Università degli Studi di Trieste

Corso di Laurea Magistrale in
INGEGNERIA CLINICA

DATA PROTECTION BASIC PRINCIPLES

Corso di Informatica Medica

Docente Sara Renata Francesca MARCEGLIA

Dipartimento di Ingegneria e Architettura

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Data protection: basic concepts

Authentication:

- Process of verifying the identity of an object/actor

Identification

- Autentication that defines univocally the identity of an object/actor

Authorization

- Process of allowing to use a specific object or accessing a specific information

Privacy

- The content of an object is known only to its creator or to whom is allowed to use it

Integrity

- Property of not being changed from its original form

Responsibility

- Signature of who is responsible for the content of an object (cannot be denied)

Data protection

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Security

- Preservation of data
- Data cannot be deleted, lost in a disaster,

Privacy

- Management of access policies
- Unwanted access have to be avoided

Why medical data are critical

UNIVERSITÀ DEGLI STUDI DI TRIESTE

BANKING

Data OWNER = Data USER

Bank account holder — Bank account holder

MEDICINE

Data OWNER ≠ Data USER

Patient — Healthcare professional

In medicine the owner of data does not have the knowledge to use it → data have to be shared with others

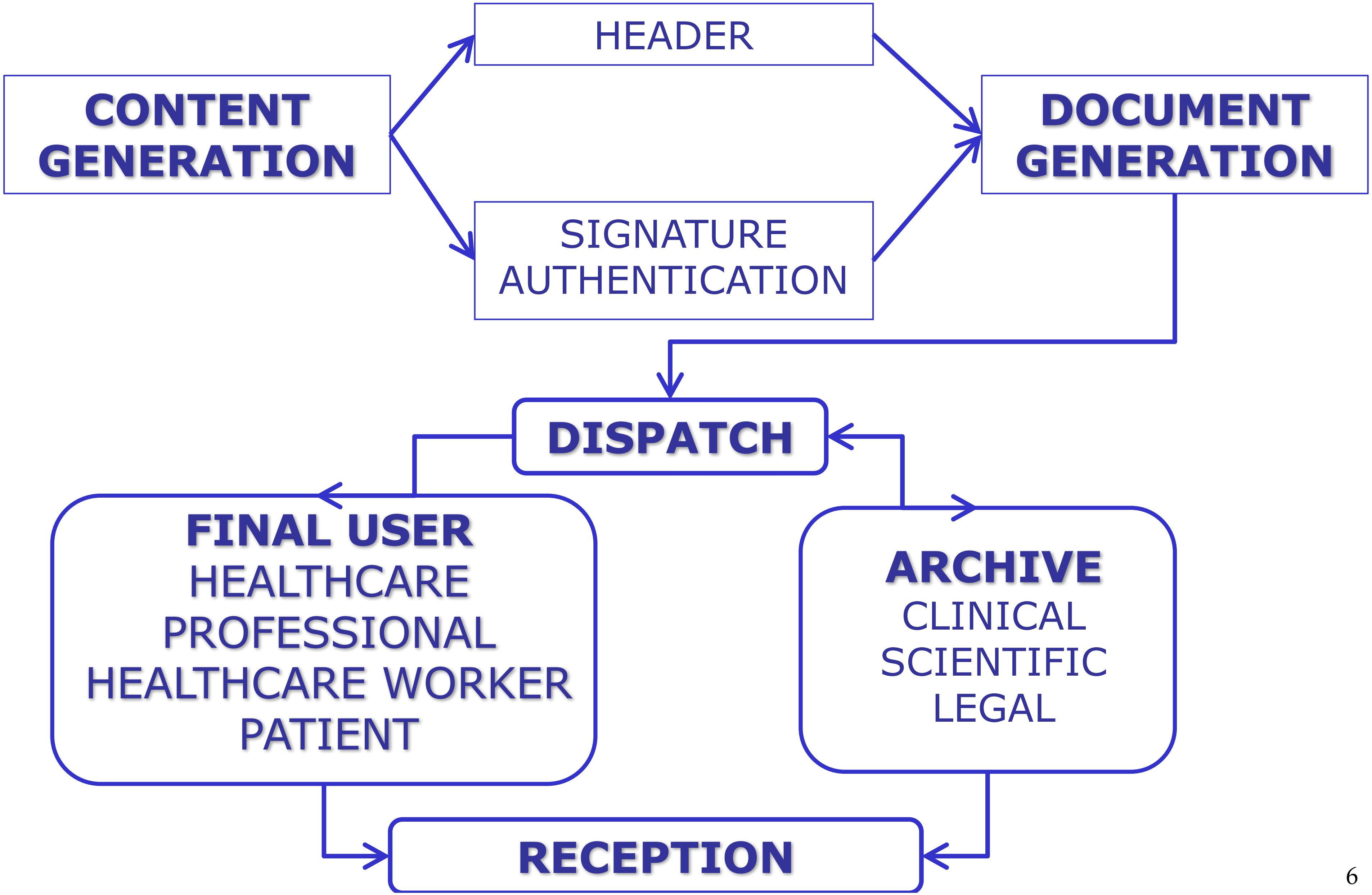
TRADE OFF

UNIVERSITÀ DEGLI STUDI DI TRIESTE

- Safety and usability → the safer the less usable

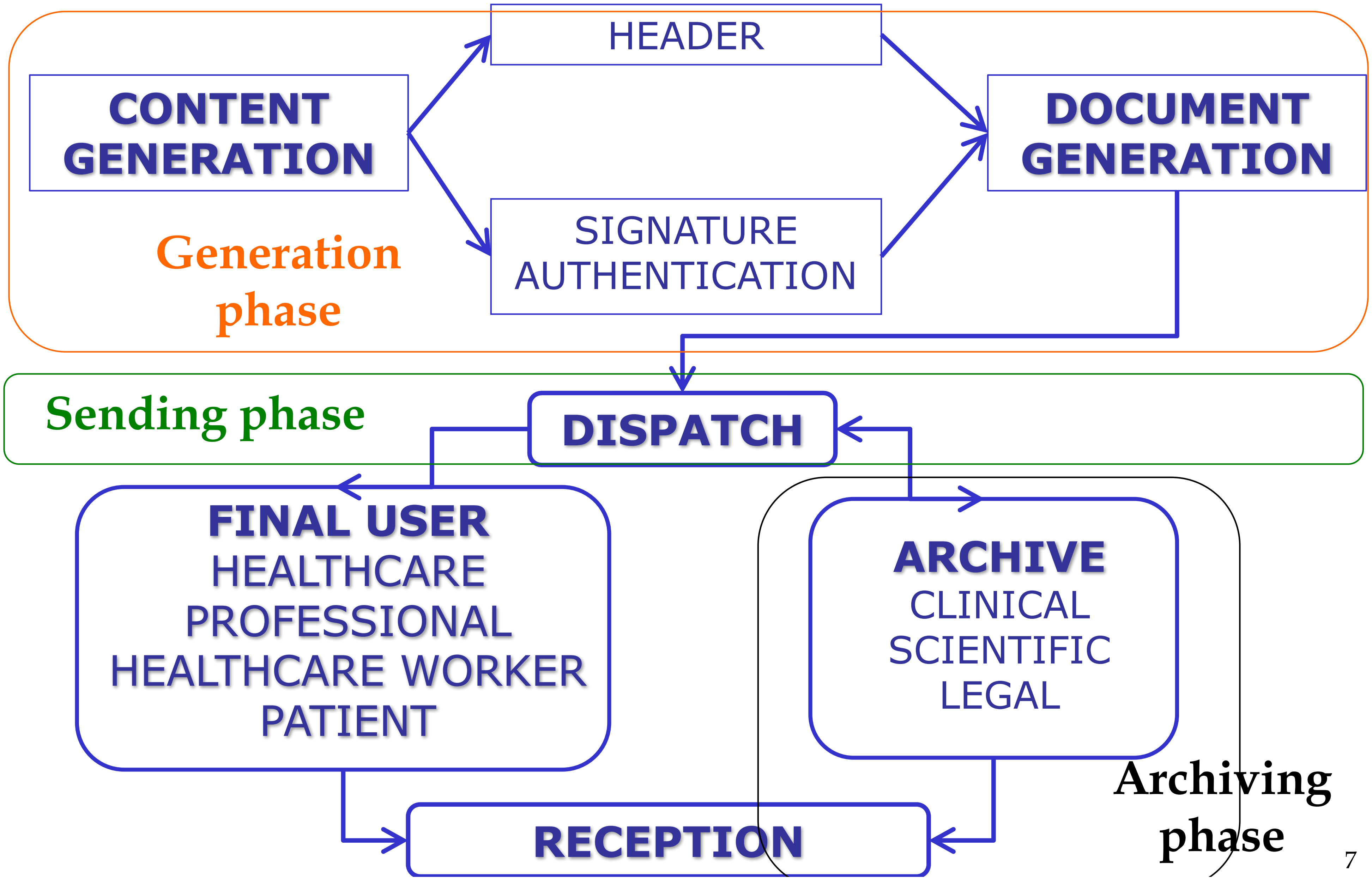
- Data sharing and system integration is required in medicine to allow all the healthcare team to ensure continuity of care.

