Symmetric encryption – Block ciphers

Crypto Engineering

Bruno Grenet

Université Grenoble-Alpes

https://membres-ljk.imag.fr/Bruno.Grenet/CryptoEng.html

Symmetric part of the course

- 3 classes – each 3h with mixed CM and TD
  - Friday, September 23. *today!*
  - Thursday, September 29.
  - Friday, September 30.

Contents and goals

- Symmetric encryption, hashing, authentication
- Goals:
  - Understanding the models $\rightarrow$ what do we want to achieve?
  - Understanding *some* designs $\rightarrow$ how are they designed and why?
  - Understanding what can *go wrong* $\rightarrow$ what should you avoid?

What is *symmetric* cryptography?

- Cryptography: we want to hide stuff
- Symmetric: we assume a shared secret between participants
- Main question: when is the hiding *good enough*?

Before we start: Encryption cannot be deterministic!

1. Block ciphers

2. Symmetric encryption

Block ciphers: what do we want to achieve?

Goal: Symmetric Encryption

- Encryption: from a plaintext and a key $\rightarrow$ ciphertext**s** *non-determinism*
- Decryption: from a ciphertext and the key $\rightarrow$ plaintext
- Security: from a ciphertext alone $\rightarrow$ (almost) nothing

Objects

- Plaintext: any message $\in \{0,1\}^*$.
- Ciphertext: word $\in \{0,1\}^*$, of length as close to the message as possible efficiency
- Key: word $\in \{0,1\}^*$ not too large, not too small

Block cipher

- Plaintext / ciphertext: fixed-length *block size*
- One-to-one mapping for each key $\rightarrow$ deterministic!

Block ciphers are (mainly) a tool to build higher-level schemes

Block cipher: definition

Definition

A block cipher is a mapping $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}'$ such that for all $k \in \mathcal{K}$, $E(k, \cdot)$ is one-to-one, with

- $\mathcal{K} = \{0,1\}^\kappa$: the *key space* $\qquad \kappa \in \{\cancel{64}, \cancel{80}, \cancel{96}, \cancel{112}, 128, 192, 256\}$
- $\mathcal{M} = \{0,1\}^n$: the *message space* $\qquad n \in \{64, 128, 256\}$
- $\mathcal{M}' = \{0,1\}^{n'}$: usually the same as $\mathcal{M}$

$\to$ a block cipher is a family of permutations, indexed by the keys

What are *good* block ciphers?

Efficiency

- Fast: e.g. *few cycles per byte* on modern CPUs
- Compact: small code / small circuit size
- Easy to implement $\to$ avoid side-channel attacks, etc.
- …

Security

- Given $c = E(k, m)$, *hard* to find m without knowing k
- Given m, *hard* to compute c without knowing k
- Given *oracle access* to $E(k, \cdot)$, *hard* to find k
- Given *oracle access* to $E^{\pm}(k, \cdot)$, *hard* to find k — $E^{\pm}$: both E and E^{-1}

$\to$ Not enough! Ex.: given E, define $E'(k, x_L \| x_R) = x_L \| E(k, x_R)$

Need a *more general* **security definition**, that encompasses all of the above (and other)

In an ideal world

Definition

Let Perm_n the set of all $(2^n)!$ permutations of $\mathcal{M} = \{0,1\}^n$. A block cipher $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ is an ideal block cipher if for all $k \in \mathcal{K}$, $E(k,\cdot) \twoheadleftarrow \mathrm{Perm}_n$.

- As random as one could hope
- All keys provide perfectly random and independent permutations
- Non-realistic world:
  - $(2^n)^{2^{n-1}} < (2^n)! < (2^n)^{2^n}$
  - Key size $\simeq \log(2^n!) \simeq n \cdot 2^n$ bits $\quad n = 32 \Rightarrow 2^{37}$-bit keys!

Why *ideal*?

