Information systems security

Lorenzo Di Maio

October 2024

Contents

Chapter 1

Introduction

Cybersecurity has become a very important issue, thus, nowadays almost every system is built with ICT (Information & Communication Technologies).

There are several ways in which the attack can create a damage:

- **Financial Loss**: direct (e.g. fund transfer) or indirect (e.g. value of share or a fine by privacy authority).
- **Recovery Cost**: costs needed to take the system back to normal operations and/or to improve its security features;
- **Productivity loss:** because the processes being stopped or delayed.
- **reputation damage:** if you are a company and yo are being attacked it's difficult to regain trust.

Another a big problem nowadays is the complexity of the scenario *Information Communication Technology* (ICT):

- **"Personal" devices:**
  - desktop, laptop, tablet, smartphone, ...
  - smart TV, fridge, car, ...
- **Communication networks:**
  - data-only network: In the past we had two independent networks, the telephone network and the data network. But nowadays is no more like that, the telephone is going on data, so if the data network not working, the analogic network not working.
  - Wired and wireless networks
  - Mobility (mobility means trustability, I can detect where you are allocated)
- **Distribute devices**: company are trying to give out computer and networks, but it's means that you are not controlling your computer (outsourcing, hosting, farming, cloud, IoT).
- **Programming become increasingly complex**: stratification, framework, libraries etc...

First Axiom of Engineering

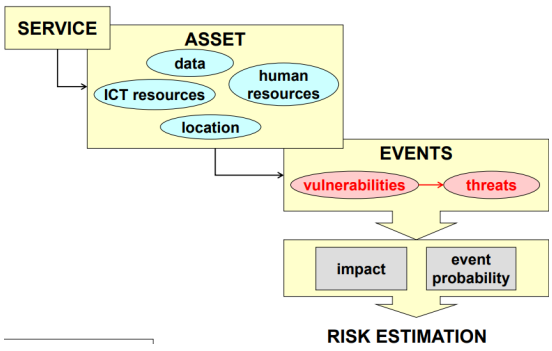
The more complex is, the more difficult its correctness verification will be (implementation, management, operation).

Example: the number of bugs in a program is proportional to its number of lines of code.

N.B. The complexity of the current information systems is in favour of the attackers that find attack paths increasingly ingenious and unforeseen by the defenders.

1.1 Risk estimation

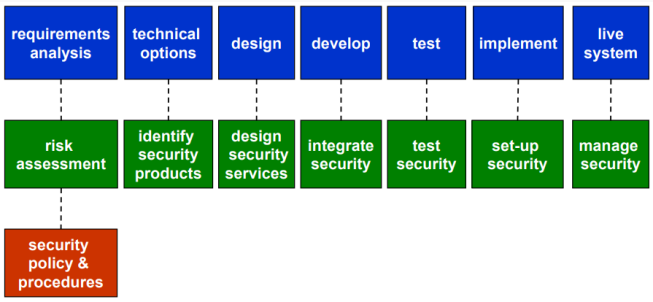
If you want to create some kind of protection, the first step is **to understand the risks**. Analyzing a system and its service in search of risk means finding the:



- **Assets**: set of ICT resources, data, people and locations needed for an IT service which needed to be protected.
- The **events**, represents the potential issues or incident that can affect the assets. Events are linked to two key concepts:
  - Vulnerabilities: intrinsic weakness of an asset (e.g.login; sensible to flooding).
  - Threats: possible deliberate action/accidental event that can produce the loss of security property by exploiting a vulnerability. (It depends upon the specific environment and/or operating conditions)

What is the correct time to think about security when we develop a system?

During the development of a system, security **must** be taken into account at every step of its lifecycle.



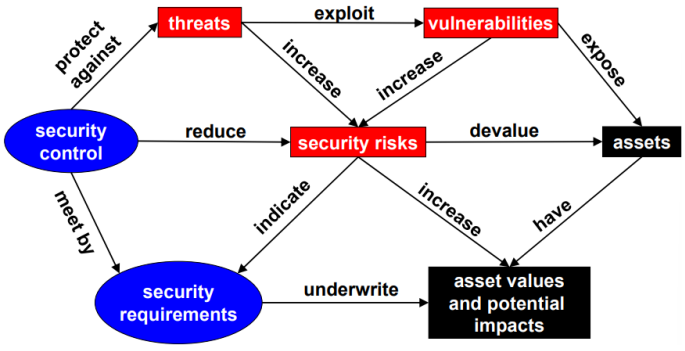
After the identification of the risk, we need to prioritize them keeping into account not only the impact but also the available time and budget. A risk assessment matrix (or risk heat map) may be useful. On the X-axis, we have the **probability** while on the Y-axis, we have the **impact**. Both of them are represented by a numerical score and each cell in the matrix is calculated by multiplying the impact by the probability:

	improbable (1)	remote (2)	occasional (3)	probable (4)	frequent (5)
catastrophic (5)	5	10	15	20	25
significant (4)	4	8	12	16	20
moderate (3)	3	6	9	12	15
low (2)	2	4	6	8	10
negligible (1)	1	2	3	4	5

1.2 Risk Management

After identifying the risks, we had to select possible countermeasures and implement them. Following this, we need to take the exam, which in real life is **Audit**: an independent person/-company that checks if the countermeasures are effective.

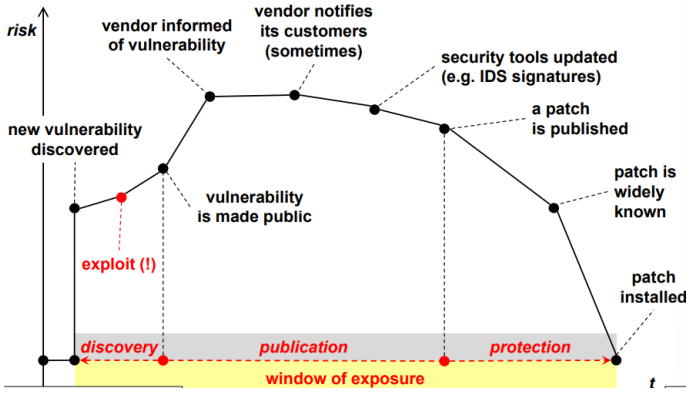
The solution in this graph is represented by the **Security control**, which is a element that you put in place to create some security features, to match the security requirement. (For example, the seatbelt is a security control for a car).



Some terminology

- **Incident**: a security event that compromises the integrity, confidentiality, or availability of an information asset.
- **(data) breach**: an incident that results in the disclosure or potential exposure of data.
- **(data) disclosure**: a breach for which it was confirmed that data was actually disclosed (not just exposed) to an unauthorized party.

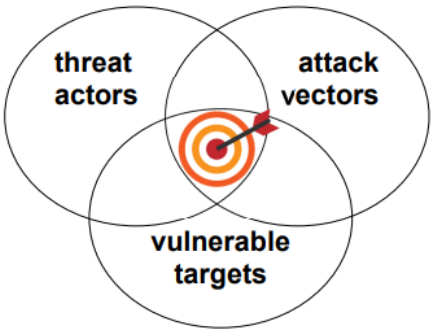
1.3 Window of exposure



1. **New vulnerability discovered**: this is the point where a vulnerability in a system is first discovered by a researcher or an attacker. The risk start to increase as this vulnerability can be exploited if it becomes known to malicious actors.
2. **Vulnerability is made public**: the vulnerability is disclosed publicly. This increases the risk further because now the attackers may attempt to exploit it.
3. **Vendor informed of vulnerability**: the vulnerability is reported to the vendor. At this point, the vendor can start working on a fix or patch.
4. **Vendor notifies its costumers (sometimes)**: Sometimes the vendor informs its customers about the vulnerability and the steps they can take to protect their systems.
5. **Security tools updated**: Security tools like Intrusion Detection Systems (IDS) or antivirus programs are updated to detect and mitigate attacks.
6. **A patch is published**: The vendor releases a patch or software update that fixes the vulnerability.
7. **Patch is widely known**: The patch becomes well-known and more organizations become aware of it. The risk continues to decrease as more systems apply the patch.
8. **Patch installed**: When the patch is finally installed across systems, the vulnerability is completely resolved, and the risk drops to nearly zero.

1.4 Cyber threats

Cyber Threat are composed by 3 main components:



- **Threat actors**: they usually follow a set of **MICE** motivations:
  - **M**oney: direct transfer, blackmail, ... or indirect (etc. data reselling).
  - **I**deology: political, religious, hacktivism,...
  - **C**ompromise: individuals with no choice due to blackmail or threat against their families or themselves.
  - **E**go: bragging around and positive reputation.
- **Attack vectors** (vulnerabilities and context)
- **vulnerable targets** (value for owner and attacker)

N.B. To successfully execute a cyber threat, there must be an intersection of all three elements.

1.5 Standardization bodies

ISO	International Organization for Standardization
ITU-T	International Telecommunication Union, Telecommunication standardization sector
ISOC	Internet Society
NIST	National Institute of Standards & Technology
ANSI	American National Standards Institute
ETSI	European Telecommunications Standards Institute
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
BSI	British Standard Institution
UNI	Italian National Body for Standards

1.6 Security

Security is **not** a product, but a **process** that recognizes the inherit insecurity in the products.
Flaws, vulnerabilities and risks are **inevitable** → we **must** try to reduce the risk of exposure regardless of the products and patches.

In order to do so, security needs to follow a series of core **principles**:

- **Security in depth**: this principle promotes for implementing multiple layers of security controls throughout an information system. The idea is that if one layer fails, additional layers can still provide protection.
- **Security by design**: means integrating security measures into the design and development process of systems and applications from the outset, rather than as an afterthought.
- **Security by default**: refers to configuring systems to a secure state out of the box, without requiring additional user intervention.
- **Least privilege**: stipulates that users and systems should be granted the minimum level of access necessary to perform their functions.
- **Need-to-know**: this principle principle restricts access to sensitive information to only those individuals who require it to perform their job functions.

C.I.A

The most important security principles can be summarized with the acronym **C.I.A.**:

- **Confidentiality**: ensuring that sensitive information is accessed only by authorized individuals.
- **Integrity**: ensuring the accuracy, consistency, and trustworthiness of data over its life cycle.
- **Availability**: ensuring that information and resources are available to authorized users when needed.

1.6.1 Other security properties

- **Authenticity**:is the property of being genuine and able to verify that you are trusted. Authenticity can be verified in two ways:
  - Simple Authentication: only one party in the communication (usually the user) is required to prove their identity to the other party (usually the system or server).
  - Mutual Authentication: both parties involved in the communication must prove their identities to each other.
- **Data Authentication**: is the process of verifying that the data received or accessed is genuine and has not been altered during transmission or storage.
- **Origin Authentication**: is the process of verifying the identity of the source or sender of the data, ensuring that the data actually comes from a legitimate or intended sender.
- **Authorization**: is a critical concept that defines what an authenticated user or system is allowed to do after they are successfully identified.
- **Non-repudiation**: Non-repudiation provides a formal proof that is legally acceptable, ensuring that a person or system cannot deny having created, sent, or received a particular message or data.
- **Accountability/Traceability**: is the security goals that generates the requirement for actions of an entity to be traced **uniquely** to that entity.

1.7 Data protection

For each security property, consider always the three cases of data protection:

1. **"data in transit"** : when data are transmitted over a communication channel.
2. **"data at rest"**: when data are stored in a memory device.
3. **"data at work"**: when data are in RAM for use by a process.

Where is the enemy?

One of the first question when we want to create some kind of protection is where is the enemy is located:

- **Outside our organization**: in this case the kind of defence is called boundary/perimeter defence (firewall).
- **Outside our organization, with the exception of our partners**: in this case we will implement extranet protection (VPN). [Extranet: when you network is extended to connect to other networks].
- **Inside our organization**: we should protect the Local Area Network, this is much more difficult because the LAN are built to cooperate.
- **Everywhere**: having a defence which is based on the position today become meaningless.

Nowadays having a defence based on the position is meaningless. The alternative is to think that the attackers are **everywhere**. That is a new modern architecture, called **ZTA (Zero Trust Architecture)**. This allows the system to be completely protected by various form of attacks, both **passive** (can **only read** the data/traffic) and **active** (can **read** but also **modify**,**delete** or **create** data/traffic). We can also classify attackers based on their position:

- MITM (Man-In-The-Middle): sitting between the two peers A and B.
- MATE (Man-At-The-End): inside one peer.
- MITB (Man-In-The-Browser): inside a component of a peer (typically, the web browser).

1.8 Technological Attacks

The networks are (almost) always insecure, since most of them use wireless connections → communication via broadcast. Moreover, most networks have access to geographical connections are that not made through end-to-end dedicated lines, but through shared lines and third-party routers. Weak user authN, lack of server side authN and software bugs are also contributing factors to the successfulness and likelihood of an attack. Technological Attacks can be divided in various classes:

- **IP spoofing (masquerading)**: IP spoofing is when someone uses the address of another host, to take its place as a client (and hide its own actions) or as a server. The spoofing usually happens at **Layer 3**, which is the **IP address**. IP spoofing can also happens at **Layer 2**. A better name would be **source address spoofing** because this kind of attack, as we said before, it can also happens at Layer 2.

  Countermeasures: Avoid address-based authentication.

- **Packet sniffing (eavesdropping)**: With a packet sniffing attack I can read packets that are not addressed to me, but to another network node. If you are on a broadcast network, this is very easy, you just need to put my network card in promiscuous mode (which accept all packet even if the address doesn't match with the destination packet).

  Countermeasures:

  - **Don't use broadcast networks**, but it's difficult because must of the networks are broadcast.
  - **Encryption of packet payload** (No the header because it contains the destination).

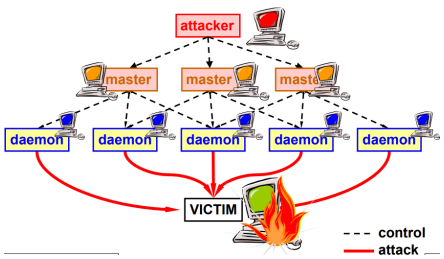
- **Denial-of-service (distributed DoS)**: Dos is a attack against availability, he keep a host busy so that it can't provide its services.

  Countermeasures: We don't have any countermeasures. Monitoring and over-sizing (in other words,making the system bigger than is needed) can mitigate the effects.

- **Distributed denial-of-service (DDOS)**:

  Attacker can spread the DoS software on many nodes (called daemons, zombies or malbots) to create a Botnet. Daemons are controlled remotely by a master node through encrypted channels (e.g. UDP packets contained inside the payload of ICMP Echo Requests),this distribution multiplies the effect of the DoS attacks by a factor equal to the number of daemons.
  DDoS attacks' effectiveness can be improved further in two ways:

  - **Reflectors**: the attacker can hide his tracks by using a legitimate third-party server.
  - **Amplification Factor N:1**: the attacker looks for a reflector server that amplifies the attack scale by a factor of N (depends on the used protocol).



- **Shadow/fake server**: The attacker manages to show itself to the victims as a service provider server without having the right to do so. To achieve this, the attacker must sniff the requests and spoof (elaborate) responses faster than the real server (e.g. can be achieved with a superior computer or via DDoS attack) or it can manipulate the routing/DNS tables. The objective of the attack is to gain knowledge of the victim's data.

  Countermeasures: possible countermeasure is to guarantee the server authentication.

- **Trojan**: A Trojan is a program that contains a dangerous payload. Typically that's done because network channels are more protected but user terminals are less protected (e.g. Smartphone, IoT, "ignorant" users [also people who want to save money]).
  Trojan are usually difficult to remove as they use **stealth techniques** to stay undetected.

- **Software bug**: In general even the best software contains bugs that can be exploited by attackers. The easiest way to exploit the bugs is DoS.

- **Malware**: Malware is a generic term that refers to all types of malicious software. It's specifically designed to harm, exploit, or otherwise compromise the security of computer systems, networks, or devices. There are many types of malware:

  - **Virus** is a program or application that damages the target and replicates itself. The virus has not the ability to propagate by itself, it's propagated by humans (involuntarily).
  - **Worm** is typically a malware that does not create damage directly, instead he damages the target by replicating itself, there by saturating the resources. It's capable to do a automatic propagation.
  - **Trojan**: malware vector.
  - **Backdoor**: it's a hidden entry point in your system. In some cases the backdoor is inserted by the police/army to foot inside an organization.
  - **Rootkit:** it's a software that it's been downloaded in your system which is providing privileged access. It remains hidden (by modifying programs, libraries, drivers, kernel modules, or hypervisors) and operates stealthily.
  - **PUA (Potentially Unwanted Applications)**: These are software programs that may not be malicious in nature, but they can negatively impact your system's performance or privacy. PUAs are often bundled with legitimate software or downloaded inadvertently by users.

