The wiretap channel applied to biometrics

Gerard Cohen, Gilles Zemor

International Symposium on Information Theory and its Applications, ISITA2004
Parma, Italy, October 10–13, 2004

The wire-tap channel applied to biometrics

Gérard COHEN and Gilles ZEMOR

ENST and CNRS
46 rue Barrault, 75013 PARIS, FRANCE
E-mail: cohen,zemor@enst.fr

Abstract

We present an application of the wire-tap channel to biometrics. Our contribution is threefold. We start with an information-theoretic approach and modelization of the biometric identification problem of the European Project (VIPBOB). Next, we generalize Wyner's coset-coding scheme and prove that linear codes achieves the Shannon-capacity. Finally, we present a solution to the original problem using LDPC codes.

1. INTRODUCTION

A traditional commitment consists of sending or publishing $y = f(\mathbf{b})$ where f is a one-way function and b is a binary vector that is supposed to remain hidden until it is disclosed. Checking that $f(\mathbf{b}) = y$ forbids one from disclosing a vector different from the one that was committed to. The *fuzzy commitment* problem arises when one wishes the protocol to accept not only the original $\mathbf{b}$, but also any vector $\mathbf{b}' = \mathbf{b} + \mathbf{e}$ where $\mathbf{e}$ is a vector of small Hamming weight. This problem appears typically in biometrical contexts when $\mathbf{b}$ encodes, for example, a fingerprint. Successive biometrical measures of the same finger will always tend to differ slightly. To solve this problem Juels and Wattenberg introduced the following fuzzy commitment scheme. Choose a random secret vector $\mathbf{s}$ and encode it as a codeword $\mathbf{c}$ of some fixed, publicly known code C. Then publish (in practice that may mean write on a smartcard)

$$\mathbf{W} = \mathbf{c} + \mathbf{b} \qquad \text{together with } H(\mathbf{c})$$

where $H(\mathbf{c})$ is the image of $\mathbf{c}$ by some cryptographic hash function. When $\mathbf{b}+\mathbf{e}$ is submitted to the protocol, it adds it to $\mathbf{W}$, yielding a noisy version $\mathbf{c}+\mathbf{e}$ of $\mathbf{c}$. The vector $\mathbf{c}+\mathbf{e}$ is then submitted to a decoding algorithm which yields $\mathbf{c}$. Validity is checked by computing $H(\mathbf{c})$ and comparing it to the stored value.

This is essentially the main protocol involved in the European Project (Virtual Identification Pin Based On Biometrics): in this project the protection of the secret vector $\mathbf{s}$ is essential. In the idealized setting of [4], the "biometric" vector $\mathbf{b}$ is assumed to be uniformly distributed among vectors of a given length. In that case, the published vector $\mathbf{W}$ yields no information on the secret codeword $\mathbf{c}$ or the original secret $\mathbf{s}$. However, in practice the distribution of $\mathbf{b}$ may be far from uniform; then $\mathbf{W}$ is liable to leak undesirable partial knowledge of $\mathbf{c}$ to an unauthorized third party (hereafter "the eavesdropper").

Our present contribution is threefold. We start with an information-theoretic approach and model the situation by involving wire-tap channel models. This means that we consider the eavesdropper to have access to a very noisy version of the secret codeword, while the correct protocol yields access to a less noisy version of the secret codeword. We would like to insure no leakage of secret information to the eavesdropper: this problem can be remodelled as that of maximizing the amount of information that can be reliably transmitted through the less noisy "channel" with maximum "equivocation", i.e. insuring that the eavesdropper gets essentially no information on the secret data.

Next, we generalize Wyner's coset coding scheme to the case when both the main channel and the wiretap channel are noisy. We then prove that this scheme, i.e. the use of linear codes, achieves the Shannon-capacity of the system. However, this goes only half-way to providing a practical "zero-leakage" fuzzy committment protocol, because to achieve capacity we need to involve random codes with an unacceptable decoding complexity.

Finally, we present a complete solution to the original problem by making use of LDPC codes, with low decoding complexity. The price to pay is a reduction of the system's original capacity to a subcapacity that we evaluate.

