Cryptography

Corso di Laurea Magistrale in Informatica

Introduction to Cryptography

Ugo Dal Lago

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

informatiques mathématiques
Inria

Academic Year 2023-2024

Classic and Modern Cryptography

- Cryptography was, up to the 80s, a form of **art**.
  - Creativity and intuition were necessary, and there were no basic principles.

Classic and Modern Cryptography

- Cryptography was, up to the 80s, a form of **art**.
    - Creativity and intuition were necessary, and there were no basic principles.
- Starting from some fundamental contributions which appeared at the end of the 70s, cryptography has become a proper **science**.
    - That permitted to go beyond classic cryptography, which was *only* about confidentiality in communication.
    - In particular, modern cryptography introduced protocolos for *message authentication*, for *key exchange*, for *digital signatures*, etc.

Classic and Modern Cryptography

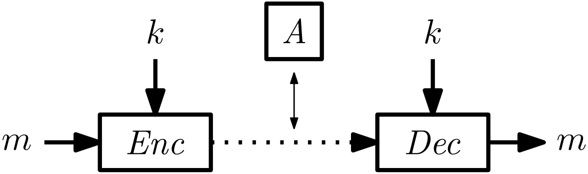
- Cryptography was, up to the 80s, a form of **art**.
  - Creativity and intuition were necessary, and there were no basic principles.
- Starting from some fundamental contributions which appeared at the end of the 70s, cryptography has become a proper **science**.
  - That permitted to go beyond classic cryptography, which was *only* about confidentiality in communication.
  - In particular, modern cryptography introduced protocolos for *message authentication*, for *key exchange*, for *digital signatures*, etc.
- Nowadays, the term **cryptography** refers to whatever technique whose objective is to guarantee *security* in communications, transactions, and distributed computation.

Classic and Modern Cryptography

- Cryptography was, up to the 80s, a form of **art**.
  - Creativity and intuition were necessary, and there were no basic principles.
- Starting from some fundamental contributions which appeared at the end of the 70s, cryptography has become a proper **science**.
  - That permitted to go beyond classic cryptography, which was *only* about confidentiality in communication.
  - In particular, modern cryptography introduced protocolos for *message authentication*, for *key exchange*, for *digital signatures*, etc.
- Nowadays, the term **cryptography** refers to whatever technique whose objective is to guarantee *security* in communications, transactions, and distributed computation.
- What does *security* actually mean, mathematically speaking?
  - In this very first part of the course, we will try to give an answer to this question, looking at *classic ciphers*.

Private-Key Encryption

- Let's start with a **specific** scenario, that of protocols aimed at ensuring *confidentiality* in the exchange of data between two parties.



- The key is shared and must be *previuosly exchanged* between the parties.
- Formally, we can see an **encryption scheme** as consisting of three spaces $\mathcal{K}$, $\mathcal{M}$ and $\mathcal{C}$ and of a triple of algorithms (Gen, Enc, Dec) where

$$Gen : 1 \to \mathcal{K} \qquad Enc : \mathcal{M} \times \mathcal{K} \to \mathcal{C} \qquad Dec : \mathcal{C} \times \mathcal{K} \to \mathcal{M}$$

- The scheme is **correct** when $Dec(Enc(x, k), k) = x$.

Kerckhoffs' Principle

- Before giving a definition of **security** for an encryption scheme, we need to understand *what the adversary A can actually do*.
- According to Kerckhoffs, A knows the inner working of Gen, Enc, Dec, so the **only element** that A *does not* know is the key k.
- Even in a scenario in which A *does not actually know* the encryption scheme, it makes sense to assume that he does know it, because the knowledge of A can change over time.
  - Changing the key is easy, changing the encryption scheme is much more complicated.
- Kerckhoffs' postulate is a *principle* and as such it **can not** be proved.
- In the past, Kerckhoffs' Principle has been repeatedly ignored, with devastating consequences.

Possible Attack Scenarios

- **Ciphertext-Only Attack**
    - The adversary A knows *only* a certain number of ciphertexts $c_1, \ldots, c_n$ so he interferes with the communication in a passive way.

Possible Attack Scenarios

- **Ciphertext-Only Attack**
  - The adversary A knows *only* a certain number of ciphertexts $c_1, \ldots, c_n$ so he interferes with the communication in a passive way.
- **Known-Plaintext Attack**
  - The adversary A knows a certain number of pairs $(m_1, c_1), \ldots, (m_n, c_n)$, where c_i is the ciphertext corresponding to m_i.
  - The attack remains passive.

