Wire-Tap Channel System and Dedicated Coding

Miodrag Mihaljevic
RCIS-AIST, Tokyo

Spring Working Meeting of Japan-India Project

Indian Statistical Institute, Kolkata, April 14, 2009
Roadmap

• Introduction – Certain References
• Motivations and the Goals
• Wire-Tap Channel Coding
• Coding for Binary Erasures Wire-Tap Channel
• Applications of Wire-Tap Channel Coding for Design of Stream Ciphers
I. Introduction

Certain References on Wire-Tap Channel

- A different approach for achieving secrecy of communication based on the noise has been reported by Wyner in 1975 assuming that the channel between the legitimate parties is with a lower noise in comparison with the channel via which a wire-tapper has access to the ciphertext.

- The proposed method does not require any secret. It is based on a specific coding scheme which provides a reliably communications within the legitimate parties and prevents, at the same time, the wire-tapper from learning the communication's contents.
Some Recent References


II. Motivations and Goals for Involvement of Wire-Tap Channel Coding into Certain Crypto Techniques

Trade-Off between Security and Communications & Implementations Overheads
Main Goals

• A framework for design of stream ciphers which provides opportunity for design the security as high as possible based on the employed secret key, i.e. complexity of recovering the key as close as possible to $O(2^K)$

• A trade-off between the security and the communications rate: Increase the security up to the upper limit at the expense of a moderate decrease of the communications rate.
The Main Underlying Ideas

• Employ physical noise which an attacker must face, in order to strengthen the stream cipher.

• Strengthen the stream cipher employing a dedicated encoding following the homophonic or wire-tap channel encoding approaches.
Notes (1): Novel Paradigm

- Traditional stream ciphers do not include any randomness: Basically, they are based on the deterministic operations which expand a short secret seed into a long pseudorandom sequence.
- This talk proposes an alternative approach yielding a novel paradigm for design of stream ciphers.
- The proposed framework employs a dedicated coding and a deliberate noise which, assuming the appropriate code and noise level, at the attacker's side provides increased confusion up to the limit determined by the secret key length.
- Decoding complexities with and without the secret key are extremely different.
In order to achieve the main security goal, the proposed stream ciphering approach includes the following two encoding schemes with impacts on the communications overhead:

• error-correction encoding of the messages;
• dedicated homophonic/wire-tap channel coding which performs expansion of the initial ciphertext.

Both of these issues imply the communications overhead: Accordingly, the proposed stream ciphers framework includes certain trade-off between the security and the communications overhead which in a number of scenarios can be considered as very appropriate.
III. Wire-Tap Channel

A. D. Wyner, “The wire-tap channel”,
Coding Strategy for the Wire-Tap Channel

• Goal of encoding paradigm for the wire-tap channel is to make the noisy data available to Eve (across the wire tap channel) useless and achieving this goal is based on adding the randomness in encoding algorithm.
Groups of the codewords: Same symbol denote different codewords belonging to the same group.
Dedicated Wire-Tap Channel Coding

Coding Method and Selection of the Code
Wire-Tap Channel Coding Preliminaries: Standard Array and Cosets

M = $2^k$

<table>
<thead>
<tr>
<th>0</th>
<th>$C_1$</th>
<th>$C_j$</th>
<th>$C_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_j$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_N$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coset $e + C$

$C_i + e_j$
We consider a generic approach for wire-tap channel coding as follows.

- To transmit $m$-bit message we first select a $(n, k)$ code $C$ such that $m \leq n - k$.

- Out of the $2^{n-k}$ cosets of $C$, we choose $2^m$ cosets and let each message correspond to a chosen coset.

- The selection of the cosets is done in a linear fashion as follows:

  (a) Suppose $G$ is a generator matrix for $C$ with rows $g_1, g_2, \ldots, g_k$.

  (b) We select $m$ linearly independent vectors $h_1, h_2, \ldots, h_m$, from $\{0, 1\}^n \setminus C$.

