On Compression of Data Encrypted with Block Ciphers

Demijan Klinč*  Carmit Hazay†  Ashish Jagmohan‡
Hugo Krawczyk‡  Tal Rabin‡

*Georgia Institute of Technology
† Bar-Ilan University
‡ IBM T.J. Watson Research Labs

March 17th, 2009
**Problem Statement**

**Conventional System**

Data is first compressed, then encrypted.
Compression and Encryption in Reverse Order

Question: Can compression and encryption be reversed?
Some Thinkable Applications

- **Sensor Networks**
  - think of a network of low-cost sensors; the sensors need to encrypt data, but do not want to compress to save resources/hardware cost
  - network operator wants to compress data to maximize the utilization of its resources, but does not have access to the key

- **Third party storage**
  - a storage provider wants to compress data to minimize storage space, but does not have access to the key
Compression and Encryption in Reverse Order

- for stream ciphers the answer is yes
- we will be interested in the scenario where the encryption scheme is based on a block cipher, like DES or AES
- to our knowledge, this is still an open problem
1. Problem Statement

2. Preliminaries

3. Compressing Block Ciphers

4. Simulation Results
Source Coding with Decoder Side-information

- $X, Y$: random variables over a finite alphabet with a joint probability distribution $P_{XY}$
- problem at hand: losslessly compress $X$ with $Y$ known only to the decoder
- asymptotically in block length, rates arbitrary close to $H(X/Y)$ are achievable [Slepian 1973]
Compressing Stream Ciphers

- Slepian-Wolf coding problem [Johnson, 2004]
- $E_K(X)$ is cast as source, $K$ is side-information
- joint probability distribution of the source and side-information is governed by the statistics of the source $X$
Compressing Stream Ciphers

- decoder knows $K$ and source statistics
- $K$ is viewed as an observation of $E_K(X)$ over a virtual correlation channel governed by the statistics of the source
- compression rate $H(X)$ is asymptotically achievable and information-theoretic security is preserved
stream ciphers are not the only form of encryption in practice
the prevalent methods of encryption in practice are based on block ciphers
a desirable extension of the technique would be to conventional encryption schemes, such as AES
problem: there is no known correlation between $E_K(X)$ and $K$
Electronic Code Book

\[
\begin{align*}
E_K(X_1) & \quad E_K(X_2) & \quad E_K(X_n) \\
\text{Block cipher} & \quad \text{Block cipher} & \quad \text{Block cipher}
\end{align*}
\]
Modes of Operation

Original

Modes of Operation

Original

Encrypted in ECB mode

Modes of Operation

Original

Encrypted in ECB mode

Encrypted in other modes of operation

Cipher Block Chaining (CBC) Mode

\[ \text{Block cipher} \]

\[ K \]

\[ \text{IV} \]

\[ E_K(\tilde{X}_1) \]

\[ \text{Block cipher} \]

\[ E_K(\tilde{X}_2) \]

\[ \text{Block cipher} \]

\[ E_K(\tilde{X}_n) \]
Block Ciphers Can Be Compressed

- the correlation between the randomization vector $E_K(X_{i-1})$ and $X_i$ is easier to characterize and can be exploited for compression

notice: last block is left uncompressed, while IV is compressed
Joint Decompression and Decoding

\[
\begin{align*}
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{n-1}) \\
&C(E_K(X_{n-1})) \\
&\vdots \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{n-2}) \\
&C(E_K(X_{n-2})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_n) \\
&C(E_K(X_n)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_1) \\
&C(E_K(X_1)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_2) \\
&C(E_K(X_2)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_3) \\
&C(E_K(X_3)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_4) \\
&C(E_K(X_4)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_5) \\
&C(E_K(X_5)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_6) \\
&C(E_K(X_6)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_7) \\
&C(E_K(X_7)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_8) \\
&C(E_K(X_8)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_9) \\
&C(E_K(X_9)) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{10}) \\
&C(E_K(X_{10})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{11}) \\
&C(E_K(X_{11})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{12}) \\
&C(E_K(X_{12})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{13}) \\
&C(E_K(X_{13})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{14}) \\
&C(E_K(X_{14})) \\
&\text{Decryption} \\
&\text{Slepian-Wolf Decoder} \\
&E_K(\tilde{X}_{15}) \\
&C(E_K(X_{15}))
\end{align*}
\]
Joint Decompression and Decoding