The medical document life cycle



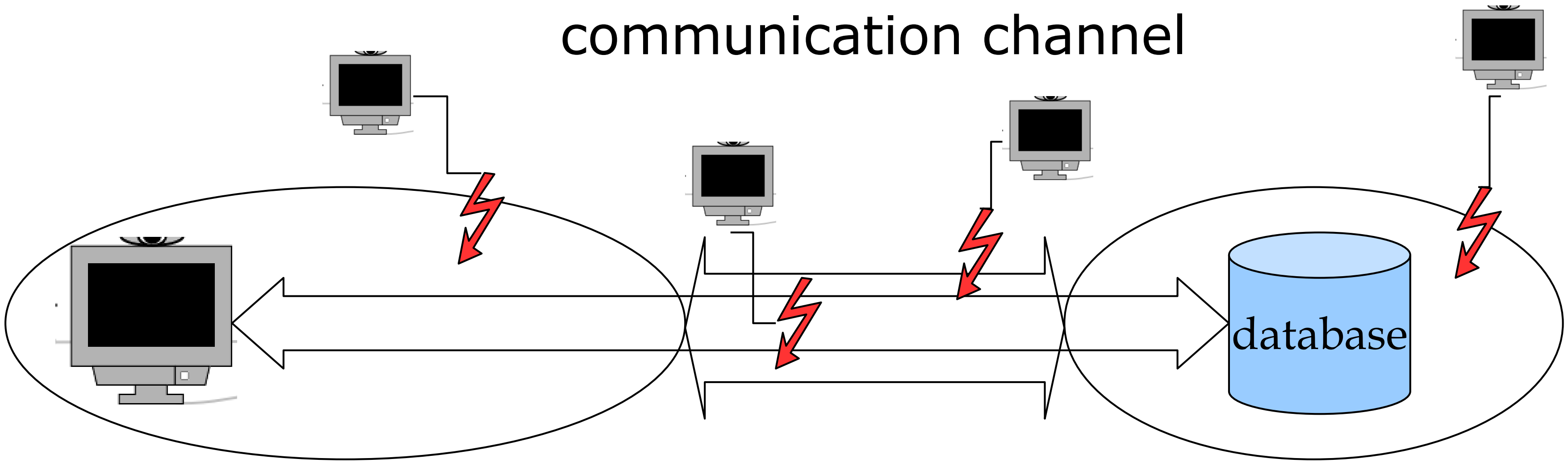
The medical document life cycle

UNIVERSITÀ DEGLI STUDI DI TRIESTE



The critical phases for data protection

INTERCEPTION
Unwanted access in the communication channel



FALSIFICATION
Unwanted access during the generation phase

UNAUTHORIZED ACCESS
Unwanted access to a data archive

FALSIFICATIONS

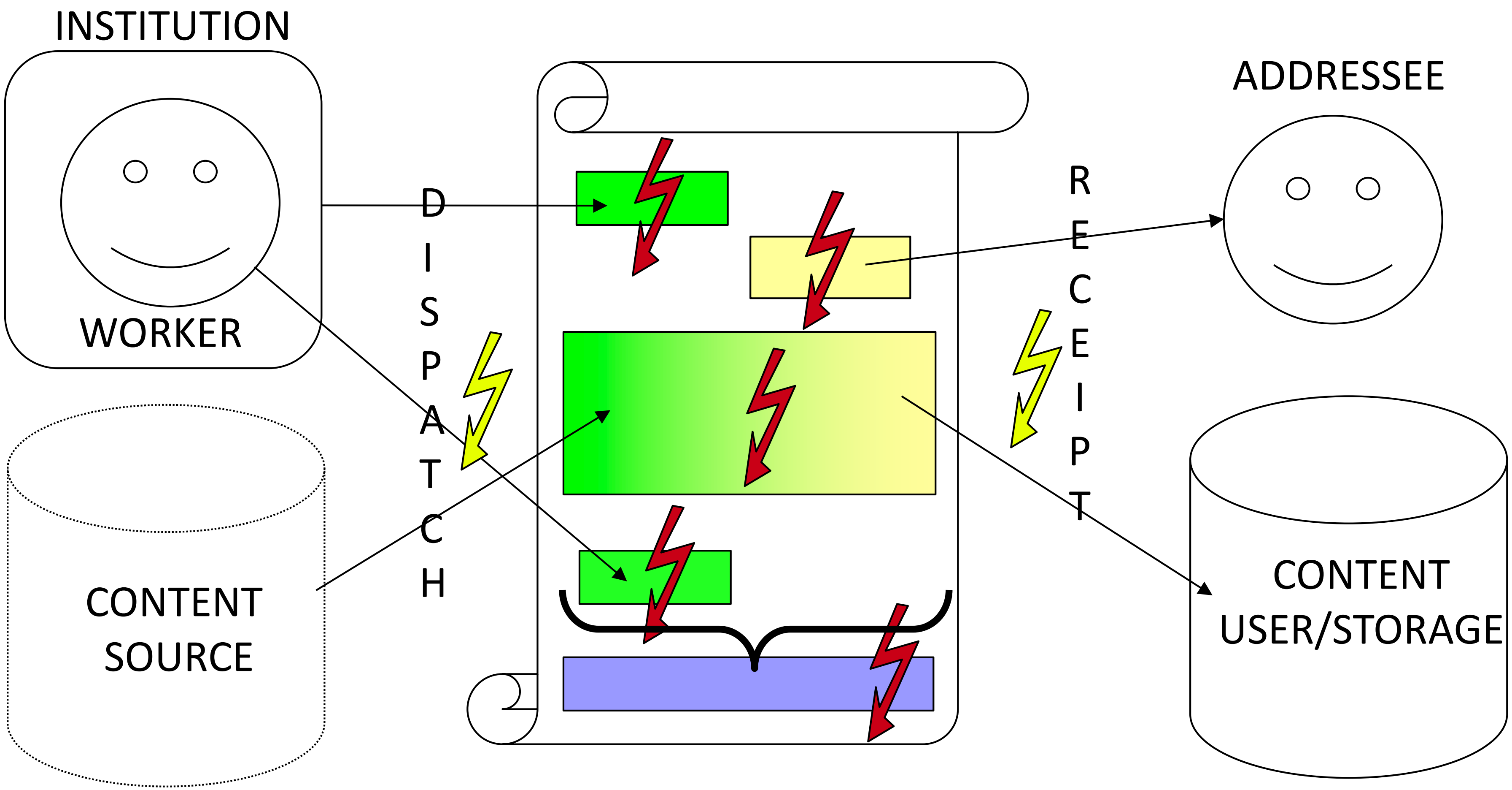
- Documents can be changed with an impact on
  - Professional ethics
  - Coherence
  - Legal implications

- All documents parts can be counterfeit

- To detect forgery, each single document has to be verified.

FALSIFICAZION TYPES

UNIVERSITÀ DEGLI STUDI DI TRIESTE



FALSIFICATION EXAMPLES

HEADER:

- Ensures that the document has been issued by an Institution who takes the responsibility for its content
- The institution can be fake
- There are lists of accredited institutions to verify

ADDRESSEE

- Ensures that the docuement is received by whom was intended to
- Privacy concerns
- Difficult to verify

CONTENT

- Information delivered in the document
- Problem of data reliability

SIGNATURE/AUTHENTICATION

- Ensures that the document has signed by someone who takes the responsibility for its content
- The person signing can be not authorized to sign
- There are lists of accredited healthcare professionals to verify

INTERCEPTION: DAMAGES

- It can be during:
  - The dispatch phase
  - The receipt phase

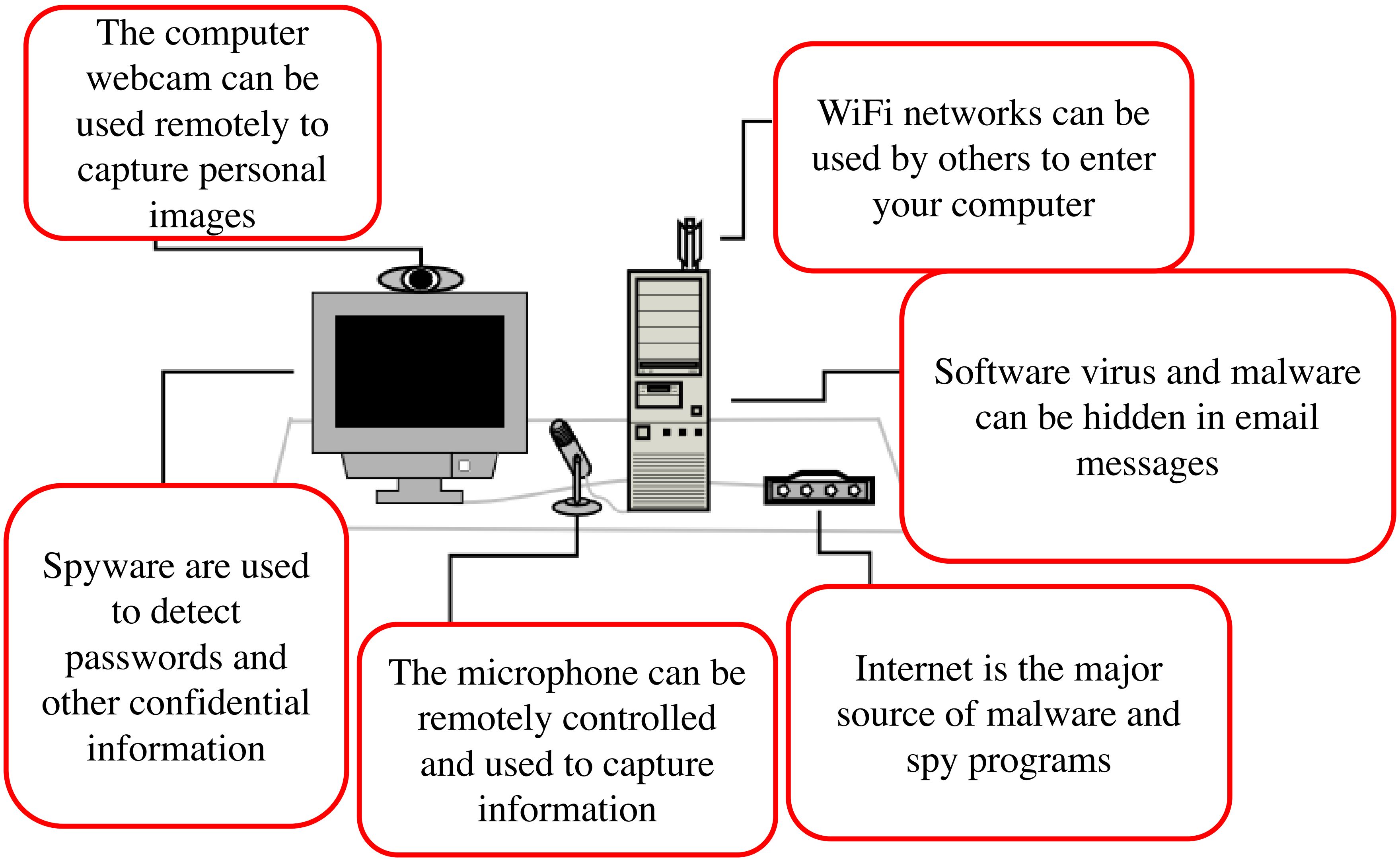
- The document can be:
  1. Stolen and lost
  2. Stolen and changed/forged
  3. Read by someone who is not authorized
  4. Copied by someone who is not authorized
  5. Redirected or used by another sender