  (c) The coset corresponding to a $m$-bit message $s = [s_1, s_2, \ldots, s_m]$ is determined as follows:

  $$ s \rightarrow s_1h_1 \oplus s_2h_2 \oplus \cdots \oplus s_mh_m \oplus C. \quad (1) $$
Coding Method (2)

The above correspondence is deterministic, but the encoding has a random component in the selection of the employed codeword. The transmitted word $c$ is specified as follows:

$$c = s_1h_1 \oplus s_2h_2 \oplus \ldots \oplus s_m h_m \oplus u_1g_1 \oplus u_2g_2 \oplus \ldots \oplus u_k g_k$$  \hspace{1cm} (1)

where $u = [u_1, u_2, \ldots, u_k]$ is an uniformly random $k$-bit vector and in a particular case $k = n - m$.

The overall encoding operation can be described as the following. Let $G^*$ be the $m \times n$ matrix with rows $h_1, h_2, \ldots, h_m$. Then

$$c = [su] \begin{bmatrix} G^* \\ G \end{bmatrix}$$  \hspace{1cm} (2)
Selection of the Code

For an arbitrary $m$-bit message $S = s$, the transmitted word belongs to $sG^* \oplus C$. Since the cosets of $C$ cover the entire space $\{0,1\}^n$, the attacker receives vector $Z$ which belongs to some coset of $C$ for example $rG^* \oplus C$. If $e$ denotes the error vector introduced by the wire-tapper’s BSC$(p)$, we have for $1 \leq i \leq 2^k$:

$$\text{Prob}\{Z \in rG^* \oplus C\} = \text{Prob}\{e \in (u \oplus s)G^* \oplus C\} = \text{Prob}\{e \in w \oplus C\}, \quad (1)$$

for some $n$-tuple $w$. Accordingly, the following criterion for selecting the code $C$ provides security of the message: Select $C$ such that for any $n$-tuple $w$, the following is valid:

$$\text{Prob}\{e \in w \oplus C\} \to 2^{-k}, \text{ as } n \to \infty. \quad (2)$$

The above condition in conjunction with (10) implies that for an attacker it is equally likely to find $Z$ in any coset of $C$ given any message $S$. Note that, assuming all $S = s$ are equally likely a priori, $\text{Prob}\{Z \in rG^* \oplus C\}$ is independent of $r$: Hence,

$$\text{Prob}\{S = s | Z \in rG^* \oplus C\} \to 2^{-k}, \quad (3)$$

implies the security.
Implications of Capacity Approaching Codes over the Wire-Tap Channel

- \( I(Z; UX) = I(U; Z) + I(X; Z|U) \)
- \( = I(X; Z) + I(U; Z|X) \)
- Since U -> X -> Z is a Markov chain, \( I(U; Z|X)=0 \)

Therefore

\[ I(U;Z)/n = I(X;Z)/n - I(X; Z|U)/n < C - (C-e) = e \]
IV. Encoding for BEC Wiretapper's Channel and Noiseless Main Channel
We consider a scenario where a wire-tapper can observe the ciphertext via a binary erasure channel (BEC) where a bit erasure appears with the probability $1 - \epsilon$.

The transmitted $n$-tuple is denoted by the random variable $\mathbf{X} = [X_1, X_2, ..., X_n]$. Note that the message $S$ can be seen as a syndrome of $C$ with respect to a carefully constructed $k \times n$ parity-check matrix $\mathbf{H}$.

Since the channel between Alice and Bob is error-free, Bob finds the message as follows: $\mathbf{S} = \mathbf{H}\mathbf{X}^T$. 
Wire-Tap Chanel Coding Preliminaries: Standard Array and Cosets

\[ N = 2^{n-k-1} \]

\[ C_1 \quad \ldots \quad C_i \quad \ldots \quad C_M \]

Coset \( e + C \)

\[ C_i + e_j \]
The eavesdropper learns $X_i$ with probability $\epsilon$. That is, the random variable $Z = [Z_1, Z_2, \ldots, Z_n]$ is such that $Z_i = X_i$ with probability $\epsilon$, and $Z_i = ?$ (unknown or erasure) with probability $1 - \epsilon$.

If a coset of $C$ contains at least one vector that agrees with $z \in \{0, 1, ?\}^n$ in the unerased positions, the \textbf{coset is consistent} with $z$.

Each consistent coset corresponds to a possible message for the eavesdropper.
Let \( v \) be a vector consistent with \( z \) in the coset \( v + C \).

