Introduction to Cryptography

INF-744: Security and Privacy for IoT

Diego F. Aranha

Institute of Computing – University of Campinas

Definitions

Classical

Cryptography is etymologically the “art of secret writing”.

Modern

Cryptography is the art and science that studies techniques for providing security properties, such as confidentiality, origin authentication, integrity and non-repudiation, primarily in computational systems.

Cryptanalysis

Techniques to analyze (and break) cryptographic methods.

Cryptology = Cryptography + Cryptanalysis

Terminology

Sets:

- Alphabet of definition $\mathcal{A}$.
- Plaintext space $\mathcal{M}$.
- Ciphertext space $\mathcal{C}$.
- Key space $\mathcal{K}$.

Algorithms:

- Bijection $E_e : \mathcal{M} \to \mathcal{C}$ parameterized by key $e \in \mathcal{K}$.
- Bijection $D_d : \mathcal{C} \to \mathcal{M}$ parameterized by key $d \in \mathcal{K}$.

Cryptosystem

A **cryptosystem** is given by
$\{\{E_e : e \in \mathcal{K}\}, \{D_d : d \in \mathcal{K}\} | \forall e \in \mathcal{K}, \exists d \in \mathcal{K}, D_d = E_e^{-1}\}$.

Consistency: $\forall m \in \mathcal{M}, D_d(E_e(m)) = m$.

Terminology

Communicating entities:

- Parties.
- Sender, receiver.
- Adversary.

Channels:

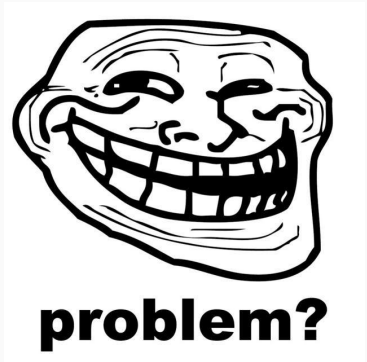
- Insecure.
- Secure.
- Physically secure.

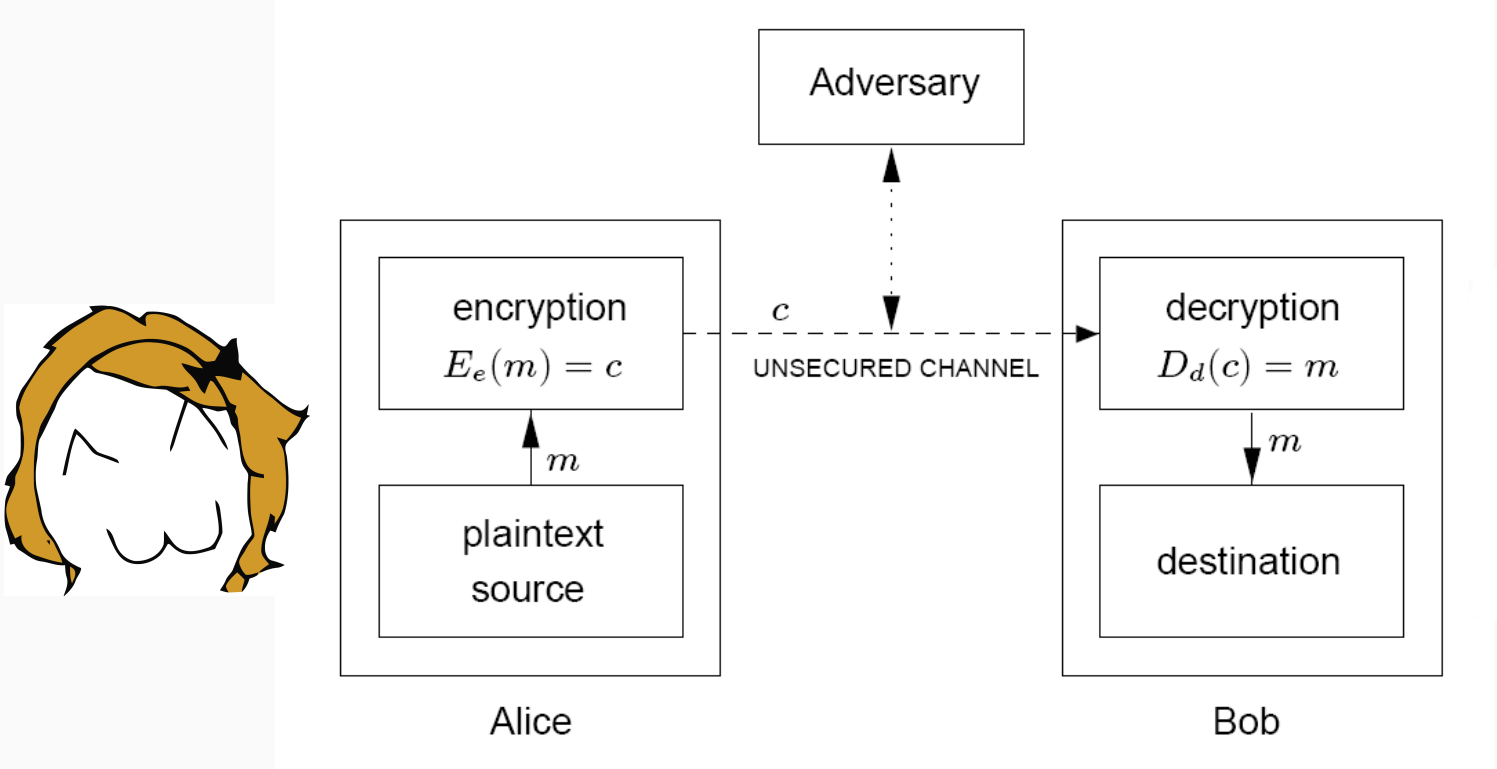
Security

A cryptography system is **breakable** is an adversary can systematically recover messages from ciphertext without knowledge of (d, e) in **reasonable** time.

Important: How much time is reasonable?

Cryptosystems

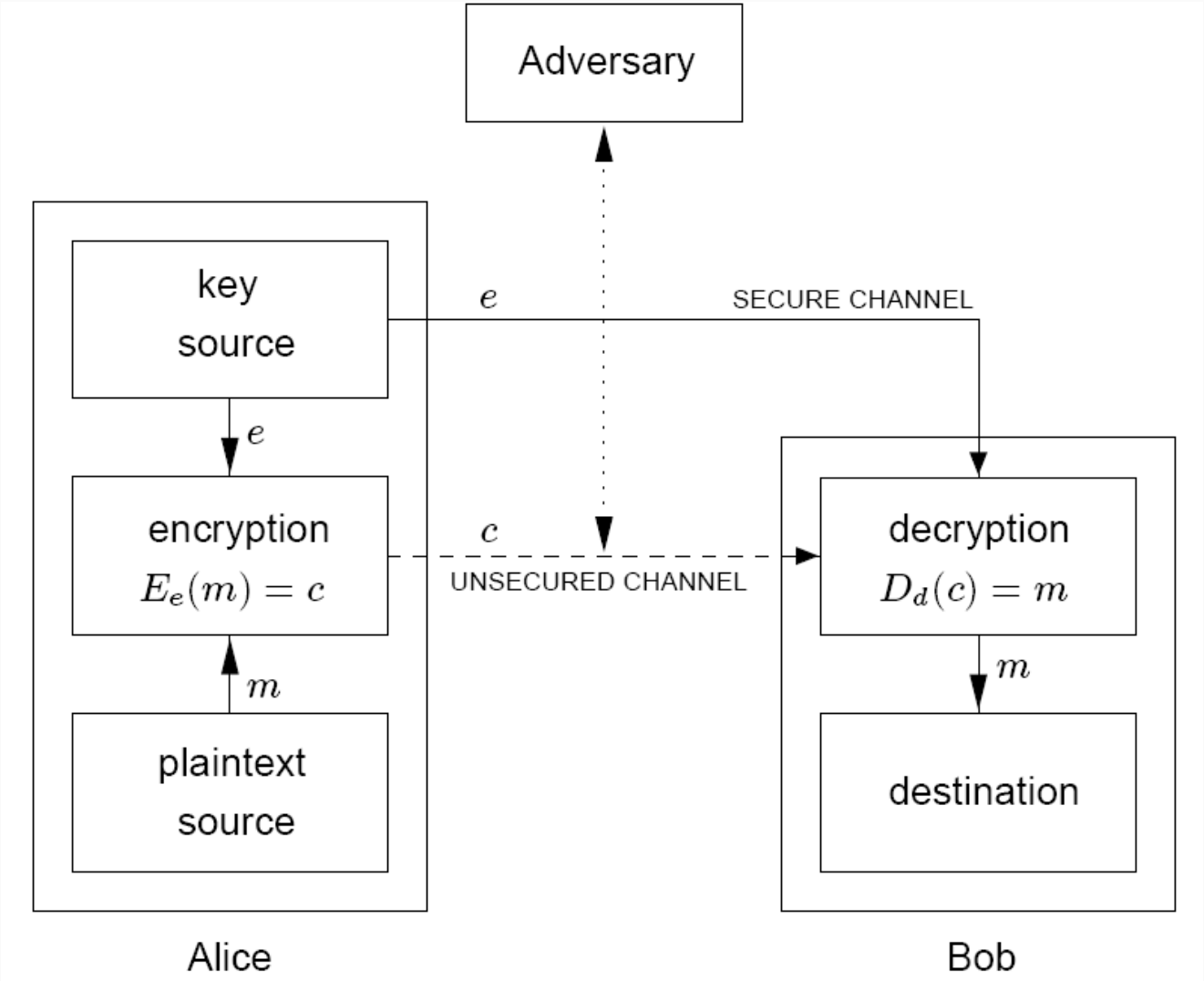




Classification: symmetric, asymmetric.

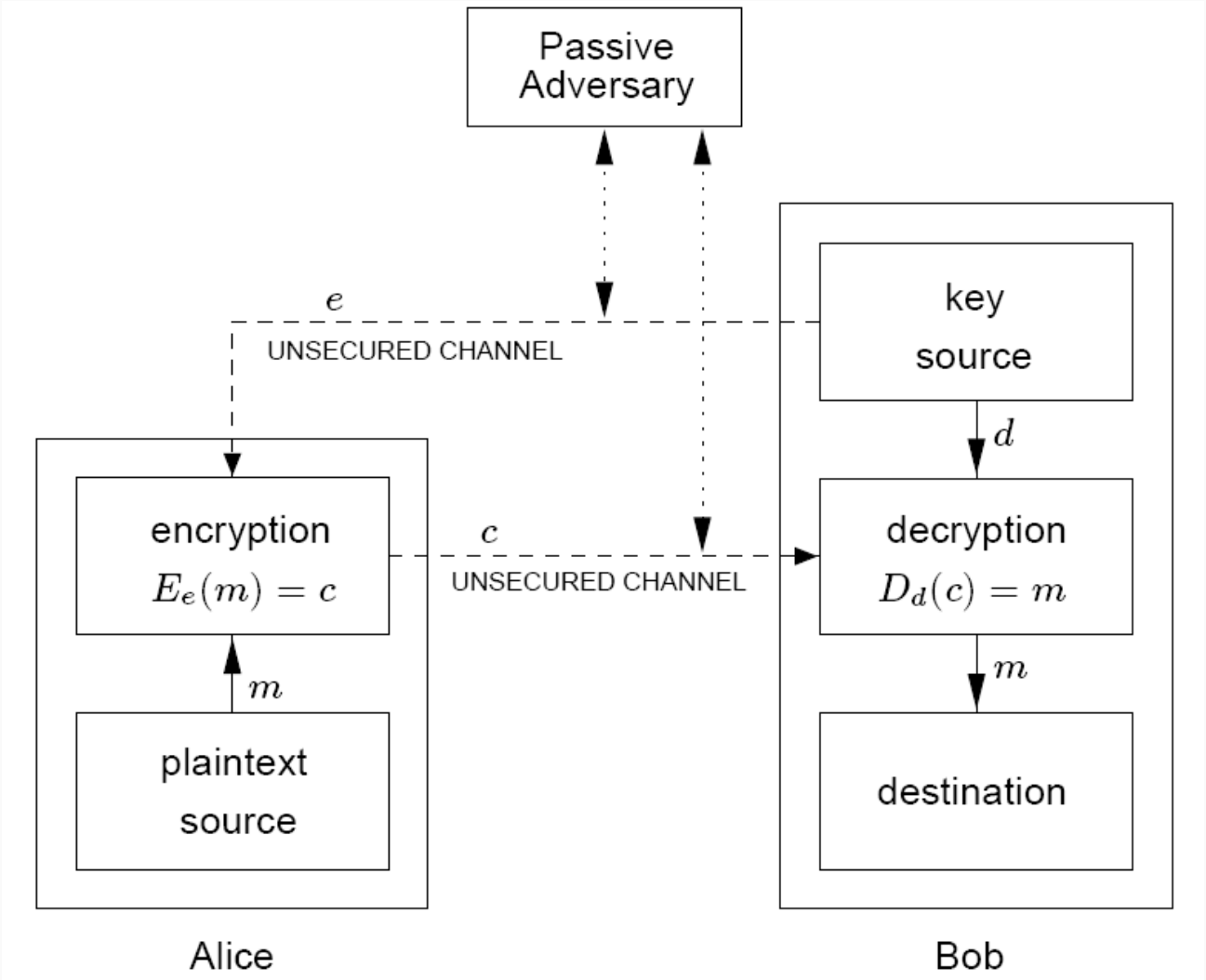
Symmetric cryptosystem

If it is possible to efficiently compute d from e:



Asymmetric system

If no method is known to efficiently compute d from e:



Requirements for cryptosystems

Kerckhoffs' Principles (*La Cryptographie Militaire*, 1883):

- The system must be practically, if not mathematically, indecipherable.
- It should not require secrecy, and it should not be a problem if it falls into enemy hands.
- It must be possible to communicate and remember the key without using written notes.
- It must be applicable to telegraph communications.
- It must be easy to use, efficient and portable.