2. FROM BIOMETRICS TO WIRE-TAP

As mentionned in the introduction, the *security problem* that we address is: What information on $\mathbf{s}$ can

be gained from $\mathbf{c}+\mathbf{b}$? This can be viewed as a communication problem, thus amenable to an information-theoretic treatement. To that end we modelize the situation as one involving two channels: one for the legitimate user and one for the pirate. To stress the parallelism with communication problems, we consider the legitimate user as split in two entities: Alice when it enrolls, Bob when it authentifies; following the tradition, the passive attacker is called Eve.

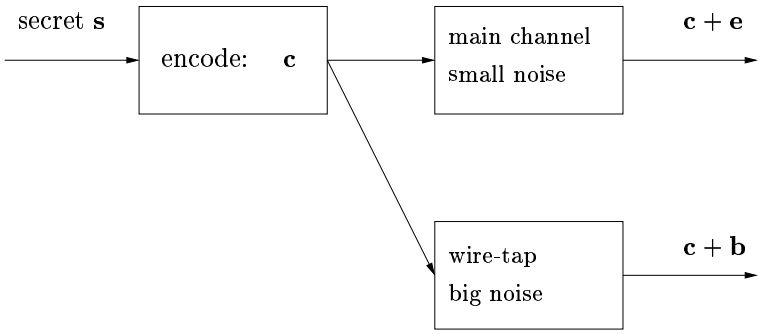
Writing on the card is analogous to transmitting on a channel, reading to receiving.

Bob incurs some noise due to biometric instability, that we denote by $\mathbf{e}$. When Eve intercepts the card, she views the biometric vector as noise.

To summarize:

Two-Channels paradigm

1. Authorized user obtains :
   $\mathbf{W}+\mathbf{b}+\mathbf{e}=\boxed{\mathbf{c}+\mathbf{e}}$.
2. Unauthorized user obtains :
   $\mathbf{W}=\boxed{\mathbf{c}+\mathbf{b}}$.



For random variables R and T, we denote the binary entropy function by $h(y) := -y\log_2 y-(1-y)\log_2(1-y)$, by $H(X)$ the entropy of a random variable X, by $H(R|T)$ the conditional entropy or equivocation of T about R; and by $I(R;T)$ the mutual information between R and T: it measures the (mutual) information leakage.

In the present work we shall focus on binary symmetric channels for Bob and Eve. That is, we assume that both Bob's and Eve's noises are independently distributed over the binary elements, with noise densities p and P respectively, i.e. $(E[w(\mathbf{e})]=np, E[w(\mathbf{b})]=nP)$.

If we perform an analysis per binary element, we get that, without any coding, when A is sent by Alice, B is received by Bob and E is received by Eve,

$I(B;A)=1-h(p)$; $\quad I(E;A)=1-h(P)$.

With an appropriate coding scheme, it was shown by Wyner [6] that a secret S can be conveyed such that

- $I(B,S)=H(S)=h(P)-h(p)$;
- $I(E;S)=0.$, i.e. $H(S|E)=H(S)=h(P)-h(p)$.

Thus, complete knowledge for Bob and complete uncertainty for Eve! The secrecy capacity ([2]) of the scheme, defined as the maximal transmission rate at which total equivocation for the enemy is maintained, is $n(h(P)-h(p))$ for n transmitted binary symbols.

3. NEW COSET-CODING SCHEME

In the special case when $p=0$, it was shown by Wyner that the secrecy capacity can be achieved through a linear coset coding scheme: to transmit the secret $\mathbf{s}$, a vector $\mathbf{x}$ is sent through the channel where $\mathbf{x}$ is randomly chosen among all vectors such that $\mathbf{s}=\mathbf{H}^t\mathbf{x}$ and where $\mathbf{H}$ is a properly chosen $nh(P)\times n$ parity-check matrix of a code C.

3.1. Description

We now make use of two codes C_1 and C_2.

Send $\mathbf{c}^1\in C_1$ (write on card $\mathbf{c}^1+\mathbf{b}$), where: $C_1[n,n(1-h(p))]$ is an np-error correcting code.

Bob gets $\mathbf{c}^1+\mathbf{e}$, and through decoding obtains $\mathbf{c}^1$. Eve gets $\mathbf{c}^1+\mathbf{b}$.

The secret is the syndrome of $\mathbf{c}^1$, $\mathbf{s}=\mathbf{H}^t\mathbf{c}^1$, where $\mathbf{H}=[\mathbf{H}_1;\mathbf{H}_2]$ is the concatenation of 2 parity-check matrices:

- $\mathbf{H}_1$ parity-check matrix of C_1, of size $nh(p)\times n$;
- $\mathbf{H}_2$ parity-check matrix of C_1/C_2, of size $n(h(P)-h(p))\times n$.