\[ \begin{align*}
    X_1 & \xrightarrow{K} \text{Decryption} & \tilde{X}_1 & \xrightarrow{E_K(\tilde{X}_{n-2})} \text{Slepian-Wolf Decoder} \\
    X_2 & \xrightarrow{K} \text{Decryption} & \tilde{X}_2 & \xrightarrow{E_K(\tilde{X}_{n-3})} \text{Slepian-Wolf Decoder} \\
    \vdots & & \vdots & & \vdots \\
    X_n & \xrightarrow{K} \text{Decryption} & \tilde{X}_n & \xrightarrow{E_K(\tilde{X}_{n-1})} \text{Slepian-Wolf Decoder} \\
    IV & \xrightarrow{K} \text{Decryption} & \tilde{X}_1 & \xrightarrow{E_K(\tilde{X}_n)} \text{Slepian-Wolf Decoder} \\
\end{align*} \]
Joint Decompression and Decoding

\[ \begin{align*}
X_n & \xrightarrow{K} E_K(X_{n-1}) \\
& \xrightarrow{K} E_K(X_{n-2}) \\
& \vdots \\
& \xrightarrow{K} E_K(\tilde{X}_1)
\end{align*} \]

where \( X_i \) are variables, \( K \) is a key, and \( E_K \) denotes encryption with key \( K \). 

Slepian-Wolf Decoder

Compressing Block Ciphers

Klinc et al.

Compression of Block Ciphers

March 17th, 2009
Joint Decompression and Decoding

\[ E_K(X_n) \sim C(E_K(X_{n-1})) \sim C(E_K(X_{n-2})) \sim C(IV) \]

\[ E_K(X_{n-1}) \sim E_K(X_{n-2}) \sim X_{n-1} \sim X_n \sim X_1 \]

Slepian-Wolf Decoder

Decryption

K

\[ C(IV) \]

\[ C(E_K(X_{n-2})) \]

\[ C(E_K(X_{n-1})) \]

\[ E_K(X_n) \]
Compressing Block Ciphers

Compression Factor

- let \( \{ C_{m,R}, D_{m,R} \} \) denote an order \( m \) Slepian-Wolf code with compression rate \( R \)
- compressor \( C_{m,R} : \mathcal{X}^m \to \{1, \ldots, 2^{mR}\} \)
- decompressor \( D_{m,R} : \{1, \ldots, 2^{mR}\} \times \mathcal{X}^m \to \mathcal{X}^m \)
- compression factor:

\[
\frac{(n + 1) \cdot m \cdot \log |\mathcal{X}|}{n \cdot m \cdot R + m \cdot \log |\mathcal{X}|} \approx \frac{\log |\mathcal{X}|}{R}
\]
for large $m$

$$R = H(E_K(\tilde{X}_{i-1})|\tilde{X}_i) = H(E_K(\tilde{X}_{i-1})|E_K(\tilde{X}_{i-1}) \oplus X_i)$$
$$= H(E_K(\tilde{X}_{i-1}) \oplus X_i|E_K(\tilde{X}_{i-1})) + H(E_K(\tilde{X}_{i-1})) - H(E_K(\tilde{X}_{i-1}) \oplus X_i)$$
$$= H(X_i) + H(E_K(\tilde{X}_{i-1})) - H(E_K(\tilde{X}_{i-1}) \oplus X_i)$$
$$\leq H(X_i).$$

equality happens when $E_K(\tilde{X}_{i-1})$ is uniformly distributed

notice: this method is asymptotically optimal; no performance loss due to the uncompressed last block, as the IV can be compressed
Short Block Lengths