Requirements for cryptosystems

Kerckhoffs' Principles (*La Cryptographie Militaire*, 1883):

- The system must be practically, if not mathematically, indecipherable.
- It should not require secrecy, and it should not be a problem if it falls into enemy hands.
- It must be possible to communicate and remember the key without using written notes.
- It must be applicable to telegraph communications.
- It must be easy to use, efficient and portable.

Modern adaptations:

- The system must be **secure**.
- Security should reside on the keys, not algorithms.
- The key must have polynomial length in the security parameter.
- The ciphertext must have polynomial length in the plaintext size.
- The encryption/decryption bijections should be computed in polynomial time.

Security of cryptosystems

Unconditional (information-theoretic) security:

- Ciphertext does not reveal any information about the plaintext.
- Perfect secrecy.

Complexity-theoretic security:

- Adversary is limited to polynomial computational power.
- Threat model for a certain security level.
- Security analysis under plausible assumptions.

In practice

A cryptosystem is considered secure if the best known attack against the system is not more efficient than **exhaustive search** in the key space.

Attacks against cryptosystems

Ciphertext-only attack:

- Attacker tries to obtain key from ciphertext only.

Known-plaintext attack:

- Adversary has access to plaintext and corresponding ciphertext.

Chosen-plaintext attack:

- Adversary **chooses** plaintext to be encrypted and receives corresponding ciphertext.

Chosen-ciphertext attack:

- Adversary **chooses** ciphertexts to be decrypted.

Monoalphabetic ciphers

Definition

A **monoalphabetic cipher** $E_\pi : \mathcal{M} \to \mathcal{C}$ is a rule to replace each letter m_i from message m with $\pi(m_i)$, where π defines a permutation in the alphabet of definition.

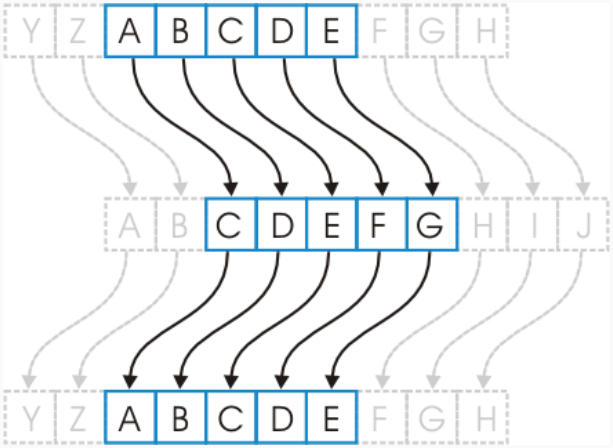
In other words:

- The key is the permutation $\pi : \mathcal{A} \to \mathcal{A}$.
- The encryption function is $E_\pi(m) = (\pi(m_1), \pi(m_2), \ldots, \pi(m_{|m|})$.
- The decryption function is $D_\pi(c) = (\pi^{-1}(c_1), \pi^{-1}(c_2), \ldots, \pi^{-1}(c_{|c|})$

Observations:

- The key space has size $(|\mathcal{A}|!)$.
- In general, the key has size $|\mathcal{A}|$.

Caesar cipher (Caesar, 58 B.C.)



Definition

It is a special case of monoalphabetic substitution where each letter is replaced by the letter after k positions.

In other words:

- The key is the number k.
- The permutation is given by $\pi(m_i) = (m_i + k) \bmod |\mathcal{A}|$.

Observations:

- What is the key size?

Transposition ciphers

Definition

A **transposition** $E_\theta : \mathcal{M} \to \mathcal{C}$ exchanges the i-th position letter m_i from message m by $\theta(i)$, where θ defines a permutation in the set $\{1, 2, \ldots, |m|\}$.

In other words:

- The key is the permutation $\theta : \{1, 2, \ldots, |m|\} \to \{1, 2, \ldots, |m|\}$.
- The encryption function is $E_\theta(m) = (m_{\theta(1)}, m_{\theta(2)}, \ldots, m_{\theta(|m|)})$.
- The decryption function is $D_\theta(c) = (c_{\theta^{-1}(1)}, c_{\theta^{-1}(2)}, \ldots, c_{\theta^{-1}(|m|)})$.

Observations:

- The key space has size $(|m|!)$.
- In the general case, the key has size $|m|$.

Scytale (Sparta, 500 B.C.)

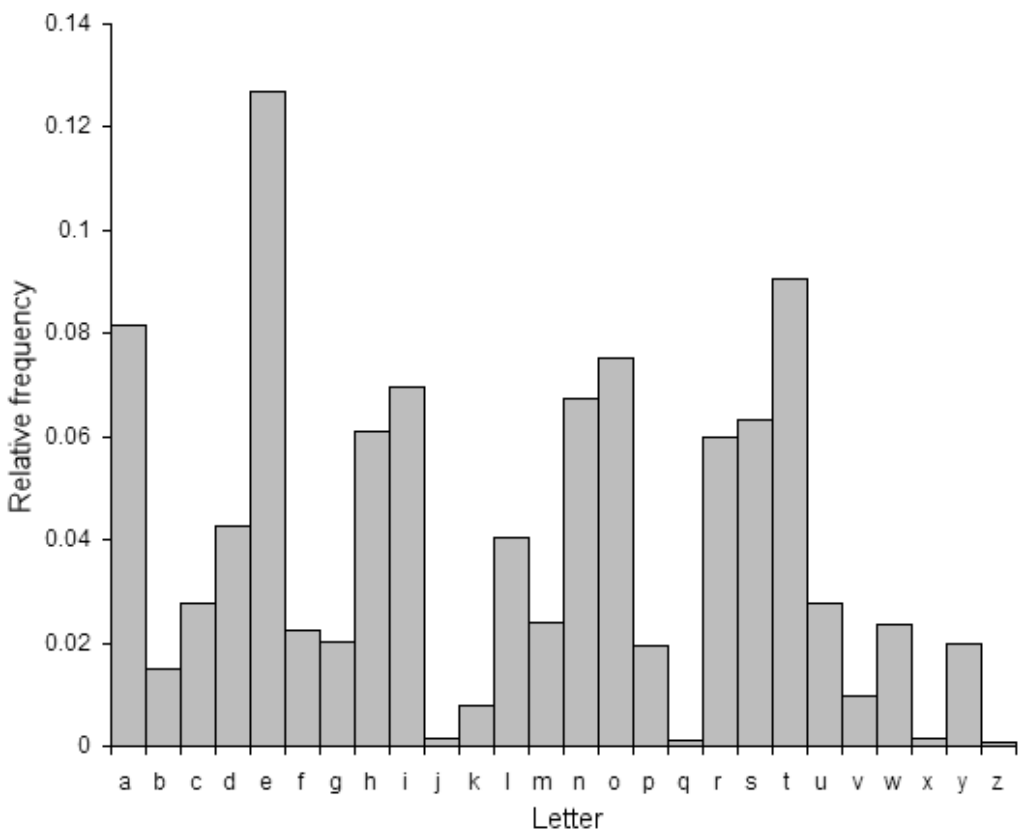
Cryptanalysis by frequency analysis (Al-Kindi, 800)

Al-Kindi writes “Manual for decrypting encrypted messages”:

Cryptanalysis by frequency analysis (Al-Kindi, 800)

Al-Kindi writes “Manual for decrypting encrypted messages”:

- Transposition preserves the exact frequencies of the letters in the original text.
- Monoalphabetic ciphers permute the frequencies of letters.



One-time pad (Vernam, 1925)

Definition

A **one-time pad** is a cipher where the key is randomly selected, never repeated and has the same length as the plaintext.

One of the biggest contributions given to Cryptography, because of the security upper bound!

Disadvantages:

- Generating truly random data is **hard**.
- Distributing large keys is **hard**.

Perfect secrecy (Shannon, 1949)

Definition

A cipher exhibits the property of perfect secrecy if a ciphertext c does not reveal any information about a message m (except its length).

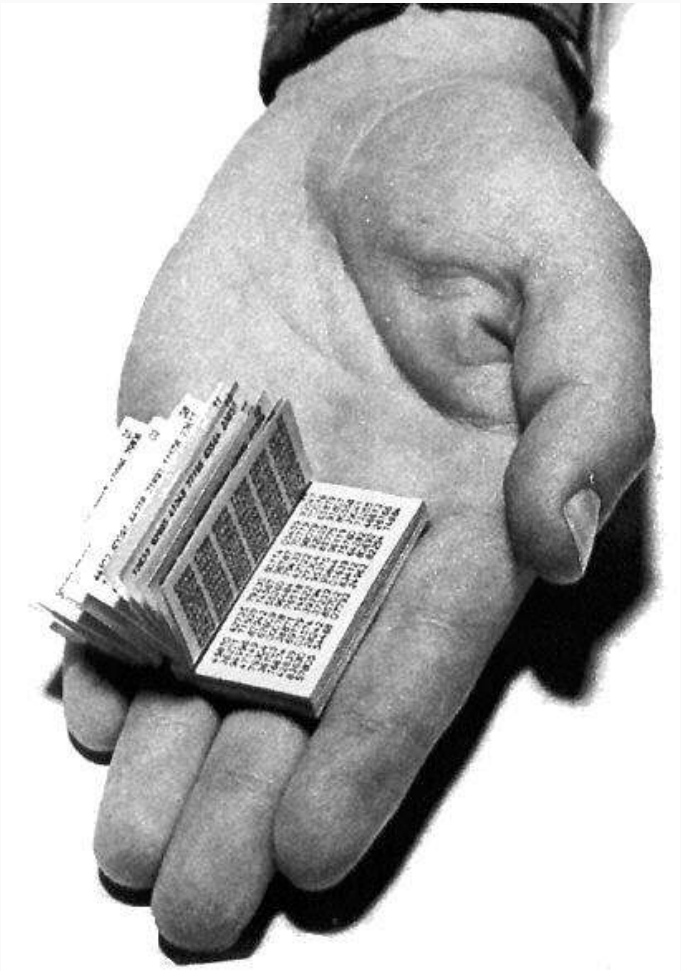
$$Pr[m = m'] = Pr[m = m'|c] \tag{1}$$

Shannon proved that the *one-time pad* provides perfect secrecy.

The intuition behind this is that an *one-time pad* ciphertext can be decrypted to any message.

Observation: One of the few cryptosystems to provide unconditional security !

One-time pad (Vernam, 1925)



Product of ciphers (Shannon, 1949)

Shannon observed that:

- Composing ciphers were already common.
- Substitution ciphers and transpositions ciphers are not secure.
- Substitution ciphers unlink message and ciphertext.
- Transposition ciphers spread the message redundancy.

Conclusion: composing these ciphers can be more secure!