- Fix a key k and a subset $\mathcal{S} \subset \mathcal{M}$ of messages
- Assume an attacker knows: $E(k', m)$ for all $k' \in \mathcal{K} \setminus k$, and $E(k,m)$ for all $m \in \mathcal{M} \setminus \mathcal{S}$
- The attacker has no information about $E(k,m)$ for m in $\mathcal{S}$

PRP and strong PRP security

Informally, a block cipher is secure if its behavior is *close enough* to the ideal world

PRP experiment

- Fix a block cipher E
- A *challenger* gives an *attacker* access to an oracle $\mathcal{O}$:
  - either $\mathcal{O} \twoheadleftarrow \mathrm{Perm}_n$
  - or $\mathcal{O} = E(k, \cdot)$ where $k \twoheadleftarrow \mathcal{K}$
- The attacker must *distinguish* between the two cases
  - Answer 1 (say) if $\mathcal{O}$ is a random permutation, 0 otherwise
- *Strong PRP* experiment: oracle access to $\mathcal{O}^{\pm}$

Why does it encompass previous tentative requirements?

- If m can be found from $c = E(k, m)$ without k
  - Take any c and compute the corresponding m
  - Query the oracle on m and compare the result with c
- …

Formalization : (strong) PRP advantage

PRP advantage

$$\mathsf{Adv}_E^{\mathsf{PRP}}(q,t) = \max_{A_{q,t}^{\mathcal{O}}} \Big| \Pr\left[A_{q,t}^{\mathcal{O}}() = 1 : \mathcal{O} \twoheadleftarrow \mathsf{Perm}_n\right] - \Pr\left[A_{q,t}^{\mathcal{O}}() = 1 : \mathcal{O} = E(k,\cdot), k \twoheadleftarrow \mathcal{K}\right] \Big|$$

where $A_{q,t}^{\mathcal{O}}$ denotes an algorithm that runs in time $\leq t$ and makes $\leq q$ queries to $\mathcal{O}$

(Similarly for $\mathsf{Adv}^{\mathsf{SPRP}}$, with $\mathcal{O}^{\pm}$ in place of $\mathcal{O}$.)

- The PRP advantage provides a *measure* on the quality of a PRP, hence a block cipher
- The PRP advantage does *not* define when it is *good*

The generic attack

Challenger: Provides oracle access to either $\mathcal{O} \twoheadleftarrow \mathrm{Perm}_n$ or $\mathcal{O} = E(k, \cdot)$ with $k \twoheadleftarrow \mathcal{K}$

Attacker: Oracle access to $\mathcal{O}$, and knows what is $E : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$

1. Draw q messages $m_1, \ldots, m_q$ from $\mathcal{M}$ and t keys $k_1, \ldots, k_t$ from $\mathcal{K}$
2. Compute $C_{k_i} = [E(k_i, m_1), \ldots, E(k_i, m_q)]$ for $1 \leq i \leq t$
3. Query $\mathcal{O}$ on $m_1, \ldots, m_q$ to get $C = [\mathcal{O}(m_1), \ldots, \mathcal{O}(m_q)]$
4. Return 1 if there exists k_i s.t. $C = C_{k_i}$, 0 otherwise

Analysis

- Number of queries: q; running time: $O(qt)$
- $\Pr\left[A^{\mathcal{O}}_{q,t}() = 1 : \mathcal{O} \twoheadleftarrow \mathrm{Perm}_n\right] = \Pr\left[\exists k_i, \forall m_j, \mathcal{O}(m_j) = E(k_i, m_j)\right] \leq t/2^{(n-2)q}$
- $\Pr\left[A^{\mathcal{O}}_{q,t}() = 1 : \mathcal{O} = E(k, \cdot), k \twoheadleftarrow \mathcal{K}\right] \geq \Pr\left[\exists k_i, k = k_i\right] = t/2^\kappa$

$$\Rightarrow \mathrm{Adv}^{\mathrm{PRP}}_E(q, qt) \geq \frac{t}{2^\kappa} - \frac{t}{2^{(n-2)q}} \simeq \frac{t}{2^\kappa}$$

So, what are *good* PRPs or block ciphers?

