MTAT.07.006 Research Seminar in Cryptography

Seminar 3: Storage media encryption

Mart Sõmermaa
Tartu University
http://blockcrypto.mrts.pri.ee

03.10.2005

Overview

- preliminaries: storage media and security notions, security goals
- overview of standard block cipher modes of operation as candidate algorithms for storage media encryption, overview of their weaknesses
- overview of new modes with better security properties
- a general model for storage media encryption schemes

Background

- *storage media*: a finite sequence of sectors, accessed randomly, assume the cardinality to be bounded by 2^{64}
- *sector*: a 512-byte/4096-bit sequence $S \in \{0,1\}^{4096}$
- *block*: a n-bit sequence that block ciphers operate on, commonly $n = 128$ for modern block ciphers

Background

- *block cipher*: a function e that maps a plaintext block $p \in \{0,1\}^n$ to a ciphertext block $c \in \{0,1\}^n$ given a key $K \in \mathcal{K}$, and its inverse d, $d(e(p)) = p$

$$e : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n, \quad d : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$$

- *tweakable sector enciphering scheme*: a function E that maps a plaintext sector $P \in \{0,1\}^{4096}$ to a ciphertext sector $C \in \{0,1\}^{4096}$, given a key $K \in \mathcal{K}$ and a tweak $T \in \mathcal{T}$, and its inverse D, $D(E(P)) = P$

$$E : \mathcal{K} \times \mathcal{T} \times \{0,1\}^n \to \{0,1\}^n, \quad D : \mathcal{K} \times \mathcal{T} \times \{0,1\}^n \to \{0,1\}^n$$

The enciphering schemes usually divide the sector into $m = 4096/n$ blocks and use a block cipher to transform the blocks.

Background

- *tweak*: tweak is an arbitrary input to the tweakable enciphering scheme. In our application domain it is the sector index, i.e. location of the sector on storage media. Using the tweak ensures that encryption depends on sector location (sectors can't be rearranged, prevents collisions).

Domain constraints

- the enciphering scheme has to be length-preserving, i.e $|P| = |E(P)|$
- a sector should be transformed independently of other sectors

It follows from these constraints that integrity preserving modes can not be used in a tweakable enciphering scheme

Goals

- the ciphertext does not leak any useful information about the plaintext
- it is impossible to manipulate plaintext in a meaningful way during decryption by modifying the corresponding ciphertext

Corresponding formal security notions:

- indistinguishability under chosen plaintext attack, IND-CPA
- indistinguishability under chosen ciphertext attack, IND-CCA

We assume that an IND-CPA/IND-CCA secure block cipher exists (AES).

Standard modes of operation

The standard modes of block cipher operation specified in a NIST standard: ECB, CBC, OFB, CTR, CFB.

We will analyse the modes in the context of tweakable sector enciphering schemes.

Electronic Codebook mode (ECB)

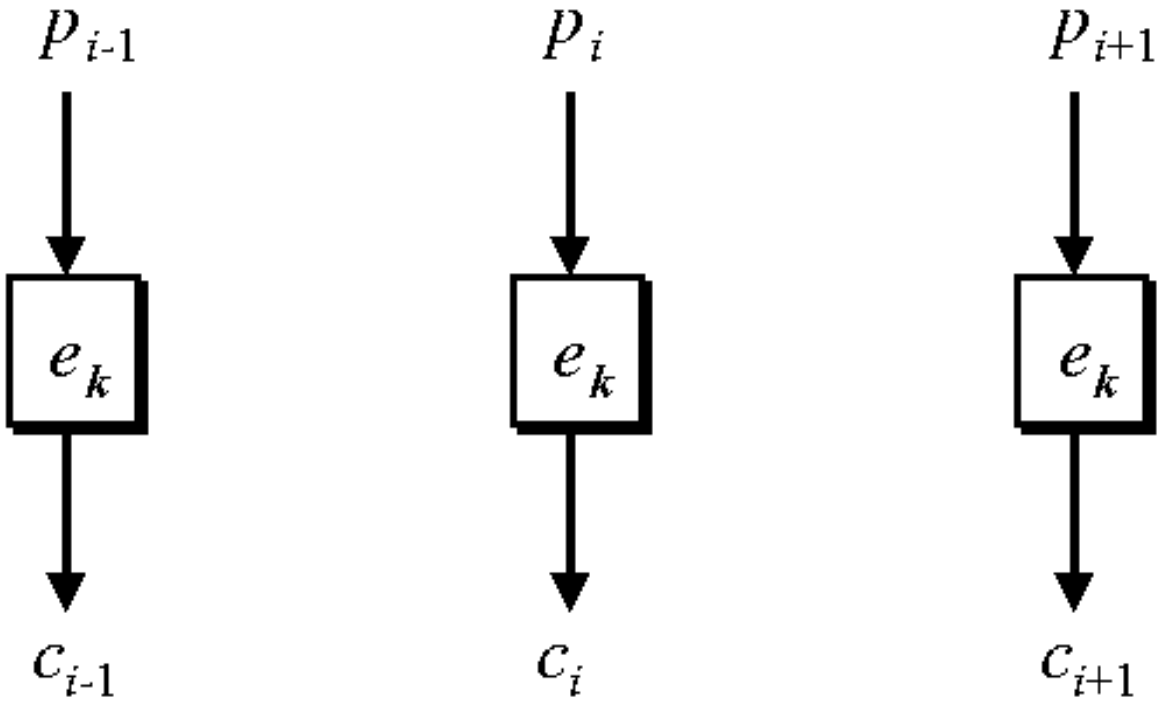
Encryption. Input: K, e, $P = p_1, \ldots, p_m$