Definitions

A **product of ciphers** is the composition of t encryption functions $E_{k_1}, E_{k_2}, \ldots, E_{k_t}$, where each function E_{k_i} is a substitution or transposition. An **iterated cipher** is a cipher represented through the repetition of a composition of elementary ciphers.

Substitution ciphers add **confusion** and transposition ciphers add **diffusion** to the encryption process.

Importance of precise definitions

Example: How to formalize security of a cryptosystem with relation to confidentiality?

Answer 1

Secure if an adversary cannot obtain the key from ciphertext.

Importance of precise definitions

Example: How to formalize security of a cryptosystem with relation to confidentiality?

Answer 1

Secure if an adversary cannot obtain the key from ciphertext.

Answer 2

Secure if an adversary cannot obtain the plaintext from ciphertext.

Importance of precise definitions

Example: How to formalize security of a cryptosystem with relation to confidentiality?

Answer 1

Secure if an adversary cannot obtain the key from ciphertext.

Answer 2

Secure if an adversary cannot obtain the plaintext from ciphertext.

Answer 3

Secure if an adversary cannot determine a single letter of the plaintext from the ciphertext.

Importance of precise definitions

Example: How to formalize security of a cryptosystem with relation to confidentiality?

Answer 1

Secure if an adversary cannot obtain the key from ciphertext.

Answer 2

Secure if an adversary cannot obtain the plaintext from ciphertext.

Answer 3

Secure if an adversary cannot determine a single letter of the plaintext from the ciphertext.

Answer 4

Secure if an adversary cannot obtain plaintext information from ciphertext only.

Importance of precise definitions

Example: How to formalize security of a cryptosystem with relation to confidentiality?

Answer 1

Secure if an adversary cannot obtain the key from ciphertext.

Answer 2

Secure if an adversary cannot obtain the plaintext from ciphertext.

Answer 3

Secure if an adversary cannot determine a single letter of the plaintext from the ciphertext.

Answer 4

Secure if an adversary cannot obtain plaintext information from ciphertext only.

Final Answer

Secure if an adversary cannot compute a function of the plaintext from ciphertext only.

Stream ciphers

There are two main types of symmetric ciphers:

1. **Block ciphers** such as AES and DES encrypt a fixed amount of plaintext (**block**) per invocation.
2. **Stream ciphers** encrypt bits individually with a **key stream**, and thus a variable amount of plaintext per invocation. The key stream is computed from the symmetric key K.

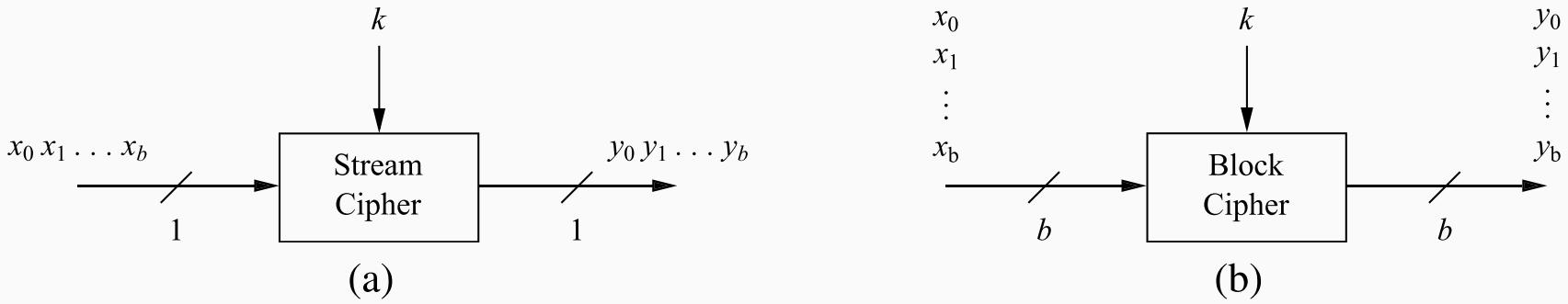


Figure 1: Operation of a stream cipher (a) and a block cipher (b).

Stream ciphers

Stream ciphers can be further classified in two types:

1. In **synchronous** stream ciphers, the key stream depends only on the key.
2. In **asynchronous** stream ciphers, the key stream also depends on the ciphertext.

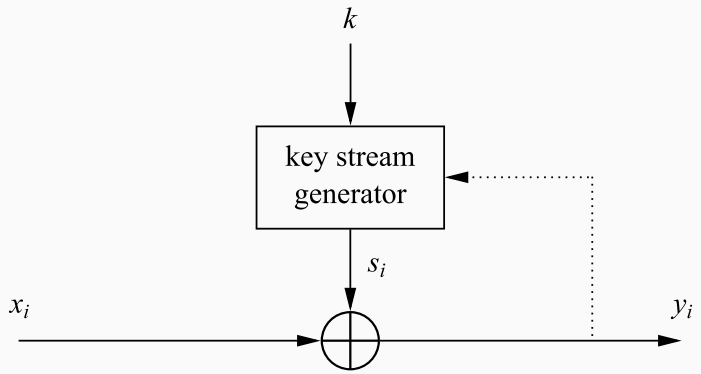


Figure 2: Synchronous and asynchronous (pointed line) stream cipher.

Stream ciphers

Features of block ciphers:

- Block ciphers are better understood in terms of security.
- Block ciphers tend to be slower in software.

Features of stream ciphers:

- Stream ciphers are more recent, and thus harder to construct securely.
- Stream ciphers tend to be more efficient, both smaller and faster in software.

Important: Beware of some **broken** stream ciphers, such as RC4.

Stream ciphers

Definition

The plaintext x, ciphertext y and key stream k consist of individual bits. Thus, $\mathcal{C} = \mathcal{M} = \mathcal{K} = \{0,1\}^*$.

The encryption function is given by $y_i = E_{k_i}(x_i) = x_i + k_i \bmod 2$.

The decryption function is given by $x_i = D_{k_i}(y_i) = y_i + k_i \bmod 2$.

Question: Why addition modulo 2?

Stream ciphers

Definition

The plaintext x, ciphertext y and key stream k consist of individual bits. Thus, $\mathcal{C} = \mathcal{M} = \mathcal{K} = \{0,1\}^*$.

The encryption function is given by $y_i = E_{k_i}(x_i) = x_i + k_i \bmod 2$.

The decryption function is given by $x_i = D_{k_i}(y_i) = y_i + k_i \bmod 2$.

Question: Why addition modulo 2?

Important: The encryption and decryption functions are essentially the same! Let's prove it.

Data Encryption Standard (IBM, 1975)

The standard was defined jointly by IBM and the NSA (*National Security Agency*), from the *Lucifer* block cipher, designed by Feistel.

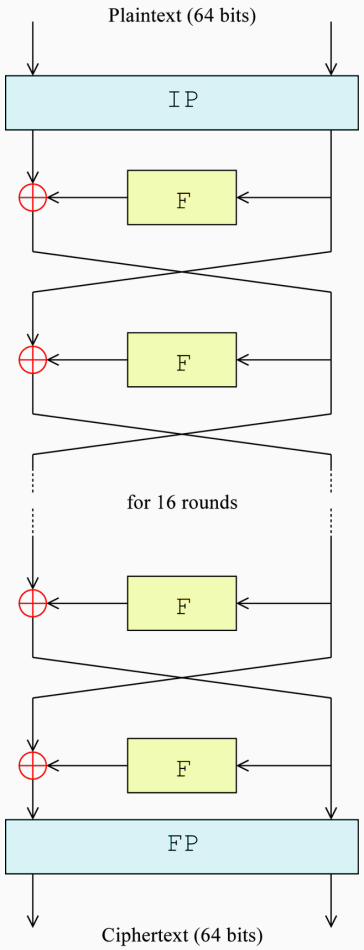
NSA wanted a secure cipher to everyone, but not *too* secure:

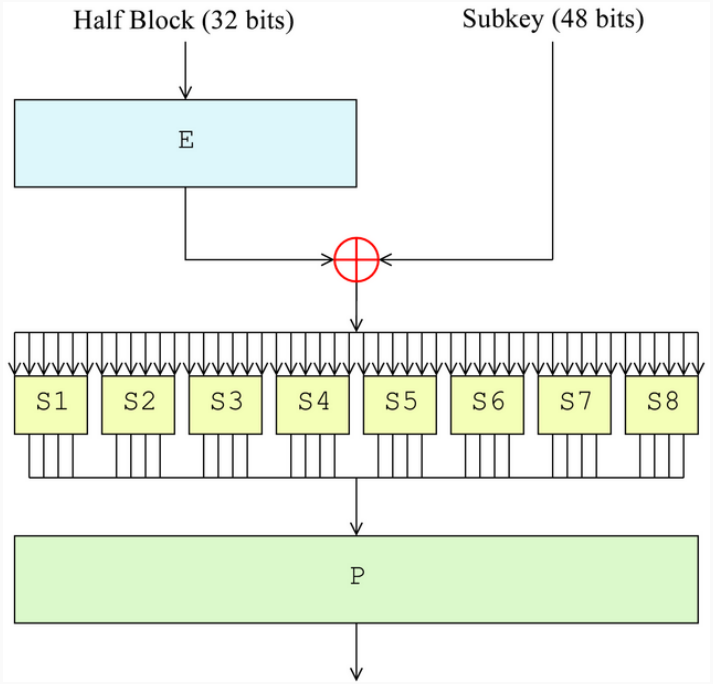
- Substitution boxes were chosen to improve resistance against differential cryptanalysis.
- Key length was reduced from 64 to 56 bits.

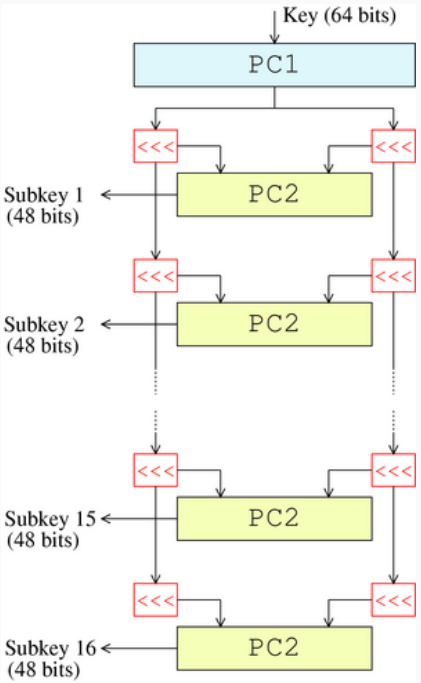
Conclusion: Never trust cryptographic standards to intelligence agencies!

Warning: Avoid DES at all costs, use AES!

Data Encryption Standard (IBM, 1975)







Advanced Encryption Standard (2001, NIST)

Definition

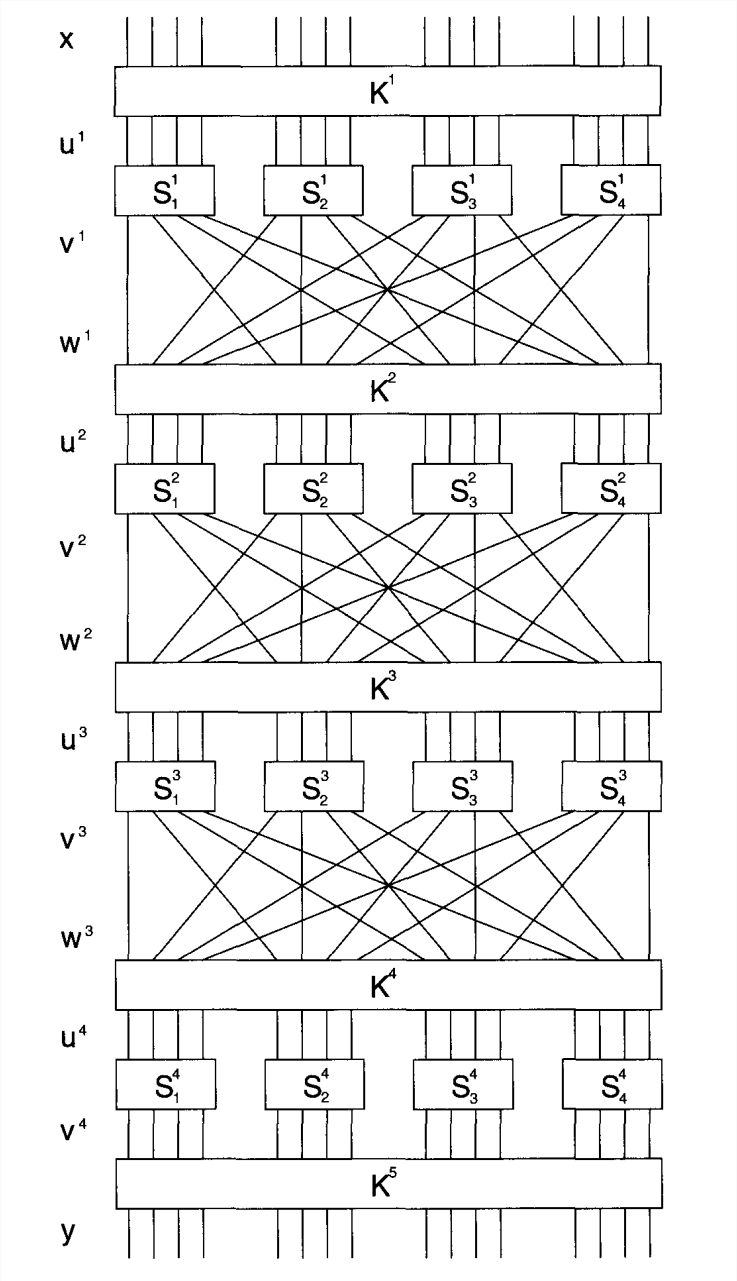
A **substitution-permutation network** (SPN) is an iterated cipher where the round function is represented by the composition of substitution and transposition ciphers.

