Cryptography

Introduction

Prof. Dr. Heiko Knospe

April 11, 2021

What is Cryptography?

- Classic cryptography: converting messages into an incomprehensible form. Classic cryptography is mainly dealing with *encryption* methods.
- Modern cryptography: applying mathematical techniques to achieve *security objectives* in the presence of adversaries. Modern cryptography goes beyond encryption schemes and includes various cryptographic primitives, e.g., pseudorandom generators, hash functions or signatures.

Some Myths about Cryptography

- *Cryptography is only about encryption methods.*
- *Every cipher can be broken with large resources.*
- *Encryption is strong if the ciphertext is not readable.*
- *Encryption also protects against the modification of data.*
- *A lot of data is now encrypted using asymmetric schemes, e.g., with RSA.*
- *Passwords can be used as secret keys.*

What do you think?

Cryptology

Cryptology has two main branches:

- *Cryptography* is the the collection of mathematical techniques (primitives, algorithms, protocols) related to information security. We distinguish between symmetric and asymmetric (or public-key) cryptography.
- *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, but also studies their strength in the presence of adversaries.

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.

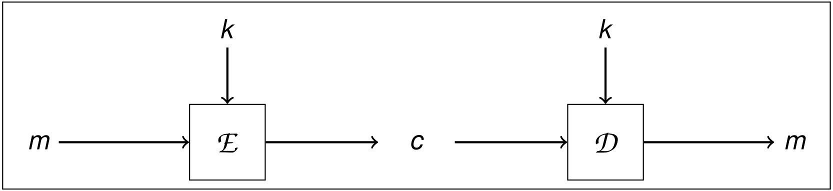
- Since the 1970s availability of modern symmetric ciphers, in particular DES (Data Encryption Standard). Development of asymmetric cryptography (Diffie-Hellman, RSA etc.).
- Since the 1990s widespread use of cryptography in computer systems and networks. Large computing resources become available. Advances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

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, 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.

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 *randomized*, i.e., the ciphertext may depend on random input data.

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) are *polynomial* with respect to the input size.

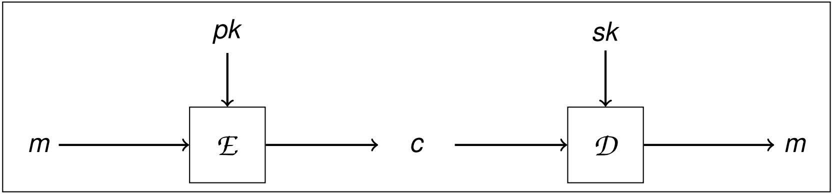
The scheme provides *correct decryption*, if for each key $k \in \mathcal{K}$ and all plaintexts $m \in \mathcal{M}$, one has

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

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.

We 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*:

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

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 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, ROM is encrypted into URP.

If we represent the letters by the residue classes 0, 1, ..., 25, then encryption corresponds to adding 3 modulo 26. For decryption, one has to subtract 3 modulo 26.

In the above example $(17, 14, 12)$ is encrypted into $(20, 17, 15)$.

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.

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. The letters are not uniquely distributed.

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 a couple of correct plaintext letters often suffices to find the complete text.

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

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. For example, if the length is 2 and the key is $k = DY$, which corresponds to $(3, 24)$, then the plaintext $m = ALFA$ is encrypted into $c = DJIY$. The *Vigenère square* may be used for encryption and decryption.

Vigenère Cipher

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

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

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.

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

Attacking the Vigenère Cipher

A special weakness of the Vigenère cipher is its *linearity*. If any plaintext m and the corresponding ciphertext c is known, then the key can be easily computed:

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

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?

Security of Polyalphabetic Substitution Ciphers

General polyalphabetic ciphers can be attacked with a frequency analysis, in a similar way as monoalphabetic ciphers. Suppose the length n is known. Then the ciphertext is grouped into n classes, i.e., positions $1, n+1, 2n+1, \ldots$ form the first class, positions $2, n+2, 2n+2$ the second class, etc. The ciphertext in each class is encrypted with a monoalphabetic cipher, and can thus be attacked with a frequency analysis. Note that this attack requires a longer ciphertext since each class must be treated separately.

How can an attacker find the key length if it is unknown? *Kasiski's* method can be used. One looks for strings of characters that are repeated in the ciphertext. The distances between these strings are likely to be multiples of the length of the keyword. We skip the details.