Cryptography

Introduction

Prof. Dr. Heiko Knospe

March 4, 2025

Some Myths about Cryptography

- *Cryptography is only about encryption methods.*
- *Encryption is strong if the ciphertext is not readable.*
- *Encryption also protects against the modification of data.*
- *Every cipher can be broken with large resources.*
- *Passwords are used as secret keys.*
- *Most data is now encrypted using asymmetric schemes, e.g., RSA.*
- *The current public-key algorithms can be used safely until large quantum computers become available.*

What do you think?

What is Cryptography?

- Classical cryptography: converting messages into an incomprehensible form. Classical cryptography is mainly dealing with *encryption* methods.
- Modern cryptography: applying mathematical techniques to achieve *security objectives* (confidentiality, integrity and availability) in the presence of adversaries.

  Modern cryptography goes beyond encryption methods and includes a variety of cryptographic primitives, algorithms, schemes and protocols.

Cryptology

Cryptology has two main branches:

- *Cryptography* is the the collection of mathematical techniques (primitives, algorithms, schemes, protocols) related to information security.
- *Cryptanalysis* is the science of analyzing cryptographic algorithms, revealing their weaknesses, launching attacks and potentially breaking them.

However, Cryptography and Cryptology are often considered to be synonymous. Modern cryptography not only defines algorithms and schemes, but also studies their security in the presence of adversaries and often provides security proofs.

History

Classic cryptography is an ancient art which has been used since ancient times.

- Egypts, Greeks and Romans used monoalphabetic substitution and transposition ciphers.
- Successful cryptanalysis with frequency analysis in medieval times.
- Systematic mathematical description of polyalphabetic ciphers and their cryptanalysis in 19th century.

History II

- Development of the Vernam One-Time-Pad and the use of statistical techniques in the beginning of 20th century.
- Cipher machines (e.g., Enigma) and advances in cipher breaking during World War II.
- Shannon (1949) provides a theoretical basis of cryptography (Communication Theory of Secrecy Systems).

History III

Modern cryptography started in the 1970s with the first commercially available secure cipher and the development of public-key mechanisms.

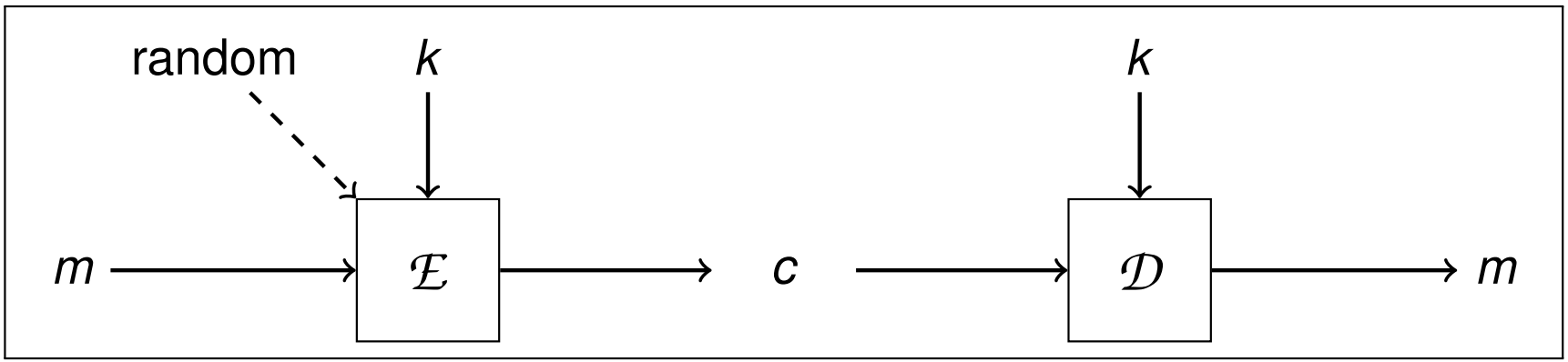
- Since the 1970s availability of modern symmetric ciphers, in particular DES (Data Encryption Standard). Development of asymmetric cryptography (Diffie-Hellman, RSA etc.).
- Since the 1990s widespread use of cryptography in computer systems and networks. Large computing resources become available. Advances in cryptanalysis.
- Successful attacks and broken schemes (e.g., DES, GSM A5, WLAN WEP, RC4, RFID Crypto1, MD5, SHA-1, ...).

