ETH zürich
Departement of Computer Science

SCB Mode: Semantically Secure Length-Preserving Encryption

Fabio Banfi

ICS, 30 November 2022

Outline

Outline

1. Motivation
2. Block Ciphers and Symmetric Encryption
3. Length-Preserving Encryption/Enciphering (LPE)
4. SCB Mode of Encryption: Semantically Secure LPE
5. Conclusions

Motivation

Usually semantically (IND-CPA) secure encryption *expands* the length of plaintexts

What if we have many *short* messages to be transmitted, and communication is expensive? E.g.:

- Each day m messages need to be transmitted
- Each message consists of b blocks (defined by the underlying block cipher)

Conventional IND-CPA scheme: $c_0 \doteq m(b+1)$ transmitted blocks

Encryption without expansion: $c_1 \doteq mb$ transmitted blocks

$\Longrightarrow$ If b small and m large: $c_0 \approx 2 \cdot c_1$!

But can we actually avoid expansion while retaining semantic security?

Outline

1. Motivation
2. Block Ciphers and Symmetric Encryption
3. Length-Preserving Encryption/Enciphering (LPE)
4. SCB Mode of Encryption: Semantically Secure LPE
5. Conclusions

Block Ciphers: Definition and Security

Definition (Block Cipher)

A pair $\mathfrak{B}$ of *deterministic* algorithms $E, D : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$ so that for any $K \in \{0,1\}^\kappa$:

- $E_K(\cdot) \doteq E(K, \cdot)$ and $D_K(\cdot) \doteq D(K, \cdot)$ are *efficiently computable permutations* on $\{0,1\}^n$
- $D_K = E_K^{-1}$, that is, for any $M \in \{0,1\}^n$, $D_K(E_K(M)) = M$

What should a *secure* block cipher $\mathfrak{B} = (E, D)$ guarantee? For uniformly random $K \in \{0,1\}^\kappa$:

E_K must be *indistinguishable* from a uniformly random permutation from $\mathrm{Perm}(\{0,1\}^n)$

Definition (PRP Security)

$\mathfrak{B} = (E, D)$ is a *secure pseudorandom permutation* if for any PRP adversary A, its advantage

$$\mathbf{Adv}_{\mathfrak{B}}^{\mathrm{prp}}(A) \doteq \Pr[A^{E_K(\cdot)} \Rightarrow 0 \,|\, K \xleftarrow{\$} \{0,1\}^\kappa] - \Pr[A^{\pi(\cdot)} \Rightarrow 0 \,|\, \pi \xleftarrow{\$} \mathrm{Perm}(\{0,1\}^n)]$$

is negligible

Encryption: Definition

Three conventional ways to define (symmetric-key) encryption:

- Probabilistic
- Deterministic (nonce-based)
- Stateful

For now let's consider *probabilistic* encryption

Definition (Probabilistic Encryption)

A pair Π of *probabilistic* algorithms $\mathcal{E}, \mathcal{D} : \mathcal{K} \times \{0,1\}^* \to \{0,1\}^*$ so that for any $K \in \mathcal{K}$:

- $\mathcal{E}_K(\cdot) \doteq \mathcal{E}(K, \cdot)$ and $\mathcal{D}_K(\cdot) \doteq \mathcal{D}(K, \cdot)$ are *efficiently computable*
- For any $t \in \mathbb{N}$ and $M \in \{0,1\}^t$:
  - $\mathcal{E}_K(M) \in \{0,1\}^{t+\lambda}$, where $\lambda > 0$ is the *expansion factor* of Π
  - "$\mathcal{D}_K = \mathcal{E}_K^{-1}$", that is, $\mathcal{D}_K(\mathcal{E}_K(M)) = M$

Encryption: Security

What should a *secure* encryption scheme $\Pi = (\mathcal{E}, \mathcal{D})$ guarantee?