  Countermeasures:

  - **User awareness**
  - **Correct configuration/Secure software**
  - **Install antivirus** (and keep updated!)

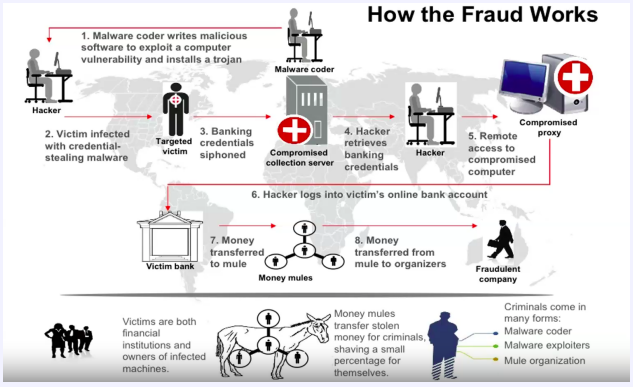
Ransomware

It's the most frequent kind of **malware** distributed nowadays. A ransowmware is a malware oriented to get a ransom (a "ransom" is a sum of money demanded for the release of someone who has been kidnapped or captured). Typically he arrived on your device and encrypt your PC with a key that you don't know, so you will not be able to read your file. You need to pay if you want your data back.
Countermeasures: A possible countermeasure consists of doing regular backups, the backup must be offline; otherwise, it can be attacked as well. However we need to pay attention to how old the backup is, because nowadays there is **silent ransomware**, this type of ransomware doesn't block your PC immediately; it encrypts your system.

Cyber Theft Ring

The cyber theft ring is a criminal organization involved in stealing money through online means.

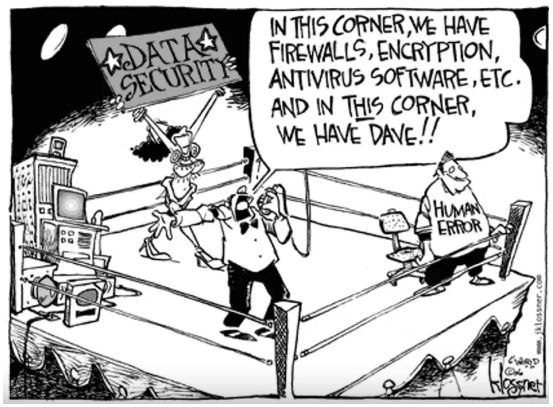
- **Malware Coders**: These individuals create malicious software (malware) that is sold on the black market. This software is used to attack victims' computers and steal their personal information, including banking credentials.
- **Malware Exploiters**: These individuals purchase malware from coders and use it to launch attacks on victims.
- **Money Mules**: These are individuals recruited to launder stolen funds. They are paid a percentage of the stolen money for their services. Money mules often transfer funds between different accounts to make it difficult for authorities to trace the money's origin.
- **Victims**: : The targets of cyber theft can include individuals, businesses, and financial institutions.



1.9 Technology and human beings

No matter how strong the technological defenses are, **human error remains a major vulnerability in cybersecurity**. Even with the best tools, human mistakes can undermine security efforts. The non technological problems are various:

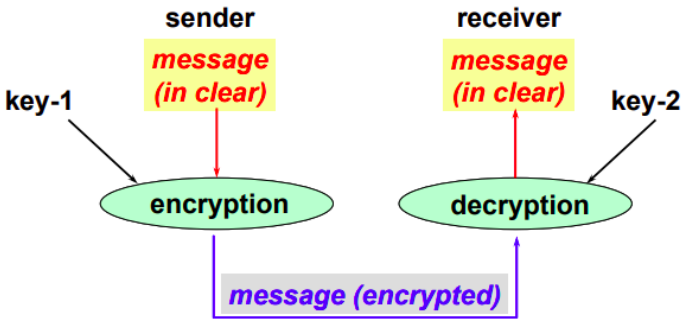
- People don't understand the problem
- Mistake of human beings (especially when overloaded, stressed, ...).
- Complex interfaces/architectures can mislead the user and originate erroneous behaviours.
- Performance decrease due to the application of security measures.



Chapter 2

Cryptography techniques

Cryptography is a mathematical technique that permits to transmit the data to a form that can't be understood by unauthorized users.



To explain the various concepts of cryptography we need to introduce some terminology:

- **Plaintext (P)**: the original message before encryption.
- **Ciphertext (C)**: the encrypted message that appears scrambled.

Kerckhoffs' Principle

This principle states that the security of a cryptographic system should rely on the **secrecy of the keys**, not the secrecy of the encryption algorithm itself. If the keys:

- are kept secure, this ensures only authorized parties can decrypt the message.
- are managed by trusted systems, this minimizes the risk of unauthorized access.
- are of adequate length, they are harder to crack with brute force methods.

It's no important that the encryption and decryption algorithms are kept secret, on the contrary it's better to make the algorithms public so that they can be widely analysed and their possible weakness identified.

Kerchoff's Principle is tightly correlated to the principle of **Security Through Obscurity (STO)**: a system is protected but the details on how it has been protected are not disclosed. This alone is not considered a valid security mechanism. STO can be used as a layer of the security system **only if** a really strong algorithm is used

2.1 Symmetric Cryptography

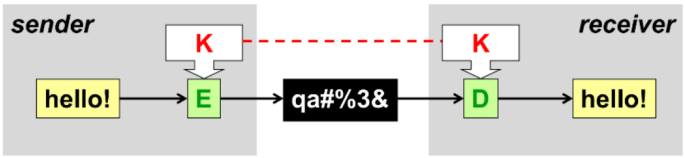
Symmetric cryptography is so called because it uses **only** a single key **shared between** the sender and the receiver.

$$key1 = key2 = K$$

The key **K** is used to generate the Ciphertext **C** by encrypting the Plaintext **P**. **C** is then sent to the receiver, which recovers **P** by applying the decryption algorithm using **K**:

$$C = enc(K, P) = \{P\}K$$

$$P = dec(K, C) = enc^{-1}(K, C)$$

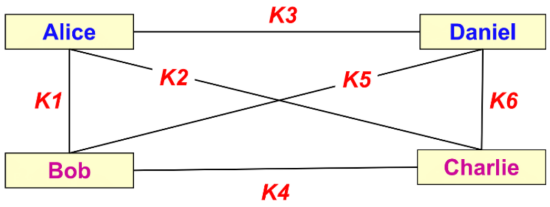


Symmetric Encryption requires the use of a unique **K** for each peer couple. A complete pairwise private communication between N parties would requires:

$$\frac{N \cdot (N-1)}{2} \text{ unique } \mathbf{K}\text{s}$$

The parties can exchanged **securely** the **K**s in two ways:

- **OOB** (**Out-Of-Band**): the parties share the **K** without using electronic channel used for transmitting the encrypted message.
- **Key Exchange Algorithms**: algorithms such as Diffie-Hellmann enable parties to securely exchange keys over a potentially insecure channel.



Symmetric Block Encryption Algorithms

The main type of symmetric is the **Block Algorithm**. Each algorithm elaborates encryption/decryption **only if** a minimum amount of data (typically 64/128 bits) is available.

name	*key (bit)*	*block (bit)*	*notes*
DES	56	64	obsolete
3-DES (2 keys)	112	64	56…112-bit strength
3-DES (3 keys)	168	64	112-bit strength
IDEA	128	64	famous for PGP
RC2	8-1024	64	usually 64-bit key
Blowfish	32-448	64	usually 128-bit key
CAST	40-128	64	usually 128-bit key
RC5	0-2048	1-256	optimal when B=2W
AES	128-192-256	128	state-of-the-art

2.1.1 DES (Data Encryption Standard)

DES uses **64 bits keys**, yet its effectiveness is only the one of a 56 bits key, since 8 bits of the key are used to check the parity of the others. This means that, when DES generates a key, every 7 bits that are generated, another is added to check the parity of the previous 7 bits. It's designed to be **efficient in hardware** because it requires:

- **XOR in details**: elementary operation on all the CPUs.

XOR

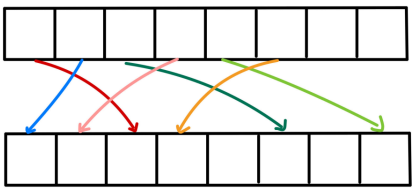
It's a fundamental operation in cryptography, it provides "confusion" by randomizing input data. ruth table:

A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Notes that it's the **inverse of itself**:

- if $A \; xor \; B \; = \; Z$ then $Z \; xor \; B \; = \; A$ (or $Z \; xor \; A \; = \; B$)

- **Shift**: elementary operation on all the CPUs.
- **Permutation**: expensive operation yet efficient if done directly in hardware.



2.1.2 3-DES or Triple DES

It was introduced as an improved version of the original DES. It is simply the repeated application of the DES. It uses **two** or **tree** keys. Usually, 3-DES is applied in the **EDE** (**Encryption-Decryption-Encryption**) **mode**:

- **With 2 keys** → Uses two 56-bit keys (K1 and K2), repeating K1 for the third operation:

$$C' \; = \; enc(K_1, P) \quad C'' = dec(K_2, C') \quad C = enc(K_1, C'')$$

Beware

If sufficient memory is available for attack (about $2^{59}B$), the effective security of this scheme is roughly **56 bits**, due to vulnerabilities. Without that much memory, the effective key size is **12 bits**.

- **With 3 keys** → Uses three distinct keys (K1, K2, and K3), providing better security with an effective key size of 112 bits:

$$C' \; = \; enc(K_1, P) \quad C'' = dec(K_2, C') \quad C = enc(K_3, C'')$$

Why 2-DES is not used?

Because the double application of any encryption algorithm is subject to a **Known-Plaintext Attack**, which is a kind of MIMT attack. This attack allows to decrypt data with **at most** 2^{N+1} attempts (if the keys are N-bit long).

During a Known-Plaintext attack both **P** and **C** ($= enc(K_2, enc(K_1, P))$ are **sniffed by** the attacker. Hypothesizing N bits keys, the attacker can then compute **K** by:

1. Compute 2^N values $X_i = enc(K_i, P)$
2. Compute 2^N values $Y_j = dec(K_j, C)$
3. Search of those values K_i and K_j such that $X_i = Y_j$
4. Discard false positive

2.1.3 AES (Advanced Encryption Standard))

AES was established through a public international competition aimed at replacing the Data Encryption Standard (DES). On October 2, 2000, Rijndael was selected as the winner of the AES competition, featuring:

- Key lengths of 128, 192, or 256 bits.
- A block size of 128 bits.
- Officially published in November 2001 as FIPS-197.
- Gradual adoption after 2010, as cryptographic algorithms need time to prove their reliability.

2.1.4 Application Modes for block algorithms

Block algorithms are designed to encrypt **fixed-size** blocks of data, so when dealing with data of varying lengths, specific techniques must be applied:

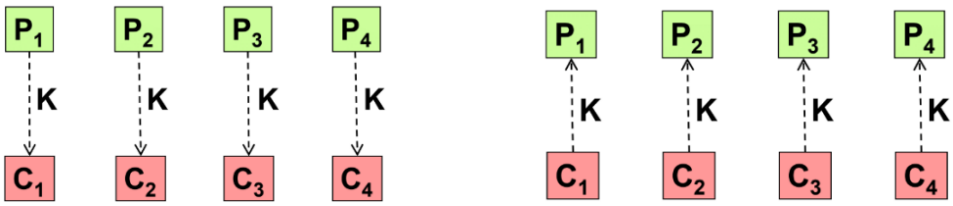
- **Size of Data to Encrypt > Algorithm's Block Size**
  - **ECB (Electronic Code Block)**: In ECB mode, each block of plaintext is encrypted independently of the others, which makes the encryption formula for the n-th block as follows:

$$C_n = enc(K, P_n)$$

    In this mode, **decryption** is the reverse of the encryption process. The formula for decrypting the n-th block is as follows:

$$P_n = enc^{-1}(K, C_n)$$

    Since there is no dependency between blocks, an error in transmission of one block generates an error at the decryption of only the affected block.



Problems with ECB mode

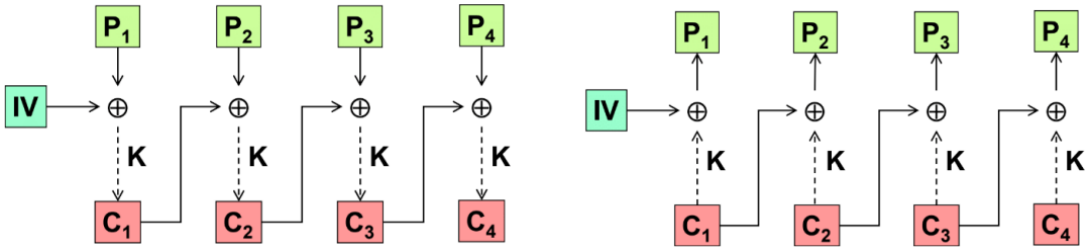
* Swapping of ciphertext blocks goes undetected: since each block is encrypted independently, if two ciphertext blocks are swapped, the decrypted message will still be valid but with blocks in the wrong order.
* Identical plaintext blocks generate identical ciphertext blocks: when two plaintext blocks are the same, their corresponding ciphertext blocks will also be identical. This lack of variation reveals patterns in the data, making it vulnerable to known-plaintext attacks

N.B. Parallel encryption and decryption are possible because ECB blocks are independent from each other.

  - **CBC (Cipher Block Chaining)**: In CBC mode, each plaintext block is XORed with the previous ciphertext block before being encrypted. The formula for encrypting the n-th block is:

$$C_n = enc(K, P_n \oplus C_{n-1})$$

    The first plaintext block is XORed with the **IV** (**Initialization vector**) instead of a previous ciphertext block because no previous ciphertext exists for the first block. The **IV** should be **random and unique** for every **encryption** session.
    The **decryption** process involves reversing the encryption steps:



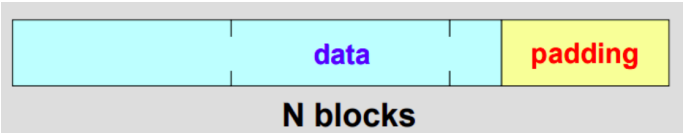
Problems with CBC mode

A single error in a ciphertext block affects only two plaintext blocks:

1. The block corresponding to the erroneous ciphertext block.
2. The next block.

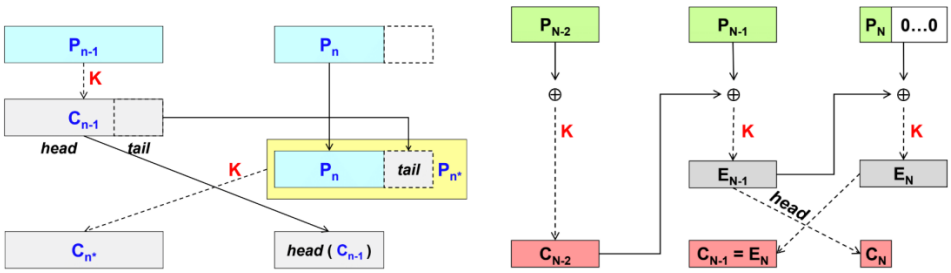
N.B. CBC allows for **parallel decryption** but not **parallel encryption**.

- **Size of data Encrypt < Algorithm's Block Size**
  - **Padding**: Padding is used to **fill the only last block** until it reaches the required block size (**C** will be longer than **P**).



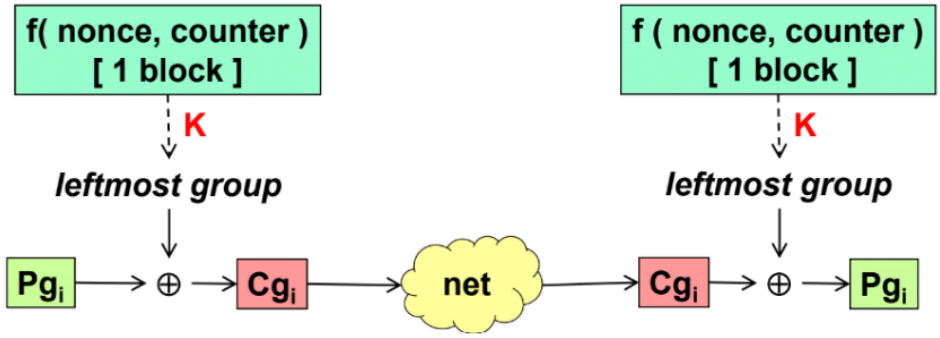
Depending on the chosen **Padding Technique**, we can allow or avoid different kind of attacks. Some techniques also offer minimal integrity control (e.g. if the key is wrong or data is manipulated, the padding bytes are incoherent).When the data size is smaller than the block size, special techniques such as **CFB** (**Cipher FeedBack**) or **OFB** (**Output FeedBack**) are the preferred.
Even if the plaintext is an exact multiple of the block size, padding must still be added to avoid errors in the interpretation of the last block.