No formal definition of a good PRP

Informal (equivalent) definitions

- $\mathrm{Adv}_E^{\mathrm{PRP}}(q, t) \simeq t/2^\kappa$
- The generic attack is almost the best possible
- The advantage is the same as for an ideal block cipher

Choice of parameter κ

- A good PRP is useless if κ is small
  - $\kappa \simeq 40$: breakable on ~ 1 day on my laptop
  - $\kappa \simeq 60$: breakable with a large CPU/GPU cluster (done in academia)
  - $\kappa \simeq 80$: breakable with an ASIC cluster (Bitcoin mining)
  - $\kappa \simeq 128$: seems hard enough
- Other considerations (application dependent, quantum computers, etc.)

Finally

In practice

- AES - Rijndael:
  - Most used block cipher nowadays
  - Standardized by the NIST, replacement of DES (considered broken: 56-bit key)
  - Block size $n = 128$ bits
  - Key size $\kappa = 128$, 196 or 256 bits
- Other (less used) possibilities:
  - Camellia: $n = 128$, $\kappa = 128$, 192 or 256
  - SHACAL-2: $n = 128$, $\kappa = 512$

In theory

- Similar notion of (strong) PRF advantage: replace Perm_n with Func_n
- *PRP-PRF switching* $\simeq$ "a good PRP is also a good PRF" *cf.* Adv. Crypto

1. Block ciphers

2. Symmetric encryption

Block ciphers are not enough

Block ciphers offer

- One-to-one (deterministic) encryption
- Fixed-size messages

We need

- One-to-many (non-deterministic) encryption
- Variable-size messages

The tool: modes of operations

- Transforms a block cipher into a *symmetric encryption scheme*

$$E : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n \rightsquigarrow \begin{cases} \mathsf{Enc} : \{0,1\}^\kappa \times \{0,1\}^\ell \times \{0,1\}^* \to \{0,1\}^* \\ \mathsf{Dec} : \{0,1\}^\kappa \times \{0,1\}^* \to \{0,1\}^* \end{cases}$$

  - For all $(k, r, m) \in \{0,1\}^\kappa \times \{0,1\}^\ell \times \{0,1\}^*$, $\mathsf{Dec}(\mathsf{Enc}(k, r, m)) = m$
  - $r \in \{0,1\}^\ell$: non-determinism
- A mode is *good* if it turns *good BCs* into *good encryption schemes*

What is a good encryption scheme?

IND-CPA security for symmetric encryption

IND-CPA experiment for $\mathsf{Enc} : \mathcal{K} \times \mathcal{R} \times \mathcal{M} \to \mathcal{M}$

Challenger	draws $k \twoheadleftarrow \mathcal{K}$
Adversary	submits queries x_i to the attacker and gets $\mathsf{Enc}(k, r_i, x_i)$
Adversary	creates two equal-length messages m_0 and m_1 and submits them
Challenger	draws $b \twoheadleftarrow \{0, 1\}$ and answers with $\mathsf{Enc}(k, r, m_b)$
Adversary	tries to guess b

(choice of r_i, r is defined by the mode, can be ignored)

IND-CPA advantage

$$\mathsf{Adv}^{\mathsf{IND-CPA}}_{\mathsf{Enc}}(q, t) = \max_{A^{\mathsf{Enc}}_{q,t}} \left| \Pr\left[A^{\mathsf{Enc}}_{q,t} \text{ succeeds}\right] - \frac{1}{2} \right|$$

where $A^{\mathsf{Enc}}_{q,t}$ is an alg. that runs in time $\leq t$ and makes $\leq q$ queries to the challenger

Comments on IND-CPA security

$$\mathrm{Adv}^{\mathrm{IND-CPA}}_{\mathrm{Enc}}(q,t) = \max_{A^{\mathrm{Enc}}_{q,t}} \left| \Pr\left[A^{\mathrm{Enc}}_{q,t} \text{ succeeds} \right] - \frac{1}{2} \right|$$

- IND-CPA: Indistinguishability under chosen plaintext attack
- $\frac{1}{2}$: stupid attacker that guesses b at random
- With q, t large enough: advantage $\frac{1}{2}$