Let \( S \) be the set of all vectors in \( v + C \) consistent with \( z \).

Then, \( v + S \) is the set of all vectors in \( C \) with zeros in the positions revealed in \( z \). That is,

\[
v + S = \{ u \in C : u_i = 0 \text{ whenever } z_i \neq ? \}.
\]
Since $|S| = |\mathbf{v} + S|$, the number of vectors consistent with $\mathbf{z}$ in each consistent coset is a constant.

Let $N(C, \mathbf{z})$ denote the total number of cosets of $C$ consistent with $\mathbf{z}$.

Since each message is equally likely a priori, we get

$$H(S|Z = \mathbf{z}) = \log_2 N(C, \mathbf{z}).$$
For an \((n, n-k)\) code \(C\), the maximum possible value for \(N(C, z)\) is the total number of cosets which is equal to \(2^k\).

If \(N(C, z) = 2^k\), we say that \(z\) is secured by \(C\) since the eavesdroppers

\[
\text{Prob}\{S = s | Z = z\} = 1/2^k
\]

for every possible message \(s\).
Groups of the codewords: *Same symbol denote different codewords belonging to the same group*

Codewords and N-dim Sphere
Theorem, [1]. Let an \((n, n-k)\) code \(C\) have a generator matrix \(G = [a_1, ..., a_n]\), where \(a_i\) is the \(i\)-th column of \(G\). Consider an instance of the eavesdroppers observation \(z \in \{0, 1, ?\}^n\) with \(\mu\) unerased positions given by \(\{i : z_i \neq ?\} = \{i_1, i_2, ..., i_\mu\}\). \(z\) is secured by \(C\) if the matrix \(G_\mu = [a_{i_1} a_{i_2} ... a_{i_\mu}]\) has rank \(\mu\).

Dedicated LDPC for Wire-Tap Channel Coding

- Just to mention -
V. Applications of Wire-Tap Channel Coding for Design of Cryptographic Primitives
Very Recent References (1)


Some Earlier Results on Crypto&Coding

(there is a number of other results achieved in the period 2005-2008)


Some Previous Results on Randomized Encryption


Algebraic Model of Stream Cipher II under Chosen Plaintext Attack

**Corollary 1.** Under the chosen plaintext attack which for each \( t \) implies \( b_t = 0 \) (i.e. the all zeros vector), Proposition 1 implies:

\[
\mathbf{z}_t = \mathbf{q}_t \mathbf{S} \oplus \mathbf{\nu}_t,
\]

(1)

where

\[
\mathbf{q}_t = (\bigoplus_{i=1}^{m} x_i h_i \oplus y_t) \mathbf{S}^{-1}, \quad \mathbf{\nu}_t = (\bigoplus_{i=1}^{n-m} u_i g_i) \oplus (\bigoplus_{i=1}^{m} v_i h_i),
\]

(2)

and where \( \mathbf{S} \) is an \( n \times n \) binary matrix determined by the length \( k \) binary secret key \( \mathbf{k} \), and \( \mathbf{S}^{-1} \) is its inverse.

**Assumption 1.** For any \( t = 1, 2, ..., \), \( \mathbf{\nu}_t \leftarrow \text{Ber}_{n, \eta} \), where \( \eta \) is the parameter.
A Statement on Stream Cipher II Security

**Theorem 1.** Assume there is an adversary $\mathcal{A}$, running in time $T$, and attacking the Stream cipher II specified by Corollary 1 and Assumption 1 with parameters $(\ell, m, k, n, \eta)$, $k = n$, in the sense of IND with advantage $\delta$ by making at most $q$ queries to the encryption oracle. Then there is an algorithm $\mathcal{M}$ making $O(q)$ oracle queries, running in time $O(T)$, and such that

$$\left| \Pr \left[ s \leftarrow \{0, 1\}^k : \mathcal{M}^\Pi_{s, \eta}(1^k) = 1 \right] - \Pr \left[ \mathcal{M}^{U_{k+1}}(1^k) = 1 \right] \right| \geq \frac{\delta}{n}.$$  

(1)
Thank You Very Much for the Attention,

and

QUESTIONS Please!