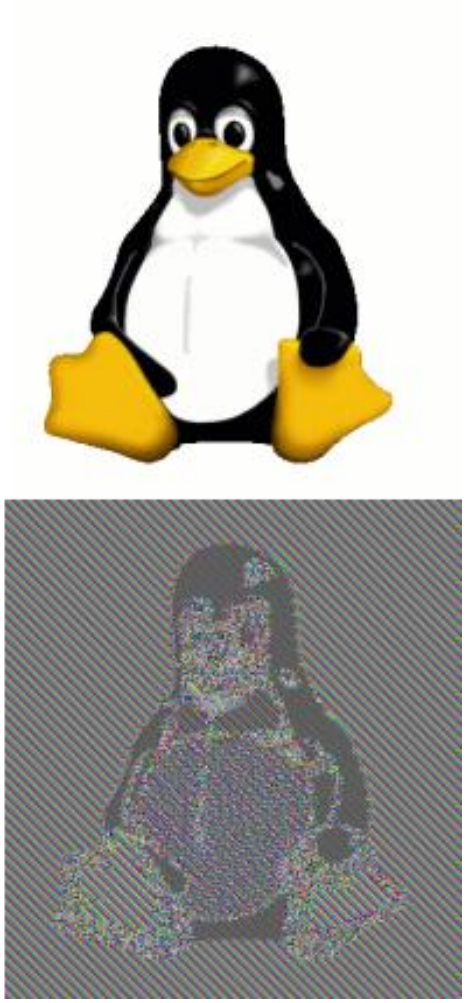
$$e_K(p_i) \rightarrow c_i, \quad i = 1, \ldots, m$$

Decryption. Input: K, d, $C = c_1, \ldots, c_m$

$$d_K(c_i) \rightarrow p_i, \quad i = 1, \ldots, m$$

Security. The mode is trivially not IND-CPA secure as equal plaintext blocks are transformed to equal ciphertext blocks — e.g. images are clearly identifiable when encrypted in ECB mode





Encryption in Electronic Codebook (ECB) mode

Cipherblock Chaining mode (CBC)

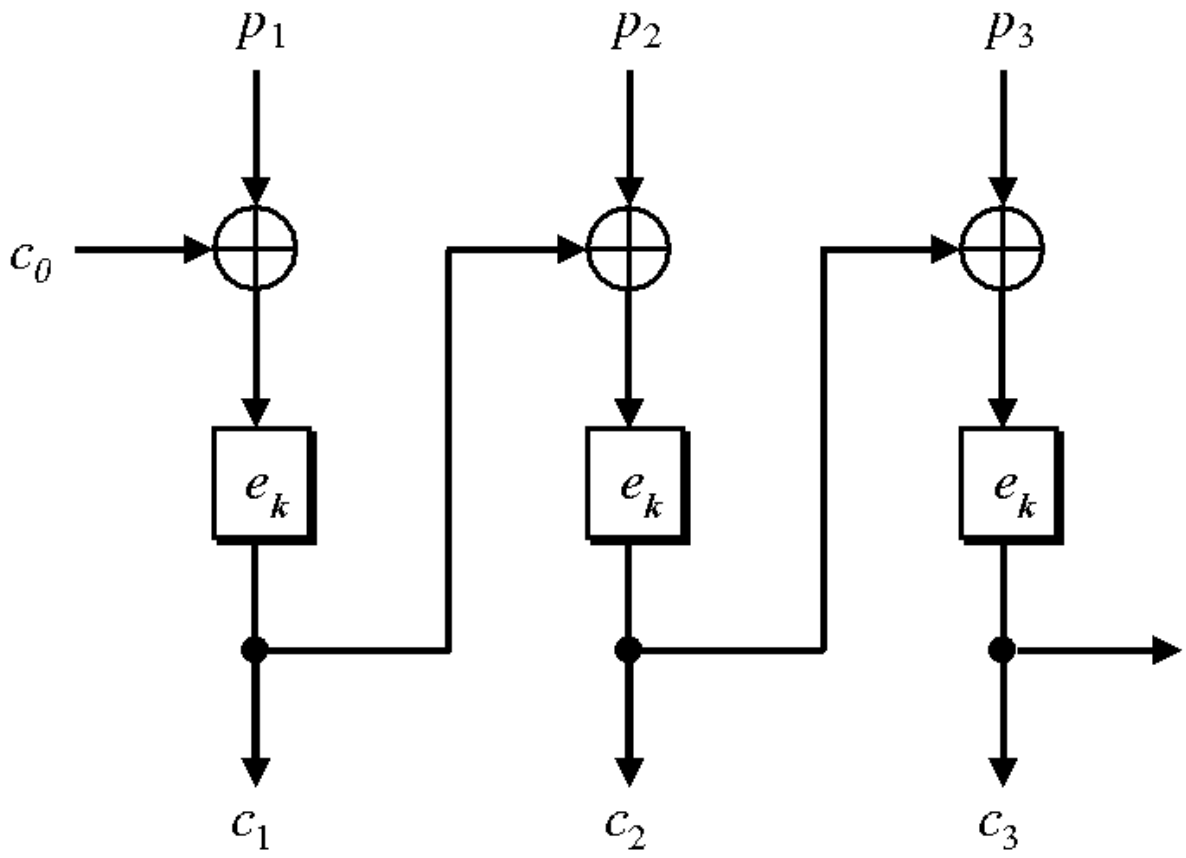
Encryption. Input: K, e, T, $P = p_1, \ldots, p_m$.

$$e_K(p_i \oplus c_{i-1}) \to c_i, \quad i = 1, \ldots, m, \; c_0 = T.$$

Decryption. Input: K, d, T, $C = c_1, \ldots, c_m$.

$$c_{i-1} \oplus d_K(c_i) \to p_i, \quad i = 1, \ldots, m, \; c_0 = T.$$

Security. The mode is malleable and vulnerable to copy-paste attacks, hence not IND-CCA secure. Also, it is vulnerable to watermarking if the adversary can compute the tweak T, in that case it is not IND-CPA secure



Encryption in Cipherblock Chaining (CBC) mode

CBC mode vulnerabilities: malleability

Modifications in the ciphertext block c_{i-1} corrupt the corresponding plaintext block p'_{i-1} and enable the adversary to fully control the contents of the next plaintext block p'_i, as $p'_i = c'_{i-1} \oplus d_K(c_i)$.