UNWANTED ACCESS: POSSIBLE DAMAGES

UNIVERSITÀ DEGLI STUDI DI TRIESTE

1. Integrity disruption → the document is totally or partially damaged
2. Data falsification → the document is changed/forged
3. Provacy violation → the document is read by someone who is not authorized
4. Knowledge theft → the document is copied by someone who is not authorized

SOME SIMPLE ATTACK TOOLS



DAMAGE TYPES (1)

UNIVERSITÀ DEGLI STUDI DI TRIESTE

1. VIRUS: data disruption

2. SPYWARE: collection of user's information that are then given to others

3. BACKDOOR: allow the unwanted access to the system or its remote control

4. AD HOC PROGRAMS: to access the system, forgery, knowledge theft, sabotage

DAMAGE TYPES (2)

UNIVERSITÀ DEGLI STUDI DI TRIESTE

MALWARE TYPE	DAMAGE TYPE (1 to 5)				SPREADING (1 to 5)
	Integrity disruption	*Privacy violation*	*Knowledge theft*	*Falsification*	
VIRUS	4	2	2	1	5
SPYWARE	2	5	3	1	4
BACKDOOR	4	5	5	4	3
AD HOC PROGRAMS	5	5	5	5	1

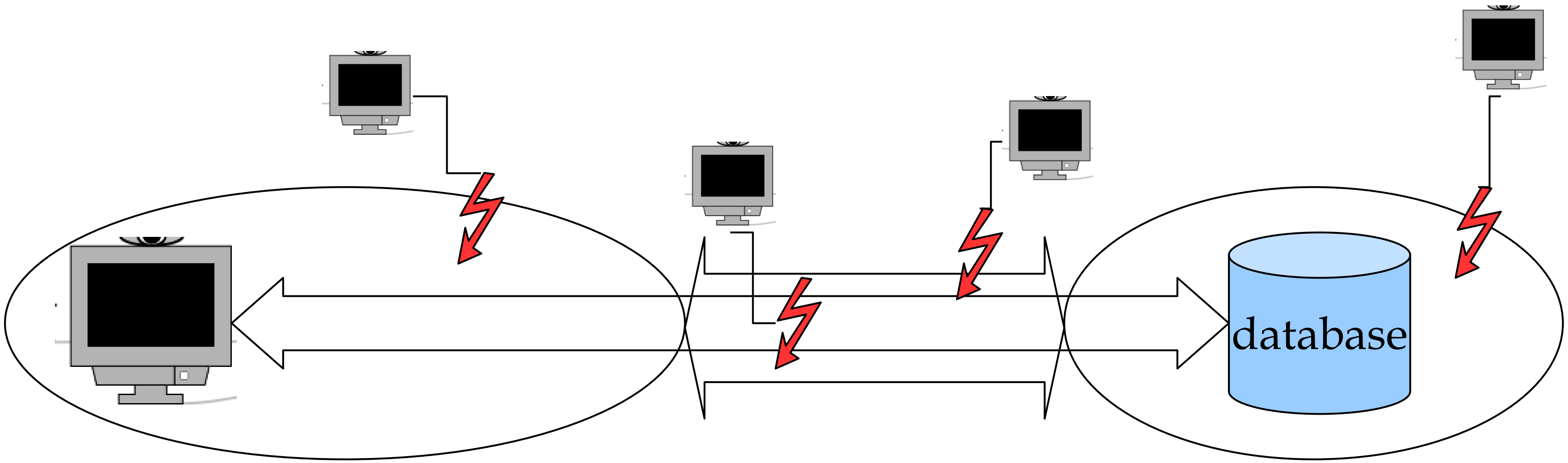
DEFENSE STRATEGY (1): BEST PRACTICES

UNIVERSITÀ DEGLI STUDI DI TRIESTE

- Preserve personal data
  - Only when necessary and using safe channels
- Be suspicious
  - Do not act when the identification of the object/person you are interacting with is not certain
- Defend the workstation
  - Firewall
  - Antivirus/Antimalware
  - Antispam
- Regular software update
  - Operative system/software update often fix security/privacy issues
- Verify attachments
  - They can include malware
- Choose the software
  - Better if open source

DEFENSE STRATEGY (2): SYSTEMIC TOOLS

INTERCEPTION
→ **CRYPTOGRAPHY**



FALSIFICATION
→ **DOCUMENT DIGEST**

UNWANTED ACCESS
→ **INFORMATICS CHECKPOINT**

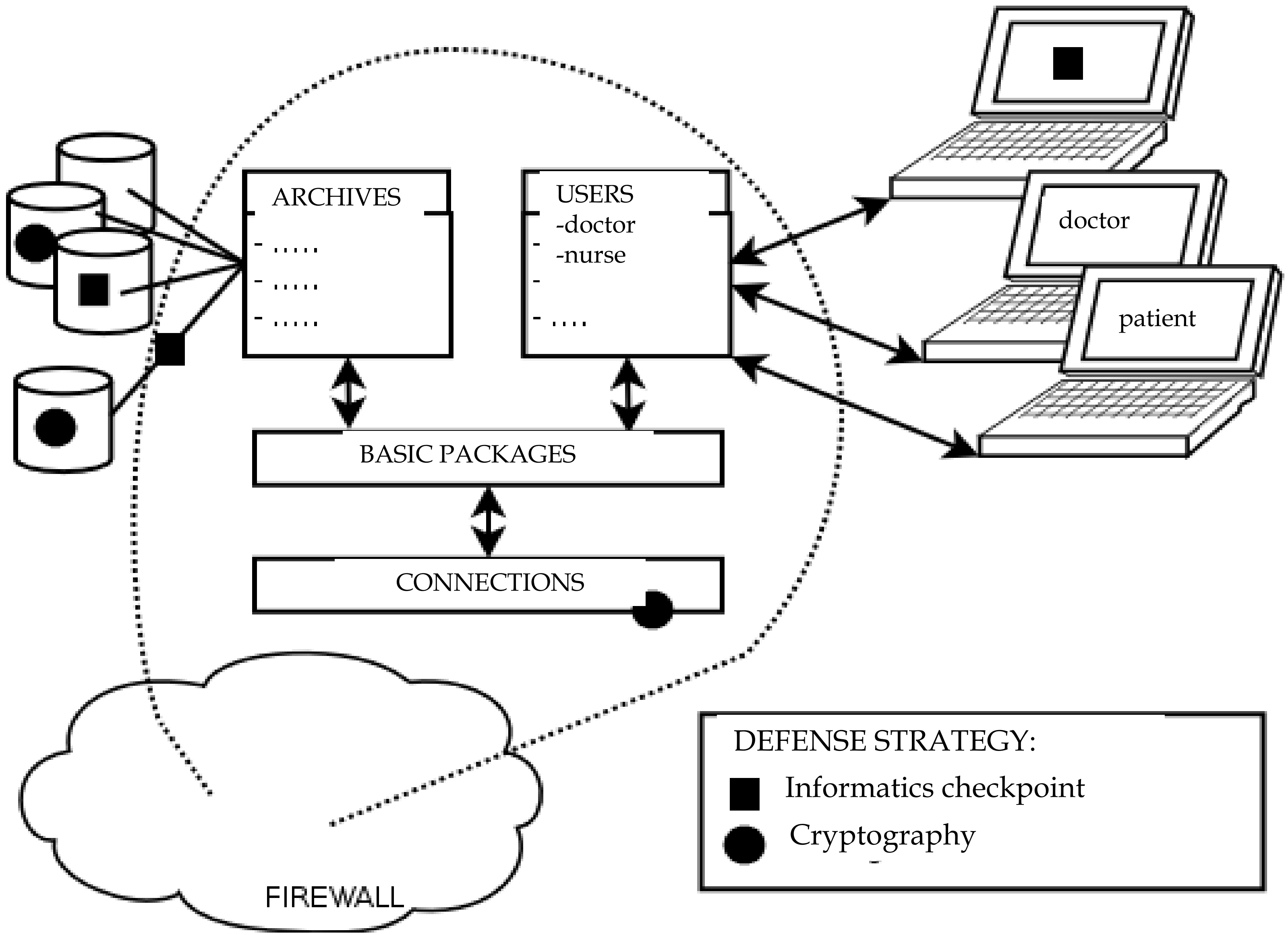
WHAT DO WE PROTECT?