Features:

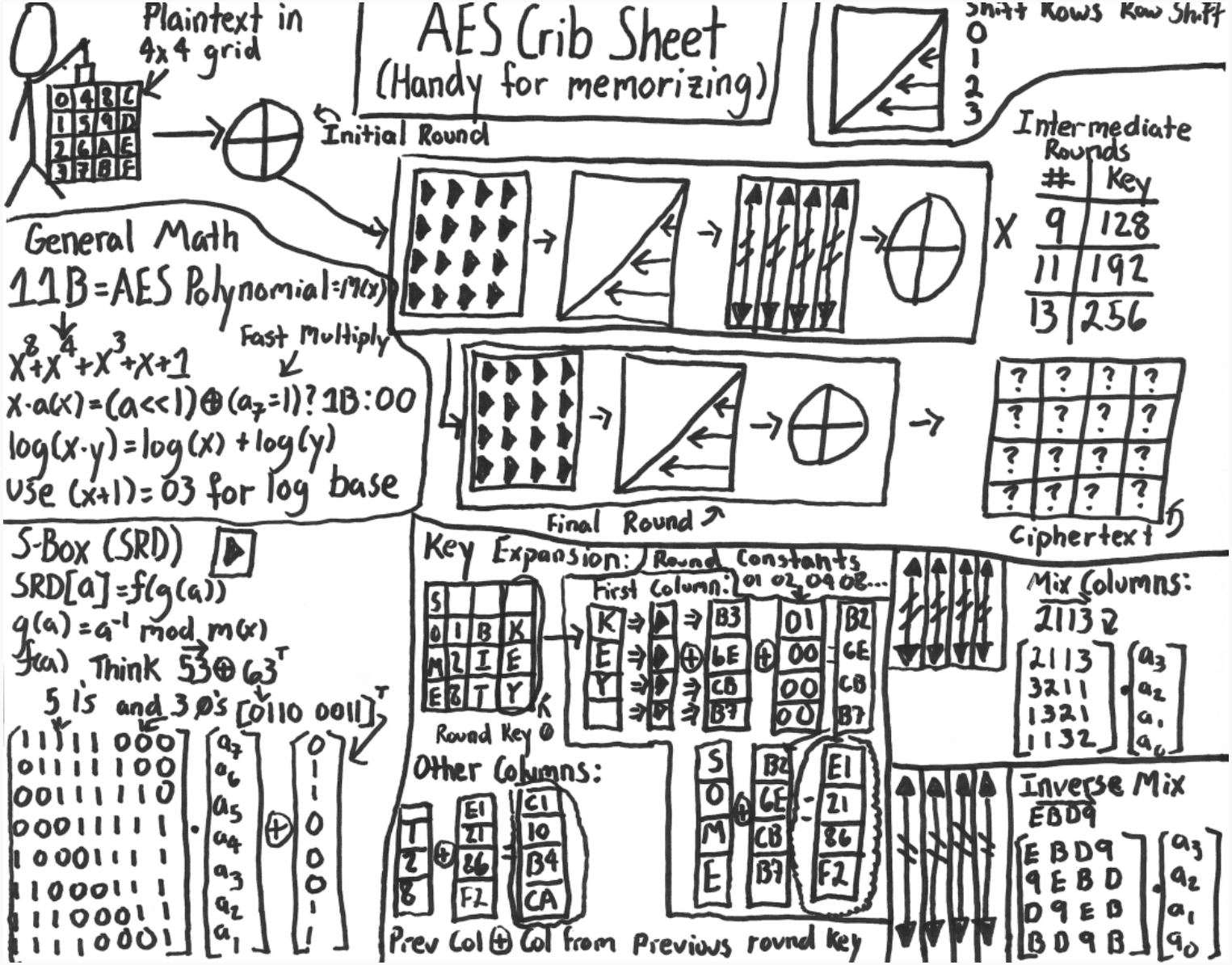
- Three security levels: 128, 192 e 256 bits.
- $\mathcal{M} = 0, 1^{128}, \mathcal{K} = \{0,1\}^{128}, \{0,1\}^{192}, \{0,1\}^{256}$.
- $Nr = 10, 12, 14$, respectively, $lm = 128$.
- Follows the SPN paradigm.

Curiosity: Implemented as native instruction in modern Intel processors!

Advanced Encryption Standard (2001, NIST)



Advanced Encryption Standard (2001, NIST)



Modes of operation

Definition

A **mode of operation** is a procedure for using a secure block cipher to encrypt messages of arbitrary lentgh with the same key.

Originally:

- *Electronic Code Book (ECB)*;
- *Output Feedback (OFB)*;
- *Cipher Feedback (CFB)*;
- *Cipher Block Chaining (CBC)*.

More recently:

- *Counter Mode (CTR)*;
- *Counter with Cipher Block Chaining (CCM)*;
- *Galois Counter Mode (GCM)*.

Modes of operation

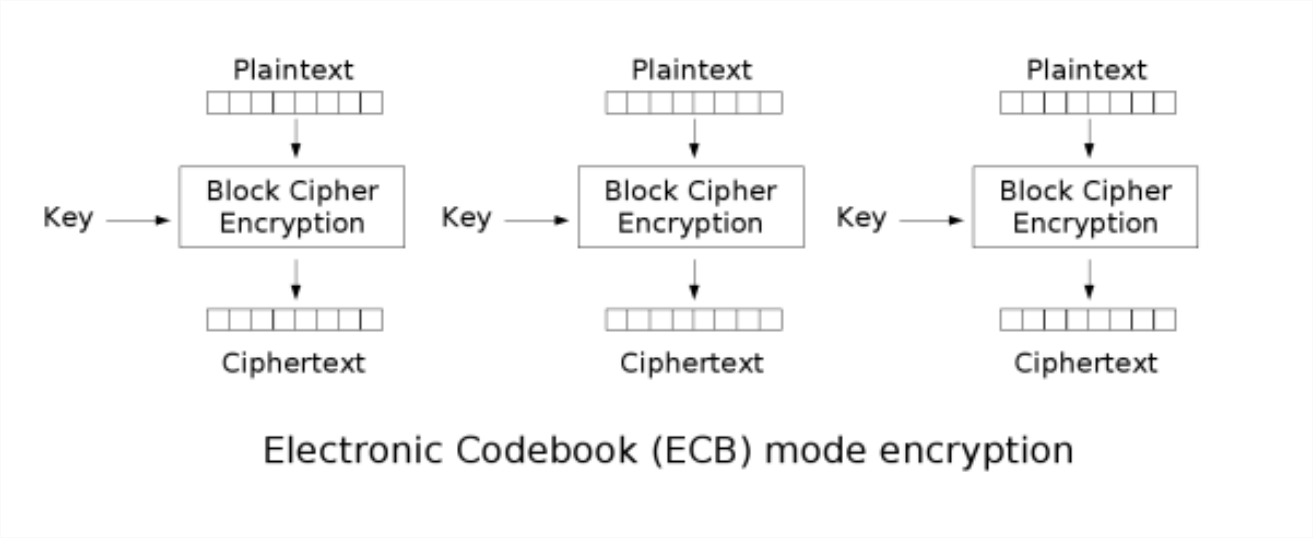
Definition

An **initialization vector** (IV) is a block employed for several modes of operation to randomize the encryption process and guarantee different output blocks even when the same block is encrypted under the same key. Security requirements for the IV are different from the key, but usually require its uniqueness and rarely its confidentiality.

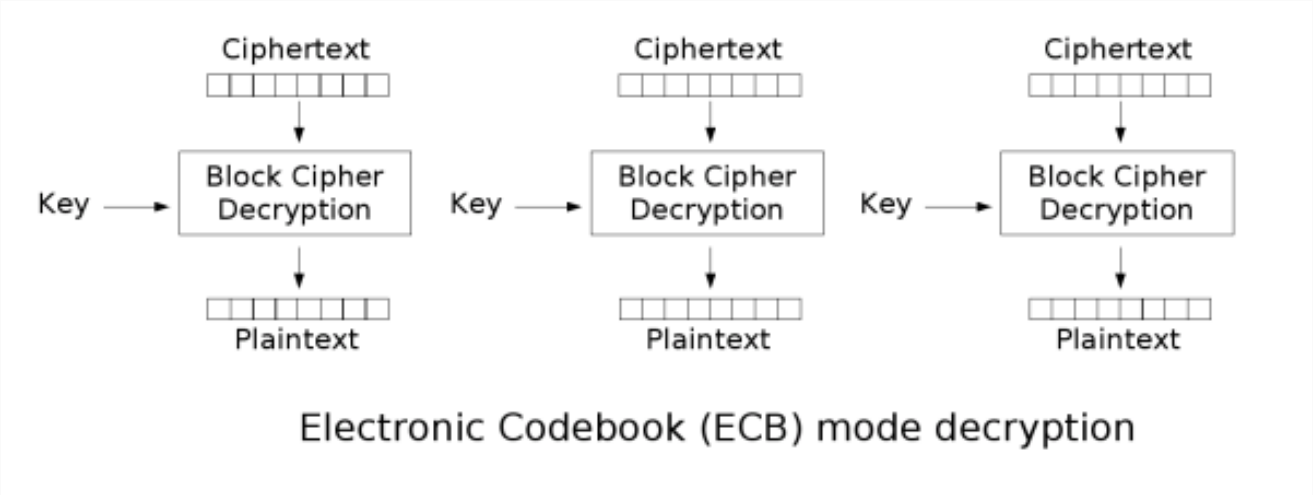
Definition

Padding is a techique for completing the last block of a message to be encrypted such that its decryption is not ambiguous.

