. *$I(B;X+B) \geq I(B;X+B|B+E)$.*

Proof. First we note that $X+B \to (X,B) \to B+E$ forms a Markov chain. Thus we have

$$I(X,B;X+B) \geq I(X,B;X+B|B+E).$$

In addition, by chain rule of the mutual information (Cover and Thomas, 2005), we have

$$I(X,B;X+B) = I(B;X+B) + I(X;X+B|B);$$

$$\begin{aligned} I(X,B;X+B|B+E) &\overset{(a)}{=} I(B;X+B|B+E) \\ &\quad + I(X;X+B|B), \end{aligned}$$

where (a) is due to the fact that E is independent of X and B. As a direct consequence, we obtain $I(B;X+B) \geq I(B;X+B|B+E)$ and thus complete the proof. □

4.3 Numerical Examples

In this subsection, we give two examples where the secrecy capacity of communication model in Fig. 4 is strictly larger than the one in Fig. 3.

Example 4.4. *Suppose that the channel shown in Fig. 4 is an additive white Gaussian noise (AWGN) wiretap channel, i.e., E and B are white Gaussian noises added into the main channel and the eavesdropper channel, respectively. Further we assume that $E \sim \mathcal{N}(0,N)$ and $B \sim \mathcal{N}(0,Q)$, where $\mathcal{N}(a,b)$ stands for a Gaussian distribution with mean a and variance b. In addition, the average power constraint on X is P.*

It is easy to calculate

$$\begin{aligned} C_s^2 &= I(B;B+E) \\ &= \begin{cases} \frac{1}{2}\log(1+\frac{P}{N}) & \text{if } P < Q \\ \frac{1}{2}\log(1+\frac{Q}{N}) & \text{if } P \geq Q \end{cases}; \\ C_s^1 &= \max_{\Pr(x)} \{I(X;X+E) - I(X;X+B)\} \\ &= \frac{1}{2}\log(1+\frac{P}{N}) - \frac{1}{2}\log(1+\frac{P}{Q}), \end{aligned}$$

where C_s^1 is in fact the difference of the capacities of the main channel and the eavesdropper channel. Easy comparison shows us that $C_s^2 > C_s^1$ holds always.

Example 4.5. *Suppose that the channel shown in Fig. 4 is a binary memoryless symmetric wiretap channel, i.e., both main channel and the eavesdropper channel are binary symmetric channels (BSCs). We further assume that E and B have the following probability distributions:* $\Pr(E=1) = 1 - \Pr(E=0) = e$ *and* $\Pr(B=1) = 1 - \Pr(B=0) = p$, *where* $0 < e, p < 1/2$.

It is easy to calculate

$$\begin{aligned} C_s^2 &= I(B; B+E) = H(B+E) - H(E) \\ &= h(p*e) - h(e); \\ C_s^1 &= \max_{\Pr(x)} \{I(X; X+E) - I(X; X+B)\} \\ &= h(p) - h(e), \end{aligned}$$

where C_s^1 is in fact the difference of the capacities of the main channel and the eavesdropper channel; $h(\cdot)$ is the binary entropy function (Cover and Thomas, 2005), and $p*e = p+e-2pe$ is the value of $\Pr(X+E=1)$. Clearly $C_s^2 > C_s^1$ holds due to the fact that $p*e > p$ and thus $h(p*e) > h(p)$.

For this specific example, it is easy to see that the gain on the secrecy capacity is $h(p*e) - h(p)$, which is increasing with respect to e while decreasing with respect to p, and equal to 0 at either $e = 0$ or $p = 1/2$.

5 CONCLUSIONS

In this paper, we reformulate the biometric authentication scheme to an extension of Shannon's secrecy system with an error-prone key, and the Juels-Wattenberg scheme with a smart encoder to a specific model of wiretap channel, where the encoder knows the noise in the eavesdropper's channel.

From the point of view of an extension of Shannon's secrecy system, we provide theoretical limits on the maximal randomness of the secret that can be concealed and revealed by an error-prone key while hidden perfectly from a third party. Since the code rate of the employed error correcting code reflects the storage efficiency, one can easily derive lower bounds and upper bounds on the storage efficiency by the numerous results on the error correcting codes. Furthermore, we show that the Juels-Wattenberg scheme is optimal if the biometric key and the biometric error pattern satisfy certain statistical distributions; otherwise, it is possible to improve the performance of the Juels-Wattenberg scheme by choosing the codeword from a prespecified error correcting code on the basis of the biometric template.

We further reformulate the Juels-Wattenberg scheme with a smart encoder to a specific model of wiretap channel with side information, where the side information is the noise of the eavesdropper's channel and known at the encoder. The idea of the smart encoder is inspired by the fact that the authorities are collecting the biometric information from people since years and this knowledge could in turn be used to better design the biometric system for people's good. From an information theoretic perspective, we demonstrate that the knowledge of the enrolled biometric template at the smart encoder does provide an advantage so as to enhance the performance of the biometric authentication scheme. As a byproduct, we derive the secrecy capacity of this special wiretap channel with side information and provide two numerical examples showing that strictly larger secrecy capacities can be achieved when both the main channel and the eavesdropper's channel are AWGN channels or BSCs.

Our work also inspires research problems on the extension of Shannon's secrecy system with an error-prone key and the general wiretap channels with side/two-sided (asymmetric) information, where the key or the side information subjects to a certain constraint in any particular metric space.