Semantic (IND\$-CPA) security: For uniformly random $K \in \mathcal{K}$, and any $t \in \mathbb{N}$ and $M \in \{0,1\}^t$:

The induced distribution $\mathcal{E}_K(M)$ must be *indistinguishable* from the uniform distribution over $\{0,1\}^{t+\lambda}$

Oracle $\$^{|(\cdot)|+\lambda}$: On input $M \in \{0,1\}^*$ with $t \doteq |M|$, output $C \xleftarrow{\$} \{0,1\}^{t+\lambda}$

Definition (Semantic Security)

$\Pi = (\mathcal{E}, \mathcal{D})$ is a *semantically secure encryption scheme* if for any IND-CPA adversary A, its advantage

$$\mathbf{Adv}_{\Pi}^{\text{ind-cpa}}(A) \doteq \Pr[A^{\mathcal{E}_K(\cdot)} \Rightarrow 0 \,|\, K \xleftarrow{\$} \mathcal{K}] - \Pr[A^{\$^{|(\cdot)|+\lambda}} \Rightarrow 0]$$

is negligible

Encryption from Block Ciphers: Modes of Operation

Turn block cipher $\mathfrak{B} = (E, D)$ into enc. scheme $\Pi = (\mathcal{E}, \mathcal{D})$: For $M = M_1 \| \cdots \| M_\ell \in \{0,1\}^{\ell n}$:

- An ***insecure*** way: Electronic Codebook (ECB) Mode ($\boldsymbol{\lambda = 0}$):

$$\mathcal{E}_K(M) \doteq E_K(M_1) \| \cdots \| E_K(M_\ell) \in \{0,1\}^{\ell n}$$

- A ***secure*** way: Cipher Block Chaining (CBC) Mode ($\boldsymbol{\lambda = n}$): Sample $R \xleftarrow{\$} \{0,1\}^n$, then

$$\mathcal{E}_K(M) \doteq R \| \underbrace{E_K(R \oplus M_1)}_{C_1} \| \underbrace{E_K(C_1 \oplus M_2)}_{C_2} \| \cdots \| E_K(C_{\ell-1} \oplus M_\ell) \in \{0,1\}^{(\ell+\mathbf{1})n}$$

Both can be adapted to handle *any* $M \in \{0,1\}^{\geq n}$ via ciphertext stealing (CTS), more on that later

Question: *Can we get the best of both (**secure** and $\boldsymbol{\lambda = 0}$)?* Seems impossible, but let's see …

Outline

1. Motivation
2. Block Ciphers and Symmetric Encryption
3. Length-Preserving Encryption/Enciphering (LPE)
4. SCB Mode of Encryption: Semantically Secure LPE
5. Conclusions

“Encryption” Schemes with $\lambda = 0$

With $\lambda = 0$, Π *cannot* be semantically secure, why?

For any $K \in \mathcal{K}$, if $\lambda = 0$ then $\mathcal{E}_K(\cdot)$ *cannot* be probabilistic, must be deterministic!

Therefore, for any $K \in \mathcal{K}$ and any $t \in \mathbb{N}$: $\mathcal{E}_K|_{\{0,1\}^t} \in \mathrm{Perm}(\{0,1\}^t)$

Known as Length-Preserving Encryption (LPE), but should be called: Length-Preserving *Enciphering*!

Alternatively, Π can be seen as a variable-input-length (VIL) *block cipher*

Definition (VIL-PRP Security)

$\Pi = (\mathcal{E}, \mathcal{D})$ is a *secure VIL pseudorandom permutation* if for any PRP adversary A, its advantage

$$\mathbf{Adv}_{\Pi}^{\mathrm{prp}}(A) \doteq \Pr[A^{\mathcal{E}_K(\cdot)} \Rightarrow 0 \,|\, K \xleftarrow{\$} \mathcal{K}] - \Pr[A^{\pi_{|(\cdot)|}(\cdot)} \Rightarrow 0 \,|\, \forall \ell \geq 1 : \pi_\ell \xleftarrow{\$} \mathrm{Perm}(\{0,1\}^\ell)]$$

is negligible

Variable-Input-Length/Length-Preserving Enciphering

Task: Turn a FIL block cipher $\mathfrak{B} = (E, D)$ into a VIL block cipher $\Pi = (\mathcal{E}, \mathcal{D})$ *preserving* PRP security

Problem first introduced and solved by Bellare and Rogaway at FSE'99:

- On input $M \in \{0,1\}^{\geq n}$, to compute $C = \mathcal{E}_K(M)$ with $|C| = |M|$, make two passes over M:

  1. Compute a tag T of M using E_K in (a variant of) CBC-MAC mode
  2. Encrypt M into C' with $|C'| = |M| + n$ in CTR mode with T as IV and drop one block of C'

- Since CBC-MAC satisfies *"parsimoniousness"*, the dropped block can be recovered!