  - **CTS (Ciphertext Stealing)**: CTS allows the use of block algorithms **without** the need of Padding. CTS works in the following way:
    1. The last (partial block) P_N is filled with a copy of the last bytes taken from the penultimate block P_{N-1}.
    2. The copied bytes are **removed** form the (now partial) penultimate block.
    3. After encryption, the last and penultimate block' positions are **exchanged**.



N.B. CTS is particularly useful when we cannot increase the size of the data after encryption, such as in storage encryption. The computation time of encryption slightly increases due to the need for additional steps, such as modifying and swapping blocks.

  - **CTR (Counter Mode)**: CTR modifies block algorithm in order to encipher N bits at a time (often N = 1). CTR can be used **only** to encipher **P** smaller than one block size. It does not require padding and allows random direct access to any ciphertext block (no dependency between blocks). Requires a **nonce** and a **counter**, combined by a suitable function (concatenation, XOR, etc.) to generate the input block. CTR works in the following way:
    1. **f(nonce, counter)** outputs exactly one block size of data.
    2. This block gets encrypted with the **K**.
    3. The **leftmost group** (meaning the first N bits) are extrapolated from the block and are **XOR**ed with the corresponding P_i's group, which in turn generates the corresponding C_i's group.



Cancellation attacks

CTR mode is vulnerable to cancellation attacks because it relies on synchronized counters between sender and receiver.

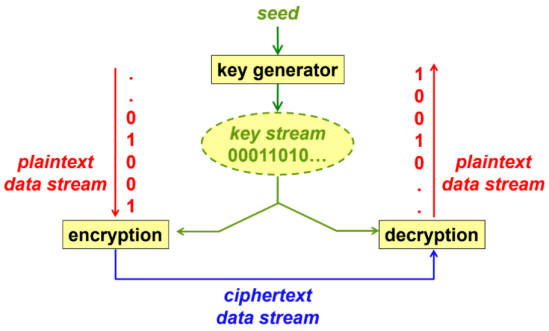
CTR as a stream ciphertext

The byte-oriented CTR mode of a block algorithm may be considered a stream algorithm.

N.B. Moreover, if a group is **modified**, the error does not propagate to the successive group since group are independent → it is possible to decrypt a random group **without** having decrypt the others.
Parallel encryption and decryption are possible.

2.1.5 Stream Encryption Algorithms

Sream Encryption Algorithms are another kind of Symmetric Encryption algorithms that **not** require the division of the **P** in blocks, as they tipically works with a **single** bit or byte at a time. This category contains the **only perfect algorithm** which is called **One-Time-Pad**: the algorithm requires a **K** as long as **P**, for this reason it's **not** of pratical use (unless **P** is very small).
Real stream algorithms use **pseudo-random key generators**, which **must** be synchronized between sender and receiver. For example, **RC4** and **SEAL** are some **old** instances of stream algorithms, while CTR can be considered a stram algorithm when **N=1**.



The **K** is **randomly generated** and is used to XOR the **P** in order to obtain the **C**. This makes stream algorithms really fast and unpredictable, which is good for security. Decryption can be done **only** by using the **same K** → if an attacker can delete part of the **C**, decryption will be wrong since the **C**'s length is not the same as the **K**.

Salsa20 and ChaCh20

Salsa20 and **ChaCha20** are stream algorithms that use **128-256 bits K**. ChaCha20 is an improvement of Salsa20: it is **faster** on some architecture and has more **Diffusion Bits** (→ changing one input bits propagates the change t more output bits).

- Base Operation: 32 bits ARX(Add-Rotate-XOR).
- Base function: $f(\text{Key}_{256}, \text{Nonce}_{64}, \text{Counter}_{64}) = 512\text{-bit keystream block}$

This allows **O(1)** decryption time for any random block. Salsa20 performs 20 random mixes of the **P**, while the **Salsa20/12** and **Salsa20/8** variants perform 12 and 8 respectively, making them faster but less secure. Meanwhile, ChaCha20 has been standardized and heavily adopted, especially in **IETF** Protocols. However, IETF uses a slightly modified version of ChaCha20 that limits the maximum size of encryptable data to **256GB**. To overcome this limit we would need to use the original ChaCha20.

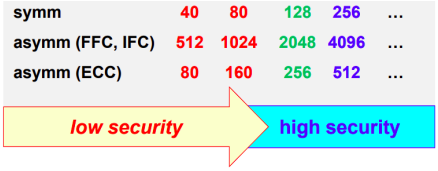
2.2 Length of Secret Keys

The **K** should be adequate, meaning that if the encryption algorithm is well designed and the **K** is kept secret, then the **only** possible attack is a **Brute Force Attack** wich requires approximately:

$$T = 2^{Nbit}$$

(Where N^{Nbit} refers to the key lenght in bits).

- **Symmetric Encryption Algorithms:** **K**s under 128 bits are **insecure**.
- **Asymmetric Encryption Algorithms:** **K**s under 2048 bits are **insecure**.

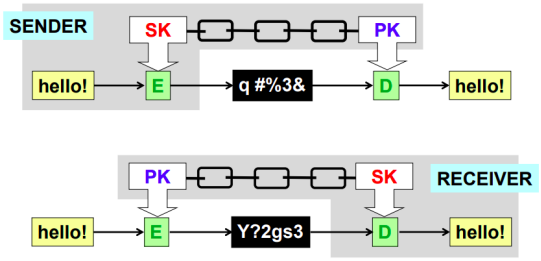


The area of low security increases with time, as computer become more powerful. Longer K means keeping data secret for longer, at the cost of an increase in time needed to encrypt data.

2.3 Asymmetric Cryptography

Asymmetric Cryptography uses two different **K**:

- **Private Key** (**SK**)
- **Public Key** (**PK**)



These keys are generated in such a way that **SK** $\neq$ **PK**. The keys have **inverse functionality** $\rightarrow$ data encrypted with one **K** can be decrypted **only** by the other **K**. This kind of cryptography has a **high computational load**, for that reason it is **only** used to transmit **K** used in Symmetric Cryptography and to create **digital signatures**.

2.3.1 Digital Signature

To perform a **digital signature** the author encrypts the message with its **SK**. The message can be then be decrypted by anyone who has access to the PK of the author. The message is **not** secret, but it is **signed by the author** ($\rightarrow$ digital signature provides **data authN**) and **integrity**.

2.3.2 Confidentiality without Shared Secrets

It is possible to generate a **secret message** for a particular **user X** just by encrypting it with its PK. By doing so, we ensure that the message can **only** be decrypted by **X**, since **X** is the **only one** who knows the **SK**.

2.3.3 Secret Key Exchange

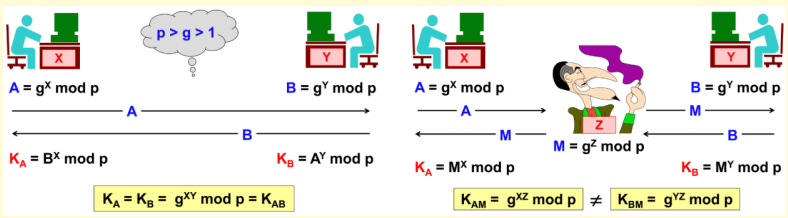
Using asymmetric algorithms for secret key exchange enables confidentiality without the need for shared secrets, and is often used to securely transmit a secret key for symmetric encryption.

Diffie-Helmann Algorithm

The Diffie-Hellman (DH) algorithm is one of the first public-key algorithms invented and is primarily used for key agreement. The process is as follows:

- Alice (A) and Bob (B) agree on two public integers:
  - p: a large prime number
  - g: a primitive root modulo p (typically $g = 2, 3$, or 5)
- The length of the DH key is determined by the number of bits in p.
- Alice randomly chooses an integer $x > 0$ and computes: $X = g^x \mod p$
- Bob randomly chooses an integer $y > 0$ and computes: $Y = g^y \mod p$
- A and B exchange X and Y.
- Alice computes the shared secret key: $K_A = Y^x \mod p$
- Bob computes the shared secret key: $K_B = X^y \mod p$
- It follows that: $K_A = K_B = g^{xy} \mod (p)$

Also this approach is not completely secure, since it can be attacked using the **Diffie-Hellman MITM attack**. The way in which the algorithm works and the attack is performed are explained by the following figures:



2.3.4 Elliptic Curve Cryptography

An **Elliptic Curve Cryptosystem (ECC)** executes operations on the surface of a 2D elliptic curve, instead of using modular arithmetic. These problems are more complex than the modular arithmetic ones, but at the same time they allow to use shorter keys (up to $\frac{1}{10}$). Most of the algorithm we have seen (except RSA) have been revisited to adapt them to the elliptic curve:

- **ECDSA** for Digital Signature;
- **ECDH** fro Key Agreement;
- **ECMQV** for AuthN Key Agreement;
- **ECIES** for Key Distribution;

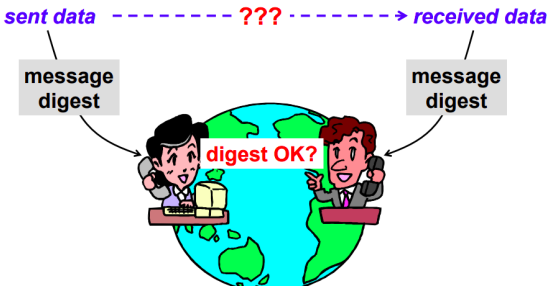
2.4 Digest for message integrity

All the previous methods can guarantee (to a certain extent and for a certain time) that a message from a sender to the receiver **cannot** be understood by a third-party. However, it is still possible for an attacker to intercept the traffic from the 2 parties and **change** the encrypted message (Spoofing attack), even in an **unpredictable** way. Upon receiving the corrupted message, the receiver discard it and (probably) asks for retransmission, causing an(indirect) DoS attack.

Beware

Integrity is **not** about **preventing** modifications of data, it actually means that when data is received, the receiver can **verify** if the data has been modified or **not**.

To **guarantee integrity** the most common technique is to compute a **Digest**: fixed length summary of the whole message. Digests must be protected, as, conceptually, they are similar to a checksum. But since checksum are easy to attack, Digests must to be calculated via a **Cryptographic Hash Function**.



2.4.1 Hash Functions

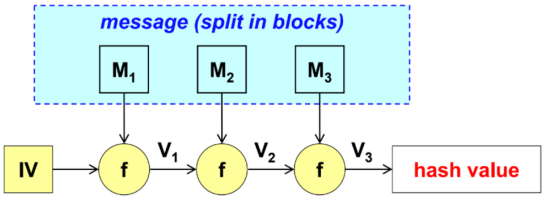
Cryptographic Hash Functions make Digests:

- **Fast to compute**
- **Impossible or very difficult to invert**

Moreover, good Cryptographic Hash Functions should make it difficult to create **Collisions**: collision happens when two different messages produce the same Digest. Generally, Cryptographic Hash Functions work like this:

1. Split the message M into N blocks: $M_1, M_2, \ldots, M_N$
2. Iteratively applying a base function f to each block.

$$V_k = f(V_{k-1}, M_k) \quad \text{with} \quad V_0 = IV \quad \text{and} \quad h = V_N$$



Common Cryptographic Hash Functions

name	block	digest	definition	notes
MD2	8 bit	128 bit	RFC-1319	obsolete
MD4	512 bit	128 bit	RFC-1320	obsolete
MD5	512 bit	128 bit	RFC-1321	obsolete
RIPEMD-160	512 bit	160 bit	ISO/IEC 10118-3	old + rare
SHA-1	512 bit	160 bit	FIPS 180-1	sufficient
SHA-224	512 bit	224 bit		
SHA-256	512 bit	256 bit	FIPS 180-2	good
SHA-384	512 bit	384 bit	FIPS 180-3	
SHA-512	512 bit	512 bit		
	SHA-2			
SHA-3	1152-576	224-512	FIPS 202 FIPS 180-4	excellent

2.4.2 SHA-2 family

Each algorithm in the **SHA-2 Family** has the same **block size (512 bits)**, but they have **increasingly higher Digest size**. The Digest size is important because a bigger Digest provoke less collisions (or aliasing). If a Cryptographic Hash Function is **well designed** and generates a N_{bits} Digest, then the **probability of aliasing** is:

$$P_A \propto \frac{1}{2^N}$$

Birthday Paradox

The **Birthday Paradox** is a statistical consideration that warns us of the risks of making Digests too big: a N_{bits} Digest is **insecure** when more than $2^{\frac{N}{2}}$ digests are generated. This is because the probability to have a collision is **PA $\sim$ 50%**.

N.B. A cryptosystem is **balanced** when the Encryption algorithm and Digest function have the same resistance. For example, SHA-256 and SHA-512 have been designed for AES-128 and AES-256 respectively.

2.4.3 SHA-3 family

SHA-3 is based on the **Keccak** algorithm, it uses an elegant design that is easy to analyse and performs very well on many different computing devices.

Function	Output	Block	Capacity	Definition	Strength
SHA3-224(M)	224 bits	1152 bits	448 bits	Keccak[448] (M || 01, 224)	112 bit
SHA3-256(M)	256 bits	1088 bits	512 bits	Keccak[512] (M || 01, 256)	128 bit
SHA3-384(M)	384 bits	832 bits	768 bits	Keccak[768] (M || 01, 384)	192 bit
SHA3-512(M)	512 bits	576 bits	1024 bits	Keccak[1024] (M || 01, 512)	256 bit
SHAKE128(M,d)	d bits	1344 bits	256 bits	Keccak[256] (M || 1111, d)	min($\frac{d}{2}$, 128) bit
SHAKE256(M,d)	d bits	1088 bits	512 bits	Keccak[512] (M || 1111, d)	min($\frac{d}{2}$, 256) bit

2.5 KDF (Key Derivation Function)

KDFs are **Cryptographic Hash Functions** used to create a random **K**:

$$K = \text{KDF}(P, S, I)$$

Where:

- P **is the password** or passphrase.
- S **is the salt**, which makes the key K harder to guess given P.
- I **is the number of iterations**, used to slow down the computation and increase security.

The main KDF is **PBKDF2**:

$$\text{DK} = \text{PBKDF2}(\text{PRF}, \text{P}, \text{S}, I, \text{dkLen})$$

Where **P**, **S** and **I** are the same as above, while:

- **PRF**: is a **Pseudo-Random Function** that outputs a value of lenght **hLen**.
- **dkLen**: is the **Desired Lenght of the DK**.

The derived key, DK, is generated by concatenating subkeys $T_1, T_2, \ldots, T_{dkLen/hLen}$, where each $|T_i| = hLen$.

2.6 MIC, MAC & MID

In the telecommunication and networking environment, the **MIC (Message Integrity Code)** is the code that is added to the message to guarantee integrity. Meanwhile, in the application field, the same code is named MAC (Message Authentication Field), since, if the message has not been modified → data authN.
Since the MIC/MAC code is an addition to the original message, a **unique MID** (**Message IDentifier**) is also concatenated with the message, in order to avoid Replay attacks.

2.7 Digest Protection

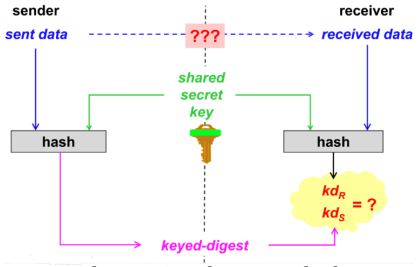
In the simple case in which a Digest is used to verify that a file on disk has not been changed, a good approach could be to store the digest in a **different and secure place**. But when the data is being **transmitted**, Digests **cannot** be exchanged OOB → Digests **must be intrinsically secure**. To achieve this, there are two main techniques:

- **AuthN by means of Keyed Digest**
- **Authenticated Encryption**

2.7.1 AuthN by means of Keyed Digest

The Digest **d** is sent along with the message **M**, but it is calculated using **both** the **M** and the **K**, which is shared **only** between sender and receiver (symmetric key).
The following operations are perfomed:

- **Sender**: $d = \text{digest}(K, M)$.
- **Transmission**: $M||d$.
- **Receiver**: $d' = \text{digest}(K, M)$.
- **Verification**: If $d == d'$, the message is authenticated, otherwise an alarm is raised.



This technique also guarantees **data authN**, since **only** who knows the **K** can compare the receive digest with the one calculated from the received data ($\rightarrow$ the Keyed Digest is the fastest way to guarantee **integrity** and **data authN**).

However, it does **not** provide **non-repudiation**, since it is **not** possible to determine which party created the data, because both have the means (data and **K**) to create the Keyed Digest.