- Hardware and software
  Technological infrastructures (network and components, hardware, software)

- Actors and roles
  Access policies to the services implemented→ data have to be shared among all the healthcare team, but only to those who are authorized

HARDWARE AND SOFTWARE TO PROTECT

UNIVERSITÀ DEGLI STUDI DI TRIESTE

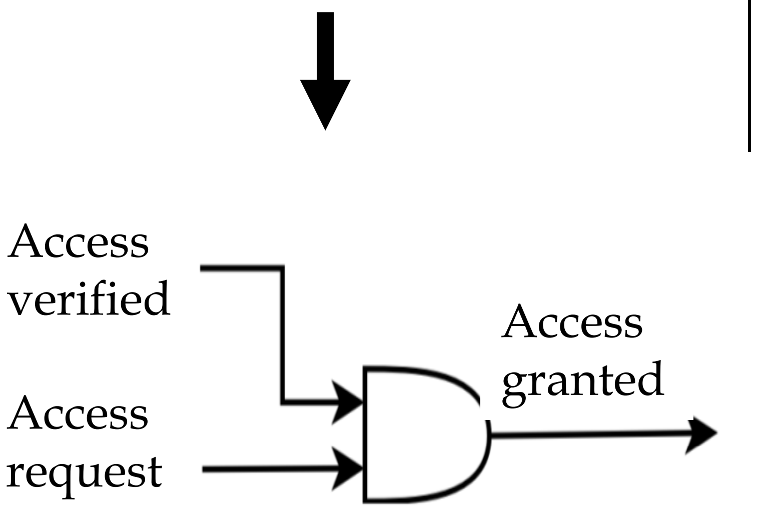


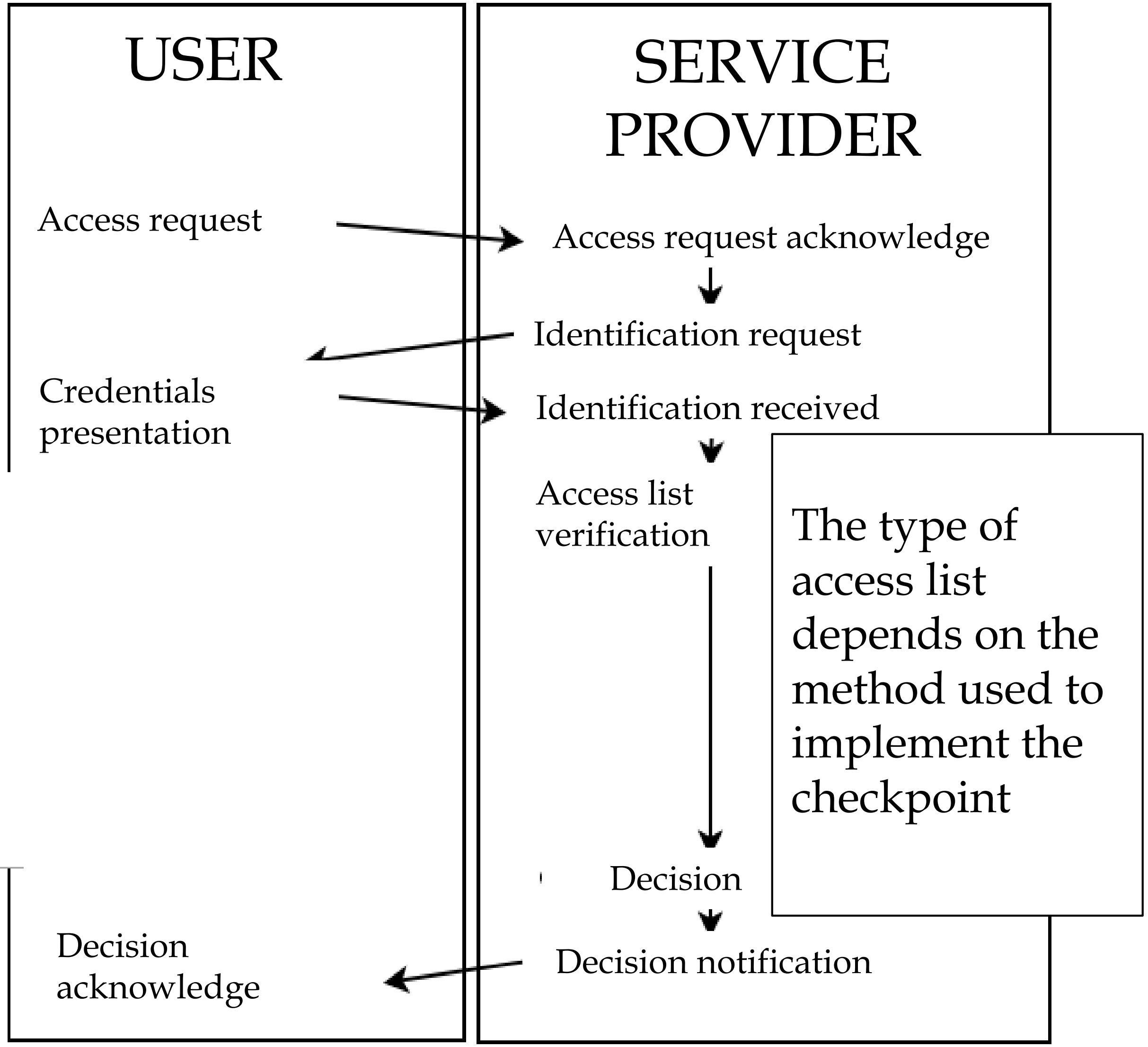
THE INFORMATIC CHECKPOINT

- It is a controlled "door"
- it requires the definition of access lists
- *Firewall* are examples of complex informatics checkpoints

INFORMATIC CHECKPOINT: basic architecrure

AND port





The type of access list depends on the method used to implement the checkpoint

A THE FIREWALL AS AN INFORMATICS CHECKPOINT

UNIVERSITÀ DEGLI STUDI DI TRIESTE

- Detection of unwanted connections from other network users or from applications running on the same computer
- Detection of dangerous web content by checking Java Applets and ActiveX controls to suggest whether or not running the application
- Port hiding → it hides the unused computer ports and monitors port scanning and access attempts
- Block of intrusions and of the attacks coming from the network

CRYPTOGRAPHY

CRYPTOGRAPHY is the practice and study of techniques for **secure communication** in the presence of third parties (called adversaries). More generally, it is about constructing and analyzing protocols that prevent adversaries to understand the content of the message

CRYPTOGRAPHY HISTORY

The word *cryptography* comes from the Greek words *kryptos* meaning hidden and *graphein* meaning writing.
Cryptography is the study of hidden writing, or the science of encrypting and decrypting text.
Nineteenth century scholars decrypted ancient Egyptian hieroglyphics when Napoleon's soldiers found the Rosetta Stone in 1799 near Rosetta, Egypt. Its inscription praising King Ptolemy V was in three ancient languages: Demotic, hieroglyphics, and Greek. The scholars who could read ancient Greek, decrypted the other languages by translating the Greek and comparing the three inscriptions.

http://www.pawlan.com/monica/articles/crypto/

THE NAVAJO CODE TALKERS

UNIVERSITÀ DEGLI STUDI DI TRIESTE

The United States Marine Corps
Navajo Code Talkers
World War II

The Navajo Nation, when called upon to serve the United States, contributed a precious commodity never before used.
In the midst of the fighting in the South Pacific, a gallant group of young men from the Navajo Reservation utilized our language in coded form to help speed the allied victory.

Equipped with the only fool proof, unbreakable code in the history of warfare, the Navajo Code Talkers confused the enemy with an earful of sounds never before heard by code experts.

The dedication and devotion to duty shown by the men of the Navajo Nation in serving as radio code talkers
in the United States Marine Corps during World War II is an example for all Americans, the Navajo Nation and graduates of WRHS.

It is fitting that at this time we also express appreciation for the Navajo Code Talkers who lived among the communities of Fort Defiance, Old Sawmill, St. Michaels and the Window Rock areas, and the families
who served the population with their children being former students and alumni of Window Rock High School.