Semantically Secure Length-Preserving *Encryption*?

Back to our question: Can we design a semantically secure encryption scheme with $\lambda = 0$?

Yes! If we relax correctness to *not* be perfect but only negligibly far from it!

Therefore, we inevitably have that $\mathcal{E}_K|_{\{0,1\}^t} \notin \mathrm{Perm}(\{0,1\}^t)$, and that $\mathcal{E}_K$ must be **stateful**

Definition (Length-Preserving *Stateful* Encryption (LPSE))

A pair Π of algs. $\mathcal{E} : \mathcal{K} \times \{0,1\}^{\geq n} \times \mathcal{S} \to \{0,1\}^{\geq n} \times \mathcal{S}$ and $\mathcal{D} : \mathcal{K} \times \{0,1\}^{\geq n} \times \mathcal{T} \to \{0,1\}^{\geq n} \times \mathcal{T}$ s.t.:
For any $K \in \mathcal{K}$, encryption state $\mathbf{S} \in \mathcal{S}$, and decryption state $\mathbf{T} \in \mathcal{T}$:

- $\mathcal{E}(K, \cdot\,; \mathbf{S})$ and $\mathcal{D}(K, \cdot\,; \mathbf{T})$ are *efficiently computable*
- For any $t \in \mathbb{N}$, and $M, C \in \{0,1\}^t$:
  - $\mathcal{E}(K, M; \mathbf{S}) \in \{0,1\}^{|M|} \times \mathcal{S}$, and $C \leftarrow \mathcal{E}_K^{\mathbf{S}}(M)$ denotes "$(C, \mathbf{S}') \leftarrow \mathcal{E}(K, M; \mathbf{S})$; $\mathbf{S} \leftarrow \mathbf{S}'$"
  - $\mathcal{D}(K, C; \mathbf{T}) \in \{0,1\}^{|C|} \times \mathcal{T}$, and $M \leftarrow \mathcal{D}_K^{\mathbf{T}}(C)$ denotes "$(M, \mathbf{T}') \leftarrow \mathcal{D}(K, C; \mathbf{T})$; $\mathbf{T} \leftarrow \mathbf{T}'$"

Note: There is no correctness requirement in the definition!

LPSE: Security and Correctness

Let $[\,] \in \mathcal{S}, \mathcal{T}$ denote the initial empty encryption/decryption state

Definition (LPSE Semantic Security)

$\Pi = (\mathcal{E}, \mathcal{D})$ is a *semantically secure LPSE scheme* if for any IND-CPA adversary A, its advantage

$$\mathbf{Adv}_{\Pi}^{\text{ind-cpa}}(A) \doteq \Pr\left[A^{\mathcal{E}_K^{\mathbf{S}}(\cdot)} \Rightarrow 0 \,\middle|\, K \xleftarrow{\$} \mathcal{K}, \mathbf{S} \leftarrow [\,]\right] - \Pr\left[A^{\$^{|(\cdot)|}} \Rightarrow 0\right]$$

is negligible

Definition (LPSE Correctness)

$\Pi = (\mathcal{E}, \mathcal{D})$ is a *correct LPSE scheme* if for any COR adversary A, its advantage

$$\mathbf{Adv}_{\Pi}^{\text{cor}}(A) \doteq \Pr\left[A^{\mathcal{D}_K^{\mathbf{T}} \circ \mathcal{E}_K^{\mathbf{S}}(\cdot)} \Rightarrow 0 \,\middle|\, K \xleftarrow{\$} \mathcal{K},\ \mathbf{S}, \mathbf{T} \leftarrow [\,]\right] - \Pr\left[A^{\mathsf{id}(\cdot)} \Rightarrow 0\right]$$

is negligible

Outline

1. Motivation
2. Block Ciphers and Symmetric Encryption
3. Length-Preserving Encryption/Enciphering (LPE)
4. SCB Mode of Encryption: Semantically Secure LPE
5. Conclusions