- **HMAC**: is the **best way** to compute a **secure MAC** starting from a hash function **H**. **H** must have a block of size **B** bytes, an output of **L** bytes, where $B > L$.
  Moreover, in order to work with HMAC we need:
  - **ipad**: $0x36$ repeated B times.
  - **opad**: $0x5C$ repeated B times.
  - A **K** bigger or equal to **L** ($|K| > L$). **Never** use a **K** smaller than **L**! (e.g using SHA-1 with HMAC means that **K** **must** be of **at least 160 bits**):
    * if $|K| > B \rightarrow K' = H(K)$.
    * else $\rightarrow K' = K$.
    * if $|K'| > B \rightarrow K'$ is 0-padded up to **B** bytes;

$$\text{hmac-H} = H\left(K' \oplus \text{opad}||H(K' \oplus \text{ipad}||\text{data})\right)$$

- **CBC-MAC**: is another way to compute a digest **without** using a hash function, but just by using a Symmetric Encryption Block Algorithm, which is applied in CBC mode and using a **null IV**. Of the whole algorithm's output, **only** the **last (encrypted) block** is used as the **MAC**.

  Process

  - The message M is split into N blocks $M_1, M_2, \ldots, M_N$.
  - The MAC is computed iteratively: $V_0 = 0$ and for each block k, compute:

$$V_k = \text{enc}(K, M_k \oplus V_{k-1})$$

  - The final MAC is the last encrypted block V_N.

  CBC-MAC does **not** guarantee **confidentiality** because the IV is fixed, but it guarantees **integrity** thanks to CBC: if the message is modified in any point, the last (encrypted) block **will** be different.
  N.B. CBC-MAC is secure **only** for fixed-lenght message $\rightarrow$ attacks that change the length of the message can be successful.

2.8 Integrity & Confidentiality Combinations

Integrity and **confidentiality** can both be achieved with the combined use of two algorithms and with **two distinct Ks**:

- Symmetric Encryption with K_1 is used to gain **confidentiality**.
- AuthN by means of Keyed Digest with K_2 is used for **integrity**.

There are various ways to combine these two algorithms:

- **A&E (Authenticate & Encrypt)**: Encrypt the message and then authenticate it using a MAC. Insecure unless performed as a single step (used by SSH).

$$enc(K_1.P) \;||\; mac(K_2, P)$$

- **AtE (Authenticate then Encrypt)**: compute the MAC on the plaintext and then encrypt both the message and MAC. Secure if used with CBC or stream encryption (used by SSL and TLS). The **Ate** is still vulnerable to DoS attacks.

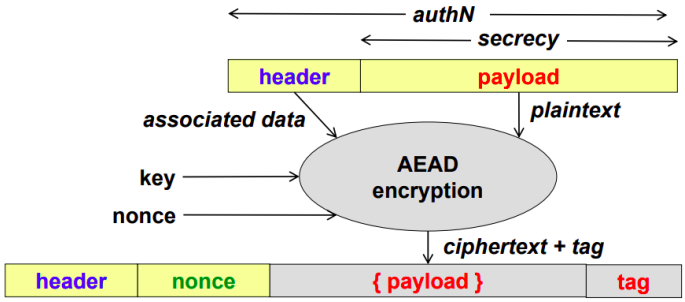
$$enc(K_1, P \;||\; mac(K_2, P))$$

- **EtA (Encrypt-then-Authenticate)**: Encrypt the message, then compute the MAC on the ciphertext. This is the most secure method (used by IPsec). With **AtE** solution is possible to avoid DoS attacks since the MAC can be calculated using **C** (fast operation) **without** the need to decrypt it (slow operation).

$$C = enc(K_1, P) \rightarrow C \;||\; mac(K_2, C)$$

2.9 AE (Authenticated Encryption)

The **AE** concept uses a **single operation** that, once performed, guarantees: **payload confidentiality**, **data authN** and **integrity**. This is achieved by using just a **single K** and a **single algorithm** → **better speed** and less likelihood of errors due to the combination of different functions.



Beware

AE does **not** provide **non-repudiation** because both peers share the same key.

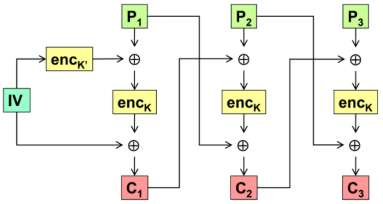
AE can work in 6 standard modes:

- **OCB 2.0** (Offset Codebook Mode): the fastest one, on-line single-pass AEAD.
- **AESKW** (AES Key Wrap).
- **CCM** (CTR mode with CBC-MAC): the slowest one, off-line double-pass (half as fast as single-pass).
- **EAX** (EtA then (X)translate): very good for constrained systems, on-line double-pass AEAD.
- **Encrypt then MAC**
- **GCM** (Galois/Counter Mode): the most popular one, on-line single-pass AEAD, it is also parallelizable.

A **lightweight environment** solution to all of this modes is **ASCON**.

IGE (Infinite Grable Extension)

IGE is an **AE algorithm** that works like CBC, but with one addition that makes any **modification** of any block **propagate to every block after that one**.
This is important, because it means that by having the last block with fixed content → in the decryption phase **integrity is verified only if** the last block has the excepted content.



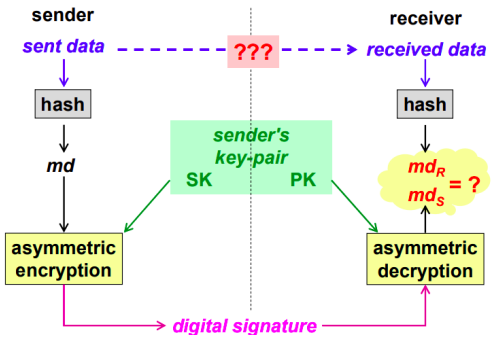
2.10 Digital Signature

To achieve **non-repudiation** we can use Digests and Asymmetric Cryptography:

- The hash of the message **M** gets encrypted with the private key **SK** of the sender **S**.
- The sender transmits (**M** || **H**) to the receiver.
- The receiver decrypts **H** with the public key **PK** of the sender:

$$X = dec(S.PK, H)$$

- if (**X** == **hash(M)**) the message is authenticated, else something has been changed during the transmission.



Changing anything in the Encrypted Data or in the Digital Signature provoke a **failure** in the verification phase (→ because it is **not** possible to say whether the modification affected the Data or the Signature).
The signature is bound to the data **M** because is calculated using its **hash(M)** → the Signature **cannot** be attached to other data.

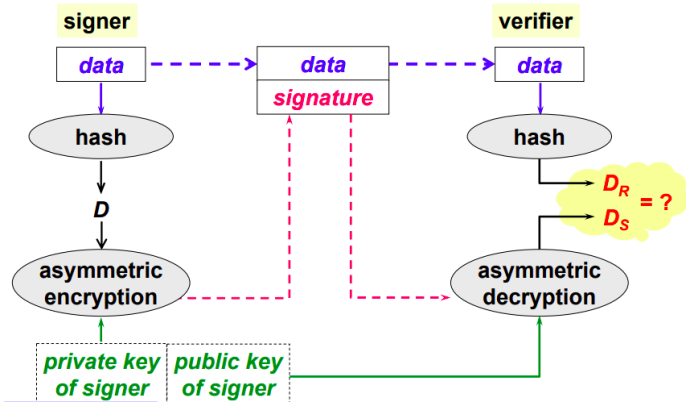


Figure 2.1: Digital Signature stored with the data

Digital vs Handwritten Signatures

- **Digital Signature** = authentication + integrity. A private key generates infinite digital signatures (one per document).
- **Handwritten Signature** = authentication only, it does not ensure that the document has not been altered after signing.

2.10.1 PKC (Public Key Certificate)

The last piece needed to provide **full non-repudiation** is a mechanism that can guarantee the association between a PK and an identity: this mechanism is called **PKC**.
A PKC is a **data structure** used to **securely bind** a **PK** to some attributes, typically an identity or an IP address. The security of this binding is provided by the Digital Signature of a **Certification Authority (CA)**, which grants, at the same time:

- A proof of **authN**: it is a valid certificate because it has been created by a CA.
- A proof of **integrity**

N.B. Like any real-life document, the PKC has a limited lifetime and it can be **revoked**.

PKCs must be encoded in one of the following formats:

- **X.509** → is a PKC format that is composed by:
  - **Version**
  - **Serial Number**
  - **Signature Algorithm**
  - **Issuer**: name of the CA that created the certificate, it uses the following fields:
    - **C** stands for country.
    - **O** stands for organization.
    - **OU** stands for organizational unit → the department who created the certificate.
  - **Validity**: period of time for which the certificate is considered valid.
  - **Subject**: entity (**identity** or **IP address**) that is controlling the **SK** corresponding to the **PK** being certified.
  - **Subject Public Key Info**: it contains the PK bit by bit, the algorithm and the key's length.
  - **CA Digital Signature**

2
1231
RSA with MD5, 1024
C=IT, O=Polito, OU=CA
1/1/97 - 31/12/97
C=IT, O=Polito,
CN=Antonio Lioy
Email=lioy@polito.it
RSA, 1024, xx...x
yy...y

- **non X.509**: PGP and SPKI.
- **PKCS#6**: RSA, its considered obsolete.

2.10.2 PKI (Public Key Infrastructure)

The **PKI** is an infrastructure which performs two kinds of tasks:

- **Technical**: creates certificates.
- **Administrative**: checks the identity of the requester of the certificate, **before** creating it.

The **revocation** can be requested by:

- The **Owner** → meaning control of the SK has been lost.
- The **Creator** of the certificate → in case the requester presented a fake identity.

When a certificate gets revoked, the CA adds the certificate to a **list** that contains every **nonvalid certificate**. The revocation mechanism can be done in 2 ways:

- **CRL (Certificate Revocation List)**: is a list of revoked certificates that is **generated and signed** by the CA (→ the CA guarantees the list's **authenticity**) and **integrity**.

  Problem

  The main problem of CRLs is that the entire CRL must be download, even if we need to check one single certificate.

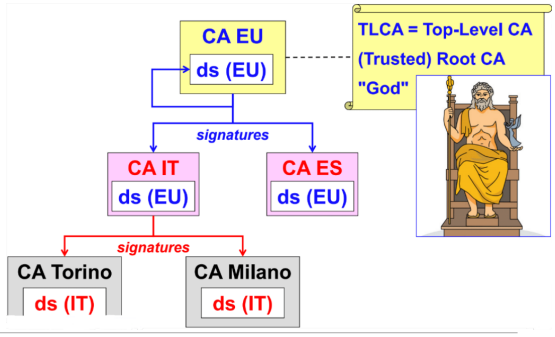
- **OCSP (On-line Certificate Status Protocol)**: is the mechanism used to perform r**real-time** checks of certificates (e.g. during a transaction). OSCP is a **client-server protocol** in which is possible to request to a specific CA if the certificate is valid or **not** at the **current time**.
  The response is **signed by the CA** (→ it is not possible for an attacker to provide fake answers).

2.10.3 Verification of a Signature/Certificate

To **verify the signature** we have to:

- Request its PKC, which is signed by a CA.
- Request CA's PKC, which is signed by another CA.
- Repeat recursively from 2 until we end up with a **self-signed certificate** form a **Root CA**.

N.B. Each device contains a list of **root CA** who's signatures are **automatically**considered valid. If a device is left **unattended**, an attacker could modify this list using admin privileges and successfully fake the whole certification tree.



2.11 Performance

Cryptographic performance does **not** depend on RAM, but on CPU and cache size: some CPU have even special hardware just for cryptographic algorithms. Performance can become a problem on servers or when implementing security on network nodes, in this case, it is possible to use:

- **Cryptographic Accelerators (HSM or Hardware Security Mode)**: it has some primitive instructions to perform some security algorithms.
- **Special Purpose Accelerators**: accelerators that not only implement in hardware cryptographic operations but also implement packet processing for some security protocols.
- **Generic Accelerators**: they speed up computation for whatever protocol implements a particular encryption algorithm;

2.12 CNSA

The **Commercial National Security Algorithm** is a suite released in 2018 by NSA. It includes the following algorithms:

- **Symmetric Encryption**: AES-256, now any key shorter than 256 bits is not considered secure enough for symmetric encryption.
  - CTR mode is suggested to protect real time data for low bandwidth applications.
  - GCM is suggested for high bandwidth applications.
- **Hash**: SHA-384.
- **Key Agreement**: ECDH and ECMQV.
- **Digital Signature**: ECDSA with EC on the curve P-384;

For old systems (a.k.a legacy systems) it is suggested to use for:

- **Key Agreement**: DH-3072.
- **Key Exchange & Digital Signature**: RSA-3072.

Chapter 3

Authentication techniques protocols, and architectures

Authentication refers to the process of verifying the identity of an entity (whether it's a human, software component, or hardware element) before granting access to resources in a system. Authentication can be applied to various type of "actors", such as:

- **Human being**
- **Software component**
- **Hardware element**

Authentication vs Authorization

- **Authentication (authC/authN)**: established the identity of an entity.
- **Authorization (authZ)**: determines where a authenticated entity has permission to access.

3.1 Authentication factors

Authentication can be based on 3 primary factors:

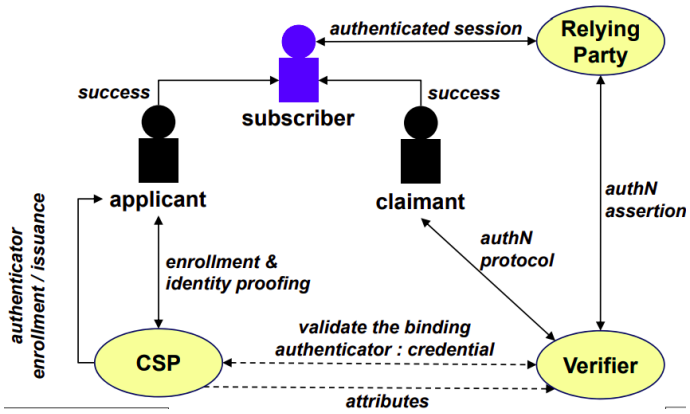
- **Knowledge**: Information that only the user knows and can provides as proof of their identity.
- **Ownership**: Physical object or device that only the user has access to.
- **Inherence**: This factor relies on unique biological traits of the user (e.g fingerprint).

N.B Authentication can be applied not just to human user, but also to processes and devices.

3.1.1 Risks

- **Knowledge:**
  - Storage → if passwords are stored improperly, they are vulnerable to thefts.
  - Demonstration → user might inadvertently reveal their password through social engineering.
  - Transmission → if passwords are sent over insecure channel, they can be intercepted by attackers.
- **Ownership:**
  - Authentication thieft
  - Cloning
  - Unathorized usage
- **Inherence:**
  - Counterfeiting → biometric data can be spoofed or replicated by attackers using sophisticated techniques.
  - Privacy → the use of biometric data raises the risk of biometric information being exposed.
  - Irreversibility → biometric traits cannot be replaced if compromised.

3.2 Digital Authentication model (NIST SP800.63B)



Entities:

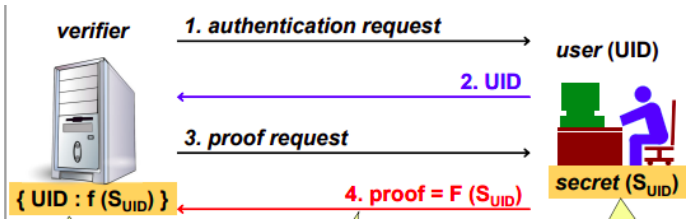
- **Subscriber**: applicant who has successfully completed identity proofing.
- **Applicant**: an individual applying to established a digital identity.
- **Claimant**: the user trying to prove their identity to access a system or service.
- **Relying Party**: entity (e.g service provider, website) that requests/receives an authN assertion from the verifier to asses user identity (and attributes).
- **Verifier**: validates the user's credential during each authentication event.
- **CSP**: an entity that issues, manages, and maintains **credentials** used by individuals to authenticate themselves.

Applicant vs Claimer

An applicant needs to enroll in the system for the first time to establish their identity. Instead, a claimant asserts their identity to gain access to the system after enrollment.

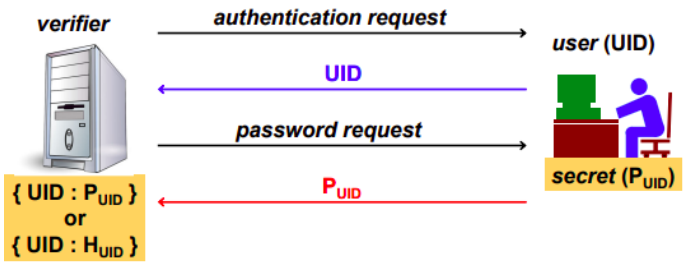
3.3 Generic authentication protocol

1. The user initiates an authentication request by sending their UID.
2. The user generates a proof based on their secret, using a secure function $F(S_{UID})$, and send this proof to the verifier.
3. The verifier checks if the received proof matches the stored representation of the secret.
4. If it matches, the user is successfully authenticated.