Electronic Code Book (ECB)

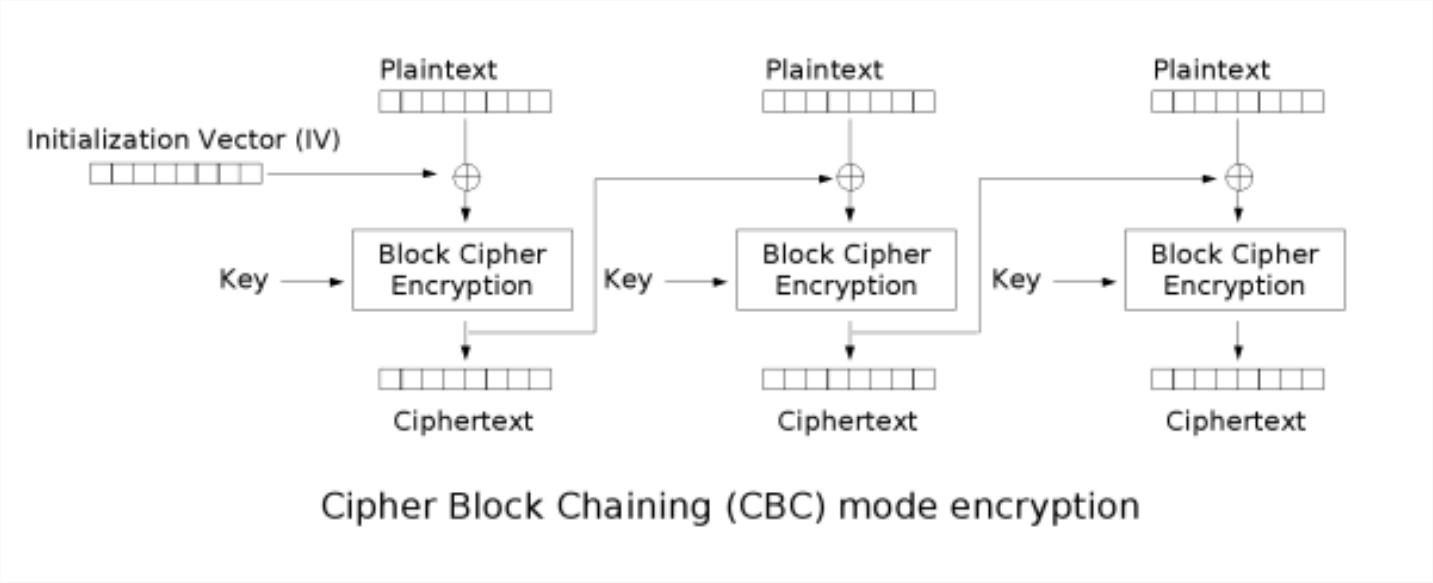


Electronic Codebook (ECB) mode encryption

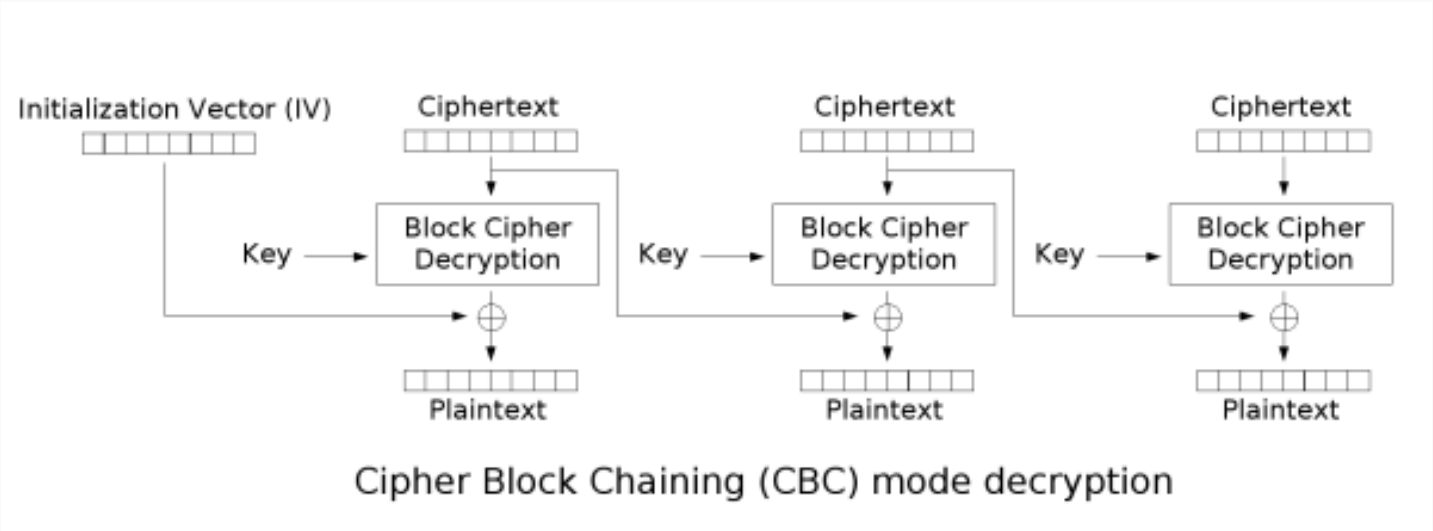


Electronic Codebook (ECB) mode decryption

Cipher Block Chaining (CBC)

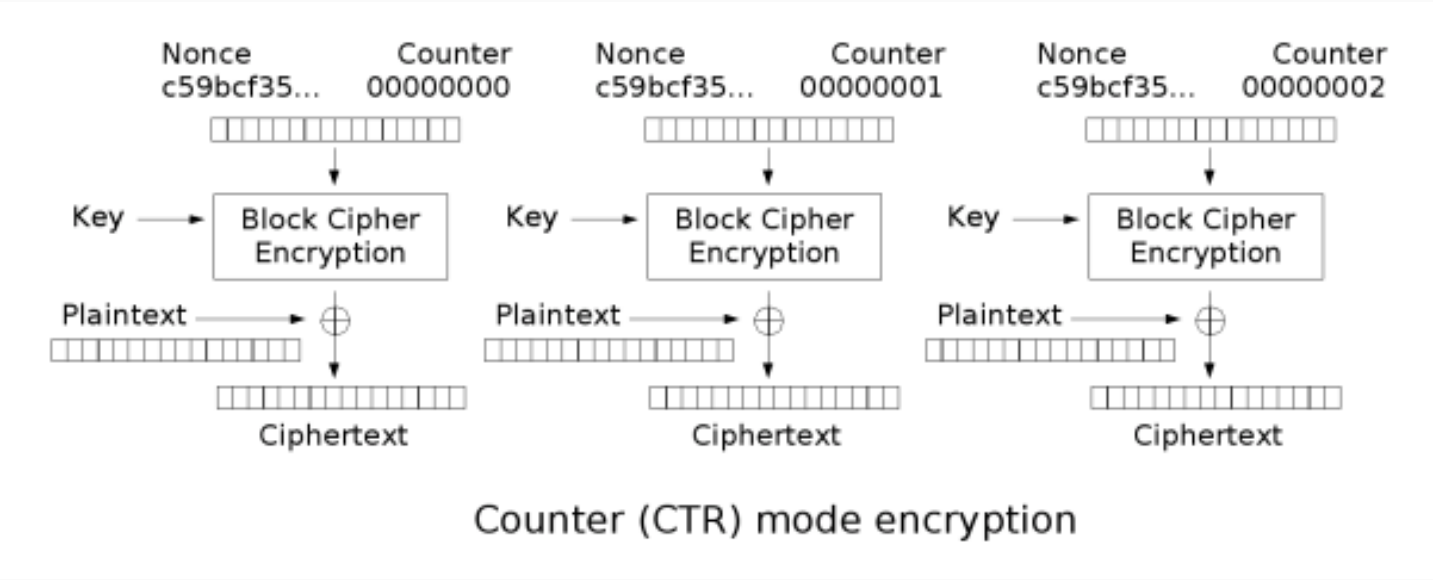


Cipher Block Chaining (CBC) mode encryption

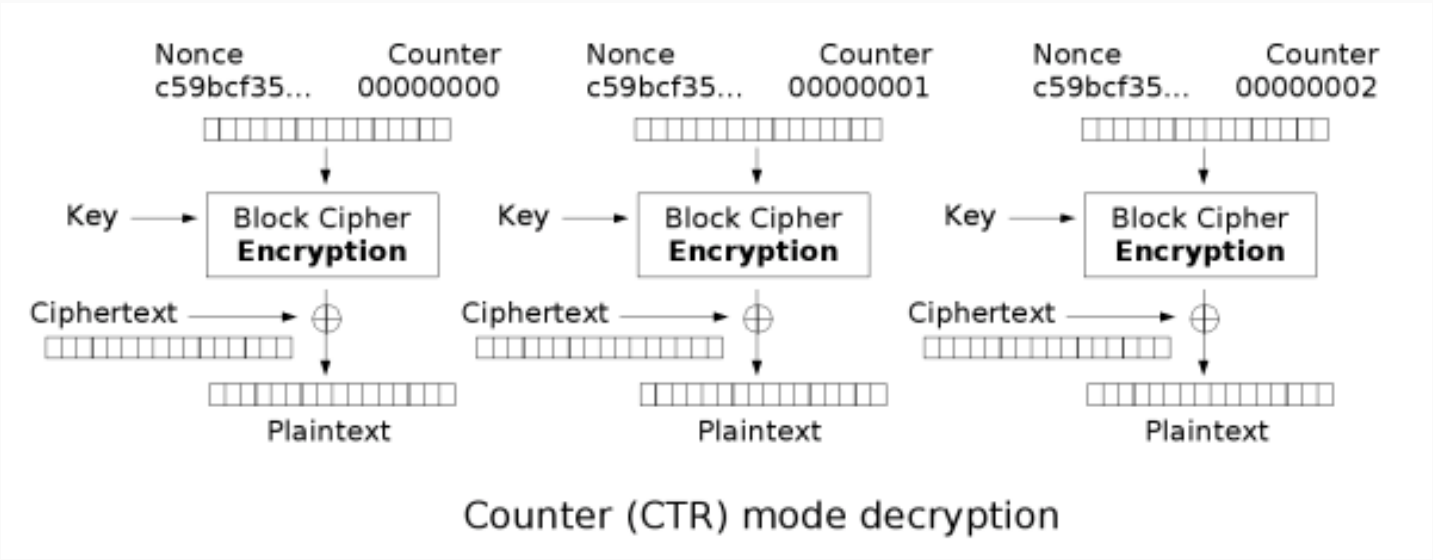


Cipher Block Chaining (CBC) mode decryption

Counter (CTR)



Counter (CTR) mode encryption



Counter (CTR) mode decryption

Message authentication code (MAC)

Important: What happens when encrypted data is manipulated?

Message authentication code (MAC)

Important: Previous modes of operation do not allow detection!

Definition

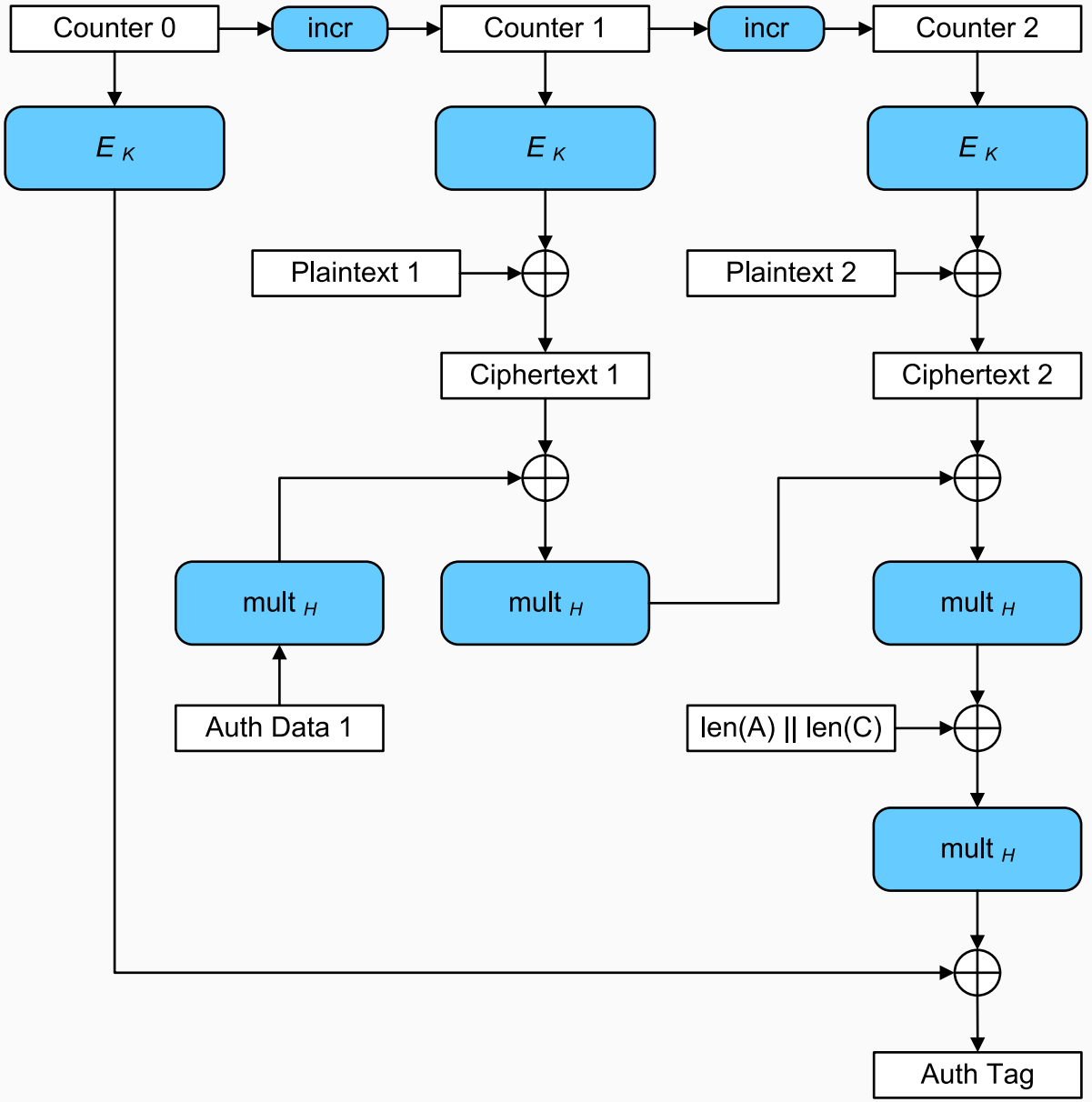
A **message authentication code** is an authentication tag produced by symmetric cryptography (block cipher or hash function). It is usually employed for constructing **authenticated encryption** modes of operation.