Attack goal: replace plaintext block p_i with block $p^{\#}$.

The adversary chooses $c'_{i-1} = p^{\#} \oplus c_{i-1} \oplus p_i$, then

$$p'_i = c'_{i-1} \oplus d_K(c_i) = c'_{i-1} \oplus c_{i-1} \oplus p_i = \\ = p^{\#} \oplus c_{i-1} \oplus p_i \oplus c_{i-1} \oplus p_i = p^{\#}.$$

CBC mode vulnerabilities: copy-paste

Sequences of ciphertext blocks can be copied and pasted to a new location. The first and the next after last block will be corrupted, the intermediate blocks will be decrypted correctly.

Attack goal: replace plaintext blocks $p_i, \ldots, p_{i+k}$, with blocks $p_j, \ldots, p_{j+k}$.

If the ciphertext blocks $c_{i-1}, \ldots, c_{i+k}$ are replaced with $c_{j-1}, \ldots, c_{j+k}$, then trivially

$$
\begin{aligned}
&p'_{i-1} = c_{i-2} \oplus d_K(c'_{i-1}) = c_{i-2} \oplus d_K(c_{j-1}) = c_{i-2} \oplus c_{j-2} \oplus p_{j-1},\\
&p'_i = c'_{i-1} \oplus d_K(c'_i) = c_{j-1} \oplus c_{j-1} \oplus p_j = p_j,\\
&\ldots,\\
&p'_{i+k} = c'_{i+k-1} \oplus d_K(c'_{i+k}) = p_{j+k},\\
&p'_{i+k+1} = c'_{i+k} \oplus d_K(c_{i+k+1}) = c_{j+k} \oplus c_{i+k} \oplus p_{i+k+1}.
\end{aligned}
$$

XOR-masking modes

The following modes, OFB, CTR and CFB, contain an intermediate cipherblock layer $z_1, \ldots, z_m$ that is XORed with plaintext to get final ciphertext. The values $z_1, \ldots, z_m$ should never repeat.

To avoid repetition, it is required that the tweak is a nonce for a given key in these modes. However, in a sector enciphering scheme the tweak is a simple integer index that is reused whenever new data is written to a particular sector.

Hence all these modes are trivially vulnerable when used in a sector enciphering scheme.

Output Feedback Mode (OFB)

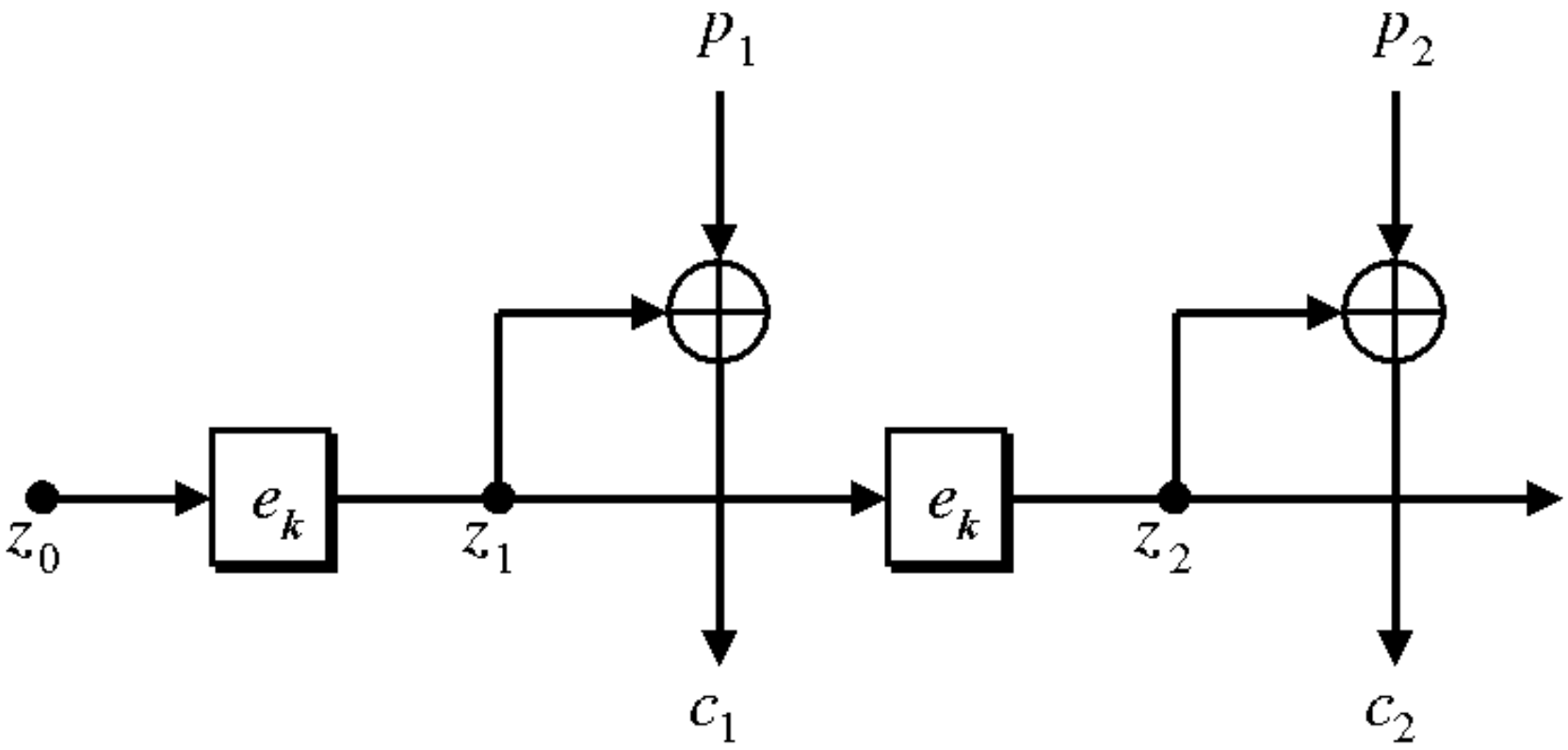
Encryption. Input: K, e, T, $P = p_1, \ldots, p_m$.