3.4 Password base authentication

1. The user sends their UID and P_{UID} (= Password) to the verifier.
2. The server verifies the proof:
   - If password are stored in cleartext, it directly compares the proof with the stored password.
   - If password are stored in hashes, it hashes the proof and compares it to the store hash H_{UID}.



Problems of reusable Passowrds

- **PWD Sniffing** (attackers intercept password during transmission)
- **PWD Database attack** (if DB contains plaintext or obfuscated PWD)
- **PWD Guessing** (very dangerous if it can be done offline, e.g against a list of PWD hashes)
- **PWD Enumeration** (PWD brute force attack)
  - If PWD is limited in length and/or character type.
  - If authN protocol does not block repeated failures.
- **PWD Duplications** (using the same PWD for one service against another one, due to user PWD reuse)
- **Cryptographic Aging** (as computing power grows, older cryptographic methods become vulnerable to new attacks)
- **PWD capture via server spoofing and phishing** (attackers deceive user into giving away their PWD by pretending to be legitimate service)

Password best practies

Suggestion to reduce password risks:

- Use alphabetical characters (upper case + lower case), digits and special characters
- Make passwords long (at least 8 character)
- Never use dictionary words
- Change password regularly, but not too frequently
- Do not reuse passwords across different services

Password storage

- **Server Side:**
  - Passwords should never be stored in cleartext.
  - Encrypted passwords aren't ideal since the server would need to know the encryption key.
  - Better to store a password digest (hashed password), though vulnerable to dictionary attacks.
  - Rainbow tables can speed up these attacks, so it's important to add a "salt" (random variation) to each password.
- **Client-side:**
  - Ideally, passwords are memorized by the user, but having many passwords makes this difficult.
  - People may resort to writing them down or using simple passwords, which is risky.
  - Using a password manager or encrypted file is a safer alternative.

3.5 The "dictionary" attack

- **Hypothesis:** The attacker knows the hash algorithm and the hashed password values.
- **Pre-computation:** For each word in a dictionary, compute and store its hash $store(DB, Word, hash(World))$
- **Attack process:**
  - Let HP (=hash password) to be the hash of an unknown password.
  - Lookup HP in the precomputed dictionary (DB) to find a matching password.
  - If found, output the password; if not, indicate it's "not in dictionary".

3.6 Rainbow Table attack

Rainbow Table is a **space-time trade-off technique** that reduces storage needs for exhaustive hash tables, making certain brute-force attacks feasible within limited space. It uses a reduction function $r : h \to p$ (which is NOT h^{-1}) to generate chains of hashes.
Example:

- For a 12-digit password, an exhaustive hash table would require $10^{12} rows(P_i : HP_i)$
- rainbow = 10^9 rows, each representing 1000 possible passwords.

Attack

```
for (k=HP, n=0; n<1000; n++)
 ■ p = r(k)
 ■ if lookup( DB, x, p ) then exit ( "chain found, rooted at x" )
 ■ k = h(p)
exit ( "HP is not in any chain of mine" )
```

3.7 Salting Passwords: A Defense Against Dictionary and Rainbow Table Attacks

Salting passwords is a security technique used to protect stored passwords from dictionary attacks and rainbow table attacks. A salt is a unique, random string added to each password before hashing. This ensures that even if two users have the same password, their hashes will be different due to the unique salt.

Steps for each user (UID):

- Generate or ask for the user's password.
- Create a unique, random slat for each user.
- Compute the salted hash: $SHP = hash(password \mid\mid salt)$
- Store the triplet $\{UID,\ SHP,\ salt\}$

Password Verification with Salt

- **Claimant:** Provides their user ID (UID) and password (PWD).
- **Verifier:**
  - Uses the UID to find the stored salted hash (SHP) and salt.
  - Computes $SHP' = hash(PWD \mid\mid salt)$.

The Linkedin attack

In 2012, LinkedIn was breached, exposing 6.5 million unsalted SHA-1 password hashes. The lack of salting allowed attackers to crack at least 236,578 passwords through crowdsources efforts before restrictions halted the exposure.

3.8 Strong authentication definitions

The concept of strong authentication (authN) is crucial in ensuring secure identity verification, but it has never been formally defined with a universal definition. Various definitions exist depending on the context, such as the European Central Bank (ECB) and PCI-DSS.

ECB definition

The ECB defines strong authentication as a process that involves at least two independent elements from **knowledge** (e.g. password), **ownership** (e.g. smartcard), and **inherence** (e.g. biometrics). The key requirement is that these elements must be mutually independent, so compromising one should not affect the others. Furthermore, at least one element should be **non-reusable** or **non-replicable** (except for inherence), with the entire process safeguarding the confidentiality of the authentication data.

PCI-DSS Definition

PCI-DSS mandates **multi-factor authentication (MFA)** for access to cardholder data, particularly for administrators and remote access from untrusted networks. Since version 3.2, MFA has become compulsory for remote access, and the use of the same factor twice (e.g., two passwords) does not qualify as MFA.

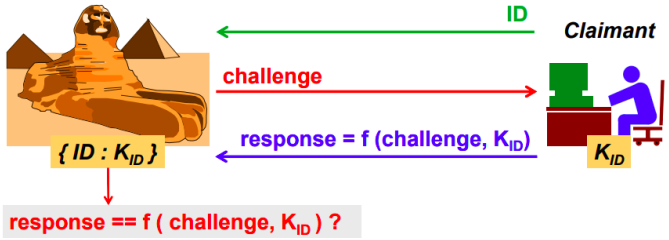
3.9 Challenge-Response Authentication (CRA)

Challenge-response authentication (CRA) is a widely used technique where a challenge is issued, and the claimant responds by solving it with a secret (shared or private). The challenge must be **non-repeatable** (usually a random nonce) to avoid replay attacks. The function used to compute the response must be **non-invertible**, otherwise, a listener can record the traffic and easily find the shared secret:

$$if(\exists f^{-1})\ then\ K_c\ =\ f^{-1}(response, challenge)$$

3.9.1 Symmetric CRA

In symmetric CRA, both the client and the server share a secret key that is used to verify the authenticity of a user or system This method is fast, often utilizing hash functions (e.g., SHA1, SHA2, SHA3).



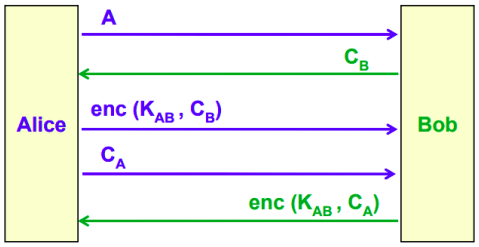
3.9.2 Mutual symmetric CRA

Mutual symmetric CRA requires both parties to authenticate each other. However, it's a old protocols so it has many vulnerabilities.

Version 1: Basic Exchange

In this case, the initiator explicitly provides its claimed identity (This version is considered outdated and insecure).
Process:

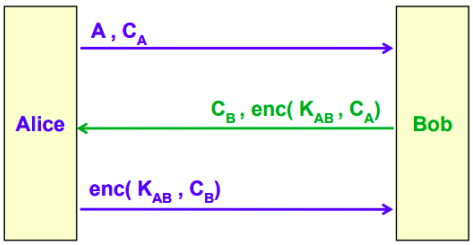
- Alice sends an encrypted challenge (C_B) to Bob using the shared key K_{AB}.
- Bob responds with an encrypted challenge (C_A) for Alice, also using K_{AB}.



Version 2: Improved Performance

Optimized by reducing the number of messages, which improves performance without compromising security.
Process:

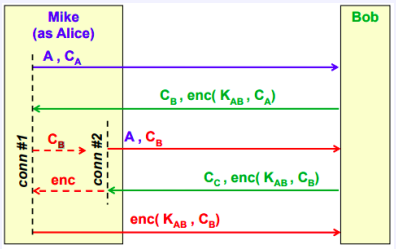
- Alice includes her identity (C_A) and sends an encrypted challenge (C_B) in the same message.
- Bob responds with his encrypted challenge C_A to complete the exchange.



Attack on Mutual Symmetric CRA

A potential attacker, "Mike" (posing as Alice), exploits the protocol by mimicking responses:

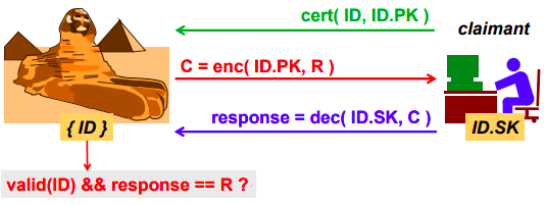
- The attacker intercepts Alice's identity (C_A) and Bob's challenge (C_B).
- The attacker uses the shared key K_{AB} to manipulate responses and mimic both parties.



3.9.3 Asymmetric CRA

Process:

- A **random nonce (R)** is generated by the Verifier.
- The verifier encrypts R using the user's public key ($ID.PK$) and sends it to the Claimant: $C = enc(ID.PK, R)$
- The Claimant decrypts C using their private key ($ID.SK$) and sends R back in cleartext: $response = dec(ID.SK, C)$
- The Verifier validates: $valid(ID)$ && $(response == R)$.



Applications

- Widely implemented in secure communication protocols like IPsec, SSH, and TLS.
- Fundamental in modern authentication frameworks such as FIDO.

Asymmetric CRA analysis

Security:

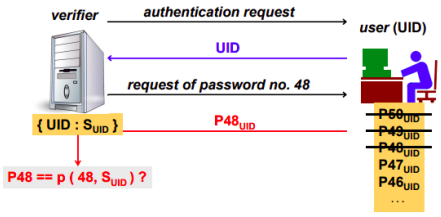
- It's the strongest mechanism.
- Does not require the Verifier to store any shared secret, reducing potential attack vectors.

Problems:

- It **slower** compared to symmetric methods.
- If designed inaccurately may lead to an involuntary signature by the Claimant.
- Trust issues managing root certificates, name constraint, and certificate revocation.

3.10 One-Time Password(OTP)

One-Time Passwords are temporary and valid for a single use in an authentication session. They mitigate risks like password reuse and passive sniffing but can still be vulnerable to man-in-the-middle (MITM) attacks. These passwords are often designed with random characters to prevent guessing, but this can make password insertion difficult for users.



3.10.1 S/KEY System

S/KEY System was the first OTP implementation by Bell Labs (1981). It pre-computes a sequence of passwords derived from a user's secret. Each password is validated and replaced with its predecessor, ensuring security without storing the secret:

$$Secret = S_{ID}$$

$$P_1 = h(S_{ID}), P_2 = h(P_1), ..., P_N = h(P_{N-1})$$

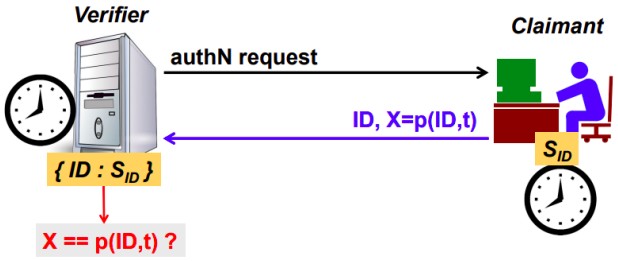
This approach minimizes verifier storage needs and offers robust protection, with users solely responsible for password retention.

One-time generation with S/KEY
In the S/KEY system, the user creates a secret passphrase (PP), which is combined with a server-provided seed to generate a 64-bit password. The passphrase is concatenated with the seed, and an MD4 hash is used to produce the password. The result is presented as six short words from a shared dictionary, making it easy to remember. This method allows secure password generation while using the same passphrase across multiple servers with different seeds. If the passphrase is compromised, security is at risk.

3.10.2 Time-based OTP (TOPT)

TOTP systems generate passwords based on the user's secret and the current time, requiring synchronization between the user and the verifier:

$$p(ID, t) = h(t, S_{ID})$$



Authentication Process:

- The claimant sends an authentication request with ID and the generated OTP x.
- The verifier checks if X matches the computed OTP for the corresponding ID and t.

Requirements:

- Local computation of OTPby the subscriber.
- Clock synchronization (or keeping track of time-shift for each subscribers).

Limitations:

- Only one authentication is allowed per time-slot, typically 30s or 60s.
- This time limit may not suit all services.

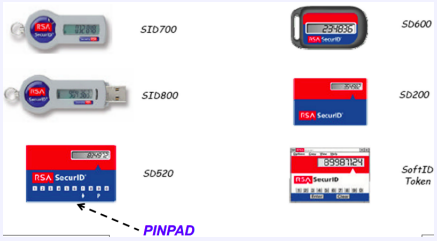
Vulnerabilities:

- Potential attacks on the subscriber and verifier:
  - Fake NTP servers or compromised mobile network femtocells.
  - Sensitive database storage at the verifier (e.g., the RSA SecurID attack).

Example: RSA SecurID

Authentication process:

- **The claimant sends to the verifier:**
  - Without PIN Pad: User ID, PIN, and Token Code (computed using seed and time).
  - With PIN Pad: User ID and a combined Token Code* (includes seed, time, and PIN).



3.10.3 Out-of-Band (OOB) OTP

Out-of-Band OTP requires a **secure channel** with server authentication to prevent MITM (Man-In-The-Middle) attacks. Traditionally, it uses text or SMS as the communication channel, but this method is increasingly vulnerable due to weaknesses in VoIP, mobile user identification, and the SS7 protocol. Nowadays, a push mechanism over a **TLS-secured channel** to a registered device is recommended for enhanced security.

3.10.4 Two-/Multi-Factor Authentication (2FA/MFA)

MFA enhances authentication (authN) by requiring multiple factors, such as a PIN, OTP, or biometrics, to verify identity. These factors can include something you know (like a PIN or password), something you have (like a token or phone), and something you are (like biometric data). MFA also protects the authenticator, for example, by using a PIN to safeguard it, but risks arise if the lock mechanism is weak or if there's no protection against multiple unlock attempts.

Importance of MFA: The iPhone Ransomware (2014)

In 2014, iCloud accounts with 1FA were hacked, allowing attackers to lock devices remotely. Victims were extorted for $100, but paying didn't help as the PayPal account was fake. This incident underscores the need for MFA to secure devices and prevent such attacks.

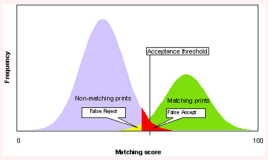
3.11 Authentication of human beings

To verify whether we're interacting with a human rather than a machine, there are two common approaches:

- **CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart):** A method where users must solve challenges like distorted characters in images to prove they are human.
- **Biometric Techniques:** These involve verifying human characteristics such as fingerprints, voice, retinal scans, iris scans, blood vein patterns in hands, heart rate, and hand geometry.

Problems of biometric systems

There are several issues with biometric systems:



- **FAR (False Acceptance Rate)** and **FRR (False Rejection Rate)**: These rates depend on the system's cost and can be adjusted, but biological factors like injuries or emotional changes can affect accuracy.
- **Psychological Acceptance:** Many people fear the "Big Brother" scenario—personal data collection and potential privacy invasions.
- **Irreplaceability:** Once compromised, biometric data cannot be changed, unlike a password or PIN. Thus, biometrics are primarily useful for local authentication but unsuitable for global identity systems.
- **Lack of Standardization:** High development costs and dependency on specific vendors are significant drawbacks in current biometric systems.

3.12 Kerberos Authentication System

Kerberos is a widely-used authentication protocol based on a Trusted Third Party (TTP) model. It's designed to ensure that user passwords are never transmitted over the network. Instead, the password is used locally for encryption.

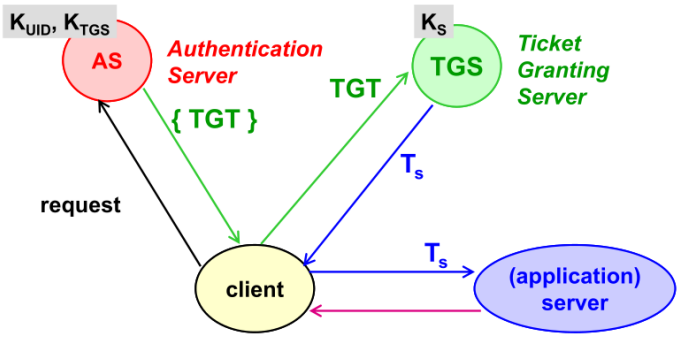
- **Realm:** Refers to a Kerberos domain, grouping together all systems that use Kerberos for authentication.
- **Credential:** A unique identifier for a user, typically in the format **user.instance@realm**.