History IV

- Since the 1990s precise definitions and formal proofs of security.
- Development and adoption of new ciphers and mechanisms, e.g., AES, SHA-3, Elliptic Curve Cryptography (ECC). Cryptography becomes feasible even for devices with very restricted resources (e.g. RFID transponder). New types of attacks, e.g., using side channel analysis.
- Quantum algorithms can break asymmetric schemes (RSA, Diffie-Hellman, ECC). However, sufficiently large and stable quantum computers are not yet available.
- Development of Post-quantum Cryptography (PQC). New public-key ciphers are standardized (ML-KEM, ML-DSA, SLH-DSA).

Encryption

An *encryption scheme* or *cryptosystem* consists of algorithms which produce keys and transform plaintext into ciphertext and conversely.



Encryption and decryption algorithms.

The encryption algorithm is either *deterministic* or *probabilistic (randomized)*, i.e., the ciphertext can also depend on random input data. In the case of probabilistic schemes, different encryptions of the same plaintext give different ciphertexts.

Definition of Encryption

Definition

An *encryption scheme* or *cryptosystem* consists of

- A plaintext space $\mathcal{M}$, the set of plaintext or clear-text messages,
- A ciphertext space $\mathcal{C}$, the set of ciphertext messages,
- A key space $\mathcal{K}$, the set of keys,
- A randomized key generation algorithm $Gen(1^n)$ that takes the security parameter n as input and returns a key $k \in \mathcal{K}$,
- An encryption algorithm $\mathcal{E} = \{\mathcal{E}_k \mid k \in \mathcal{K}\}$, which is possibly randomized.
- A deterministic decryption algorithm $\mathcal{D} = \{\mathcal{D}_k \mid k \in \mathcal{K}\}$. An error symbol $\perp$ is returned if the ciphertext is invalid.

Definition of Encryption II

We require that all algorithms (key generation, encryption, decryption) have *polynomial running time* with respect to their input size.

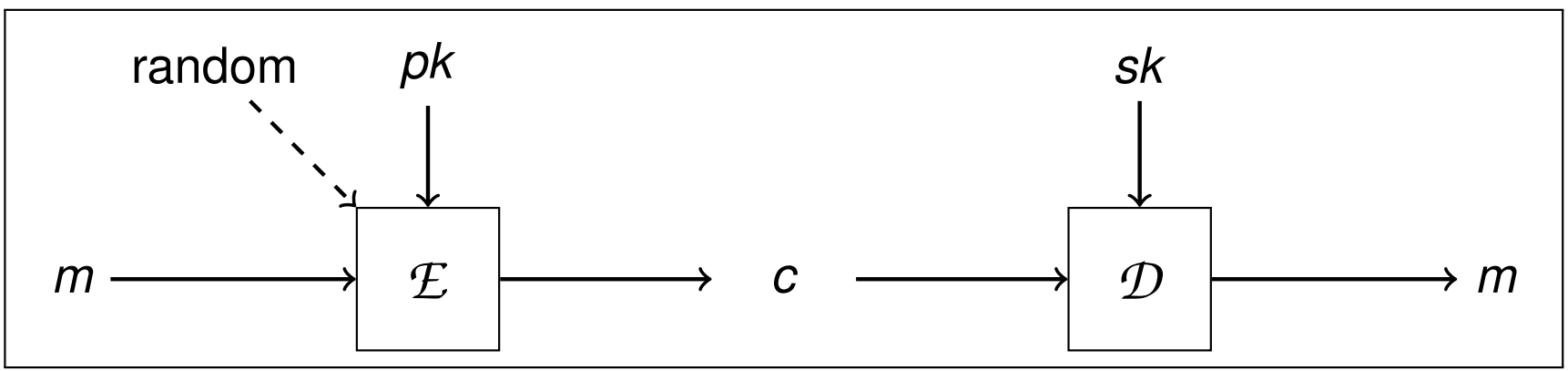
The scheme provides *correct decryption*, if for each key $k \in \mathcal{K}$ and all plaintexts $m \in \mathcal{M}$, one has