SCB: The Idea

We introduce a new mode of operation that turns a block cipher $\mathfrak{B} = (E, D)$ into an LPSE $\Pi = (\mathcal{E}, \mathcal{D})$

Secure Codebook (SCB): Can be interpreted as a secure variant/patch of ECB

Observation: ECB insecure as soon as a block $\hat{M} \in \{0,1\}^n$ is repeated *within* or *across* plaintexts

$\implies$ Use state to keep track of blocks seen so far, and *on repeated blocks do something different!*

But what to do exactly? We need to signal to the receiver that this block is a repetition of $\hat{M}$

This inevitably would introduce errors, since a subspace of $\{0,1\}^n$ must represent such signals!

But we can be clever about the choice of such subspace :)

SCB: Encryption

Idea: Let σ and τ be such that $\sigma + \tau \leq n$, $K_1 \in \{0,1\}^\kappa$ (for $\mathfrak{B}$), and $K_2 \in \{0,1\}^n$ (pad), and consider:

- A compression function $H : \{0,1\}^n \to \{0,1\}^\tau$
- A look-up table $\mathbf{S} : \{0,1\}^\tau \to \{0,1\}^\sigma$ (for $h \in \{0,1\}^\tau$, $\mathbf{S}[h] \in \{0,1\}^\sigma \cup \{\perp\}$)

Then for each block M_i:

1. Get $h \leftarrow H(M_i)$, and check whether h is in $\mathbf{S}$, i.e., $\mathbf{S}[h] \neq \perp$ (approximates "M_i is a repetition")
2. If **not** (M_i is a *new* block), then compute $C_i \leftarrow \mathfrak{B}.E_{K_1}(M_i)$ (plain ECB) and set $\mathbf{S}[h] \leftarrow 0^\sigma$
3. If **yes** (M_i is *probably a repeated* block, but **might be wrong**), then:
   - Let $R \leftarrow (0^{n-\sigma-\tau} \,\|\, \mathbf{S}[h] \,\|\, h) \in \{0,1\}^n$, and compute $C_i \leftarrow \mathfrak{B}.E_{K_1}(K_2 \oplus R)$
   - Set $\mathbf{S}[h] \leftarrow (\mathbf{S}[h] + 1) \bmod 2^\sigma$

SCB: Decryption

But how do we decrypt now?

We need to distinguish between normal blocks and repetition signals!

Let $\sigma, \tau, K_1, K_2, H$ as before, and consider look-up table $\mathbf{T} : \{0,1\}^\tau \to \{0,1\}^n$ (approximates "H^{-1}")

Then for each block C_i:

1. Get $M_i \leftarrow \mathfrak{B}.D_{K_1}(C_i)$ (plain ECB)
2. Compute $R \leftarrow K_2 \oplus M_i$ and $h \leftarrow R \bmod 2^\tau$, and check whether $R < 2^{\sigma+\tau}$ and $\mathbf{T}[h] \neq \perp$
3. If **not** (C_i is a *not* a repetition signal), then keep M_i and set $\mathbf{T}[H(M_i)] \leftarrow M_i$
4. If **yes** (C_i is *probably a repetition signal*, but **might be wrong**), then set $M_i \leftarrow \mathbf{T}[h]$

SCB: The Scheme

$\mathsf{SCB}[\mathfrak{B}, H].\mathcal{E}^{\mathbf{S}}_{K_1,K_2}(M_1 \parallel \cdots \parallel M_\ell)$

$\textbf{for } i = 1, \ldots, \ell \textbf{ do}$
- $h \leftarrow H(M_i)$
- $\textbf{if } \mathbf{S}[h] = \perp \textbf{ then}$
  - $C_i \leftarrow \mathfrak{B}.E_{K_1}(M_i)$
  - $\mathbf{S}[h] \leftarrow 0^\sigma$
- $\textbf{else}$
  - $R \leftarrow 0^{n-\sigma-\tau} \parallel \mathbf{S}[h] \parallel h$
  - $C_i \leftarrow \mathfrak{B}.E_{K_1}(K_2 \oplus R)$
  - $\mathbf{S}[h] \leftarrow (\mathbf{S}[h] + 1) \bmod 2^\sigma$