$$e_K(z_{i-1}) \rightarrow z_i, \quad p_i \oplus z_i \rightarrow c_i, \quad i = 1, \ldots, m, \ z_0 = T.$$

Decryption. Input: K, d, T, $C = c_1, \ldots, c_m$.

$$e_K(z_{i-1}) \rightarrow z_i, \quad c_i \oplus z_i \rightarrow p_i, \quad i = 1, \ldots, m, \ z_0 = T.$$

Security. The mode is not IND-CPA secure, if the tweak T is reused as the XOR operand will be used repeatedly. Even if the tweak is not reused, the mode is still malleable by essentialy the same attack as against CBC, hence not IND-CCA secure.



Encryption in Output Feedback (OFB) mode

OFB vulnerability: repeating XOR operand

If a block z_i from the cipherblock layer is used twice, then the adversary gains information about the plaintext.

Attack goal: retrieve $p_i^A \oplus p_i^B$, given ciphertext blocks $c_i^A = p_i^A \oplus z_i$ and $c_i^B = p_i^B \oplus z_i$.

Trivially,

$$c_i^A \oplus c_i^B = p_i^A \oplus z_i \oplus p_i^B \oplus z_i = p_i^A \oplus p_i^B.$$

If either p_i^A or p_i^B is zero, the other plaintext block will be revealed to the adversary.

Counter mode (CTR)

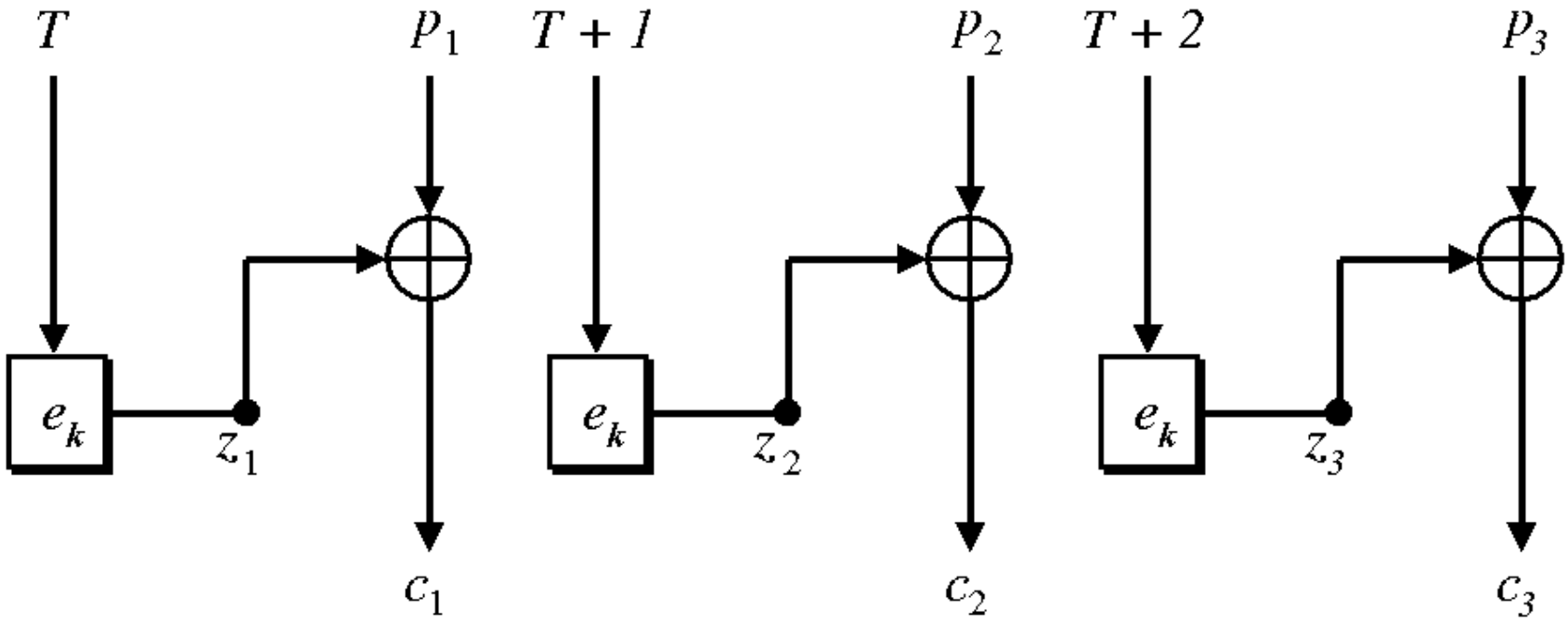
Encryption. Input: K, e, T, $P = p_1, \ldots, p_m$.

$$e_K(T + i - 1) \rightarrow z_i, \quad p_i \oplus z_i \rightarrow c_i, \quad i = 1, \ldots, m,$$

Decryption. Input: K, d, T, $C = c_1, \ldots, c_m$.

$$e_K(T + i - 1) \rightarrow z_i, \quad c_i \oplus z_i \rightarrow p_i, \quad i = 1, \ldots, m,$$

Security. The mode is not IND-CPA secure, if the tweak T is reused (see OFB vulnerability)



Encryption in Counter (CTR) mode

Cipher Feedback mode (CFB)