$$\mathcal{D}_k(\mathcal{E}_k(m)) = m.$$

Symmetric and Asymmetric Encryption Schemes

A scheme is said to be *symmetric-key* if encryption and decryption use the same secret key k. In contrast, *public-key* (*asymmetric-key*) encryption schemes use *key pairs* $k = (pk, sk)$, where pk is public and sk is private; encryption uses pk and decryption sk.

Public-key encryption can also be *randomized*. We will discuss symmetric-key schemes first and deal with public-key schemes later.



Asymmetric encryption and decryption.

Historical Ciphers

Ciphers have been used since ancient times. Historical ciphers use letters, i.e., the plaintext and ciphertext space consists of *strings of letters* A to Z:

$$\mathcal{M} = \mathcal{C} = \Sigma^* \text{ , where } \Sigma = \{A, B, \ldots, Z\}.$$

The * denotes strings of arbitrary length. Basically, such a cipher transforms plaintext words and sentences into ciphertext, and vice versa. Special characters are omitted or transcribed.

Of course, modern ciphers instead use the binary alphabet $\{0, 1\}$.

Caesar's Cipher

Caesar encrypted by shifting letters three places forward: *A* is replaced with *D*, *B* with *E*, and so on. At the end of the alphabet, *X* is replaced with *A*, *Y* with *B* and *Z* with *C*. For decryption, the ciphertext is shifted three places backward.

For example, *TOY* is encrypted into *WRB*.

Representing the letters by the residue classes $0, 1, \ldots, 25$ modulo 26, encryption corresponds to adding 3 modulo 26. For decryption, one has to subtract 3 modulo 26. Caesar's cipher (encryption and decryption) is *affine*, i.e. is the composition of a linear map and a constant translation.

In the above example, the plaintext $(19, 14, 24)$ is encrypted into the ciphertext $(22, 17, 1)$. What is the corresponding linear map and what the constant translation? What about the key of this cipher?

Monoalphabetic Substitution Ciphers

Monoalphabetic substitution ciphers replace one letter of the alphabet with another one. The key is a *permutation* of the letters. Note that general permutations are *not affine*.

Example:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Q	W	E	R	T	Z	U	I	O	P	A	S	D	F	G	H	J	K	L	Y	X	C	V	B	N	M

For example, *ROM* is mapped to *KGD*.

There are $26! \approx 2^{88}$ different keys. Is the cipher secure?

Frequency Analysis

Substitution ciphers can be attacked with a frequency analysis. One exploits the fact that letters are not uniquely distributed in natural languages.

Frequency of letters in English texts:

E	T	A	...	Z
12.7%	9.06%	8.17%	...	0.07%

For longer English texts, the most frequent ciphertext letter corresponds to the plaintext letter *E*, the second most frequent letter is *T* etc. Note that some correct plaintext letters are often sufficient to guess the complete text.

This method can be refined by looking at pairs or triples of letters. For example, combinations such as *IN*, *OF*, *AND* are frequent in English texts.

Attacks against Monoalphabetic Ciphers

The frequency analysis is a *ciphertext-only attack* and works well if the ciphertext is long and the plaintext comes from a known language.