Important: At least two entities have access to encryption key!

Authenticated encryption modes:

- *Counter with Cipher Block Chaining (CCM)*: CTR + CBC;
- *Galois Counter Mode (GCM)*.

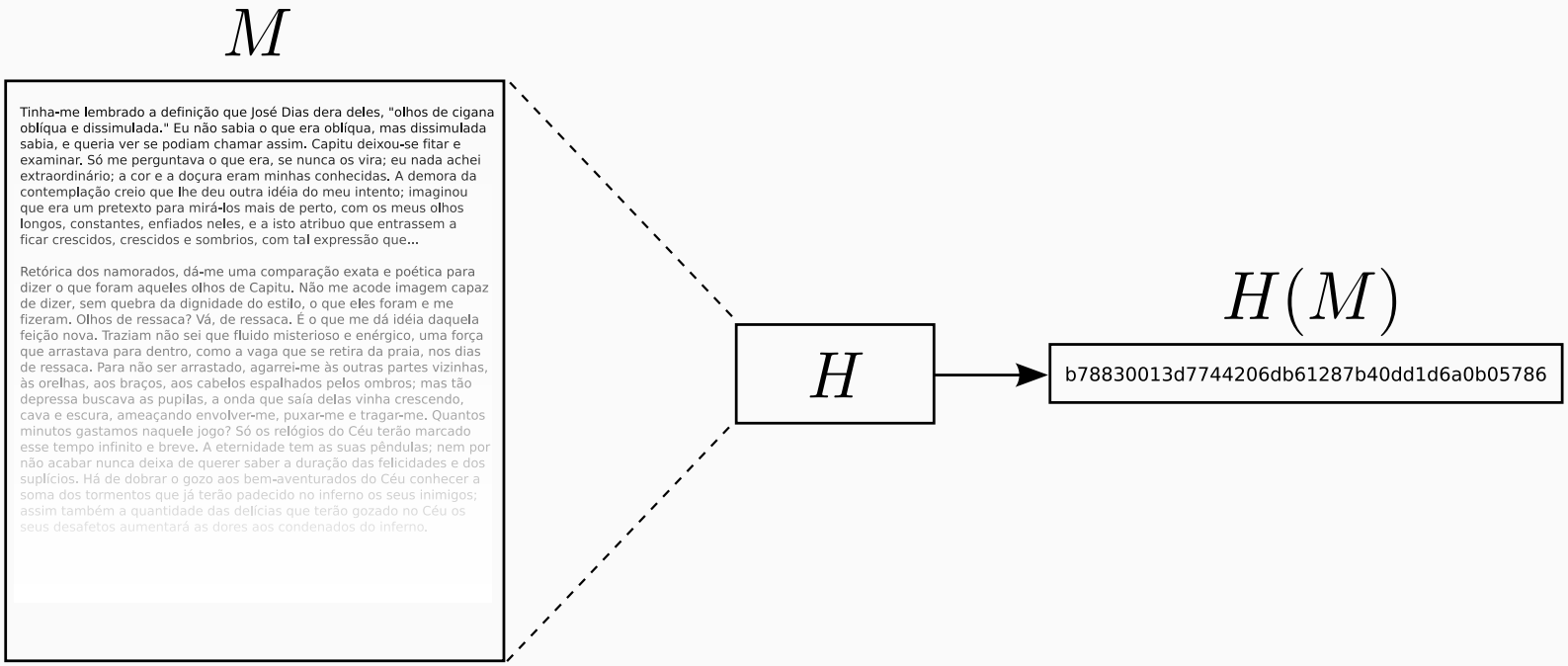
Galois Counter Mode (GCM)



Cryptographic hash functions

Informal definition

Cryptographic hash functions are employed to produce a short descriptor of a message. Informally, this descriptor is analogous to a fingerprint for human identification.



Cryptographic hash functions

Formal definition

A **cryptographic hash function** maps messages from a set $\mathcal{X}$ to hash values or authenticators in a set $\mathcal{Y}$. In this first case, it is denoted by $h : \mathcal{X} \to \mathcal{Y}$. In the second, it is parameterized by a key $K \in \mathcal{K}$ and represented by $h_K : \mathcal{X} \to \mathcal{Y}$. If $\mathcal{X}$ is finite h is also called a **compression function**.

Many different applications:

- Password storage (store $h(s)$ instead of s).
- Key derivation ($k = h(g^{xy} \bmod p)$, $k_i = h(k_{i-1})$).
- Integrity verification ($y = h(x)$).
- Digital signatures (sign $h(m)$ instead of just m).
- Message Authentication Codes (MACs) ($y = h_K(x)$).

Hash functions design

- **Merkle-Damgård paradigm**: MD4, MD5, SHA-1, SHA-2.
- Block cipher-based: Matyas-Meyer-Oseas, David-Meyer.
- New paradigms: Sponge (SHA3/Keccak).
- Number theory: VHS (integer factoring), ECOH (elliptic curves).

Properties of hash functions

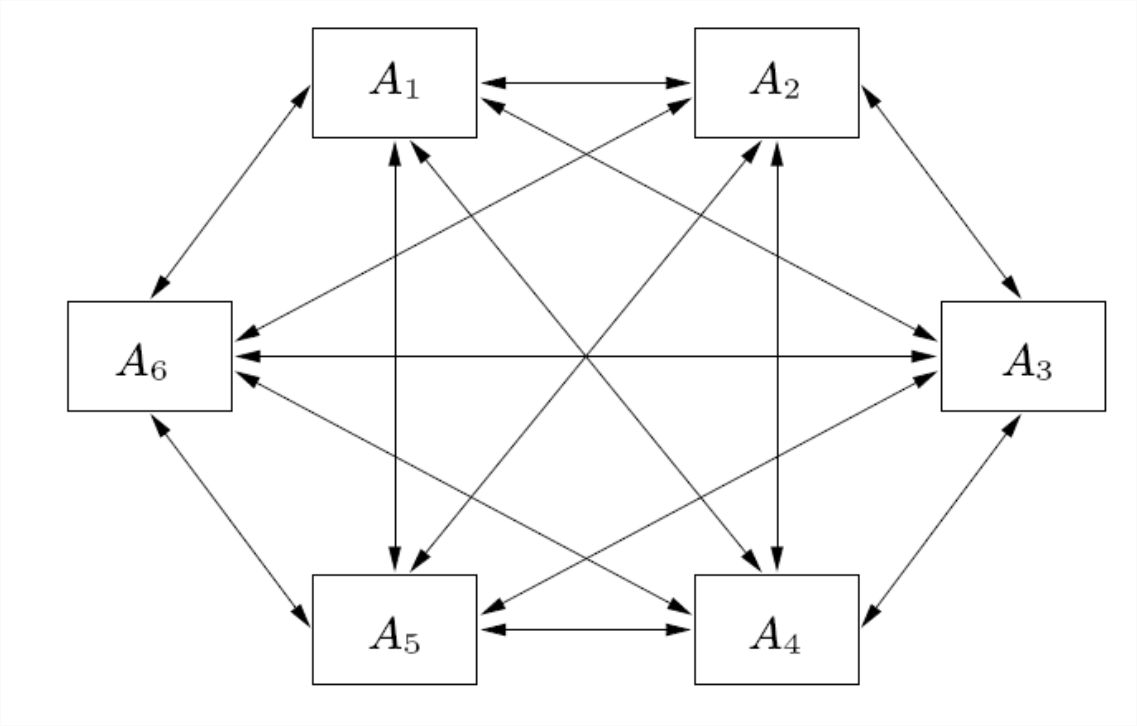
- **Preimage resistance**: Given hash y, it should be computationally infeasible to find x such that $y = h(x)$.
- **Second preimage resistance**: Given hash y and a message x such that $y = h(x)$, it should be computationally infeasible find $x' \neq x$ such that $h(x') = h(x) = y$.
- **Collision resistance**: It should be computationally infeasible to find x, x' such that $h(x) = h(x')$.

Important: Each property implies the previous one (in the first case, conditionally).

Key distribution problem

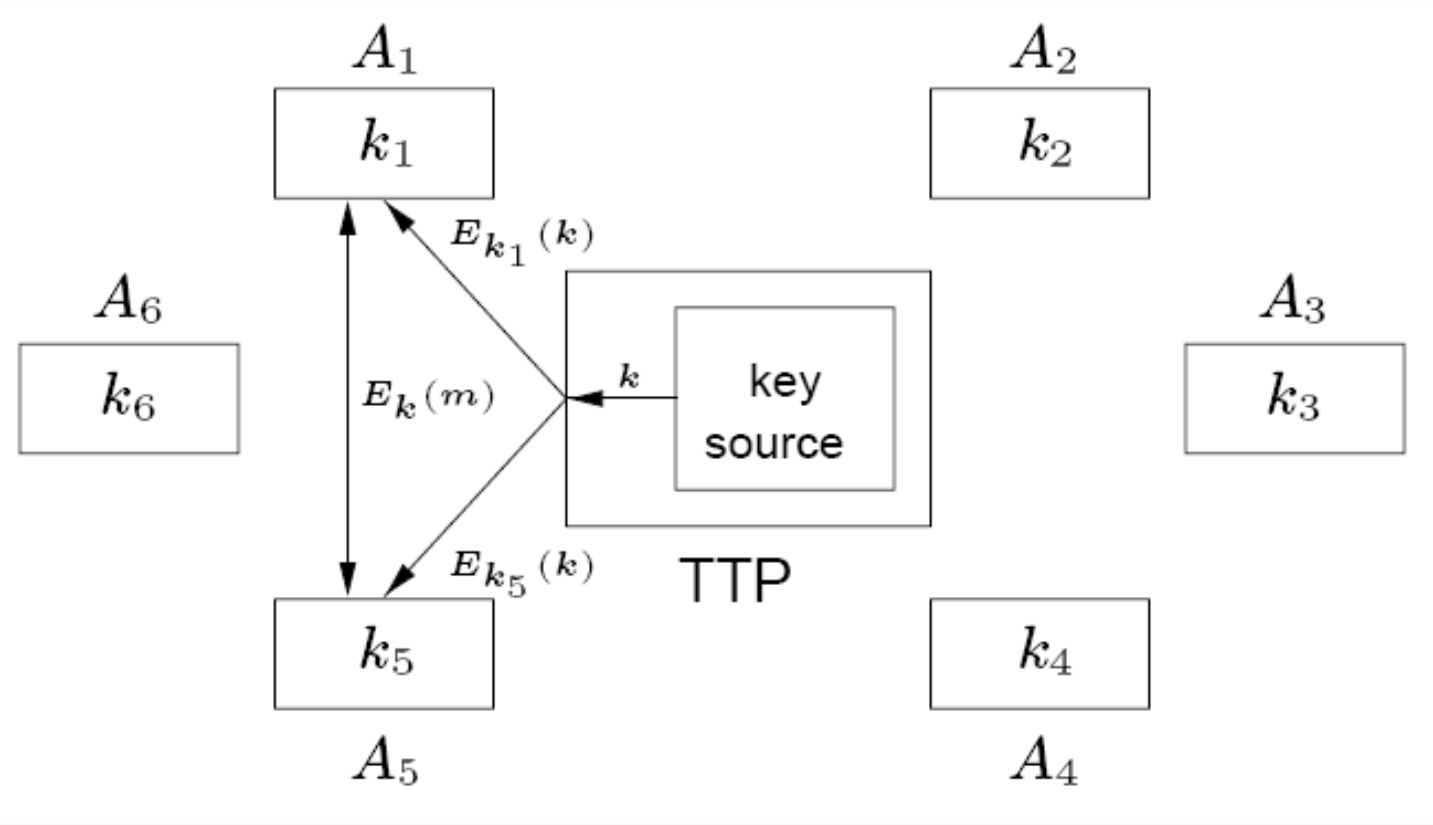
Symmetric ciphers require secrets to be shared.