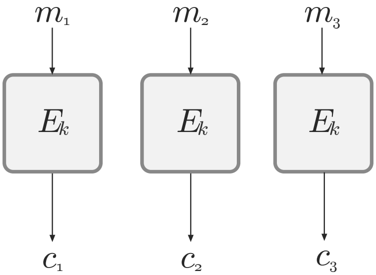
computational security

- IND-CPA $\Rightarrow$ non-determinism
- IND-CPA $\Rightarrow$ the attacker cannot find a single bit of the message

Stronger notions: IND-CCA and IND-CCA2

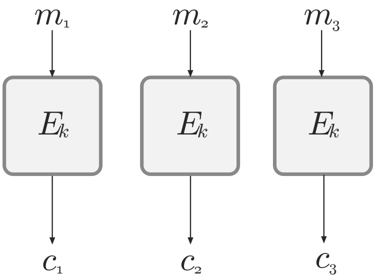
- Indistinguishability under chosen **cipher**text attack
- Access to both an encryption oracle and a decryption oracle
- 2 variants: non-adaptative (IND-CCA) or adaptative (IND-CCA2)

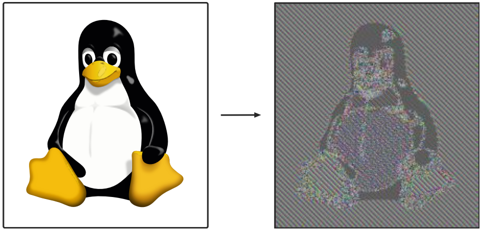
First (bad) example of mode of operation: Electronic Code Book (ECB)



Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

First (bad) example of mode of operation: Electronic Code Book (ECB)

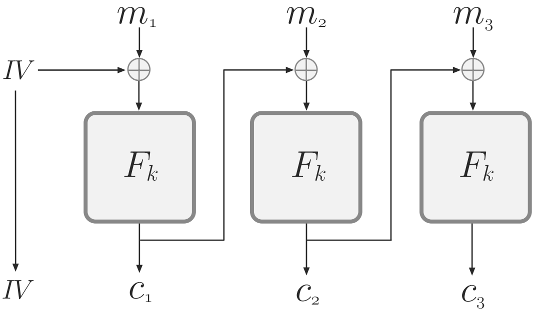




Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

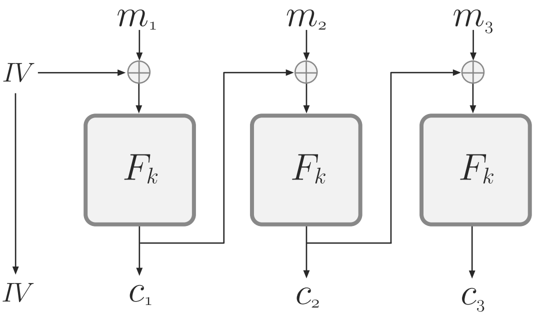
Source : Wikipédia (modif.)

Second (real) example of mode of operation: Cipher Block Chaining (CBC)



Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

Second (real) example of mode of operation: Cipher Block Chaining (CBC)



- IND-CPA security if E is a good PRP and IV truly random
- Assume IV not random:
  - Adversary sends a query m and gets first IV r and $c = E(k, m \oplus r)$
  - Assume adversary knows that for next IV r', $\Pr[r' = x]$ is *large*
  - Adversary sends challenges $m_0 = m \oplus r \oplus x$ and $m_1 = m_0 \oplus 1$
  - Gets back $r' \| c_b = \mathrm{Enc}(m_b)$ with $b \twoheadleftarrow \{0, 1\}$
  - If $c_b = c$, guess $b = 0$, else $b = 1$

Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

Generic CBC collision attack

Observation

- For fixed k, $E(k, \cdot)$ is a permutation $\rightarrow E(k, x) = E(k, y) \iff x = y$
- In CBC, inputs to E are of the form $m_i \oplus y$ with
  - m_i a message block,
  - y either an IV or a ciphertext block
- In particular: $E(k, m_i \oplus c_{i-1}) = E(k, m'_j \oplus c'_{j-1}) \iff m_i \oplus c_{i-1} = m'_j \oplus c'_{j-1}$

Consequence

- Assume we get two identical ciphertext blocks $c_i = c'_j$
  $\iff E(k, m_i \oplus c_{i-1}) = E(k, m'_j \oplus c'_{j-1})$
  $\iff m_i \oplus c_{i-1} = m'_j \oplus c'_{j-1}$
  $\iff c_{i-1} \oplus c'_{j-1} = m_i \oplus m'_j$
- That is: c_{i-1} and c'_{j-1} reveal information about m_i and m'_j
  $\Rightarrow$ breaks IND-CPA security (no matter how good E!)