A more powerful attacker has access to plaintext-ciphertext pairs (*known-plaintext attack*). He can then set up a *dictionary* (*codebook*) of plaintext and associated ciphertext letters. The dictionary can then be used to decrypt new ciphertexts.

Polyalphabetic Substitution Ciphers

Polyalphabetic ciphers are based on substitution, but the substitution depends on the *position* of the plaintext or ciphertext character. A polyalphabetic cipher has length n if the substitution key repeats every n characters.

The best-known example is the *Vigenère cipher*, which shifts the plaintext by the number of positions given by the key. The Vigenère cipher is affine.

Example: Suppose the key-length is 2 and the key is $k = DY$, which corresponds to $(3, 24)$, then the plaintext $m = ALFA$ $(0, 11, 5, 0)$ is encrypted into $c = DJIY$ $(3, 9, 8, 24)$.

$$(0, 11, 5, 0) + (3, 24, 3, 24) = (3, 9, 8, 24) \mod 26$$

Vigenère Cipher

The *Vigenère cipher* of length n is a classical example of a polyalphabetic substitution cipher. We have

$$\mathcal{M} = \mathcal{C} = \Sigma^* \quad \text{and} \quad \mathcal{K} = \Sigma^n, \text{ where } \Sigma = \{A, B, \ldots, Z\}.$$

For encryption and decryption, the message and the ciphertext is split into blocks of length n; the last block can be shorter. Encryption adds the key to each plaintext block, decryption subtracts the key.

$$\begin{aligned} c = \mathcal{E}_k(m) &= \mathcal{E}_k(m_1 \| m_2 \| \ldots) = (m_1 + k \parallel m_2 + k \parallel \ldots) \mod 26 \\ m = \mathcal{D}_k(c) &= \mathcal{D}_k(c_1 \| c_2 \| \ldots) = (c_1 - k \parallel c_2 - k \parallel \ldots) \mod 26 \end{aligned}$$

Attacking the Vigenère Cipher

A special weakness of the Vigenère cipher is the simple linear relation between plaintext, ciphertext and key. If a plaintext m and the corresponding ciphertext c is known, the key can be easily computed:

$$c - m = (c_1 \parallel c_2 \parallel \dots) - (m_1 \parallel m_2 \parallel \dots) = (k \parallel k \parallel \dots) \mod 26$$

This is a *known-plaintext* attack.

Vigenère Cipher: Exercises

- Let $k = FZC$ be the key of a Vigenère cipher. Encrypt $m = GKLM$ and decrypt $c = MZNQN$.
- Let $m = ZHUK$ be a plaintext and $c = BGFM$ a ciphertext of a Vigenère cipher of length 3. Find the key.
- In the above example of a Vigenère cipher of length 3, can you write down any four-letter plaintext and ciphertext? Why or why not? Give examples of infeasible plaintext-ciphertext combinations.

Security of Polyalphabetic Substitution Ciphers

Since the Vigenère cipher is affine, known plaintext attacks are easy.

General polyalphabetic ciphers can be attacked by a frequency analysis, in a similar way as monoalphabetic ciphers. Suppose the length n is known. Then the ciphertext is grouped into n classes, where the positions $1, n+1, 2n+1, \ldots$ form the first class, the positions $2, n+2, 2n+2$ the second class, etc. Since the ciphertext in each class was encrypted with a monoalphabetic cipher, it can be attacked by a frequency analysis. However, this attack requires long ciphertexts, and it can be infeasible for very long keys since each class must be treated separately and contain a sufficient number of characters.

If some plaintext and the corresponding ciphertext is known, a dictionary can be set up as for monoalphabetic ciphers, but for each class of positions separately.

Key-length of Polyalphabetic Substitution Ciphers

How can an attacker find the length of the key of a polyalphabetic cipher?

Kasiski's method can be used. One looks for strings of characters that occur repeatedly in the ciphertext. The distances between these strings are likely to be multiples of the key length (why?). Then take the greatest common divisor of all distances. We skip the details.