Theorem 1 *The above scheme achieves the secrecy capacity per binary element* $(h(P)-h(p))$ *for randomly chosen codes* C_1 *and* C_2.

The proof of this result is presented in the Appendix.

3.2. Protocol

to transmit $\mathbf{s} = (\mathbf{0} : \mathbf{s}_2)$

1. Pick an "easy" vector $\mathbf{y}_2$ with syndrome $\mathbf{s}_2$; For example, with $\mathbf{H}$ in **systematic** form,
   $[\mathbf{H} = I : P]$, I the identity of order $nh(P)$:
   $\mathbf{y}_2 = \Sigma_{i \in supp(\mathbf{s}_2)} e^i$,
   with $\{e^i\}$ the natural basis.
2. Add a **random** $\mathbf{c}_2 \in C_2$ to $\mathbf{y}_2$
3. Transmit $\mathbf{x}_1 := \mathbf{y}_2 + \mathbf{c}_2 \in C_1$.

4. USING LDPC CODES

Problem: To implement the scheme, we need that C_1 should be **easy** to decode (for Bob), e.g., by making use of LDPC codes that we now discuss.

The incertainty for Eve can be rewritten $H(S|E) = \kappa(p) - \kappa(P)$,

where $\kappa(x) = 1 - h(x)$ is the capacity of the BSC with crossover probability x. If the code is LDPC, capacity **cannot** be reached. Instead ([3]), if rows of $\mathbf{H}_1$ have weight w, an upper bound on capacity is given by a **subcapacity**

$\kappa(p, w) := 1 - h(p)/h(p_w)$,
where $p_w := (1 - (1-2p)^w)/2$.
Thus we get:

$$H(S|E) \leq h(P) - h(p)/h(p_w) \leq h(P) - h(p).$$

5. APPENDIX

5.1. Random Coding

Here we prove Theorem 1.

Let C denote a **random linear** $[n, n(1-h(P))]$ code. Almost surely, C is a covering code of radius nP. In other words, considering $\mathbf{H}$, a $nh(P) \times n$ parity-check matrix of C, every syndrome, i.e. vector of $\mathbb{F}^{nh(P)}$, can be written as combination of at most (in fact exactly) nP columns of $\mathbf{H}$.

Let the secret $\mathbf{s}$ be chosen in some way in the syndrome space, i.e. according to some distribution. Let the vector x be chosen **uniformly** among the vectors of syndrome $\mathbf{s}$ (in a given coset of C).

Remark We have, for clarity, described the situation when the noise is binomially distributed (the binary symmetric channel case): but really, what we need to suppose is only that there is a set T of typical noise vectors of cardinality $|T| = 2^{nh(P)}$, each vector $\mathbf{b}$ occuring with probability $2^{-nh(P)}$.

We have the immediate

Lemma 1 *For* $\mathbf{b}, \mathbf{s}$ *fixed,* $\mathbf{H}$ *uniformly distributed:* $Pr\{\mathbf{H}^t\mathbf{b} = \mathbf{s}\} = 2^{-nh(P)}$.

Set $X_{\mathbf{b},\mathbf{s}} = 1$ if $\mathbf{H}^t\mathbf{b} = \mathbf{s}$,
$X_{\mathbf{b},\mathbf{s}} = 0$ otherwise;
also $X_{\mathbf{s}} := \Sigma_{\mathbf{b} \in T} X_{\mathbf{b},\mathbf{s}}$. Then

Lemma 2 $E[X_{\mathbf{b},\mathbf{s}}] = 2^{-nh(P)}$, $E[X_{\mathbf{s}}] = 1$.

We now need a technical result, whose proof is omitted here.

Lemma 3 *For* m *integer,*
$E[X_{\mathbf{s}}^m] \leq (2m)^m$.

Then, we invoke the "Markov Inequality of order m", stating that for a positive random variable Y and real number λ: $Pr\{Y > \lambda\} \leq E[Y^m]/\lambda^m$.

We apply it to $Y = X_{\mathbf{s}}, \lambda = 2^{n\alpha}$, which yields:
$Pr\{X_{\mathbf{s}} > 2^{n\alpha}\} \leq (2m)^m / 2^{n\alpha m}$.