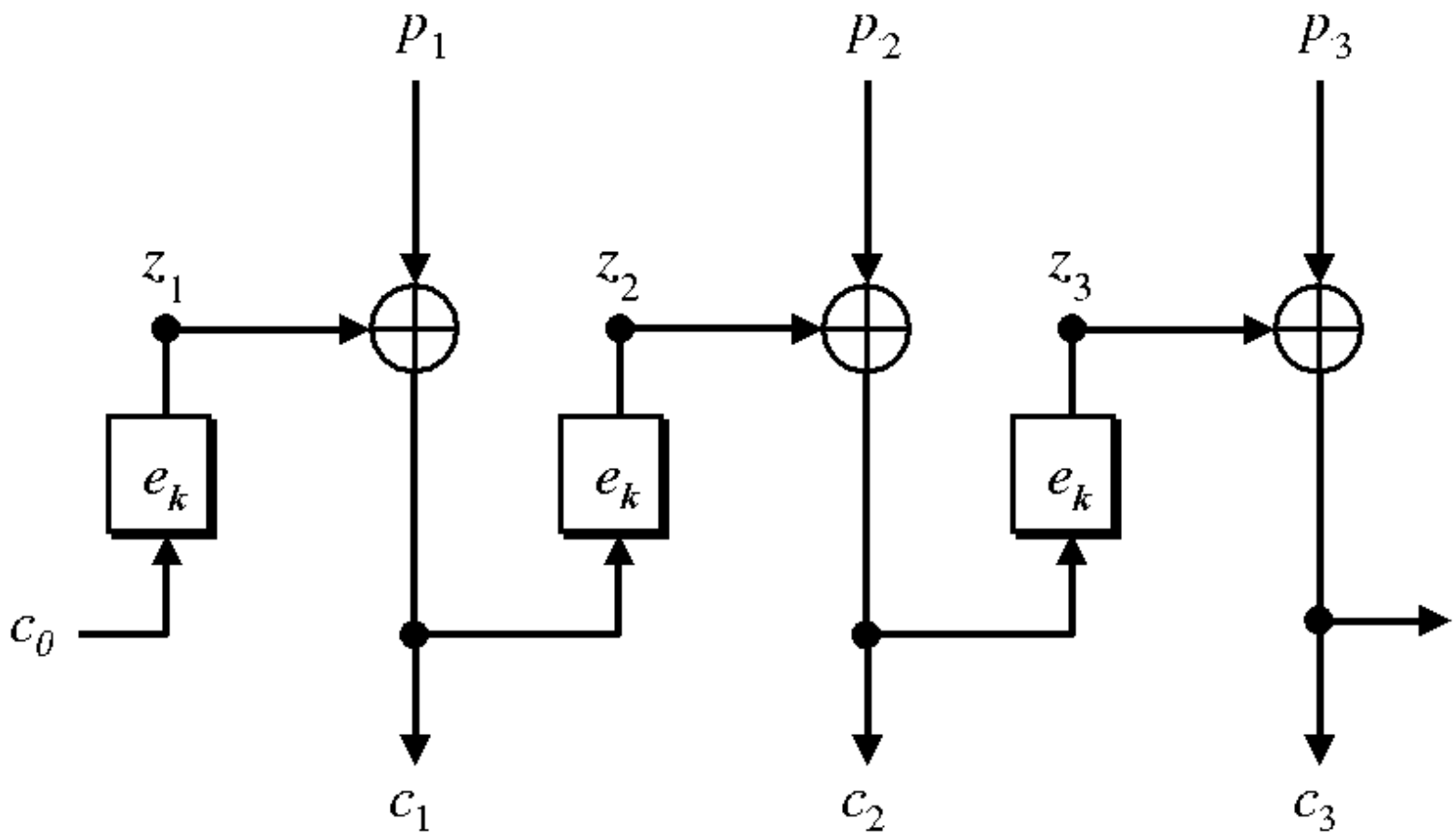
Encryption. Input: K, e, T, $S^p = p_1, \ldots, p_m$.

$$e_K(c_{i-1}) \to z_i, \quad p_i \oplus z_i \to c_i, \quad i = 1, \ldots, m, \ c_0 = T.$$

Decryption. Input: K, d, T, $S^c = c_1, \ldots, c_m$.

$$e_K(c_{i-1}) \to z_i, \quad c_i \oplus z_i \to p_i, \quad i = 1, \ldots, m, \ c_0 = T.$$

Security. The first block is subject to the same vulnerability as OFB and CTR if the tweak T is reused. The mode is not IND-CPA secure in this case. Additionally, like CBC, the mode is malleable and vulnerable to copy-paste attacks, hence not IND-CCA secure even if the tweak is a nonce.



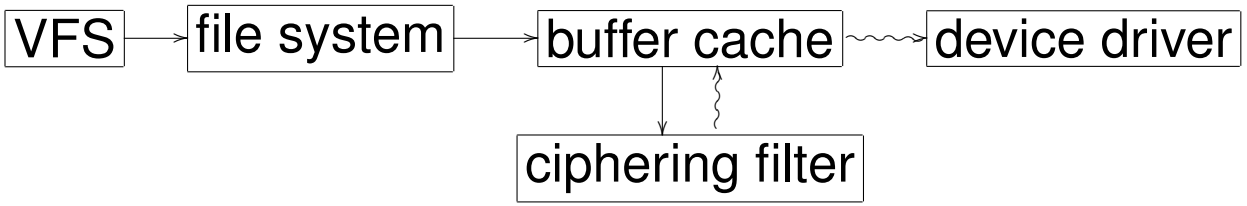
Encryption in Cipher Feedback (CFB) mode

Conclusion

None of the standard modes is suitable for constructing a length-preserving sector enciphering scheme.

Case study: a CBC-based sector level cryptosystem

The sector level cryptosystem described below was used in the Linux kernel block device encryption modules *loopAES* and *dm_crypt*. The enciphering scheme described below is considered to be deprecated, but serves well as a case study. We will refer to it as cryptosystem $\mathcal{A}$.



The encryption was (and is) implemented as a ciphering loop filter between filesystem and device driver layers in the storage stack.