How to establish shared keys with all users you want to communicate?



Key distribution problem

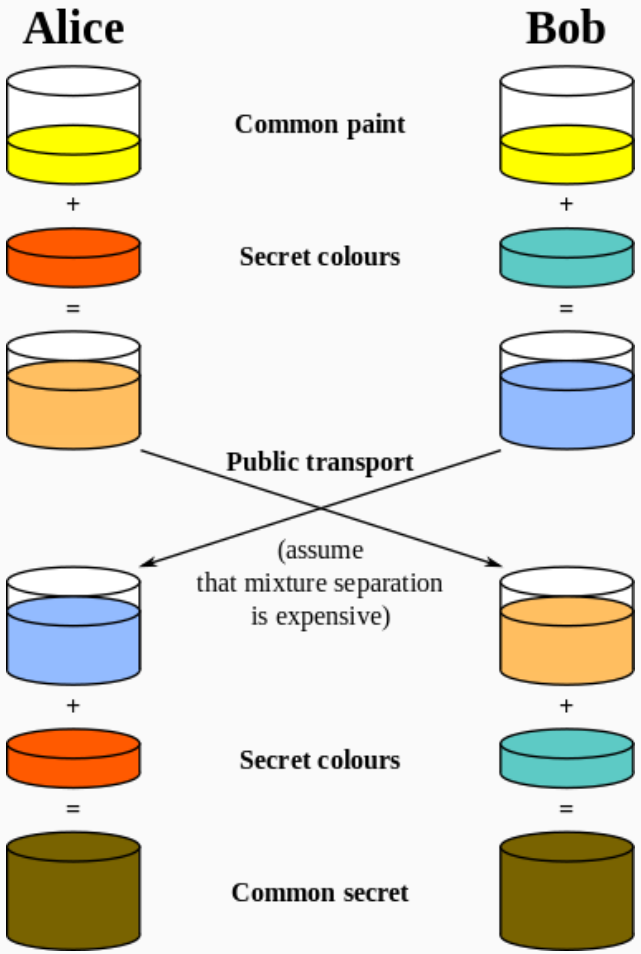
We can rely on an entity capable of producing ephemeral keys.



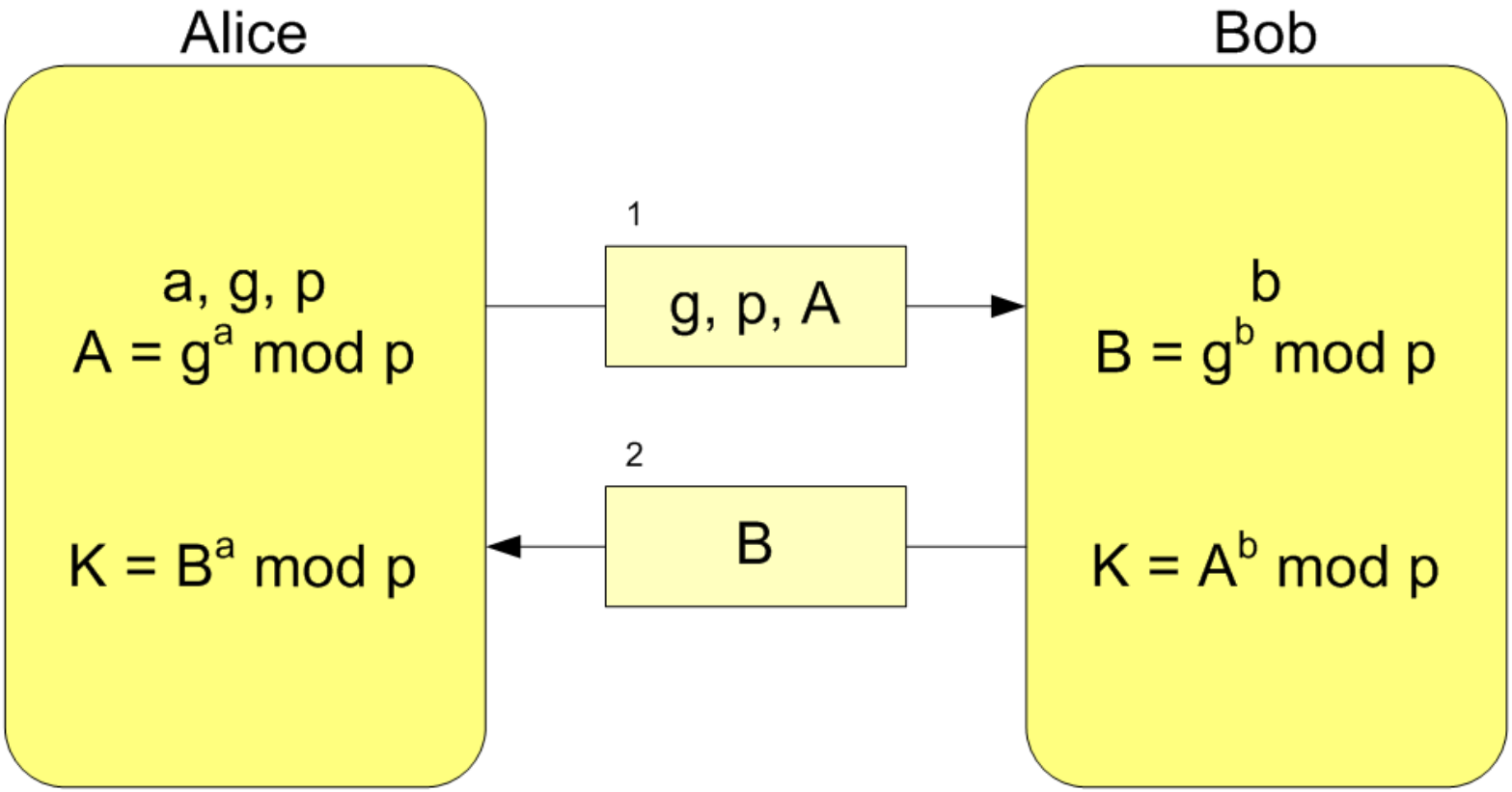
Problem: What if we don't have an entity to trust?

Asymmetric cryptography (Diffie, Hellman, 1976)

Another of the biggest contributions given to Cryptography!



Asymmetric cryptography (Diffie, Hellman, 1976)



$K = A^b \bmod p = (g^a \bmod p)^b \bmod p = g^{ab} \bmod p = (g^b \bmod p)^a \bmod p = B^a \bmod p$

Assumption: Recover x from $g^x \bmod p$ is **hard**!

Asymmetric cryptography (Rivest,Shamir,Adleman, 1977)

First practical realization of asymmetric encryption that allows encryption and signatures!

Key generation:

1. Choose two large distinct primes p and q.
2. Compute the modulus $n = pq$ and $\phi(n) = (p-1)(q-1)$.
3. Choose an integer $1 < e < \phi(n)$ as the public key.
4. Compute the private key $d = e^{-1} \bmod \phi(n)$.

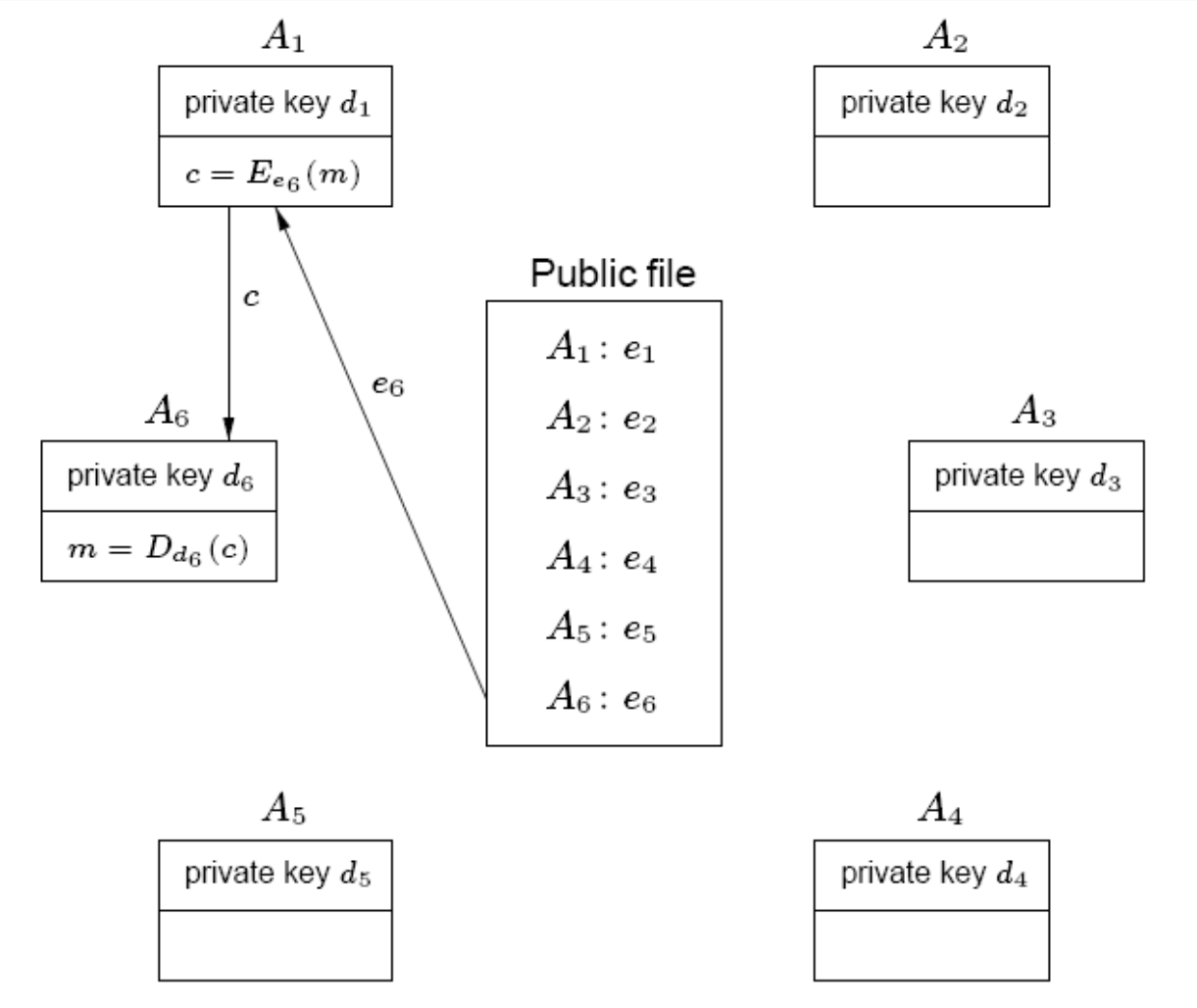
Encryption: Compute $c = m^e \bmod n$.

Decryption: Compute $m = c^d \bmod n$.

Consistency: $c^d \equiv (m^e)^d \equiv m^{ed} \equiv m \bmod n$.