http://www.wrscouts.com/code_talkers.htm

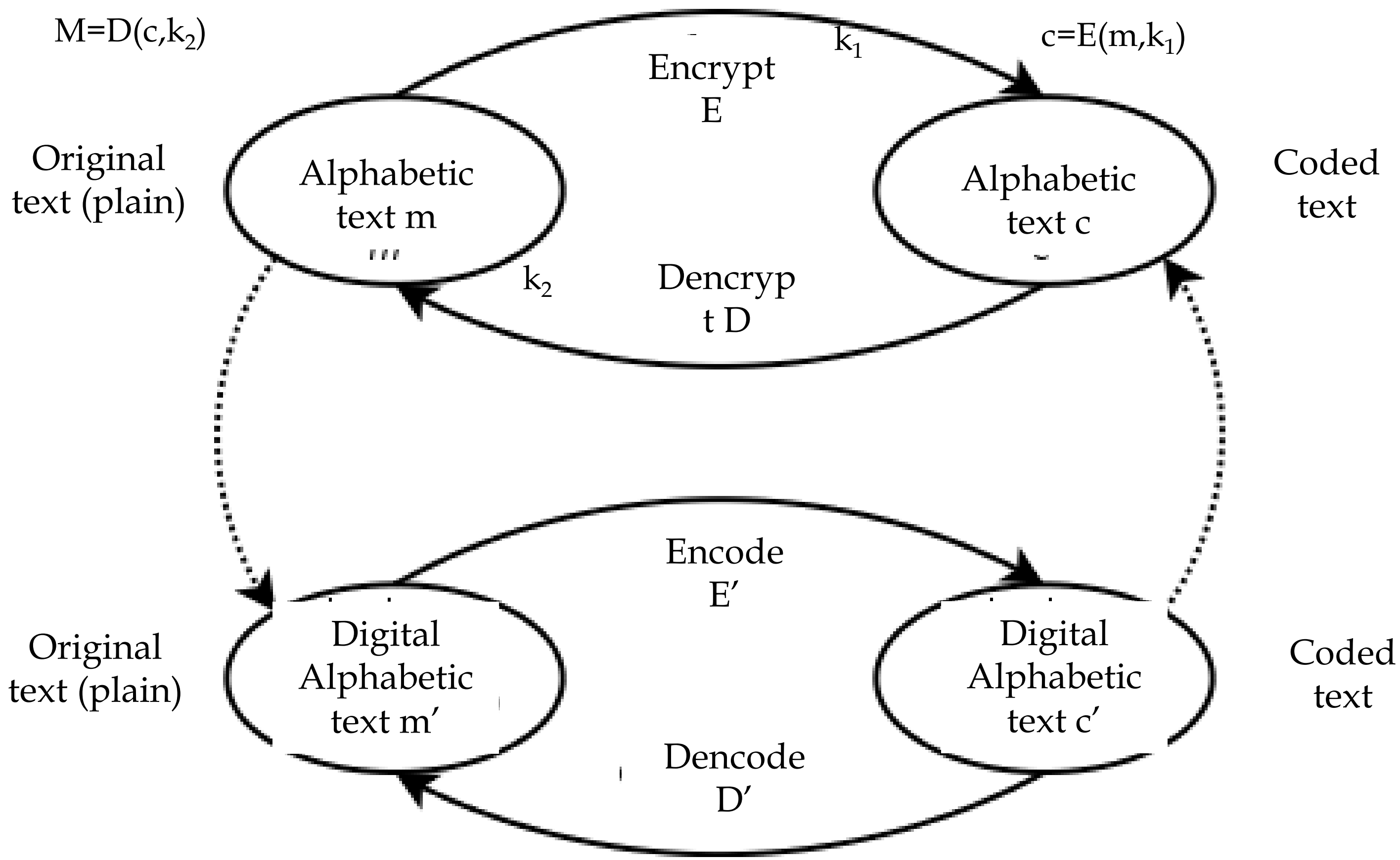
Navajo Code Talkers' Dictionary

http://www.history.navy.mil/faqs/faq61-4.htm

CRYPTOGRAPHY ARCHITECTURE

ALGORITHM → Mathematical process or method that transforms a plain text into a non-readable text

KEY (k_i) → Information (usually alphanumeric) that is able to modify te behaviour of the cryptographyc algorithm.



D e k_2 devono essere corretti rispetto ai corrispondenti E e k_1, per cui avremo M=m

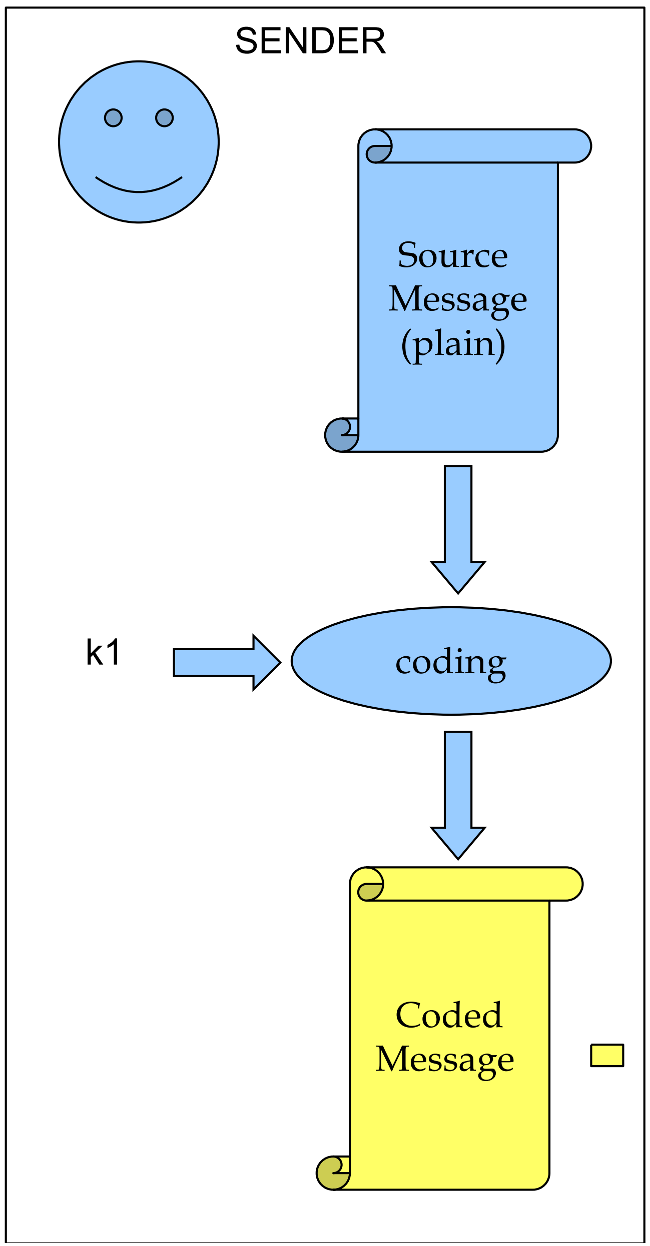
UNIVERSITÀ DEGLI STUDI DI TRIESTE

KERCKHOFFS PRINCIPLE

- The security of a cryptosystem should depend solely on the secrecy of the key and the private randomizer.
- A method of secretly coding and transmitting information should be secure even if everyone knows how it works