Case study: continued

The following CBC-based sector enciphering scheme was utilized in the cryptosystem:

$$E(K, i, P_i) \rightarrow C_i, \quad D(K, i, C_i) \rightarrow P_i,$$
$$E : e_K(p_j \oplus c_{j-1}) \rightarrow c_j, \quad j = 1, \ldots, m,$$
$$D : c_{j-1} \oplus d_K(c_j) \rightarrow p_j, \quad j = 1, \ldots, m,$$
$$c_0 = i.$$

The enciphering scheme is vulnerable to the same attacks as CBC mode. Also, the sector index is directly used as the tweak, which opens up another vulnerability.

Cryptosystem $\mathcal{A}$ vulnerability: watermarking

It is possible to create collisions in subsequent sectors in cryptosystem $\mathcal{A}$ with a chosen plaintext attack.

The attack is based on the following property of any CBC-based sector level cryptosystem: if the values of tweaks are known to the adversary, she can choose p_1^i, p_1^j, given tweaks (sector indexes) $i, j \in \mathcal{T}$, such that $T_i \oplus T_j = p_1^i \oplus p_1^j$, then

$$T_i \oplus p_1^i = T_j \oplus p_1^j \Rightarrow e_K(T_i \oplus p_1^i) = e_K(T_j \oplus p_1^j) \Rightarrow c_1^i = c_1^j.$$

New IND-CPA/IND-CCA secure modes

Let $K \in \mathcal{K}$ be a random secret key, $T \in \mathcal{T}$ be a tweak and $P \in \{0,1\}^l$ be a l-bit plaintext for some $l \in \mathbb{N}$. A transform $\widetilde{E} : \mathcal{T} \times \mathcal{K} \times \{0,1\}^l \rightarrow \{0,1\}^l$ and its inverse $\widetilde{D} : \mathcal{T} \times \mathcal{K} \times \{0,1\}^l \rightarrow \{0,1\}^l$ is a *tweakable strong pseudorandom permutation* ($\pm\widetilde{prp}$), if an oracle that maps (T,p) into $\widetilde{E}_K^T(P) = C$ and maps (T,C) into $\widetilde{D}_K^T(C) = P$ is indistinguishable from an oracle that realizes an T-indexed family of random permutations and their inverses.

$\pm\widetilde{prp}$ security is equivalent to IND-CPA/IND-CCA security.

IEEE Security in Storage working group is currently standardizing an architecture for encrypted storage media. The following provably $\pm\widetilde{prp}$ secure transforms were proposed as candidate algortithms for tweakable sector enciphering schemes.

New modes: continued

Wide-block mode: can operate with at least sector granularity,

Narrow-block mode: operates with cipher block granularity.

Wide-block modes are preferrable as it is possible to detect e.g. database write patterns to the storage media when a narrow block mode is in use.

We use $\otimes$ to signify multiplication in the field $GF(2^n)$. Note that multiplication by 2 is much easier to implement and computationally less costly than general multiplication in $GF(2^n)$.

EME

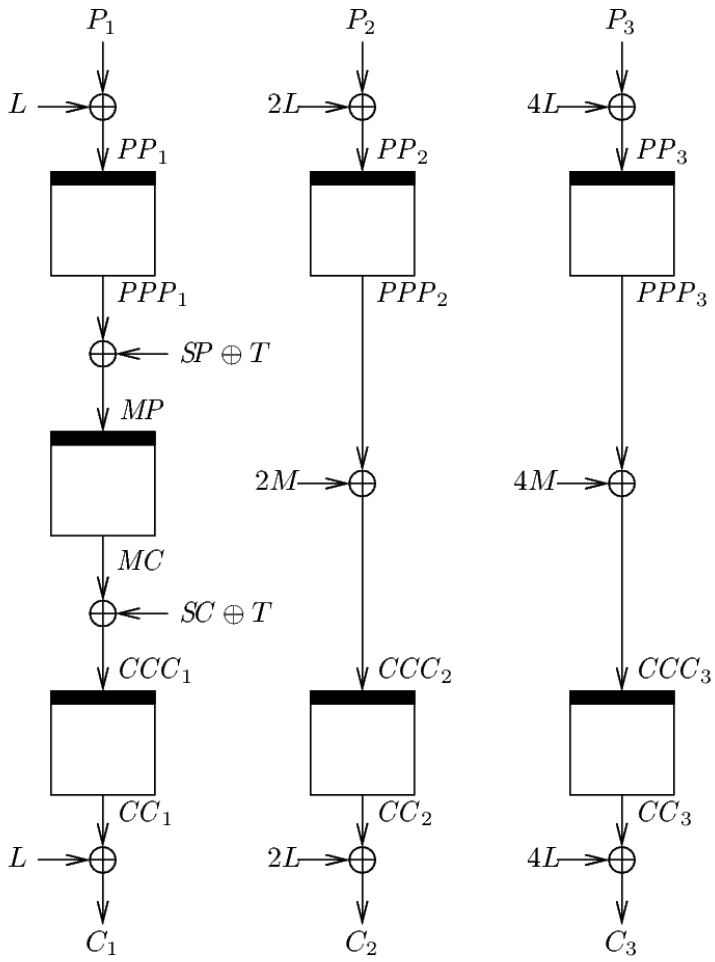
EME stands for *ECB-mix-ECB*, the algorithm entails two layers of ECB encryption and a “lightweight mixing” in between.