Key distribution problem

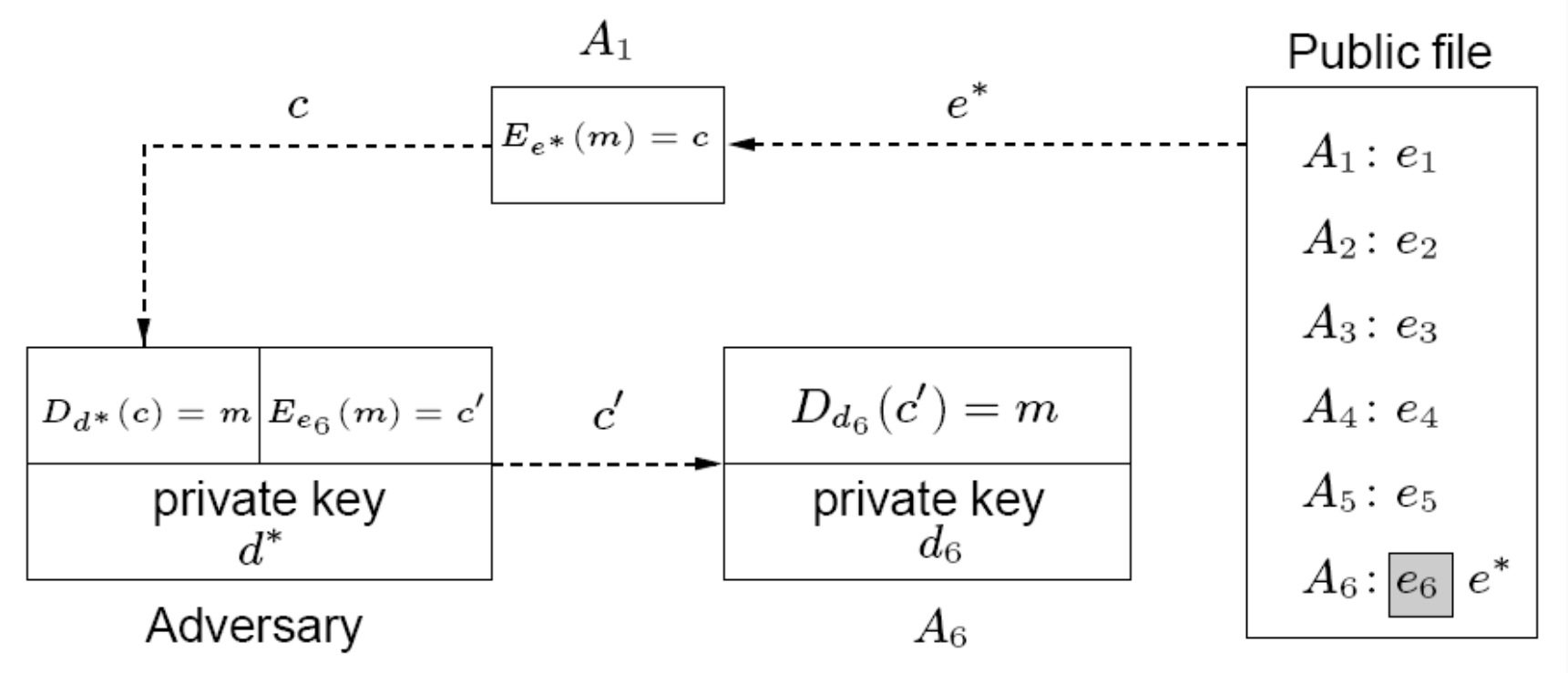
Asymmetric cryptography solves the problem!



Problem: The repository needs to be trusted!

Public key authentication

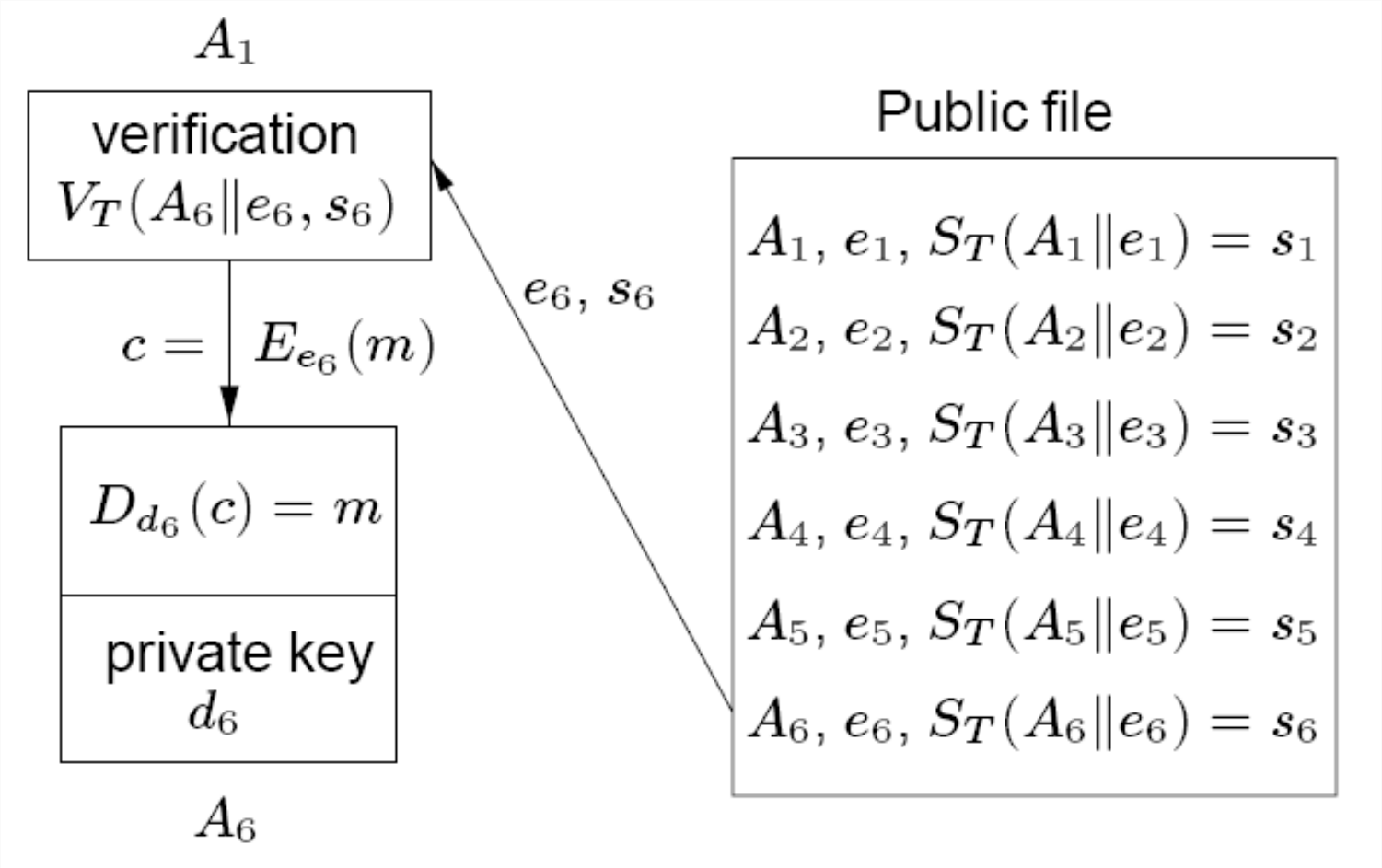
Asymmetric cryptography creates a new problem!



Problem: We need someone to trust again!

Certificates (Kohnfelder, 1979)

Solution: Central authority authenticated public keys!



Digital signatures

Definition

A **digital signature** is a cryptographic technique to add a publicly verifiable non-repudiable authenticator to a message.

Sets:

- Message space $\mathcal{M}$.
- Signature space $\mathcal{S}$.
- Key space $\mathcal{K}$.

Algorithms:

- Signature function $S_A(m) = D_d(h(m)) = s$.
- Verification function $V_A(s) = 1$ iff $E_e(s) = h(m)$.

Digital signatures (Rivest,Shamir,Adleman, 1977)

Key generation:

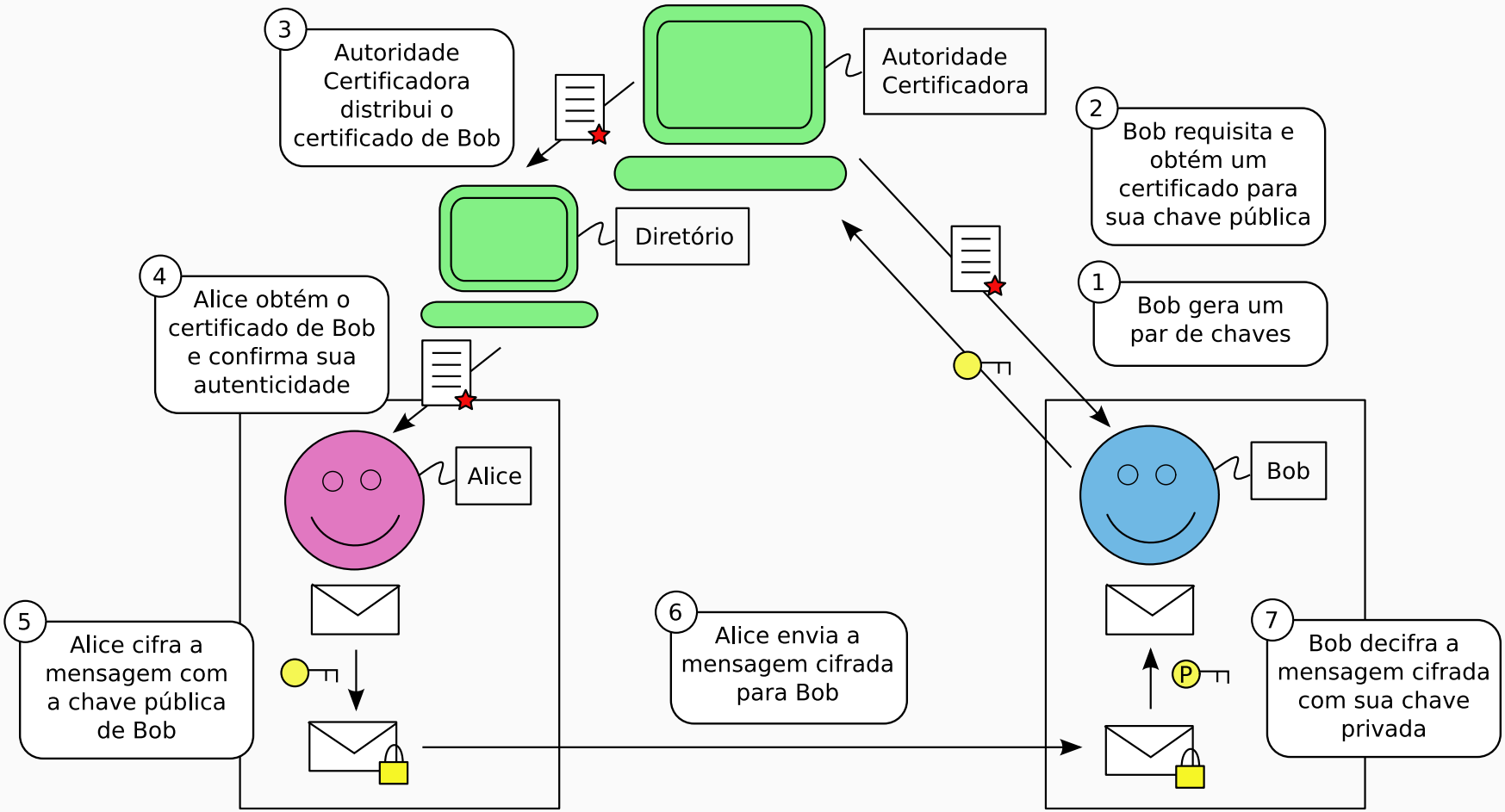
1. Choose two large distinct primes p and q.
2. Compute the modulus $n = pq$ and $\phi(n) = (p-1)(q-1)$.
3. Choose an integer $1 < e < \phi(n)$ as the public key.
4. Compute the private key $d = e^{-1} \bmod \phi(n)$.

Signature: Compute $s = h(m)^d \bmod n$.

Verification: Compute $h' = s^e \bmod n$ and check if $h' = h(m)$.

Assumption: Factoring n into p and q is **hard**!

Public-Key Infrastructures (PKIs)



Comparison

Advantages of symmetric cryptography:

- High performance and shorter keys.
- Well-studied.

Disadvantages of symmetric cryptography:

- Key sharing.
- Key distribution problem.
- Impossible to provide non-repudiation.

Advantages of asymmetric cryptography:

- Only private key needs to be securely stored.
- Trusted authority is less powerful.
- Efficient digital signatures.

Disadvantages of asymmetric cryptography:

- Low performance due to longer keys.
- Based on the apparent hardness of computational problems.

References

1. *The Code Book*
2. *Understanding Cryptography*