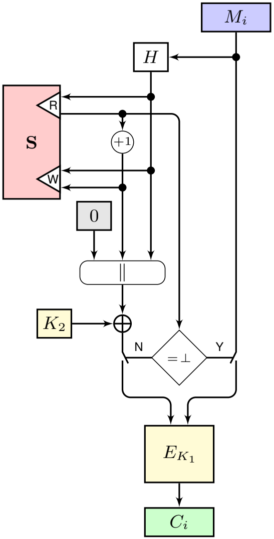
$\textbf{return } C_1 \parallel \cdots \parallel C_\ell$

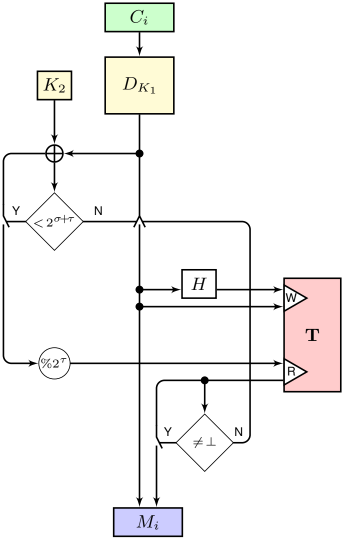
$\mathsf{SCB}[\mathfrak{B}, H].\mathcal{D}^{\mathbf{T}}_{K_1,K_2}(C_1 \parallel \cdots \parallel C_\ell)$

$\textbf{for } i = 1, \ldots, \ell \textbf{ do}$
- $M_i \leftarrow \mathfrak{B}.D_{K_1}(C_i)$
- $R \leftarrow K_2 \oplus M_i$
- $h \leftarrow R \bmod 2^\tau$
- $\textbf{if } R < 2^{\sigma+\tau} \wedge \mathbf{T}[h] \neq \perp \textbf{ then}$
  - $M_i \leftarrow \mathbf{T}[h]$
- $\textbf{else}$
  - $h \leftarrow H(M_i)$
  - $\mathbf{T}[h] \leftarrow M_i$

$\textbf{return } M_1 \parallel \cdots \parallel M_\ell$

SCB: Schematic Representation





SCB: Security

We show that SCB is secure if the underlying block cipher $\mathfrak{B} = (E, D)$ is a secure PRP

Theorem (Security)

For any $\mathrm{IND\text{-}CPA}$ *adversary* A *querying* $\beta \leq 2^\sigma$ *blocks we can construct a* PRP *adversary* B *such that*

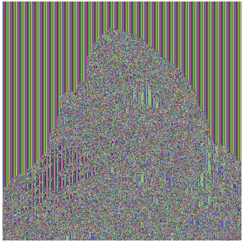
$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathsf{SCB}[\mathfrak{B},H]}(A) \leq \mathbf{Adv}^{\text{prp}}_{\mathfrak{B}}(B) + \frac{\beta^2}{2^n}$$

The additional term comes from:

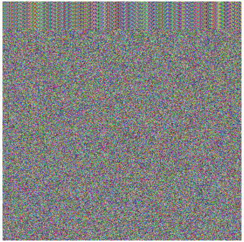
- The PRP/PRF switching lemma: $\beta(\beta - 1) \cdot 2^{-(n+1)}$
- The probability that a repetition signal collides with a previous block: $\beta \cdot 2^{-n}$

SCB: Visualizing Security

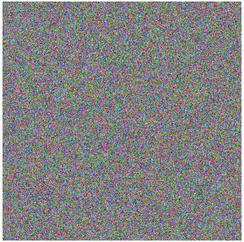
Original



ECB



SCB: $\sigma = 8, \tau = 32$



SCB: $\sigma = 16, \tau = 32$

$\beta = 512 \times 512 \times 3 \div 16 = 49\,152 \leq 2^{\sigma}$ only for $\sigma = 16$