Probability to get collisions?

Assumption

The distribution of the $(m_i \oplus c_{i-1})$ is approx. uniform

- If c_0 is the IV, it has to be approx. uniform
- If c_{i-1} is a ciphertext, non (approx.) uniformity would imply an attack

Birthday bound

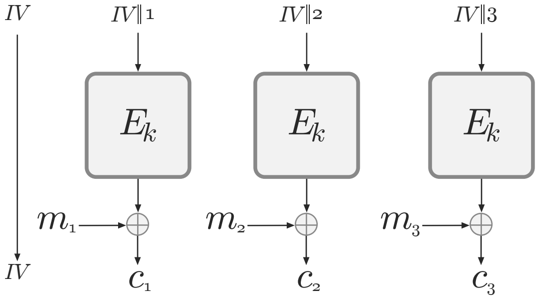
Draw $y_1, \dots, y_q$ uniformly from a size-N set, with $q \leq \sqrt{2N}$. Then

$$\frac{q(q-1)}{4N} \leq 1 - e^{-q(q-1)/2N} \leq \boxed{\Pr\left[\exists i \neq j, y_i = y_j\right]} \leq \frac{q(q-1)}{2N}$$

Consequence

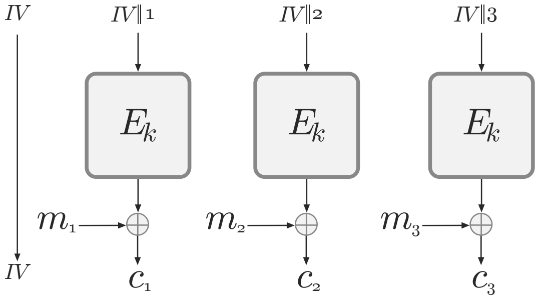
- Collision found w.h.p. if $q \simeq \sqrt{N}$
- For CBC: Collision w.h.p. after observing $\simeq 2^{n/2}$ ciphertext blocks
- Note: does not depend on key size κ

Last (classic) mode of operation: Counter (CTR)



Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

Last (classic) mode of operation: Counter (CTR)



- Parallel encryption (fast!)
- Looks like a stream cipher
- Sensitive to birthday bound

Security

If E is a *good* PRF, IND-CPA security

Source : J. Katz, Y. Lindell. Introduction to modern cryptography. 3rd ed, CRC Press, 2021. (modif.)

Finally

Modes of operations

- A *good* mode of operation turns a *good* block cipher into a *good* symmetric encryption scheme
- Different mode of operations require different quality for the block cipher
  - *Good* PRP
  - *Good* PRF
  - Ideal Block Cipher
- Proofs of security $\rightarrow$ reductions between problems
- Usually: need more $\rightarrow$ *ad hoc* analysis of the resulting system

Other symmetric encryption schemes

- Other modes of operations
- Stream ciphers (Wifi, 5G, ...)

Conclusion

Symmetric encryption, as we saw it

- Two ingredients:
  - a block cipher — fixed-size, deterministic
  - a mode of operation — variable-size, non-deterministic
- Security notions:
  - PRP advantage — block cipher
  - IND-CPA advantage — symmetric encryption
- More advanced security definitions:
  - strong PRP adv., (strong) PRF adv., ideal block cipher
  - IND-CCA, IND-CCA2

In practice

- Block cipher: mainly AES, with key size 128 bits
- Modes of operations: *e.g.* extension of CTR in TLS

Final words: **Definitions and proofs are important!**