

On Compression of Data Encrypted with Block Ciphers

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

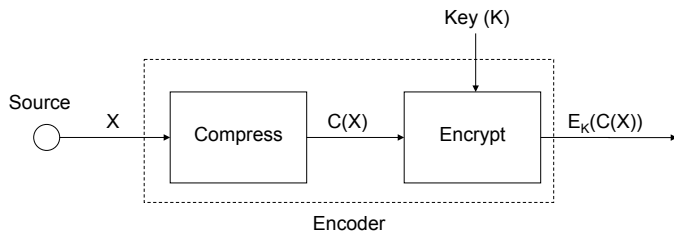
*Georgia Institute of Technology

† Bar-Ilan University

‡IBM T.J. Watson Research Labs

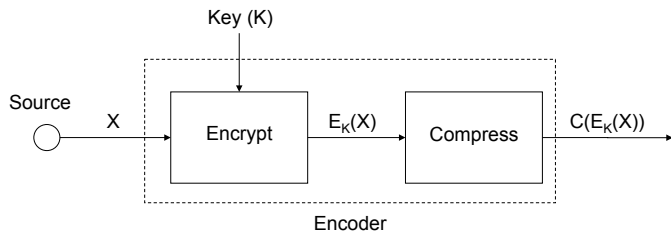
March 17th, 2009

Conventional System



Data is first compressed, then encrypted.

Compression and Encryption in Reverse Order



Question: Can compression and encryption be reversed ?

Some Thinkable Applications

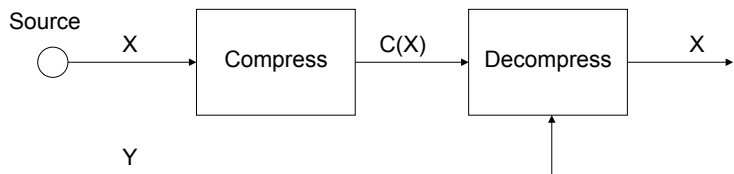
- Sensor Networks
 - think of a network of low-cost sensors; the sensors need to encrypt data, but do not want to compress to save resources/hardware cost
 - network operator wants to compress data to maximize the utilization of its resources, but does not have access to the key
- Third party storage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

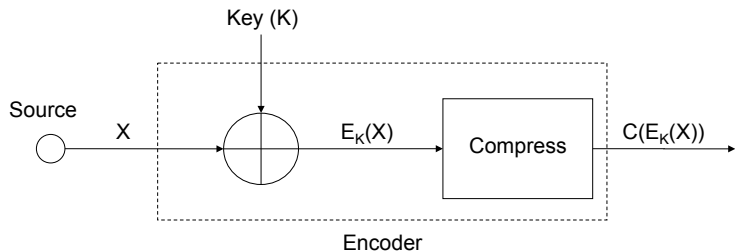
- 1 Problem Statement
- 2 Preliminaries
- 3 Compressing Block Ciphers
- 4 Simulation Results

Source Coding with Decoder Side-information



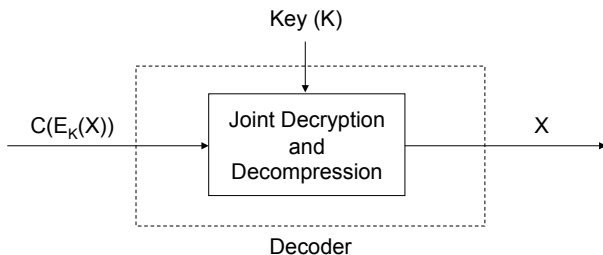
- X, Y : random variables over a finite alphabet with a joint probability distribution P_{XY}
- problem at hand: losslessly compress X with Y known only to the decoder
- asymptotically in block length, rates arbitrary close to $H(X/Y)$ are achievable [Slepian 1973]

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

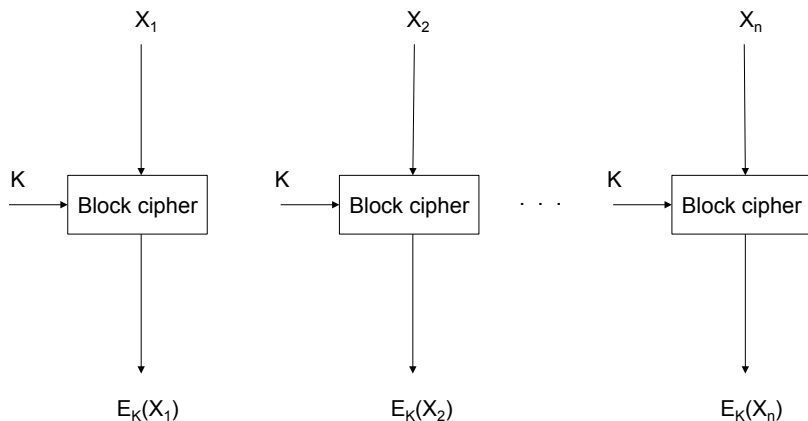


- decoder knows K and source statistics
- K is viewed as an observation of $E_K(X)$ over a virtual correlation channel governed by the statistics of the source
- compression rate $H(X)$ is asymptotically achievable and information-theoretic security is preserved