3.12.1 Players

There are 4 key players:

- **Client (User/Application:)** the person or application trying to access a resource.
- **Authentication Server (AS):** Confirms who you are and issues a Ticket Granting Ticket (TGT).
- **Ticket Granting Server (TGS):** Issues tickets for specific services after seeing the TGT.
- **Service (Server):** The resource you want to access.

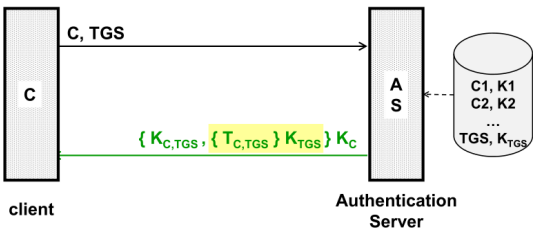
3.12.2 How it Works?



1. **TGT Request**: Client authenticates with the Authentication Server (AS) to obtain a Ticket Granting Ticket (TGT).

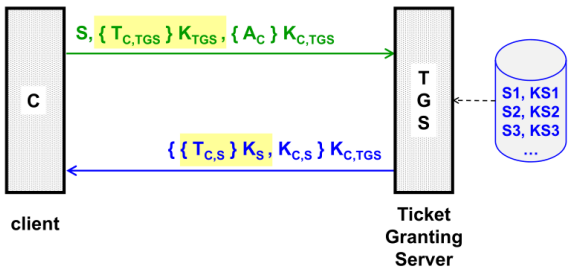
   The expression $KC, TGS, TC, TGS\ KTGSKC$ represents a **TGT**:

   - The enitre structure is encrypted using the client's secret key (KC) (This ensure that only the intended client can decrupt and use the TGT).
   - $TC, TGS\ KTGS$ is the **TGS** that contains the client's identity and other information. It's encrypted using the TGS's secret key ($KTGS$), ensuring only TGS can read and verify the ticket.



2. **Service Ticket Request**: TGT is sent to the Ticket Granting Server (TGS) to request a service-specific ticket.

   - The client sends: (S) $\rightarrow$ the indetifier of the target device, $TG, TGS\ \ KTGS$ $\rightarrow$ TGS ticket and ($AC\ KC, TGS$) $\rightarrow$ the authenticator.
   - TGS Verifies and Responds:
     - The TGS decrypts the TGS ticket using its secret key (KTGS).
     - It verifies the client's identity using the authenticator.
     - If successful, the TGS generates a Service Ticket ($TC, S\ KS$) encrypted with the secret key (KS), ensuring only the service can read it; and a new session key (KC, S), shared between the client and the service.

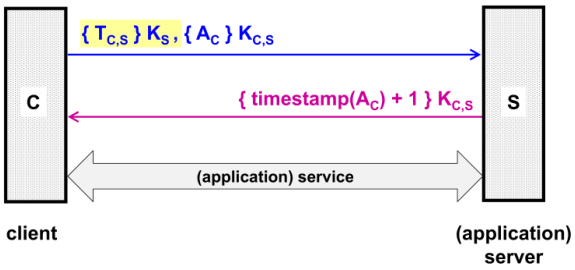


3. **Access Service**: Client uses the service ticket to authenticate and access the resource.

   The client send:

   - The service ticket encrypted using the service's secret key (KS), so only the service can decrypt and validate it.
   - The authenticator encrypted using the session key (KS, S) shared bewteen the client and the service.

   The service responds with a response in which confirms successfully authentication. It encrypts the message with the session key (KC, S) to ensure it's secure and readable only by the client.



3.12.3 Single Sign-On (SSO)

SSO allows users to authenticate once and access multiple services without repeated logins. Types of SSO:

- **Fictitious SSO:**
  - Relies on tools like password synchronization or management (e.g., password wallets).
  - Limited to specific applications.
- **Integral SSO:** Uses advanced multi-application methods like Asymmetric Challenge-Response Authentication (CRA) or Kerberos. Often requires application changes.
- **Multi-Domain SSO:** Expands SSO across domains using technologies like SAML tokens, which generalize Kerberos tickets.

N.B. Single Sign-On (SSO) is not exclusive to Kerberos, but Kerberos is one of the prominent technologies that implements SSO capabilities.

3.13 Authentication Interoperability

Authentication interoperability define methods, standard, and protocol for performing authentication securely and efficiently. Let's start to analyze some framework.

3.13.1 OATH

The Open Authentication (OATH) framework provides standards for one-time password (OTP) and symmetric key management.

- **HOTP:** Uses a shared secret key (K) and a counter (C) to generate an OTP.
  Function:
  $$HOTP(K, C) = sel(HMAC - h(K, C)) \ \& \ 0x7FFFFFFF$$
  The result is truncated and transformed into an N-digit code (e.g., 6 digits)[1].
- **TOTP:** Similar to HOTP but uses time intervals (TS) instead of counters.
  Function:
  $$C = (T - T0)/TS$$
  With default values: $T0$ -> Unix epoch, TS -> 30-second intervals, T -> unixtime (now).
- **OCRA** (OATH Challenge-Response Algorithm)
- **PSKC** (Portable Symmetric Key Container): XML-based format for symmetric key transport.
- **DSKPP** (Dynamic Symmetric Key Provisioning Protocol): A client-server protocol for securely provisioning symmetric keys.

3.13.2 Google Authenticator

Supports HOTP and TOTP with adjustments for usability:

- **K:** Base-32 encoded.
- **C:** 64-bit unsigned integer.
- **sel(X):** Uses the 4 least-significant bits of X to locate a portion of the result.
- **Defaults:** TS = 30 seconds, N = 6 digits (zero-padded if necessary).

[1] K-> shared key, C -> counter, sel -> function to select 4 bytes out of a byte string

3.13.3 FIDO (Fast Identity Online)

FIDO, developed by the FIDO Alliance, improves authentication by offering secure, password-less, and multi-factor methods. It uses biometric data locally to unlock cryptographic keys and employs asymmetric cryptography for signing challenges or transactions. FIDO prevents phishing by ensuring that authentication responses cannot be reused. Each response is a unique signature created over various data, including the Relying Party (RP) identity, making it specific to the service being accessed. Additionally, a new key pair is generated during each registration, which prevents the association of a user's identity across different services or accounts.

FIDO's frameworks include:

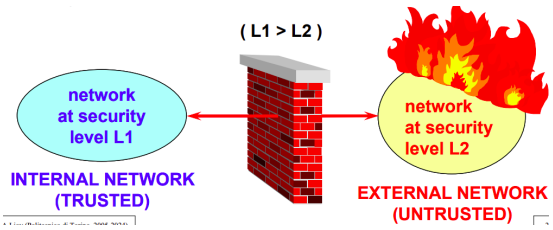
- **UAF** (The Universal Authentication Framework): for password-less login.
- **U2F** (Universal 2nd Factor) for device-based two-factor authentication.
- **FIDO2**: integrates U2F with WebAuthn for robust web authentication.

Chapter 4

Firewall and IDS/IPS

4.1 What is a Firewall?

A firewall acts as a protective barrier, much like a physical wall against fire. It is primarily a controlled connection point between networks with differing security levels. Its key functions include:



- **Boundary Protection:** Serving as a network filter between trusted (internal) and untrusted (external) networks.
- **Compartmentalization:** Dividing network zones based on security levels to enforce stricter control.

4.2 Ingress vs Egress firewall

Firewalls can be classified based on the direction of traffic they manage:

- An **ingress firewall** focuses on incoming connections, typically regulating access to public services provided by the network or supporting exchanges initiated by internal users.
- An **egress firewall** monitors outgoing connections. It's typically used to check the activity of internal personnel [Works well for channel-based services (e.g., TCP applications) but faces challenges with stateless, message-based services (e.g., ICMP, UDP)].

4.3 The three principles of the firewall

1. The firewall should be the only connection between the internal and external networks.
2. Only the "authorized" traffic is allowed to pass the firewall.
3. The firewall itself must be secure against potential vulnerabilities.

These principles were outlined by *D.Cheswick, S.Bellovin.*

4.4 Authorization policies

- **Permitlist/allowlist**: All that is not explicitly permitted, is forbidden.
  - It offers higher security but it's difficult to manage.
- **Blocklist/denylist**: All that is not explicitly forbidden, is permitted.
  - It's less secure but it's more easy to manage.

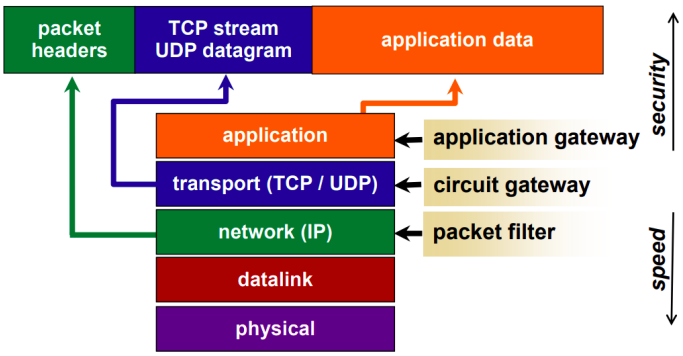
4.5 Basic components

Firewalls include several key components that work together to provide security:

- **Packet Filter/ Screening Router / Choke**: Filters traffic at the network level based on packet attributes such as IP headers and transport headers.
- **Bastion Host**: A secure system with auditing capabilities, often positioned to handle critical traffic.
- **Application Gateway (Proxy)**: Acts on behalf of an application, managing access control and providing detailed packet inspection at the application level.
- **Dual-Homed Gateway**: A system with two network interfaces and disabled routing, isolating internal and external networks. (In the way we can decide which packets are sent from a network to another).

4.6 A which level the controls are made?

Firewalls at higher levels provide more security but tend to be slower, while firewalls at lower levels offer less security but greater speed.



To undersand:

security is better up → speed is better down
security is better down → speed is better up

4.7 Network-Level Controls

Firewalls operate at various levels of the network stack applying different controls based on the type of firewall used:

- **Packet Filters**: Operate at the network level, inspecting packet headers.
- **Stateful Packet Filters**: Track connection states for more dynamic filtering.ù
- **Circuit-Level Gateways / Proxies**: Work at the transport level, ensuring secure connection establishment.
- **Application-Level Gateways / Proxies**: Provide inspection and control at the application layer, often with more granular rules.

4.7.1 Packet filter

A packet filter inspects packets based on headers, such as IP and transport headers. These filters are traditionally available on routers and now in most OS. They allow for rules like permitting incoming connections to specific services (e.g., a web server) or limiting DNS queries from specific internal servers.

- **Pro:**
  - Independent of applications, scalable, and cost-effective (available in many OS and routers).
  - Good performance and low cost.
- **Cons:**
  - Vulnerable to attacks like IP spoofing or fragmented packets.
  - Difficult to support services using dynamically allocated ports (e.g. FTP).
  - Complex configuration and hard to implement user authentication.

4.7.2 Circuit-Level Gateway

A circuit-level gateway acts as a transport-level proxy, creating secure communication channels between client and server without inspecting the payload. It protects against Layer 3/Layer 4 attacks like IP fragmentation or TCP handshake exploits.

- **Pro:**
  - Provides protection by isolating the server from attacks.
  - Offers client authentication and eliminates many low-level attack vectors.
- **Cons:**
  - Still shares many limitations of packet filters and requires modifications to the application for full functionality.

4.7.3 Application-Level Gateway

An application-level gateway (or proxy) operates at the application layer, inspecting the payload of packets. These proxies often require modifications to client applications and can enhance security by checking the semantics of application data (e.g., HTTP methods) or performing peer authentication.

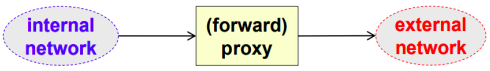
- **Pro:**
  - Strong security against application vulnerabilities (e.g., buffer overflow attacks).
  - Fine-grained access controls and the ability to mask internal IP addresses.
  - May provide protection against attacks like buffer overflows.
  - Not transparent to the client and may break the client-server model.
- **Cons:**
  - Requires specific proxies for each application and can introduce delays when supporting new applications.
  - High resource usage and lower performance due to user-mode operations.

Variants of application-level proxies include **transparent proxies**, which are less intrusive to clients, and **strong application proxies**, which focus on checking data semantics, not just syntax.

4.7.4 HTTP Proxies

An HTTP forward proxy is a server that acts as an intermediary or front-end for client requests. It receives requests from internal users and forwards them to the real external server (So it's a egress control). **Benefits:**

- **Shared Cache**: External web pages are cached, reducing load times for all internal users.
- **Authentication and Authorization**: Enforces user authentication and controls access for internal users.



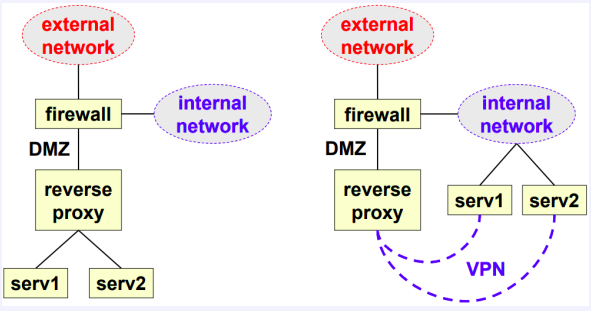
Type of Proxies

- **Forward Proxy (HTTP Proxy)**: This type of proxy controls outgoing traffic from the internal network to the external one, enforcing access controls and caching content. It provides features like user authentication and authorization, content filtering, and bandwidth control.
- **Reverse Proxy (HTTP Reverse Proxy)**: A reverse proxy sits in front of web servers, providing additional services like load balancing, content inspection, TLS acceleration, and caching. It can obfuscate the internal server structure, improving security, and support performance optimizations like dynamic page feeding based on client speed.

Reverse proxy: possibile configuration

Key components:

- **External Network (Red Cloud)**
- **Firewall**
- **DMZ (Demilitarized Zone)**: A buffer zone between external and internal networks where systems that interact with the external network are placed for added security.
- **Reverse Proxy**
- **Internal Network**: The secure area where your actual servers (e.g., serv1, serv2) reside.



How configurations work?

Left configuration: Direct reverse proxy setup

- **Flow:** External user → Reverse Proxy → Internal Servers (serv1, serv2).
- **Benefits:** Internal servers are hidden and protected, while performance and security are improved.

Right Configuration: Reverse Proxy with VPN

- **Flow**: External user → Reverse Proxy → Encrypted VPN connection → Internal Servers (serv1, serv2).
- **Benefit**: Adds an extra layer of security through encryption for communication between the reverse proxy and internal servers.

4.7.5 WAF (Web Application Firewall)

A WAF is a module installed at a proxy (forward and/or reverse) to filter the application traffic. Filters the followings types of traffic:

- HTTP commands.
- HTTP request/response headers.
- HTTP request/response content.

Popular WAF example: ModSecurity

ModSecurity is a widely-used WAF plugin for web servers like Apache and NGINX, offering protection through predefined rules like the OWASP ModSecurity Core Rule Set (CRS).

4.8 Firewall's Architectures

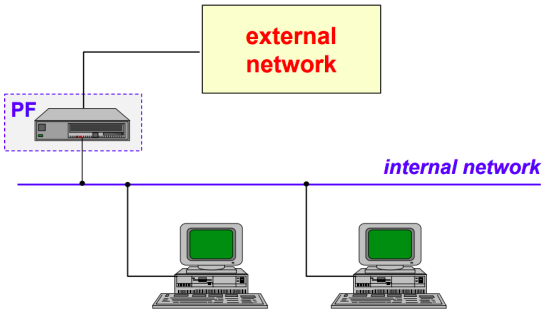
4.8.1 "Packet Filter" architecture

What it does?

This architecture uses a simple packet filter to screen traffic at both the IP and higher protocol levels (like TCP)

Key Points:

- If implemented with a router then it's a "screening router" and there's **no need for extra dedicated hardware**.
- The packet filter element represents a **single point of failure**.
- **There's no need for proxies** or application modifications.



Beware

Simple, cost-effective, but insecure!

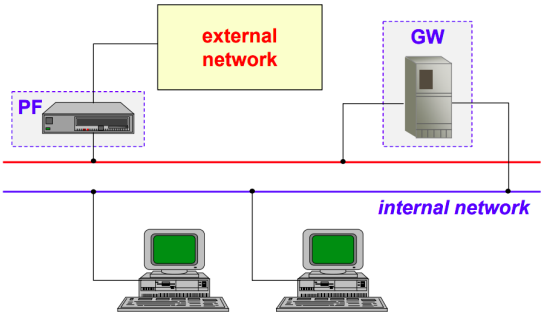
4.8.2 "Dual-homed Gateway" architecture

What it does?