REFERENCES

A. Pankanti, S. P. and Jain, A. K. (2002). On the individuality of fingerprints. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society.

Cavoukian, A. and Stoianov, A. (2009). Biometric encryption. In *Encyclopedia of Biometrics*. Springer.

Chen, Y. and Vinck, A. J. H. (2008). Wiretap channel with side information. In *IEEE Transactions on Inforamtion Theory*. IEEE Computer Society.

Cohen, G. and Zémor, G. (2004). The wire-tap channel applied to biometrics. In *Proc. Int. Symp. Inf. Theory & its Apps*.

Cover, T. M. and Thomas, J. A. (2005). *Elements of Information Theory*. John Wiley & Sons, Inc., New Jersey, 2nd edition.

J.MacWilliams, F. and Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes*. North-Holland, New York, 1st edition.

Juels, A. and Sudan, M. (2006). A fuzzy vault scheme. In *Design, Codes and Cryptography*. Kluwer Academic Publishers.

Juels, A. and Wattenberg, M. (1999). A fuzzy commitment scheme. In *Proc. 6th ACM Conf. Computer and Comm. Security*. ACM.

Liu, W. and Chen, B. (2007). Wiretap channel with two-sided channel state information. In *The 41 Asilomar Conference on Signals, Systems & Computers*.

Shannon, C. E. (1949). Communication theory of secrecy systems. In *Bell Sys. Tech. J.*

V. B. Balakirsky, A. R. G. and Vinck, A. J. H. (2009). Permutation block coding from biometrical authentication. In *Proc. 7th Int. Conf. on Computer Science and Info. Technologies*.

Vinck, A. J. H. and Balakirsky, V. (2010). Template protection and biometric verification. In *Proc. 6th Asia-Europe Workshop on Information Theory*.

Wyner, A. D. (1975). The wire-tap channel. In *Bell Sys. Tech. J.*

Y. Dodis, R. Ostrovsky, L. R. and Smith, A. (2008). Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *SIAM Journal on Computing*. Society for Industrial and Applied Mathematics (SIAM).

APPENDIX

Fano's Inequality

Let $P_e = \Pr\{\hat{\mathbf{S}}(\mathbf{Z}) \neq \mathbf{S}\}$, where $\hat{\mathbf{S}}(\mathbf{Z})$ is an estimate of $\mathbf{S}$ based on the observation $\mathbf{Z}$. Then

$$H(P_e) + P_e \log|\mathcal{S}| \geq H(\mathbf{S}|\mathbf{Z}), \tag{8}$$

where $\mathcal{S}$ is the set of $\mathbf{S}$ and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$.

Upper Bound on FAR

Fano's inequality gives a lower bound on the 'error' probability of decoding P_e on secret $\mathbf{S}$. However, in the context when we consider FAR, we talk about the probability of 'correct' decoding on $\mathbf{S}$ given the eavesdropper's observation $\mathbf{Z}$, where

$$\text{FAR} = \Pr\{\hat{\mathbf{S}}(\mathbf{Z}) = \mathbf{S}\} = 1 - P_e. \tag{9}$$

The information leakage from $\mathbf{Z}$ on $\mathbf{S}$ is by definition $I(\mathbf{S};\mathbf{Z})$. By Fano's inequality, we have

$$H(P_e) + P_e \log|\mathcal{S}| \geq H(\mathbf{S}|\mathbf{Z}),$$

i.e.,

$$H(1-\text{FAR}) + (1-\text{FAR})\log|\mathcal{S}| \geq H(\mathbf{S}) - I(\mathbf{S};\mathbf{Z}).$$

Note that $H(1-\text{FAR}) \leq 1$. Easy calculation gives us

$$\text{FAR} \leq \frac{1 + \log|\mathcal{S}| + I(\mathbf{S};\mathbf{Z}) - H(\mathbf{S})}{\log|\mathcal{S}|}. \tag{10}$$

In particular, if the secret $\mathbf{S}$ is uniformly distributed over $\mathcal{S}$, then $H(\mathbf{S}) = \log|\mathcal{S}|$. The above upper bound can be simplified to

$$\text{FAR} \leq \frac{1 + I(\mathbf{S};\mathbf{Z})}{\log|\mathcal{S}|}. \tag{11}$$

List of Notations

- $\mathbf{b}$ biometric template at enrollment phase
- $\mathbf{b}'$ biometric template at the authentication phase
- $\mathbf{e}$ variation in biometric readings
- $\mathbf{c}$ a codeword
- R transmission rate
- R_l^n information leakage rate
- P_e^n average probability of error
- R_s^2 secrecy rate for model in Fig. 4
- C_s^1 secrecy capacity for model in Fig. 3
- C_s^2 secrecy capacity for model in Fig. 4
- $(2^{nR}, n)$ a secrecy code of 2^{nR} codewords of length n
- (n,k,d) a linear code, of length n, dimension k and minimum distance d
- $hw(\cdot)$ Hamming weight function, numbers of non-zeros in a sequence
- $\Pr\{\cdot\}$ probability function
- $h(\cdot)$ binary entropy function
- $H(\cdot)$ entropy function, uncertainty associated with a variable
- $I(\cdot)$ mutual information function, amount of information shared by two variables