SYMMETRIC AND ASYMMETRIC CRYPTOGRAPHY

UNIVERSITÀ DEGLI STUDI DI TRIESTE

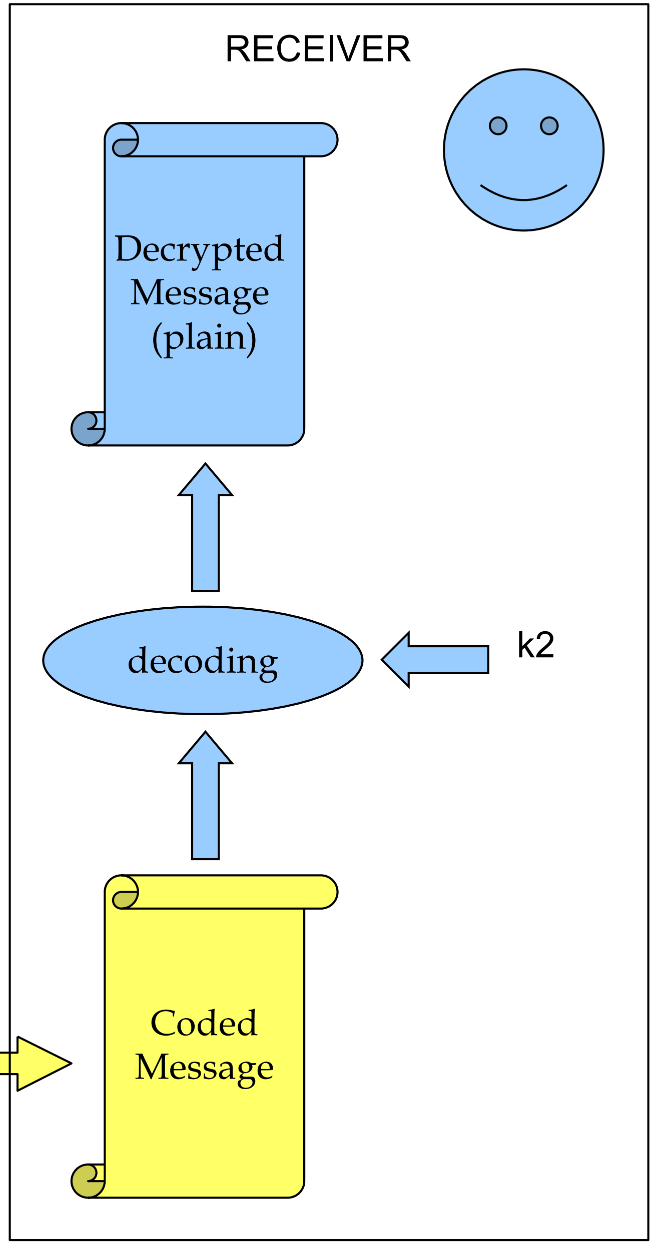


k1=k2 → symmetric key cryptography

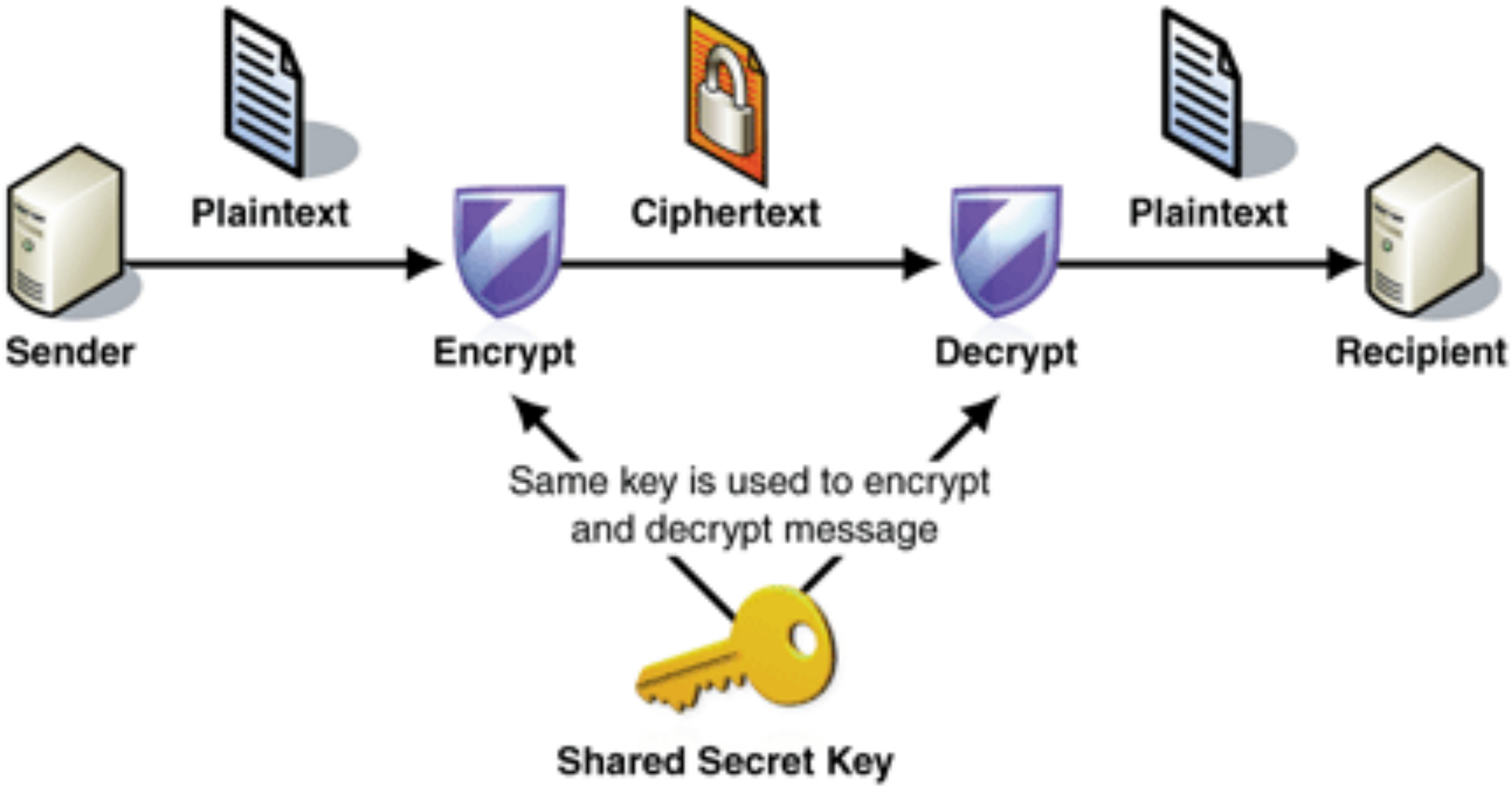
k1<>k2 → asymmetric key cryptography:

k1 receiver public key

k2 receiver private key

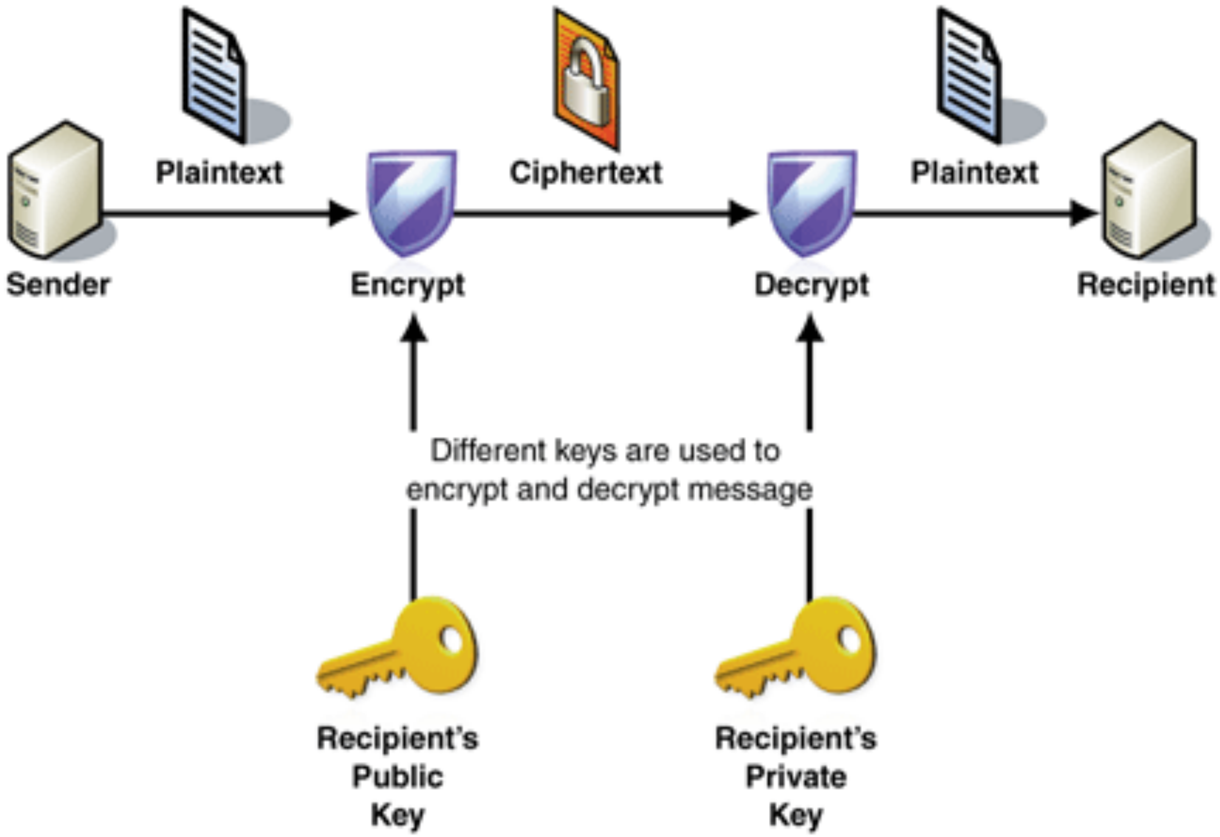


SYMMETRIC ENCRYPTION



- The sender and the recipient have to share the key
- The key is used both to encrypt and to decrypt

ASYMMETRIC ENCRYPTION



- The public key of the recipient is used only to encrypt data (cannot decrypt). It can be openly distributed to those who want to encrypt a message to the recipient.
- The private key of the recipient is used to decrypt messages, and only the recipient must be able to access it.

ASYMMETRIC ALGORITHMS: THE TWO LOCKERS MECHANISM

SENDER A

RECEIVER B

Prepares the message *m*, puts it in a chest, closes the chest with an α locker (he owns the only key), sends the locked chest to B

1

Receives the chest, closes it with a second locker β (he owns the only key), and sends it back to A.

2

Recieves the chest, opens the α locker, and sends the chest to B.

3

Riceves the chest, opens the locker β, opens the chest, and reads the message *m*.

KEY GENERATION: THE RSA SYSTEM

UNIVERSITÀ DEGLI STUDI DI TRIESTE

It is based on a factorization problem in prime numbers of a big number

1. Choose two distinct prime numbers p and q.
   - For security purposes, the integers p and q should be chosen at random, and should be of similar bit-length. Prime integers can be efficiently found using a primality test.
2. Compute $n = pq$.
   - n is used as the modulus for both the public and private keys. Its length, usually expressed in bits, is the key length.
3. Compute $\phi(n) = \phi(p)\phi(q) = (p - 1)(q - 1) = n - (p + q - 1)$, where ϕ is Euler's totient function. This value is kept private.
4. Choose an integer e such that $1 < e < \phi(n)$ and $\gcd(e, \phi(n)) = 1$; i.e., e and $\phi(n)$ are coprime.
   - e is released as the public key exponent.
   - e having a short bit-length and small Hamming weight results in more efficient encryption – most commonly $2^{16} + 1 = 65{,}537$. However, much smaller values of e (such as 3) have been shown to be less secure in some settings.[5]
5. Determine d as $d \equiv e^{-1} \pmod{\phi(n)}$; i.e., d is the modular multiplicative inverse of e (modulo $\phi(n)$).
   - This is more clearly stated as: solve for d given $d \cdot e \equiv 1 \pmod{\phi(n)}$
   - This is often computed using the extended Euclidean algorithm. Using the pseudocode in the *Modular integers* section, inputs a and n correspond to e and $\phi(n)$, respectively.
   - d is kept as the private key exponent.

The *public key* consists of the modulus n and the public (or encryption) exponent e. The *private key* consists of the modulus n and the private (or decryption) exponent d, which must be kept secret. p, q, and $\phi(n)$ must also be kept secret because they can be used to calculate d.

- An alternative, used by PKCS#1, is to choose d matching $de \equiv 1 \pmod{\lambda}$ with $\lambda = \text{lcm}(p - 1, q - 1)$, where lcm is the least common multiple. Using λ instead of $\phi(n)$ allows more choices for d. λ can also be defined using the Carmichael function, $\lambda(n)$.

The Advanced Encyption System (AES)

- AES is a block cipher with a block length of 128 bits.

- AES allows for three different key lengths: 128, 192, or 256 bits. Most of our discussion will assume that the key length is 128 bits.

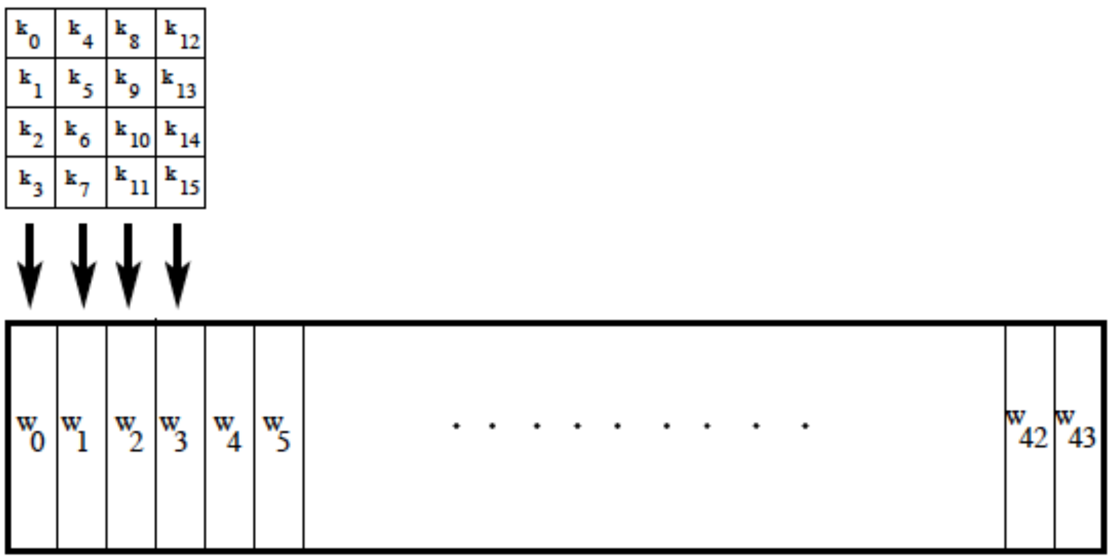
- Encryption consists of 10 rounds of processing for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys.