Uses a system with two network interfaces (hence "dual-homed") to provide a basic firewall between external and internal networks.

Key Points:

- Easy to implement with small hardware requirements.
- The internal network can be masquerade.



Beware

Inflexible, with higher management overhead and reduced flexibility in handling complex network setups.

4.8.3 "Screened host" architecture

What it does?

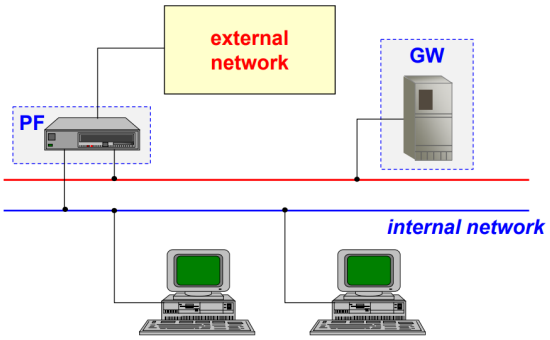
Uses two primary components:

- **Router**
- **Bastion host**

This setup aims to provide enhanced security by controlling the flow of traffic between the internal and external networks.

Key Points:

- Traffic from the internal network (INT) to external (EXT) is blocked unless it's from the bastion host. Similarly, external traffic is blocked unless directed to the bastion.
- More flexible (skip control over some services / hosts)



Beware

- **More Expensive and Complex** to manage: Requires two systems (router + bastion host) instead of just one.
- **Limited Masking**: Only the host and protocols that go through the bastion host can be masked for security (such as hiding internal IP address or data). However, if the packet filter (PF) uses NAT, it can mask additional traffic and hide internal network details.

4.8.4 "Screened subnet" architecture

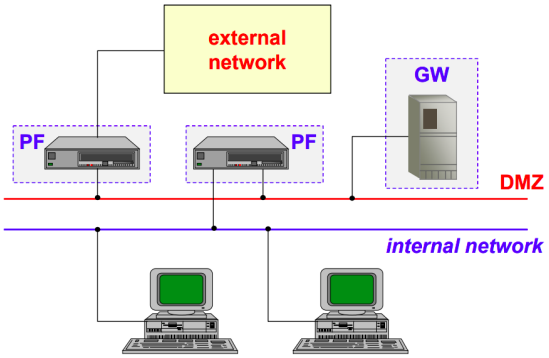
What it does?

A DMZ (De-Militarized Zone) is created between an internal network and a external network. The DMZ acts as a buffer zone to isolate external entities from the internal network, such as web servers or remote access points.

Traffic flows:

- The first firewall controls access from the external network to the DMZ, ensuring only authorized traffic can reach the public-facing systems.
- The second firewall restricts traffic from the DMZ to the internal network, allowing only specific and necessary connections.

It the most secure solution because this setup hides the internal network from external users, reducing exposure to attacks.



Beware

- **Higher cost** due to the additional hardware.
- **The DMZ is home not only to the gateway** but also to other host (typically the public servers).

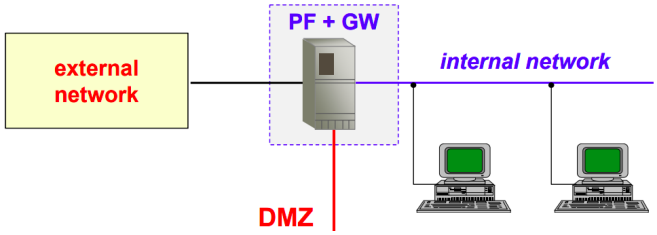
4.8.5 "Screened subnet" architecture (version 2)

What it does?

In this version of the Screened Subnet architecture, the packet filters (**PF**) and gateway (**GW**) are merged into a single device that is called **AKA** ("three-legged firewall").

Three-Legged design:

- One interface connecting to the **External network** (untrusted).
- One interface connecting to the **Internal network** (trusted).
- One interface connecting to the DMZ (where public-facing services reside).

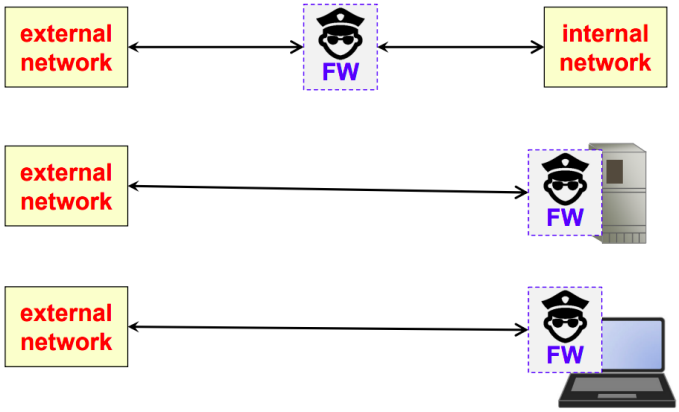


Beware

It's **more simple** than the previous version but presents a **single point of failure**.

4.9 Local/Personal Firewall

This image represents three different firewall configurations: **Network firewall**, **Local firewall** and **Personal firewall**. In the previous pages we have considered the network firewall, now let's start to analyze the Local/Personal firewall.



A local or personal firewall is installed directly on a device to protect it from unauthorized access or attacks. Unlike a normal network firewall, it may limit the PROCESSES that are permitted:

- To open network channels towards other nodes (client behavior).
- To answer network requests (server behavior).

These features make local firewalls essential for containing malware, blocking trojans, or preventing accidental misconfigurations. **However, in order to be effective, the firewall management MUST be separated from system management.**

4.10 Protection offered by a firewall

A firewall is 100% effective only for attacks over/against blocked channels. The other channels require other protection technique:

- **VPNs**
- **intrusion detection systems (IDS)**
- **application-level protections**

4.11 Intrusion Detection System (IDS)

Definition

An **Intrusion Detection System (IDS)**text is a security mechanism designed to:

- identify actors using a system or a network without authorization (and their actions).
- identify authorized actors who violate their privileges.

Hypothesis

- The behavior "pattern" of non-authorized users differs from that of the authorized ones

4.11.1 IDS functional features

- **Passive IDS:** Observes and detects issues but doesn't act to stop them.
  Techniques:
  - Cryptographic checksums to detect changes in files.
  - Pattern matching to identify attack signatures (e.g., malware signatures).
- **Active IDS:** Monitors activity dynamically and reacts when thresholds are exceeded.
  Techniques:
  - Tracks normal behavior to identify anomalies (**"learning"**).
  - Gathers detailed statistics on traffic and actions (**text"monitoring"**).
  - Triggers alerts or responses when unusual patterns occur (**"reaction**).

4.11.2 IDS topological features

- **HIDS (Host-Based IDS):** Monitors individual devices (hosts).
  - Analyzes logs from the OS, services, or applications.
  - Uses built-in OS tools to track internal activity.
- **NIDS (Network-Based IDS):** Monitors traffic at the network level.
  - Uses network traffic monitoring tools.

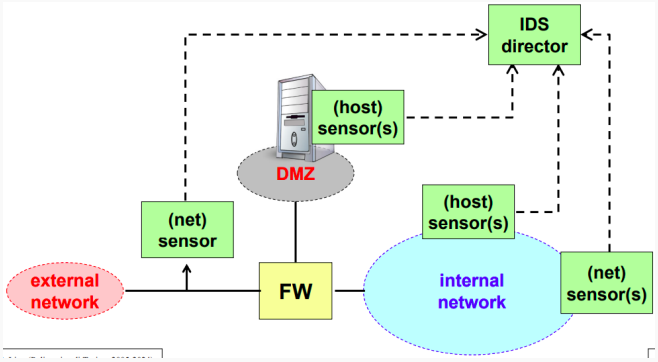
NIDS Architecture

Components:

- **Sensor:** The frontline tool that analyzes traffic or logs looking for suspect patterns. Generates security alerts and can modify access controls (e.g., block traffic).
- **Director:** Coordinates sensors and manages a central database.
- **Message System:** Ensures secure communication among IDS components.

Workflow:

1. Traffic from the external network enters through the firewall.
2. (Net) sensors monitor traffic before it reaches the DMZ or internal network.
3. Within the DMZ, host sensors monitor server activities.
4. Traffic entering the internal network is further monitored by both host and network sensors.
5. All sensors report suspicious activities to the IDS director.
6. The IDS director consolidates this data and raises alerts or takes action if any patterns of intrusion or policy violations are detected.



4.12 Intrusion Prevention System (IPS)

Definition

An **Intrusion Prevention System (IPS)** is an advanced security technology (not a product) that identifies unauthorized activities like an IDS but actively blocks or mitigates them in real time (= IDS + distributed dynamic firewall).
Often, Intrusion Prevention Systems (IPS) are integrated with Intrusion Detection Systems (IDS) into a unified solution called **Intrusion Detection and Prevention System (IDPS)**

Why is it useful to merge IPS with IDS?

IPS already has some of IDS's features, such as detection capabilities. But merging IPS with IDS brings many advantages because:

- **IDS** excels in providing deep visibility into all network activity, including suspicious behavior that might not warrant immediate action.
- **IPS** actively blocks threats, but in doing so, it may not focus as much on monitoring benign-but-suspicious activities.

N.B. IPS requires careful configuration and monitoring to balance its protective capabilities with the risk of blocking innocent traffic.

4.13 Next-Generation Firewall (NGFW)

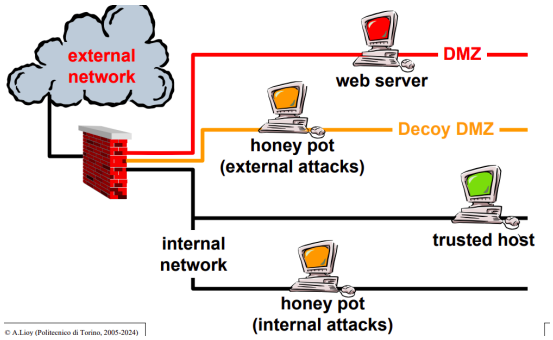
A Next-Generation Firewall (NGFW) enhances traditional firewalls by offering features like **application identification**, regardless of the network port used, and traffic decryption. It can integrate with **authentication systems** such as Active Directory, enabling per-user and per-application security policies. Additionally, NGFWs provide advanced protection by **filtering traffic based on known vulnerabilities**, threats, and malware.

4.14 Unified Threat Management (UTM)

Unified Threat Management (UTM) integrates multiple security features such as firewalls, VPNs, anti-malware, content inspection, and IDPS into a single device. The goal is to simplify security management by consolidating various solutions, reducing complexity and costs, though the specific capabilities depend on the manufacturer.

4.15 Honey pot / Honey net

A honeypot is a decoy system designed to attract attackers and analyze their behavior, while a honeynet is a network of these decoy systems used to study both external and internal attacks. These tools help in identifying and understanding attack methods.



Chapter 5

Security of network applications

5.1 Standard situation

The standard situation for an ordinary network application is very **negative**, since most of the systems implement very **weak authN mechanism** that typically rely on `username` and `password`, which can lead to:

- Password Snooping
- IP Spoofing

Even if **stronger authN mechanism** is implemented, there are still problems with data:

- Data Snooping/Forging
- Shadow Server/MITM
- Replay and Filtering attacks

So to resolve these problems, we can use two different approaches: **Channel Security** and **Message/Data Security**.

5.2 Channel Security

Channel Security implements a secure connection between two nodes. To achieve this, before the start of the communication, the two nodes need to negotiate the `algorithms`, `parameters` and `keys` to protect the whole traffic that will be sent through the communication channel.

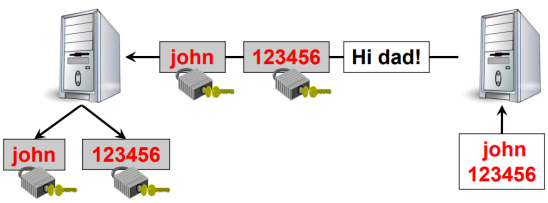
Since all this features are negotiated **before transmitting data**, we ensure:

- **Single** or **mutual authN**
- **Data integrity**
- **Data confidentiality**

Channel Security is very easy to implement because it requires no (or small) modification of applications. Since this features are also negotiated **automatically** we cannot have **non-repudation**.

N.B. The **main issue** with Channel Security is that its security properties are provided **only** during the transit **inside** the communication channel. So data in **not** protected when it's located at the end-user.

5.3 Message/Data security



It applies protection **only** when it's needed. This means that data is **individually** protected by wrapping it into a **secure container**. Data (not the channel) ensures the following security properties:

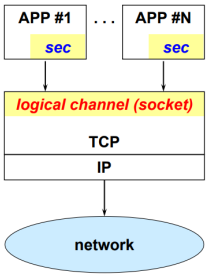
- **Single authN, not mutual** because features are **not negotiated**
- **Data integrity**
- **Data confidentiality**

If protection is applied **voluntary** and **explicity** by the user, we can have **non-repudation**. Message/Data Security requires modification to application.

5.4 Different implementation

`Channel Security` and `Message/Data Security` can be **combined**. Security to be applied to applications. These security concepts can be implemented in two different way:

Security Internal to Applications

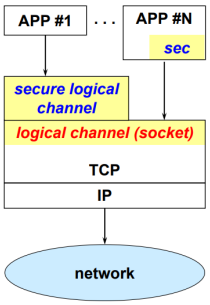


In this approach **each application implements security internally**. The common part is limited to the communication channels (socket). They could be:

- **Possible implementation errors** (implementing security protocols is not simple)
- Does not guarantee interoperability (-> security specification could be different between applications)

N.B. **Communication channels** are the **only** common part between the various application application and are used to exchange information.

Security External to Applications



A **Secure Socket** is created on top of the already existing "normal" socket, effectively creating a **new level** called **Secure Session Level (SSL)**. The Secure Socket implements all the security functio3.12.1 Web-Based Payment Systems The steps are: 1. The merchant M presents the products or services on their website for the carholder C to browser. 2. C places an order through the merchant's website, initiating the pay- ment process. 3. M redirects C to a payment gateway PG. 4. PG will create a Secure TLS channel between the PG itself and C. 5. C provide the credit card data. 6. Inside the PG, a Virtual POS (Point Of Sale) will handle the data and will ask to the payment network (of the specific credit card) if they are valid. 7. If the provided credit card data are correct, the payment network will return a positive answer. 8. Finally M isns and can be used by any application that wants to communicate in a secure way. Thanks to **SSL** we can:

- Simplify the work of application developer
- Avoid implementation errors

5.5 TLS (Transport Layer Security)

Originally known as **SSL**, **TLS** is a **network/session level protocol** which is capable of creating **secure transport channles** that grants:

- peer authentication
- message confidentiality
- message authentication and integrity
- protection against replay and filtering attacks

It's easily applicable to all protocols based on **TCP**, such as HTTP, SMPT, NNTP, FTP, TELNET, ... (e.g. HTTP (https://....) = 443/TCP).

Beware

HTTPS is not a protocol, it's a combination of HTTP over TLS.

Beware

The current version of TLS is **TLS-1.3** and nowadays everything that uses a version **older than TLS-1.2** is considered **insecure and deprecated**.

```
nsiiops       261/tcp # IIOP Name Service over TLS/SSL
https         443/tcp # http protocol over TLS/SSL
smtps         465/tcp # smtp protocol over TLS/SSL (was ssmtp)
nntps         563/tcp # nntp protocol over TLS/SSL (was snntp)
imap4-ssl     585/tcp # IMAP4+SSL (use 993 instead)
sshell        614/tcp # SSLshell
ldaps         636/tcp # ldap protocol over TLS/SSL (was sldap)
ftps-data     989/tcp # ftp protocol, data, over TLS/SSL
ftps          990/tcp # ftp protocol, control, over TLS/SSL
telnets       992/tcp # telnet protocol over TLS/SSL
imaps         993/tcp # imap4 protocol over TLS/SSL
ircs          994/tcp # irc protocol over TLS/SSL
pop3s         995/tcp # pop3 protocol over TLS/SSL (was spop3)
msft-gc-ssl   3269/tcp # MS Global Catalog with LDAP/SSL
```

Figure 5.1: Official Ports for TLS/SSL Applications

5.5.1 TLS - AuthN and Integrity

Peer AuthN is performed **always at channel setup**:

- The server must be authenticated (mandatory). Sends its public key (X.509 certificate) and responds to an implicit asymmetric challenge.
- The client can authenticate itself (optional): with public key, X.509 certificate and explicit challenge.

For **authN** and **integrity** of **data**, TLS uses:

- A Keyed-Digest (SHA-1 or better).
- An implicit MID to **avoid** Replay and Filtering attacks.