Possible Attack Scenarios

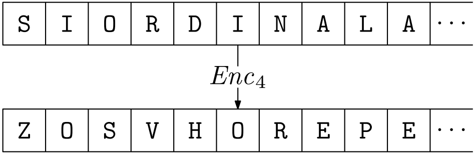
- **Ciphertext-Only Attack**
  - The adversary A knows *only* a certain number of ciphertexts $c_1, \ldots, c_n$ so he interferes with the communication in a passive way.
- **Known-Plaintext Attack**
  - The adversary A knows a certain number of pairs $(m_1, c_1), \ldots, (m_n, c_n)$, where c_i is the ciphertext corresponding to m_i.
  - The attack remains passive.
- **Chosen-Plaintext Attack**
  - The adversary A begins to play an active role.
  - In particular, he can compute $Enc_k(m) = Enc(m, k)$ for messages of his own choice.

Possible Attack Scenarios

- **Ciphertext-Only Attack**
  - The adversary A knows *only* a certain number of ciphertexts $c_1, \ldots, c_n$ so he interferes with the communication in a passive way.
- **Known-Plaintext Attack**
  - The adversary A knows a certain number of pairs $(m_1, c_1), \ldots, (m_n, c_n)$, where c_i is the ciphertext corresponding to m_i.
  - The attack remains passive.
- **Chosen-Plaintext Attack**
  - The adversary A begins to play an active role.
  - In particular, he can compute $Enc_k(m) = Enc(m, k)$ for messages of his own choice.
- **Chosen-Ciphertext Attack**
  - The adversary A participates in an even more active way in the communication, having access to an "oracle" $Dec_k(\cdot)$ for decryption.
  - Obviously without having access to k!

Caesar's Cipher

- It is worth taking a look at some historical ciphers, in order to understand the limitations of the classical approach to cryptography.
- In Caesar's Cipher:
  - Messages in $\mathcal{M}$ are simply texts in any language.
  - The set $\mathcal{K}$ of keys is very simple $\mathcal{K} = \{4\}$.
  - *Enc* builds the ciphertext "shifting" each character four places down the alphabet



S	I	O	R	D	I	N	A	L	A	⋯

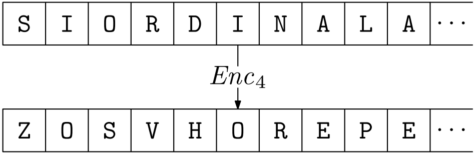
Enc_4

Z	O	S	V	H	O	R	E	P	E	⋯

  - *Dec* works in a perfectly dual way.

Caesar's Cipher

- It is worth taking a look at some historical ciphers, in order to understand the limitations of the classical approach to cryptography.
- In Caesar's Cipher:
  - Messages in $\mathcal{M}$ are simply texts in any language.
  - The set $\mathcal{K}$ of keys is very simple $\mathcal{K} = \{4\}$.
  - *Enc* builds the ciphertext "shifting" each character four places down the alphabet

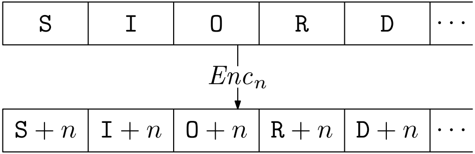


S	I	O	R	D	I	N	A	L	A	···
Z	O	S	V	H	O	R	E	P	E	···

  - *Dec* works in a perfectly dual way.
- The key is unique, so anyone who knows *Enc* will easily decode any message.

Shift Cipher

- It is an obvious generalization of Caesar's cipher, where $\mathcal{K}$ becomes $\{1, \ldots, |\Sigma| - 1\}$ and Σ is the underlying alphabet.



S	I	O	R	D	$\cdots$

Enc_n

$\mathtt{S} + n$	$\mathtt{I} + n$	$\mathtt{O} + n$	$\mathtt{R} + n$	$\mathtt{D} + n$	$\cdots$

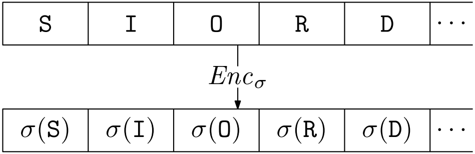
- There are many more keys, but they still remain too few.
- *A* can *try* to decrypt any ciphertext with all possible keys and after at most $|\Sigma| - 1$ attempts he obtains the plaintext.
- It is obvious that this attack only works only when the message is meaningful. How to formalize the fact that the cipher is trivially insecure?