Encryption. Input: K, e, T, $P = p_1, \ldots, p_m$.

$$
\begin{aligned}
&L \leftarrow 2 \otimes e_K(0^n) \\
&PP_i \leftarrow 2^{i-1} \otimes L \oplus p_i, \\
&\quad PPP_i \leftarrow e_K(PP_i), \quad i = 1, \ldots, m \\
&SP \leftarrow PPP_2 \oplus \cdots \oplus PPP_m \\
&MP \leftarrow PPP_1 \oplus SP \oplus T \\
&MC \leftarrow e_K(MP) \\
&M \leftarrow MP \oplus MC \\
&CCC_i \leftarrow PPP_i \oplus 2^{i-1} \otimes M, \quad i = 2, \ldots, m \\
&SC \leftarrow CCC_2 \oplus \cdots \oplus CCC_m \\
&CCC_1 \leftarrow MC \oplus SC \oplus T \\
&CC_i \leftarrow e_K(CCC_i), \\
&\quad CC_i \oplus 2^{i-1} \otimes L \rightarrow c_i, \quad i = 1, \ldots, m
\end{aligned}
$$

Decryption. Input: K, d, T, $C = c_1, \ldots, c_m$.

$$
\begin{aligned}
&L \leftarrow 2 \otimes e_K(0^n) \\
&CC_i \leftarrow 2^{i-1} \otimes L \oplus c_i, \\
&\quad CCC_i \leftarrow d_K(CC_i), \quad i = 1, \ldots, m \\
&SC \leftarrow CCC_2 \oplus \cdots \oplus CCC_m \\
&MC \leftarrow CCC_1 \oplus SC \oplus T \\
&MP \leftarrow d_K(MC) \\
&M \leftarrow MP \oplus MC \\
&PPP_i \leftarrow CCC_i \oplus 2^{i-1} \otimes M, \quad i = 2, \ldots, m \\
&SP \leftarrow PPP_2 \oplus \cdots \oplus PPP_m \\
&PPP_1 \leftarrow MP \oplus SP \oplus T \\
&PP_i \leftarrow d_K(PPP_i), \\
&\quad PP_i \oplus 2^{i-1} \otimes L \rightarrow p_i, \quad i = 1, \ldots, m
\end{aligned}
$$



Encryption in EME mode

CMC

CMC stands for *CBC-mix-CBC*, the algorithm makes a pass of CBC encryption, XORs in a mask, and then makes a pass of CBC decryption. The layered structure is similar to EME.

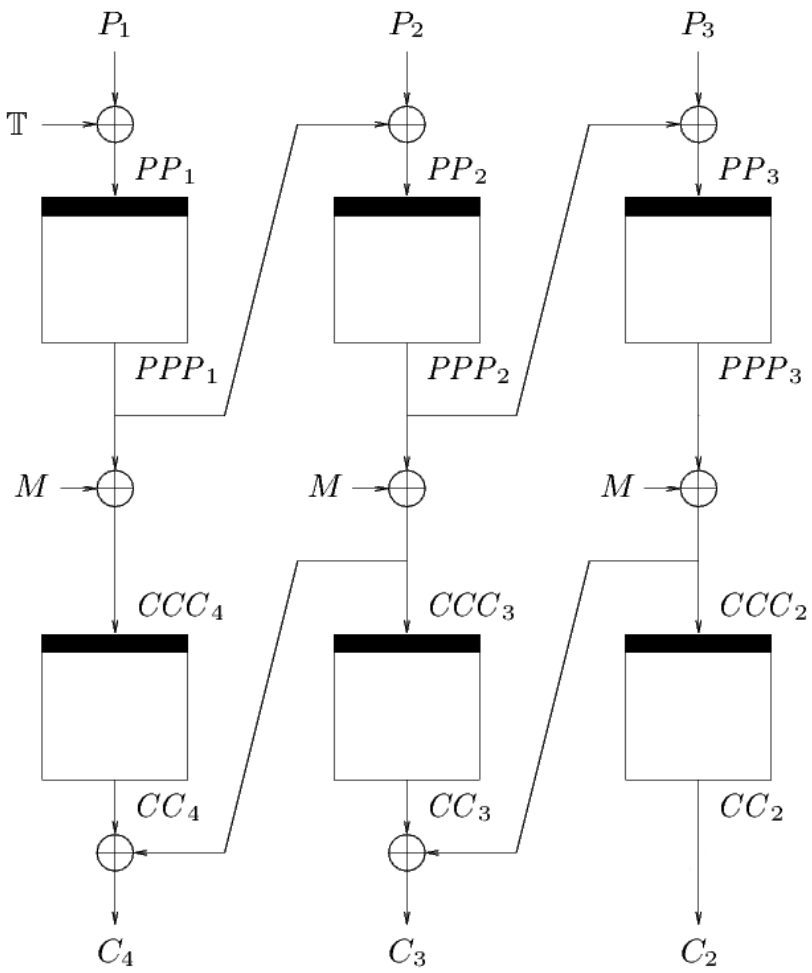
The authors recommend EME over CMC as it is as secure but has several advantages: it is parallelizable, only one key required, utilizes only e in encryption and d in decryption, remains secure for variable length input.

Encryption. Input: K, $\tilde{K}$, e, T, $P = p_1, \ldots, p_m$.

$$
\begin{aligned}
&\mathsf{T} \leftarrow e_{\tilde{K}}(T)\\
&PPP_0 \leftarrow \mathsf{T}\\
&PP_i \leftarrow p_i \oplus PPP_{i-1},\\
&\quad PPP_i \leftarrow e_K(PP_i), \quad i = 1, \ldots, m\\
&M \leftarrow 2 \otimes (PPP_1 \oplus PPP_m)\\
&CCC_i \leftarrow PPP_{m+1-i} \oplus M, \quad i = 1, \ldots, m\\
&CCC_0 \leftarrow 0^n\\
&CC_i \leftarrow e_K(CCC_i),\\
&\quad CC_i \oplus CCC_{i-1} \rightarrow c_i, \quad i = 1, \ldots, m\\
&c_1 \oplus \mathsf{T} \rightarrow c_1.
\end{aligned}
$$

Decryption. Input: K, $\tilde{K}$, e, d, T, $C = c_1, \ldots, c_m$.

$$
\begin{aligned}
&\mathsf{T} \leftarrow e_{\tilde{K}}(T)\\
&CCC_0 \leftarrow \mathsf{T}\\
&CC_i \leftarrow c_i \oplus CCC_{i-1},\\
&\quad CCC_i \leftarrow d_K(CC_i), \quad i = 1, \ldots, m\\
&M \leftarrow 2 \otimes (CCC_1 \oplus CCC_m)\\
&PPP_i \leftarrow CCC_{m+1-i} \oplus M, \quad i = 1, \ldots, m\\
&PPP_0 \leftarrow 0^n\\
&PP_i \leftarrow e_K(PPP_i),\\
&\quad PP_i \oplus PPP_{i-1} \rightarrow p_i, \quad i = 1, \ldots, m\\
&p_1 \oplus \mathsf{T} \rightarrow p_1.
\end{aligned}
$$



Encryption in CMC mode

LRW

The name is based on the first letters of the surnames of original authors. It is a simple mode utilizing multiplication in $GF(2^n)$.