Finally, we choose $m \geq 2/\alpha$ to get the following theorem and corollary, which in essence state that syndrome distribution is uniform among typical noise vectors.

Theorem 2 *For any* $0 < \alpha < 1$, *for any* $\mathbf{s}$*:*
$Pr\{X_{\mathbf{s}} > 2^{n\alpha}\} \to 0$.
Moreover,
$Pr\{\exists \mathbf{s} : X_{\mathbf{s}} > 2^{n\alpha}\} \to 0$.

Corollary 1 *For a random code* $[n, n(1-h(P))]$
(in fact for almost all codes)
$\forall \mathbf{b} \in T, \forall \mathbf{s} Pr\{\mathbf{H}^t\mathbf{b} = \mathbf{s}\} \leq 2^{-n(h(P)-\alpha)}$.

Remark. Defining the **min-entropy** ([5]) of a random variable R as:
$H_\infty := \text{Max}\ \{i : \forall r : Pr\{R = r\} \leq 2^{-i}\}$,
we get (per binary element):

$$H(S|E) \geq H_\infty(S|E) \geq h(P) - \alpha.$$

The consequence of Theorem 2 is that with this scheme, the distribution of $\mathbf{H}^t\mathbf{b}$ (the syndrome of $\mathbf{b}$) is uniform. Once again, this holds under very general conditions on the noise (essentially that typicallity

can be defined, which encompasses all classical additive channels).

Informally, this can be reinterpreted as saying that the eavesdropper is submitted to a one-time pad in the syndrome space: hence her uncertainty on the secret $\mathbf{s} = [0 : \mathbf{s}_2]$ is the same before and after reception, namely $H(\mathbf{s}_2) = n(h(P) - h(p))$.

More formally, start by noticing that

$H(\mathbf{s} \mid \mathbf{x}+\mathbf{b}) - H(\mathbf{s} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b}) \leq 0$.

This is because, since $\mathbf{s}+\mathbf{H}^t\mathbf{b}$ is a function of $\mathbf{x}+\mathbf{b}$, knowledge of $\mathbf{x}+\mathbf{b}$ can only yield more knowledge (and less uncertainty) than $\mathbf{s}+\mathbf{H}^t\mathbf{b}$. Let us now prove the reverse inequality: we have

$$
\begin{aligned}
& H(\mathbf{s} \mid \mathbf{x}+\mathbf{b}) - H(\mathbf{s} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
= \;& H(\mathbf{s}, \mathbf{x}+\mathbf{b}) - H(\mathbf{x}+\mathbf{b}) \\
& -H(\mathbf{s}, \mathbf{s}+\mathbf{H}^t\mathbf{b}) + H(\mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
= \;& H(\mathbf{s}, \mathbf{x}+\mathbf{b}, \mathbf{s}+\mathbf{H}^t\mathbf{b}) - H(\mathbf{s}, \mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
& -[H(\mathbf{x}+\mathbf{b}, \mathbf{s}+\mathbf{H}^t\mathbf{b}) - H(\mathbf{s}+\mathbf{H}^t\mathbf{b})] \\
= \;& H(\mathbf{x}+\mathbf{b} \mid \mathbf{s}, \mathbf{s}+\mathbf{H}^t\mathbf{b}) - H(\mathbf{x}+\mathbf{b} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
\geq \;& H(\mathbf{x} \mid \mathbf{s}, \mathbf{b}) - H(\mathbf{x}+\mathbf{b} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
= \;& H(\mathbf{x} \mid \mathbf{s}) - H(\mathbf{x}+\mathbf{b} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b}) \\
\geq \;& 0,
\end{aligned}
$$

where the last inequality is due to $\mathbf{x}$ being uniformly distributed among vectors with syndrome $\mathbf{s}$, hence the maximality of $H(\mathbf{x} \mid \mathbf{s})$.

We have therefore proved that

$H(\mathbf{s} \mid \mathbf{x}+\mathbf{b}) = H(\mathbf{s} \mid \mathbf{s}+\mathbf{H}^t\mathbf{b})$,

meaning that there is no advantage for the eavesdropper in possessing $\mathbf{x}+\mathbf{b}$ on top of its syndrome. Since in the syndrome space $\mathbf{s}$ is submitted to a one-time pad we obtain Theorem 1. Note that we did not need to suppose anything on the distribution of $\mathbf{s}$.