Block Ciphers

- 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



Modes of Operation



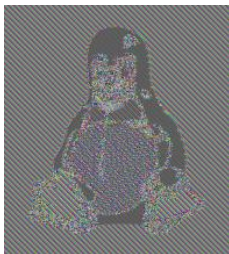
Original

Source: Wikipedia

Modes of Operation



Original



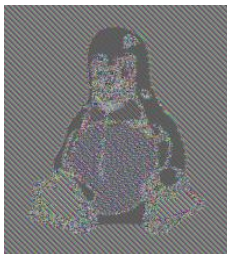
Encrypted in ECB
mode

Source: Wikipedia

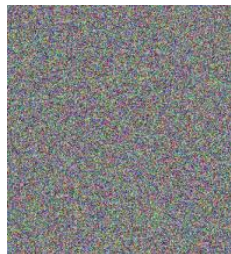
Modes of Operation



Original



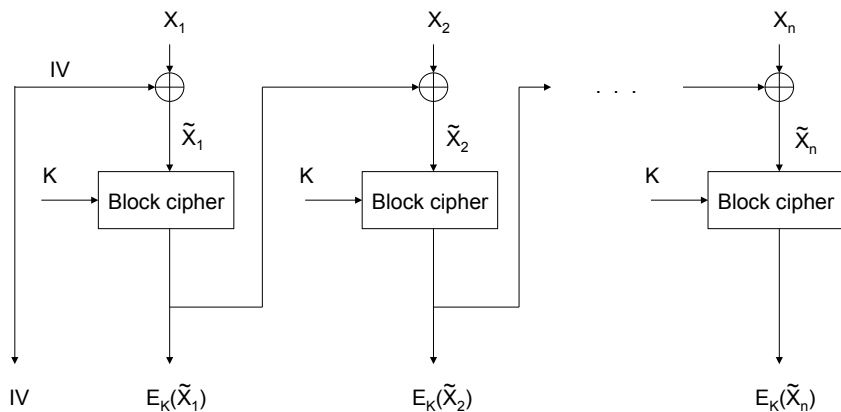
Encrypted in ECB
mode



Encrypted in other
modes of operation

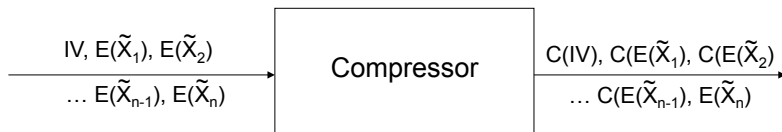
Source: Wikipedia

Cipher Block Chaining (CBC) Mode



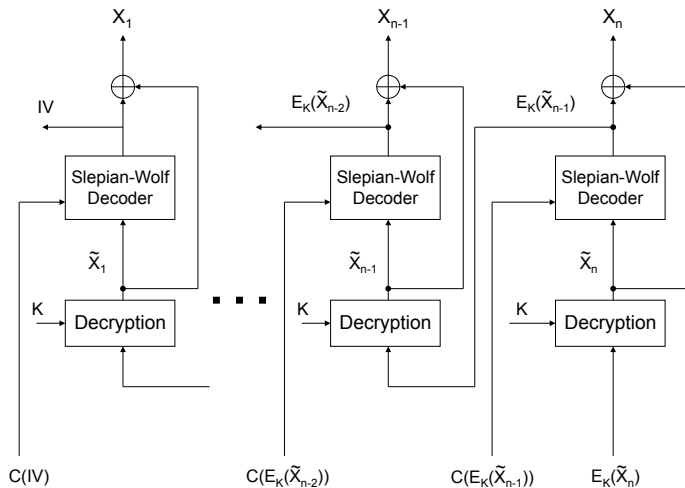
Block Ciphers Can Be Compressed

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

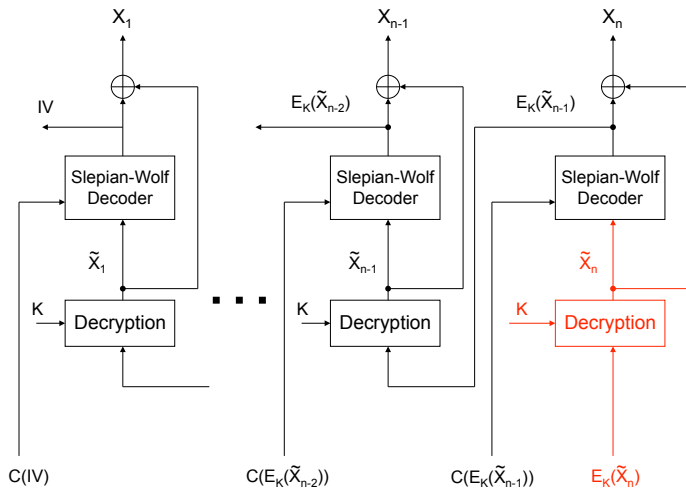


- notice: last block is left uncompressed, while IV is compressed

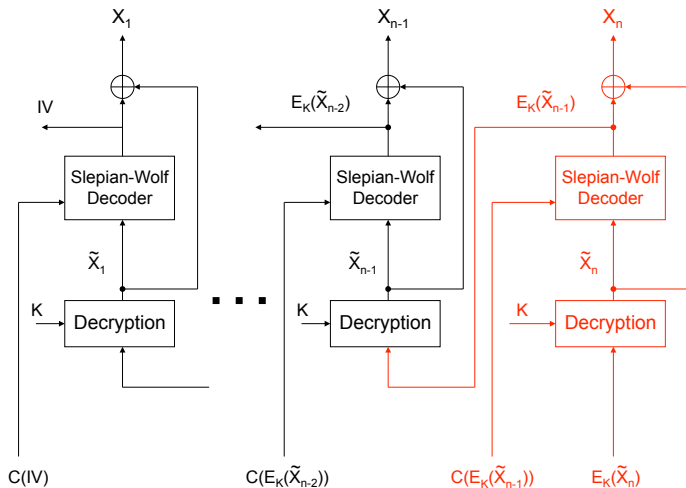
Joint Decompression and Decoding



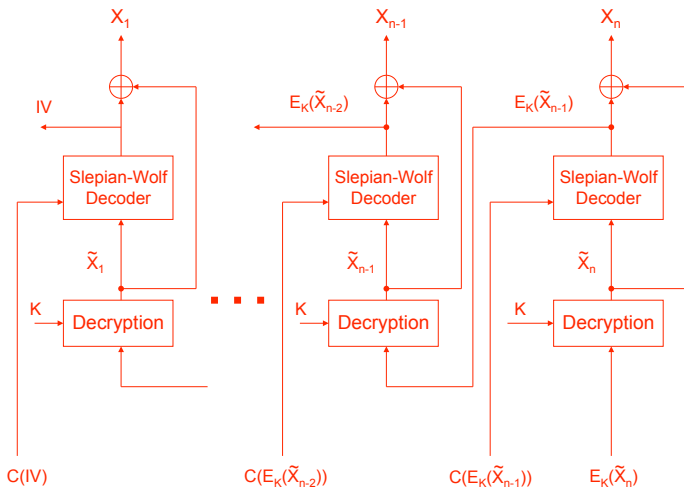
Joint Decompression and Decoding



Joint Decompression and Decoding



Joint Decompression and Decoding



Compression Factor

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

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

Rate

- for large m

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

- 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

Compression Results

- irregular LDPC codes were used in our performance evaluation

| p | Target FER | Compression Rate | Source Entropy |