Encryption. Input: K, $\tilde{K}$, e, T, $P = p_1, \ldots, p_m$.

$$\mathsf{T}_i \leftarrow \tilde{K} \otimes (T + i), \quad e_K(p_i \oplus \mathsf{T}_i) \oplus \mathsf{T}_i \rightarrow c_i, \quad i = 1, \ldots, m$$

Decryption. Input: K, $\tilde{K}$, d, T, $C = c_1, \ldots, c_m$.

$$\mathsf{T}_i \leftarrow \tilde{K} \otimes (T + i), \quad d_K(c_i \oplus \mathsf{T}_i) \oplus \mathsf{T}_i \rightarrow p_i, \quad i = 1, \ldots, m$$

XEX

A simple, efficient mode utilizing the same construction as LRW.

Encryption. Input: K, e, T, $P = p_1, \ldots, p_m$.

$$\mathsf{T}_i \leftarrow 2^i \otimes e_K(T), \quad e_K(p_i \oplus \mathsf{T}_i) \oplus \mathsf{T}_i \rightarrow c_i, \quad i = 1, \ldots, m$$

Decryption. Input: K, d, T, $C = c_1, \ldots, c_m$.

$$\mathsf{T}_i \leftarrow 2^i \otimes e_K(T), \quad d_K(p_i \oplus \mathsf{T}_i) \oplus \mathsf{T}_i \rightarrow p_i, \quad i = 1, \ldots, m$$

XCB

XCB stands for *Extended Codebook*.

The mode entails two layers of hashing and a CTR-like layer in between that utilises the hash. GHASH is used as the hash function.

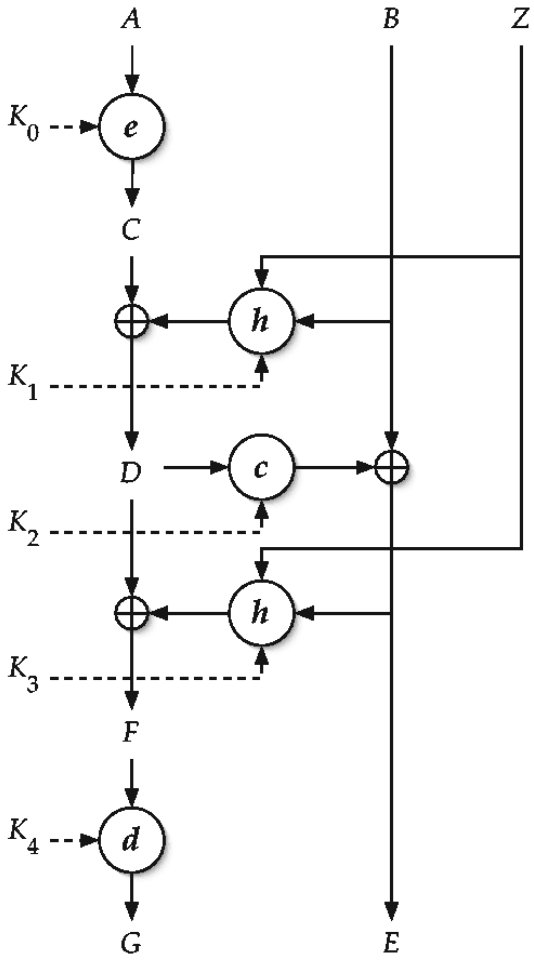
The mode seems to be a direct follow-up to the officially unpublished ABL mode.

Encryption. Input: K, e, d, $h = GHASH$, T, $P = p_1, \ldots, p_m$.

$$
\begin{aligned}
&K_i \leftarrow e_K(i), \quad i = 0, \ldots, 4 \\
&B \leftarrow p_2, \ldots, p_m \\
&D \leftarrow e_{K_0}(p_1) \oplus h_{K_1}(B, T) \\
&E \leftarrow B \oplus [e_{K_2}(D + 0) || e_{K_2}(D + 1 \bmod 2^n) || \cdots \\
&\quad || e_{K_2}(D + m - 2 \bmod 2^n)] \\
&F \leftarrow D \oplus h_{K_3}(E, T) \\
&G \leftarrow d_{K_4}(F) \\
&G || E \rightarrow C
\end{aligned}
$$

Decryption. Input: K, e, d, h, T, $C = c_1, \ldots, c_m$.

$$
\begin{aligned}
&K_i \leftarrow e_K(i), \quad i = 0, \ldots, 4 \\
&E \leftarrow c_2, \ldots, c_m \\
&D \leftarrow e_{K_4}(c_1) \oplus h_{K_3}(E, T) \\
&B \leftarrow E \oplus [e_{K_2}(D + 0) || e_{K_2}(D + 1 \bmod 2^n) || \cdots \\
&\quad || e_{K_2}(D + m - 2 \bmod 2^n)] \\
&A \leftarrow D \oplus h_{K_1}(B, T) \\
&p_1 \leftarrow d_{K_0}(A) \\
&p_1 || B \rightarrow P
\end{aligned}
$$



Encryption in XCB mode

General model for secure sector level cryptosystems

There are other considerations apart from specifying a mode of operation when implementing a secure sector level cryptosystem — cryptoanalysis will be harder if per-sector unique keys are used and if the ordering of sectors is changed. The following components can be identified in the system:

1. a block cipher (IND-CPA/IND-CCA secure)
2. a mode of operation (tweakable IND-CPA/IND-CCA secure)
3. a function for generating sector keys (pseudorandom)
4. a function for sector reordering