Mono-alphabetic Substitution

- It can be seen as a further generalization of the previous two ciphers.
- The message space $\mathcal{M}$ remains the same, while the space of the keys becomes:

$$\mathcal{K} = \{\sigma \mid \sigma : \Sigma \to \Sigma \text{ is a permutation}\}.$$

- It then becomes:



S	I	O	R	D	$\cdots$

Enc_σ

σ(S)	σ(I)	σ(O)	σ(R)	σ(D)	$\cdots$

- Now $|\mathcal{K}|$ is the factorial of $|\Sigma|$, a very large number. Brute-force attack is no longer possible, at least in a reasonable time.

Statistical Attacks

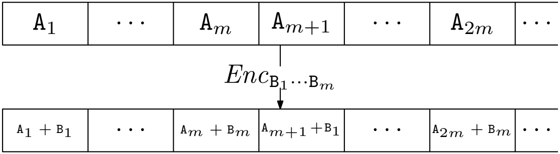
- If the messages in $\mathcal{M}$ have some statistical property that makes them distinguishable from random strings, then A could **analyze the frequencies** $q_1, \ldots, q_{|\Sigma|}$ of each symbol in Σ in the ciphertext c, comparing them with that of the same symbols in the messages from $\mathcal{M}$.
  - For each natural language, there are tables through which one can derive the probabilities p_i of the i-th symbol in the sentences of that language.
  - As $|c|$ increases, the probability of success converges to 1.
- Similarly, a statistical attack on the Shift Cipher could consist of computing the following quantities

$$K = \sum_{i=1}^{|\Sigma|} p_i^2 \qquad\qquad I_j = \sum_{i=1}^{|\Sigma|} (p_i \cdot q_{i+j})$$

  and in determining for what value of j the quantities K and I_j are most similar.

Vigenère Cipher

- Also called Poly-alphabetic Shift Cipher.
- The space of keys is the set of strings of finite length in Σ, i.e. $\mathcal{K} = \Sigma^*$.



$\mathtt{A}_1$	$\cdots$	$\mathtt{A}_m$	$\mathtt{A}_{m+1}$	$\cdots$	$\mathtt{A}_{2m}$	$\cdots$
$\mathtt{A}_1 + \mathtt{B}_1$	$\cdots$	$\mathtt{A}_m + \mathtt{B}_m$	$\mathtt{A}_{m+1} + \mathtt{B}_1$	$\cdots$	$\mathtt{A}_{2m} + \mathtt{B}_m$	$\cdots$

- Again, the key space appears to be sufficiently large, thus preventing brute force attacks.

Vigenère Cipher — Statistical Attacks

- If the length t of the key $k \in \Sigma^*$, called the *period*, is known, then the techniques already seen can be used.
- How can we determine the period t?
  - If the maximum period T is not too large (i.e., if $\mathcal{K} = \Sigma^T$), we could for example try to force the cipher simply **by trial and error**, assuming that t takes progressively greater values in $\{1, \ldots, T\}$.
  - We could also use the so-called **Kasiski's method**,

Plaintext:	the	man	and	the	woman	retrieved	the	letter	from	the	post	office
Key:	bea	dsb	ead	sbe	adsbe	adsbeadsb	ead	sbeads	bead	sbe	adsb	eadsbe
Ciphertext:	ULE	PSO	ENG	LII	WREBR	RHLSMEYWE	XHH	DFXTHJ	GVOP	LII	PRKU	SFIADI

  - Finally, there is also the method based on the **index of coincidence**. For increasing values of τ natural number, we tabulate the characters of the ciphertext in position $1, 1+\tau, 1+2\tau, 1+3\tau, \ldots$, obtaining the frequencies q_i^τ. At this point we compute

$$K = \sum_{i=1}^{|\Sigma|} p_i^2 \qquad\qquad S_\tau = \sum_{i=1}^{|\Sigma|} (q_i^\tau)^2$$

and we check for which values of τ, S_τ and K are close.

Historical Ciphers — The Take-Home Messages

- Historical ciphers were never actually used, but their study leaves us three important messages.
- On the one hand the key space must be *large enough* to prevent brute-force attacks, but the latter is a condition **necessary but not sufficient** to guarantee security.
  - Cf. The substitution cipher.
- On the other hand, the **descriptive complexity** of a cipher gives us no guarantee of its security.
  - For the Kerckhoffs' principle, the adversary knows *Gen*, *Enc*, *Dec*.
- Finally, **formally defining** the security of a cipher is not at all obvious.