5.2. Practical Decoding

In the preceding section the parity check matrix $\mathbf{H}_1$ was chosen randomly, with uniform distribution among all possible matrices of a given size. This gives a random code C_1 for which no effective decoding algorithm is known and makes the coset coding scheme pretty much impractical for the legitimate user Bob. To make it practical let us now suppose that C_1 is a code for which a decoding algorithm exists, for example an LDPC code. The price to pay for this will be that its redundancy $h(p)n$ will have to be increased to some value $r_1 n$.

Recall from section 3.2 that the secret syndrome $\mathbf{s}$ is chosen of the form $\mathbf{s} = (0 : \mathbf{s}_2)$. The syndrome computed from the received vector $\mathbf{s} + \mathbf{H}^t\mathbf{b}$ can be decomposed as $(\mathbf{H}_1\,{}^t\mathbf{b} : \mathbf{s}_2 + \mathbf{H}_2\,{}^t\mathbf{b})$ where $\mathbf{b}$ is the noise vector. By the same argument as before, we may suppose that the eavesdropper makes only use of the computed syndrome since knowledge of $\mathbf{x}+\mathbf{b}$ does not give him any extra information. However, since we have lost control of the nature of the matrix $\mathbf{H}_1$, we may not hope to argue as before that $\mathbf{H}^t\mathbf{b}$ has a uniform distribution. To fix this we focus on the distribution of $\mathbf{H}_2\,{}^t\mathbf{b}$ conditionnal on $\mathbf{H}_1\,{}^t\mathbf{b}$. Using Markov Inequality, we have that only an exponentially small proportion of the subsets $T_i \subset T$ of noise vectors with given $\mathbf{H}_1\,{}^t\mathbf{b} = i$ will have cardinality exponentially smaller than the mean value $E[|T_i|] = 2^{n(H(P)-r_1)}$. Since $n(H(P) - r_1)$ is exactly the length of the subsyndrome $\mathbf{s}_2$, Theorem 2 gives us that the distribution of $\mathbf{H}_2\,{}^t\mathbf{b}$ conditionnal on $\mathbf{H}_1\,{}^t\mathbf{b}$ is again uniform. Therefore the subvector $\mathbf{s}_2$ is again submitted to a one-time pad with perfect secrecy.

To summarize, the only loss incurred, when moving from the existential approach to the actually implementable codes, is in the size of the secret, but not in the **security per binary element**, which remains **maximal**.

Thus, referring to Section 4, Eve's global equivocation about the secret is

$nH(S|E) = n(h(P) - r_1)$,
with $r_1 \geq h(p)/h(p_w)$.

5.3. Generalizations

The previous analysis extends *mutatis mutandis* to

- Parameters p_i and P_i depending on the location:
  $nH(S|E) \approx \Sigma_i(h(P_i) - h(p_i))$.
- Correlated noise. If the noise is additive, with average entropies per bit for Bob and Eve
  $h(\theta)$ and $h(\Theta)$, then
  $H(S|E) = h(\Theta) - h(\theta)$.

Comments. To see why the last extension holds, denote $T(\mathbf{0})$ the set of **typical** noise vectors around $\mathbf{0}$, i.e. of probability 1 conditional to $\mathbf{0}$ being emitted.

The additivity property means that

$T(\mathbf{c}) = \mathbf{c} + T(\mathbf{0})$.

The existence of good linear coverings by **tiles** $T(\mathbf{0})$ ([1]) implies constructed maximum incertainty for Eve:

Setting for Bob and Eve respectively
$|T_B(\mathbf{0})| = 2^{n\theta}$; $|T_E(\mathbf{0})| = 2^{n\Theta}$, we have:
$H(S|E) = h(\Theta) - H(\theta)$.

References

[1] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, "Covering codes", North-Holland Mathematical Library 54 (1997).

[2] I. Csiszár and J. Körner "Information Theory", Academic Press (1982).

[3] R. Gallager "Low density parity-check codes", Cambridge MIT Press (1963).

[4] A. Juels and M. Wattenberg, "A fuzzy commitment scheme", in 6th ACM Conference on Computer and Communications Security, pp. 28–36, ACM Press, 1999.

[5] N. Nisan and D. Zuckerman "Randomness is linear in space", J. Computer and System Sciences 52, 43-52 (1996).

[6] A. Wyner "The wire-tap channel", BSTJ 54 , 1355-1387 (1975).