SCB: Correctness

We show that SCB is correct if the underlying compression function H is collision resistant

Theorem (Correctness)

For any COR *adversary* A *querying* β *blocks we can construct a* CR *adversary* B *such that*

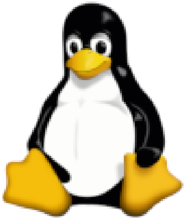
$$\mathbf{Adv}^{\mathrm{cor}}_{\mathsf{SCB}[\mathfrak{B},H]}(A) \leq \mathbf{Adv}^{\mathrm{cr}}_{H}(B) + \frac{2^{\sigma}\beta^2}{2^n}$$

The additional term comes from the probability that a new block looks like a repetition signal, that is:

1. When XORed with K_2 it has $n-\sigma-\tau$ leading zeros ($R < 2^{\sigma+\tau}$): $2^{-(n-\sigma-\tau)}$
2. Its last τ bits correspond to the hash of a previous block ($\mathbf{T}[h] \neq \perp$): $2^{-\tau}$

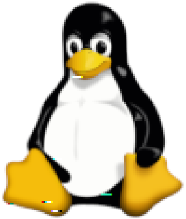
By the union bound: $\beta^2 \cdot 2^{-(n-\sigma-\tau)} \cdot 2^{-\tau} = \beta^2 \cdot 2^{\sigma-n}$

SCB: Visualizing Correctness

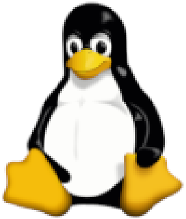


Original

SCB: $\sigma = 16, \tau = 8$



SCB: $\sigma = 16, \tau = 16$



SCB: $\sigma = 16, \tau = 24$

Only for $\tau = 24$ the original image was successfully recovered without any errors

SCB: Security-Correctness Trade-Off

Bounding the parameters σ and τ:

- From security we have $\beta \leq 2^\sigma$ and from correctness we have $\frac{2^\sigma \beta^2}{2^n} \ll 1$, hence:

  $$\log \beta \leq \sigma \ll n - 2 \log \beta$$

  Note: We should *minimize* σ

- From the birthday bound we have $\beta \ll 2^{\frac{\tau}{2}}$, and since $\sigma + \tau \leq n$:

  $$2 \log \beta \ll \tau \leq n - \sigma$$

  Note: We should *maximize* τ

Setting $\tau = n - \sigma$ would imply that the condition $R < 2^{\sigma+\tau}$ would always be true

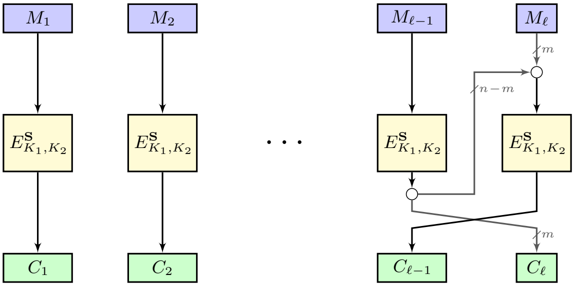
But for efficiency reasons, it might be still better *not* to set $\tau = n - \sigma$ (less look-ups in $\mathbf{T}$ on average)

VIL-SCB: Handling *any* Input Length

Recall: So far, SCB only handles messages over $(\{0,1\}^n)^+$

We extend SCB into a VIL-LPSE handling messages over $\{0,1\}^{\geq n}$ using **ciphertext stealing** (CTS)

Let $E^{\mathbf{S}}_{K_1,K_2}$ be the stateful *block enciphering* of SCB (the code inside the $\mathbf{for}$ loop)



Outline

Conclusions

We introduced the first semantically secure length-preserving encryption scheme

In the paper we also consider a variant that is secure and correct even if ciphertexts are *reordered*

We also identify possible improvements for future work:

- Checking counters upon decryption to remove factor 2^σ in correctness
- Is it possible to have better *state size growth*? (probably can't be zero)
- Are there other schemes with better security/correctness bounds?

Thank you for your attention!