|-------|------------|------------------|----------------|
| 0.026 | 10^{-3} | 0.50 | 0.1739 |
| 0.018 | 10^{-4} | 0.50 | 0.1301 |
| 0.068 | 10^{-3} | 0.75 | 0.3584 |
| 0.054 | 10^{-4} | 0.75 | 0.3032 |

Table: Attainable compression rates for $m = 128$ bits.

Compression Results

- irregular LDPC codes were used in our performance evaluation

| p | Target FER | Compression Rate | Source Entropy |
|-------|------------|------------------|----------------|
| 0.026 | 10^{-3} | 0.50 | 0.1739 |
| 0.018 | 10^{-4} | 0.50 | 0.1301 |
| 0.068 | 10^{-3} | 0.75 | 0.3584 |
| 0.054 | 10^{-4} | 0.75 | 0.3032 |

Table: Attainable compression rates for $m = 128$ bits.

Compression Results

- irregular LDPC codes were used in our performance evaluation

| p | Target FER | Compression Rate | Source Entropy |
|-------|------------|------------------|----------------|
| 0.058 | 10^{-3} | 0.50 | 0.3195 |
| 0.048 | 10^{-4} | 0.50 | 0.2778 |
| 0.134 | 10^{-3} | 0.75 | 0.5710 |
| 0.126 | 10^{-4} | 0.75 | 0.5464 |

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

Compression Results

- irregular LDPC codes were used in our performance evaluation

| p | Target FER | Compression Rate | Source Entropy |
|-------|------------|------------------|----------------|
| 0.058 | 10^{-3} | 0.50 | 0.3195 |
| 0.048 | 10^{-4} | 0.50 | 0.2778 |
| 0.134 | 10^{-3} | 0.75 | 0.5710 |
| 0.126 | 10^{-4} | 0.75 | 0.5464 |

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