- compression efficiency depends on the performance of underlying Slepian-Wolf codes
- Slepian-Wolf compression approaches entropy with speed \( O(\sqrt{\frac{\log n}{n}}) \) [He 2006]
- problem: many contemporary block ciphers have short block sizes (AES: 128 bits)
- in the proposed approach the block length of Slepian-Wolf codes must be equal to the block size of a block cipher
irregular LDPC codes were used in our performance evaluation

<table>
<thead>
<tr>
<th>$p$</th>
<th>Target FER</th>
<th>Compression Rate</th>
<th>Source Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.026</td>
<td>$10^{-3}$</td>
<td>0.50</td>
<td>0.1739</td>
</tr>
<tr>
<td>0.018</td>
<td>$10^{-4}$</td>
<td>0.50</td>
<td>0.1301</td>
</tr>
<tr>
<td>0.068</td>
<td>$10^{-3}$</td>
<td>0.75</td>
<td>0.3584</td>
</tr>
<tr>
<td>0.054</td>
<td>$10^{-4}$</td>
<td>0.75</td>
<td>0.3032</td>
</tr>
</tbody>
</table>

Table: Attainable compression rates for $m = 128$ bits.
irregular LDPC codes were used in our performance evaluation

<table>
<thead>
<tr>
<th>$p$</th>
<th>Target FER</th>
<th>Compression Rate</th>
<th>Source Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.026</td>
<td>$10^{-3}$</td>
<td>0.50</td>
<td>0.1739</td>
</tr>
<tr>
<td>0.018</td>
<td>$10^{-4}$</td>
<td>0.50</td>
<td>0.1301</td>
</tr>
<tr>
<td>0.068</td>
<td>$10^{-3}$</td>
<td>0.75</td>
<td>0.3584</td>
</tr>
<tr>
<td>0.054</td>
<td>$10^{-4}$</td>
<td>0.75</td>
<td>0.3032</td>
</tr>
</tbody>
</table>

Table: Attainable compression rates for $m = 128$ bits.
irregular LDPC codes were used in our performance evaluation

<table>
<thead>
<tr>
<th>$p$</th>
<th>Target FER</th>
<th>Compression Rate</th>
<th>Source Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.058</td>
<td>$10^{-3}$</td>
<td>0.50</td>
<td>0.3195</td>
</tr>
<tr>
<td>0.048</td>
<td>$10^{-4}$</td>
<td>0.50</td>
<td>0.2778</td>
</tr>
<tr>
<td>0.134</td>
<td>$10^{-3}$</td>
<td>0.75</td>
<td>0.5710</td>
</tr>
<tr>
<td>0.126</td>
<td>$10^{-4}$</td>
<td>0.75</td>
<td>0.5464</td>
</tr>
</tbody>
</table>

*Table:* Attainable compression rates for $m = 1024$ bits.
irregular LDPC codes were used in our performance evaluation

<table>
<thead>
<tr>
<th>$p$</th>
<th>Target FER</th>
<th>Compression Rate</th>
<th>Source Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.058</td>
<td>$10^{-3}$</td>
<td>0.50</td>
<td>0.3195</td>
</tr>
<tr>
<td>0.048</td>
<td>$10^{-4}$</td>
<td>0.50</td>
<td>0.2778</td>
</tr>
<tr>
<td>0.134</td>
<td>$10^{-3}$</td>
<td>0.75</td>
<td>0.5710</td>
</tr>
<tr>
<td>0.126</td>
<td>$10^{-4}$</td>
<td>0.75</td>
<td>0.5464</td>
</tr>
</tbody>
</table>

**Table:** Attainable compression rates for $m = 1024$ bits.
Concluding Remarks

- data encrypted with block ciphers are practically compressible, when chaining modes are employed.
- Notable compression factors were demonstrated with binary memoryless sources.
- Short block sizes limit the performance, but that could change in the future.