The Three Principles of Modern Cryptography

1. **Use of Rigorous and Precise Definitions of Security of Primitives and Protocols**.
   - One cannot simply give informal definitions. It is necessary to be formal, this way being able to avoid any ambiguity.
   - The risk is to become too restrictive, but it is a risk that must be taken.
2. **Accuracy in Specifying the Underlying Assumptions**.
   - Without assumptions, unfortunately, not much can be proved about the security of primitives and protocols.
   - And even in this case, we have to be rigorous and precise.
3. **Proof of Security Written in the Language of Mathematics**.
   - When formal definitions and assumptions are free from ambiguity, the temptation is to infer directly the security of the scheme from the assumptions.
   - History teaches us not to take anything for granted

Formulate Exact Definitions — When Crucial?

- When **Designing**.
  - If you do not know what you are aiming for, how is it possible for you to go in the right direction? And would it be possible for you to realize that the goal has been achieved?
- When **Using**.
  - A precise definition of security can be used to *prove* that an existing scheme has (or does not have) the desired security properties.
- When **Studying and Comparing**.
  - There may be *distinct* (strong or weak) definitions of security for *the same* type of scheme.
  - One way to *choose* between different schemes is to compare the security guarantees offered by each of them.

Formulate Exact Definitions — An Example

- In the context of private-key cryptography and of encryption schemes, we should ask ourselves: *when can such a scheme be considered secure*?
- Any idea?

Formulate Exact Definitions — An Example

- In the context of private-key cryptography and of encryption schemes, we should ask ourselves: *when can such a scheme be considered secure*?
- Any idea?
    1. When no adversary A can determine the key, given the informations available to him.
        - So, a scheme such that $Enc(k, x) = x$ is safe?!?

Formulate Exact Definitions — An Example

- In the context of private-key cryptography and of encryption schemes, we should ask ourselves: *when can such a scheme be considered secure*?
- Any idea?
  1. When no adversary A can determine the key, given the informations available to him.
     - So, a scheme such that $Enc(k, x) = x$ is safe?!?
  2. When no adversary A can reconstruct m from $Enc(k, m)$.
     - So, is a scheme that allows the adversary to reconstruct *the last* 10 *bits* of m from $enc(k, m)$ safe?

Formulate Exact Definitions — An Example

- In the context of private-key cryptography and of encryption schemes, we should ask ourselves: *when can such a scheme be considered secure*?
- Any idea?
  1. When no adversary A can determine the key, given the informations available to him.
     - So, a scheme such that $Enc(k, x) = x$ is safe?!?
  2. When no adversary A can reconstruct m from $Enc(k, m)$.
     - So, is a scheme that allows the adversary to reconstruct *the last* 10 *bits* of m from $enc(k, m)$ safe?
  3. When no adversary A can determine any bit of m from $Enc(k, m)$.
     - So, should a scheme that allows the adversary to determine whether some bits in m are *in a certain relation* safe?

Formulate Exact Definitions — An Example

- In the context of private-key cryptography and of encryption schemes, we should ask ourselves: *when can such a scheme be considered secure*?
- Any idea?
  1. When no adversary A can determine the key, given the informations available to him.
     - So, a scheme such that $Enc(k, x) = x$ is safe?!?
  2. When no adversary A can reconstruct m from $Enc(k, m)$.
     - So, is a scheme that allows the adversary to reconstruct *the last* 10 *bits* of m from $enc(k, m)$ safe?
  3. When no adversary A can determine any bit of m from $Enc(k, m)$.
     - So, should a scheme that allows the adversary to determine whether some bits in m are *in a certain relation* safe?
  4. When no adversary A can determine any (decidable?) property of m from $Enc(k, m)$.
     - Here we are very close to a reasonable definition, but how to make it *accurate* and *formal*?

Why Specifying the Assumptions Accurately?

- **Inaccurate assumptions cannot be validated nor refuted**.
    - Although the assumptions are neither proven nor refuted, they are certainly studied, and this study leads to the formulation of *conjectures about their truth*.
    - In the lack of a precise specification, the study becomes difficult, fundamentally impossible.
- **Different schemes can be compared**.
    - Schemes whose security is based on weak assumptions are obviously preferable to schemes in which the underlying assumptions are very strong.
    - Comparing different assumptions is only possible in the presence of a rigorous formalization.
- **Sufficiently weak assumptions may not be affected by an attack.**