- Except for the last round in each case, all other rounds are identical.

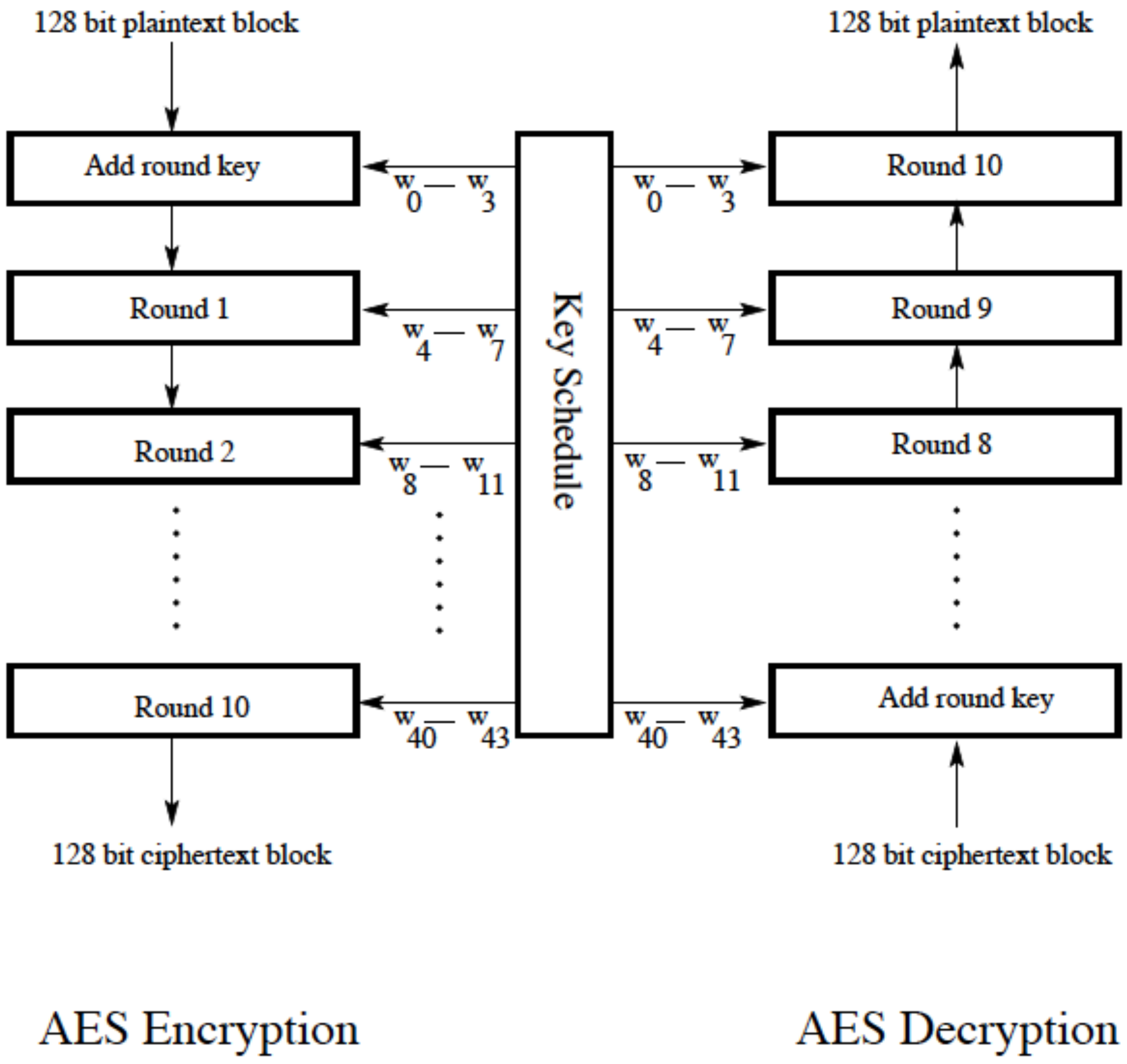
AES key

- Assuming a 128-bit key, the key is also arranged in the form of a matrix of 4×4 bytes. As with the input block, the first word from the key fills the first column of the matrix, and so on.

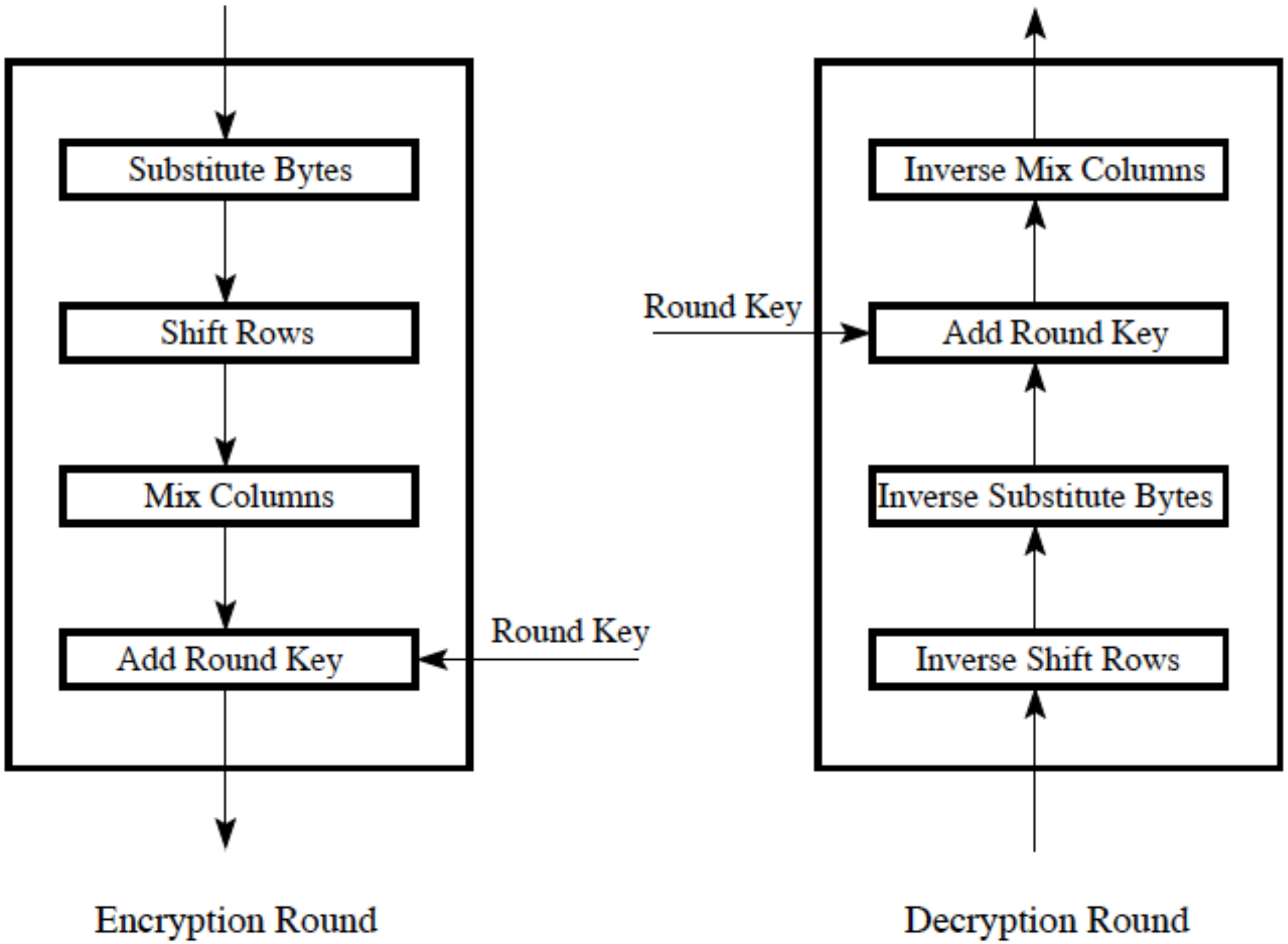
- The four column words of the key matrix are expanded into a schedule of 44 words. (As to how exactly this is done, we will explain that later in Section 8.8.) Each round consumes four words from the key schedule.



AES overall structure



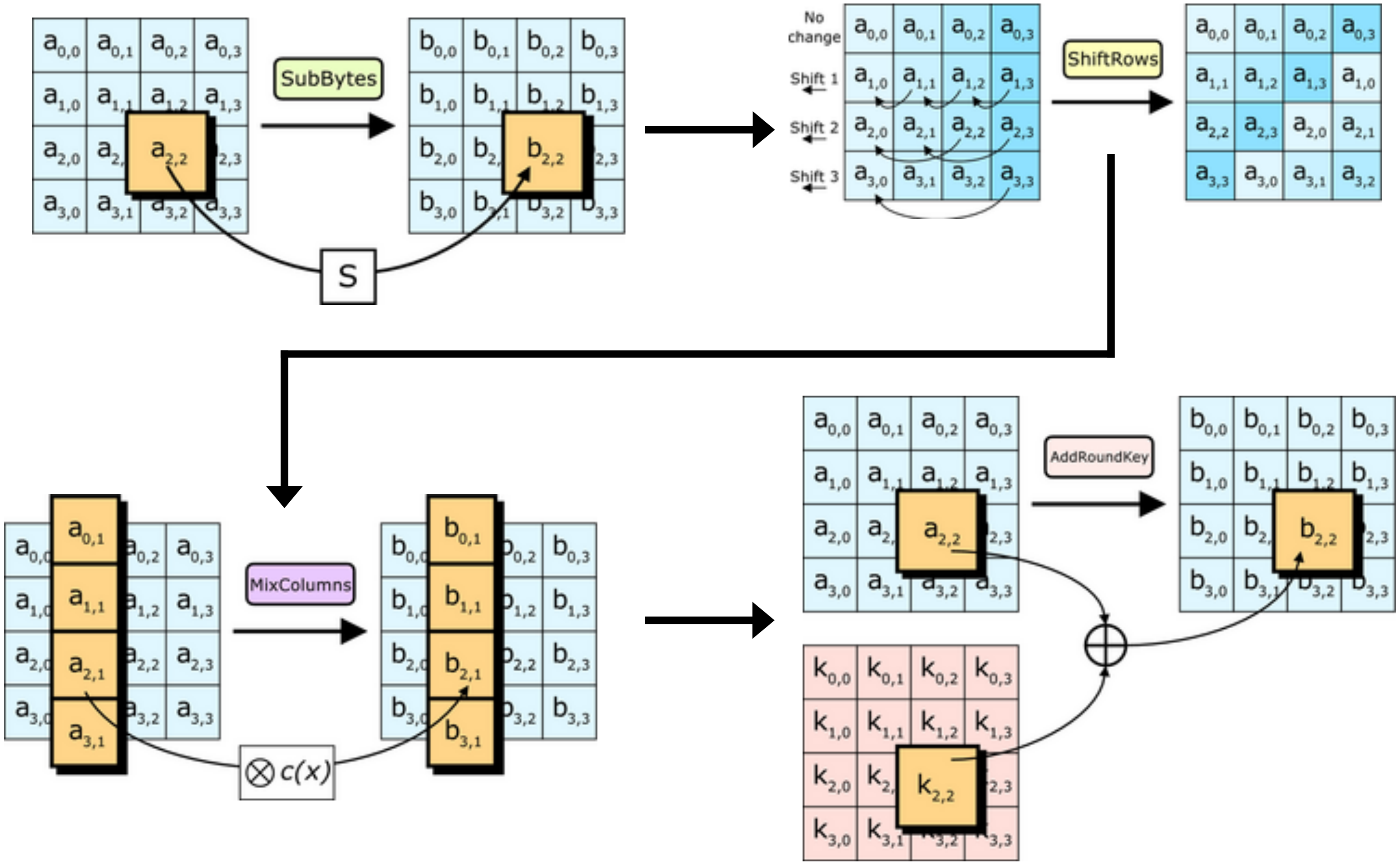
AES single round



THE FOUR STEPS

1. SubBytes — non linear substitution of all the bytes according to a specific table
2. ShiftRows — byte shifting of some positions on their row
3. MixColumns — Byte combination using a linear operation on columns.
4. AddRoundKey — each byte in the table is combined with the round key.

The 4 steps


SubBytes
S
No change
Shift 1
Shift 2
Shift 3
ShiftRows
MixColumns
⊗ c(x)
AddRoundKey

SYMMETRIC VS ASYMMETRIC CRYPTOGRAPHY

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Algorithm	Advantages	Disvantages
Symmetric key	▪Easy to implement ▪Low computational requirements → speed execution	▪Need to share the key
Asymmetric key	▪Different keys for the sender and the receiver ▪Knowing the public key does not allow decrypting the message	▪More difficult to implements ▪High computational requirements → slow execution

Secure Socket Layer (SSL)

- The Secure Sockets Layer (SSL) is a computer networking protocol that manages server authentication, client authentication and encrypted communication between servers and clients.
- SSL uses a combination of public-key and symmetric-key encryption to secure a connection between two machines, typically a Web or mail server and a client machine, communicating over the Internet or an internal network.

How SSL works

UNIVERSITÀ DEGLI STUDI DI TRIESTE

- The SSL protocol includes two sub-protocols: the record protocol and the "handshake" protocol.
- These protocols allow a client to authenticate a server and establish an encrypted SSL connection: a server that supports SSL presents its digital certificate to the client to authenticate the server's identity.
- The authentication process uses public-key encryption to validate the digital certificate and confirm that a server is in fact the server it claims to be.
- Once the server has been authenticated, the client and server establish cipher settings and a shared key to encrypt the information they exchange during the remainder of the session.
- The handshake also allows the client to authenticate itself to the server. In this case, after server authentication is successfully completed, the client must present its certificate to the server to authenticate the client's identity before the encrypted SSL session can be established.

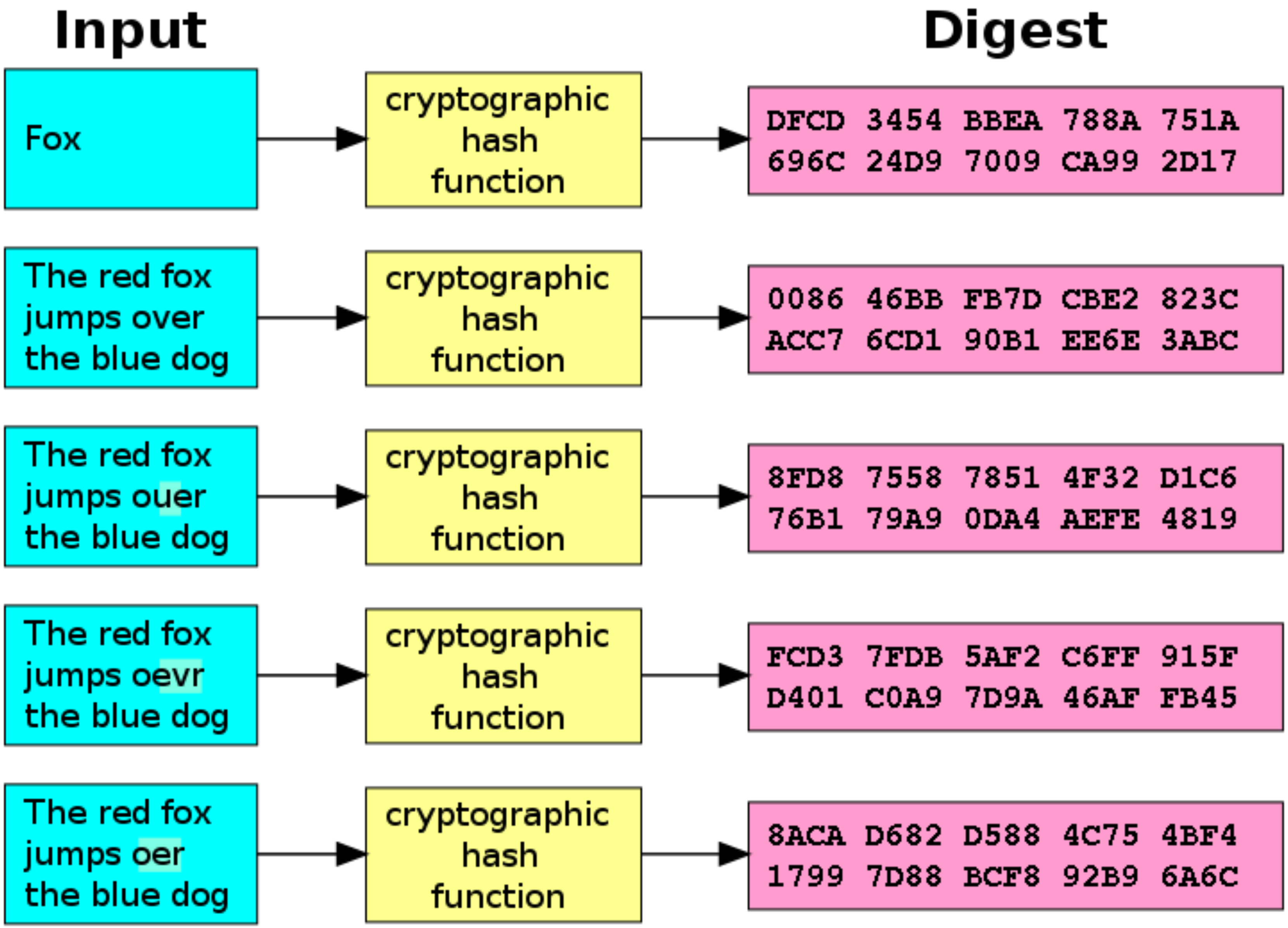
MESSAGE DIGEST

- **MESSAGE DIGEST**
  - Short string of predefined length
  - Characterizes the document
  - Verify the integrity of the document itself
  - Calculated by the sender, sent to the receiver, calculated by the receiver and compared to the one that the receiver received → if the two match → the integrity of the document is preserved

- Created through hash functions

- The ideal cryptographic hash function has **four main properties:**
  - it is easy to compute the hash value for any given message
  - it is infeasible to generate a message from its hash
  - it is infeasible to modify a message without changing the hash
  - it is infeasible to find two different messages with the same hash.

MESSAGE DIGEST EXAMPLE



SECURITY: DATA BACKUP

UNIVERSITÀ DEGLI STUDI DI TRIESTE

POSSIBLE CAUSES

- Service interrupted(heartquakes, fire, energy, malware)
- Distruction (natural events)
- Theft (or delete)

BACKUP LEVELS

- Local backup (immediate, RAID, mirror disks)
- Remote backup with short recovery time (depends on the system and the network)
- Remote backup with long recovery time (>30 km, non continuous)