5.5.2 TLS - Confidentiality

Data confidentiality is granted by TLS in the following way:

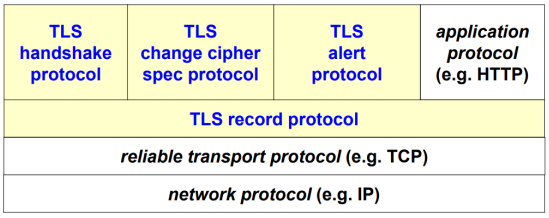
1. The client generates a session key used for symmetric encryption of data (using RC4, 3DES, IDEA, AES, or ChaCha20).
2. The session key is **exchanged** with the server via Asymmetric cryptography (using RSA or DH).

N.B. Authentication is available in TLS 1.2, while is mandatory in TLS 1.3.

5.5.3 TLS Architecture

TLS is positioned **on top** of the transport layer (e.g TCP) and network layer (e.g. IP).

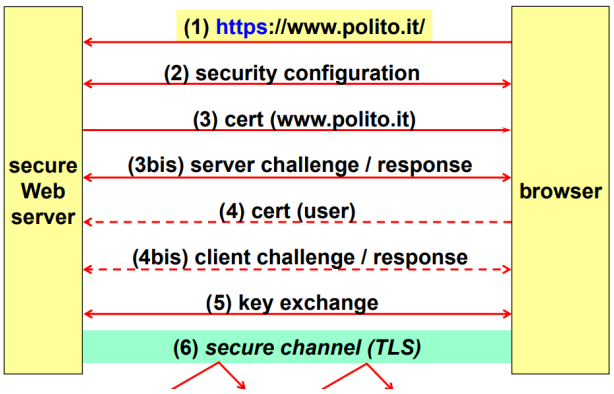
- **Record protocol**: is positioned between TCP and the other TLS protocols and the various application protocol (e.g HTTP).
- **Handshake protocol**
- **Change Cipher Spec Protocol**: used to change the `algorithms`, `parameters` or `keys` on an already established TLS channel.
- **Alert Protocol**: used to send error information on the TLS channel before permanently closing it.



5.5.4 TLS Handshake Protocol

The **Handshake Protocol** is used by TLS to perform **channel setup**, let's see an example:

1. The browser (**client C**) initiates a connection to the **web server** (**S**) by requesting the website (www.polito.it) over HTTPS.
2. **C** and **S** discuss about the security configuration. They should both agree to use the **strongest common algorithms**.
3. The **S** sends its certificate to **C** to prove its identity. The certificate includes the server's public key and the domain name.
   3b) **S** will then respond to an Asymmetric CRA sent by the **C** (This is part of the handshake to confirm (implicitly) the server's identity.)
4. (Optional) **S** may require (explicitly) a client certificate to verify the user's identity.
5. **Key exchange** is performed. During this phase **C** and **S** exchange **master keys** that are linked to the session-ID. Every time a new **TLS Connection** is opened with that **session-ID**, a new connection oriented key must be generated by combining the master key and one of the previously exchanged random numbers.
6. Finally the **Secure TLS Channel** is opened and available to exchange data.

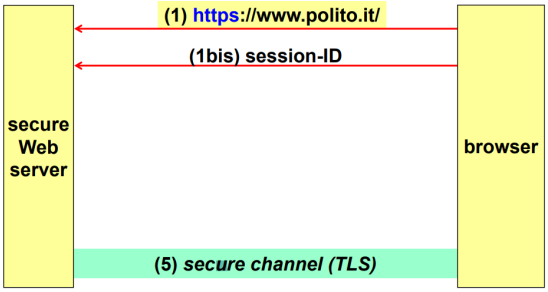


After the Secure TLS Channel has been established, client and server will:

- Negotiate the **Session-ID**
- Exchange **random numbers** to be used for the subsequence generation of keys.

5.5.5 TLS Session ID

A typical web transaction involves the transmission of multiple elements between the client (e.g., a web browser) and the server. To avoid the overhead of negotiating a new session for each element (i.e. Many connections can be part of the same logical session), TLS uses a session ID. **The session ID is a unique identifier that allows the client and server to resume a previous session.** If the client, when opening the TLS connection, sends a valid session-id, the negotiation phase is skipped, and data can be immediately exchanged over the secure channel.
N.B. However, the client must start communicating in encrypted form; otherwise, the server will reject the connection.



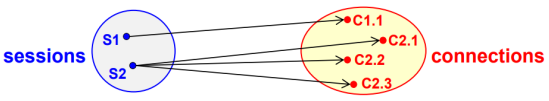
Beware

The server can reject the use of session-id (always or after a time passed after its issuance)

5.5.6 TLS Session & Connections

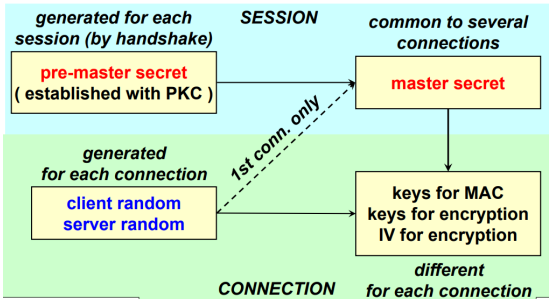
In TLS we have to distinguish the two:

- **TLS Session**: it's a **logical association between client and server** created by the Handshake protocol. It defines a set of cryptographic parameters which are used during communication and can be shared by one or more TLS Connections (1:N).
- **TLS Connection**: it's a **transient** TLS Channel between client and server that is associated to one specific TLS Session (1:1).



5.5.7 Relationship among Keys and Sessions

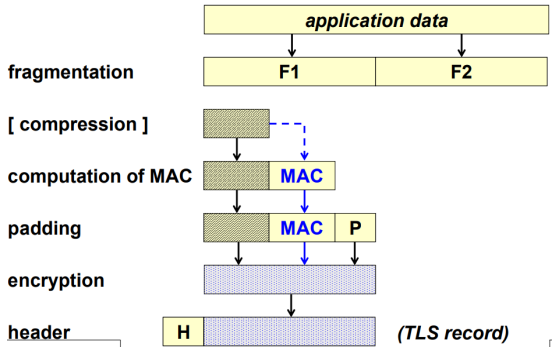
Every time we create a TLS session: we generate (with PKC) a **pre-master secret**. By generating the client random/server random value, we can mix them with the pre-master secret to generate the **master secret**, which is common to **all** the Connections of the Session. Each time we create a TLS Connection: we generate a **unique** client random/server random value. Combining them with the master secret we create **for this specific connetction and direction**:



- Keys for MAC
- Keys for encryption
- IV for encryption

5.5.8 TLS Record Protocol (authenticate-then-encrypt)

The TLS Record Protocol secures application data during transmission.The steps:



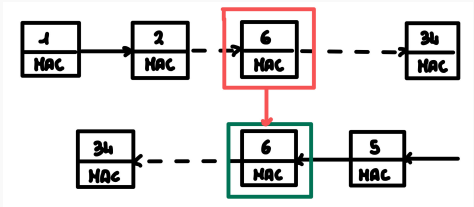
1. **Fragmentation**: Divides the data (from the application layer or higher-level TLS protocols) into smaller blocks or "records" that can be transmitted efficiently.
2. The fragmented data is **compressed** (**Optionally**).
3. Ma MAC is computed over the data (→ to verify integrity and authenticity of the data).

$$MAC = message\_digest(key, seq_number \,||\, type \,||\, version \,||\, length \,||\, fragment)$$

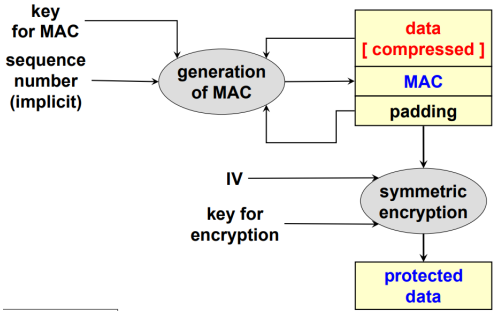
There are two different keys: the sender-write-key and the receiver-write-key. This separation prevents replay attack.

Replay attack

- The sender (Client) sends an encrypted command, like Transfer $100 to the receiver (Server).
- An attacker intercepts this message and replays the exact encrypted data back to the sender (Client).
- Since the sender and receiver share the same key, the sender would decrypt the message and potentially act on it (depending on the system design), causing confusion or security breaches.



4. **Padding** is added to make data encryptable.
5. Data is **encrypted**. Steps:
   - **Generation of MAC**: the (compressed) data is combined with the MAC key and sequence of number as explained before.
   - **Symmetric Encryption**: takes (data || MAC || padding), and, by using the IV and the encryption key, generates the protected data.



6. The encrypted data is packaged into a TLS record (header+payload).

5.5.9 Perfect Forward Secrecy

If a server has a certificate valid for both signature and encryption, it can be used bot for authN (via Signature) and Key Exchange (Asymmetric Encryption of the Session Key). If an attacker copies all the encrypted traffic and later discovers the server's SK, the attacker can decrypt all traffic (**past**, **present** and **future**).
Perfect Forward Secrecy is a technique in which: if a PK/SK pair is compromised, only the **present** and **future** traffic can be decrypted.
This mechanism (implemented in TLS) is achieved through the use of **ephemeral keys** (temporary keys) that are generated for each session and are not stored. For **authenticity**, the one-time key must be signed (the server's SK is only used for

signing, granting server authN). However, it cannot have an associated X.509 certificate, as the CA process is slow and often not available online. "Ephemeral" mechanisms are:

- Suitable for DH (Diffie-Helmann).
- Slow for RSA → a compromised would be to re-use N times the same key.

This means that with Perfect Forward Secrecy:

- If the **short term SK** is compromised → the attacker can decrypt **only the present and future** (typically only for the session) traffic.
- If the **long term SK** is compromised → the server can **no longer perform authN** but at least the attacker cannot decrypt any traffic.

Example

ECDHE (Elliptic Curve Diffie-Helmann Ephemeral) is a key exchange protocol that provides PFS.

5.5.10 TLS Downgrade Attack

When the client negotiates with the server which version of TLS to use:

- The client always send (in `Client Hello`) the **highest supported version**.
- The server notifies (in `Server Hello`) the version to be used, which must be the highest in common.

An attacker could send fake server response, forcing a version downgrade until a vulnerable one is reached, and then execute an attack.

Beware

Initial messages, before the setup of the secure channel, are not protected by the security protocol.

This behaviour does not (always) mean that we are under attack, it could be simply a **Insecure downgrade** (→ Some servers do not send the correct response, rather they close the connection. Then the client has no choice to try again with a lower protocol version).

5.6 Virtual Servers and TLS

Problem: the **Virtual Servers problem** is very frequent with **web hosting**: different logical names are associated to the **same IP address** (e.g.: `home.myweb.it=1.2.3.4` and `food.myweb.it=1.2.3.4`).
This problem is easily solved in **HTTP/1.1**: the client uses **Host header** to specify the Logical Name of the server it want to connect to.
But it is a problem for HTTPS, where the domain name is encrypted (because TLS, transport layer security protocol, is activated before HTTP, application layer protocol).

Solutions:

- Collective (wildcard) certificates: the private key is shared by all servers. e.g. `*.myweb.it`. Different browsers may react differently to this solution.
- Certificate with a list of servers in the SAN (subjectAltName) field: The private key is shared among all servers, and the certificate needs to be reissued whenever a server is added or removed.
- The SNI (Server Name Indication) extension: The client sends the domain name in the initial message. The server can then choose the correct certificate. Limited support by browsers and servers.

5.7 ALPN extension (Application-Layer Protocol Negotiation)

The **ALPN extension** was created to **speed up** connection creation by avoiding negotiation `round-trips`. ALPN works in the following way:

- (ClientHello) ALPN = true + list of supported application protocols;
- (ServerHello) ALPN = true + selected application protocols;

N.B. It is particularly important to negotiate HTTP/2 and QUIC, since (for example) Chrome and Firefox support HTTP/2 only over TLS.

ALPN is useful also for those servers that use different certificates for the different application protocols.

5.8 TLS Fallback - Signaling Cipher Suite Value

The **SCSV estension** is used to prevent protocol downgrade attacks. SCSV works in the following way:

- When a connection to a server is closed during the Handshake protocol negotiation phase, the client (upon opening downgraded connection) should send to the server a new (dummy) ciphersuite called **TLS_FALLBACK_SCSV**;
- Upon receiving TLS_FALLBACK_SCSV and a version lower then the highest one supported, the server must send a new Fatal Alert value: "inappropriate_fallback".

  Immediately after sending this value, the **TLS channel is closed** → the client should retry with its highest protocol version;

N.B. However, many servers do **not** support SCSV, but at least **most** of them have fixed this bad behaviour → browsers can **disable insecure downgrade**.

5.9 DTLS (Datagram Transport Layer Security)

DTLS tries to apply the concepts of TLS to **datagram security** (e.g. UDP). It does not offer the same security properties as TLS and is in competition with IPsec and Application Security.
For example, if we use **SIP** (**Session Initiation Protocol**) we could implement security with:

- IPsec
- TLS (SIP over TCP)
- DTLS (SIP over UDP)
- Secure SIP

5.10 HTTP Security

In**HTTP/1.0** some **bad security mechanisms** are defined:

- **Address-based Access Control**: performed by the server on the IP address of client.
- **Password-based Access Control**: **username** and **password** are sent encoded (base64).

Both of them are **highly insecure** since HTTP assumes we are using a Secure Channel.

> From HTTP/1.1 **digest authN** based on Symmetric CRA was introduce.

5.10.1 HTTP Basic Authentication

1. After having created the channel using **HTTP/1.0**, the server automatically responds with an "authentication failed" message but does **not** close the channel, because the client (maybe) did not know that authN was needed → the servers sends an **authN request** to the client;
2. The client's browser will open a pop-up that asks the client to enter its credentials in order to perform authN.
3. The client submits its credentials encoded in Base64 (**not encrypted → no secure**).

```
#encode
echo username:password | openssl enc -a

#decode
echo YWRtaW46cGFzczEyMw== | openssl enc -a -d
```

4. The server can **verify** the credentials and (eventually) send the information requested by the client.

```
C>S: GET /path/to/protected/page HTTP/1.0
S>C: HTTP/1.0 401 Unauthorized - authentication failed
     WWW-Authenticate: Basic realm="POLITO - didattica"
C>S: Basic czEyMzQ1NjpTZWdyZXRpc3NpbWE=
S>C: HTTP/1.0 200 OK
     Server: Apache/1.3
     Content-type: text/html
     <html> ... protected page ... </html>
```

5.10.2 HTTP Digest Authentication

The keyed-digest introduced in **HTTP/1.1** is computed in the following way:

- The sever may even insert an **"opaque"** field to send some state information to the client.
- The server uses a nonce to prevent replay attacks.

$$HA1 = \text{md5 (A1)} = \text{md5(user":"realm":"pwd)}$$
$$HA2 = \text{md5 (A2)} = \text{md5(method":"URI)}$$
$$response = \text{md5 (HA1":"nonce":"HA2)}$$

1. After having created the channel using **HTPP/1.1**, the server **automatically** responds with a "`Authentication failed`" message but **not** close the channel, because the client maybe did **not** know that authN was needed, so the server sends an **authN request** to the client containing the **nonce** and the **opaque** values encoded in Base64.
2. The clint's browser will open a pop-up that asks the client to perform authN → the client computes the keyed-digest and sends it, together with the **opaque**, to the server.
3. The server can then **verify** the credentials and (eventually) send the information request by the client.

```
C>S: GET /private/index.html HTTP/1.1
S>C: HTTP/1.0 401 Unauthorized - authentication failed
    WWW-Authenticate: Digest realm="POLITO",
    nonce = "dcd98b7102dd2f0e8b11d0f600bfb0c093",
    opaque = "5ccc069c403ebaf9f0171e9517f40e41"
C>S: Authorization: Digest username="lioy",
    realm="POLITO",
    nonce="dcd98b7102dd2f0e8b11d0f600bfb0c093",
    uri="/private/index.html",
    response="4c9d010fac372e048297160bff991883",
    opaque="5ccc069c403ebaf9f0171e9517f40e41"
S>C: HTTP/1.0 200 OK
    Server: NCSA/1.3
    Content-type: text/html
    <html> ... protected page ... </html>
```

5.10.3 HTTP & TLS/SSL

We can use both HTTP and TLS/SSL by activating them in one of the following orders:

- **TLS then HTTP**: the client connects to the server using TLS and then sends the HTTP request (→ Content inspection is not possible!).
- **HTTP then TLS**: the client connect to the server using HTTP and then upgrades the connection to TLS (→ We see the HTTP methods until the upgrade!).

Beware

They are **not** equivalent as the activation order has an impact over `applications`, `firewall` and IDS.

In general the most used approach is "TLS then <proto>" (e.g. HTTPS).

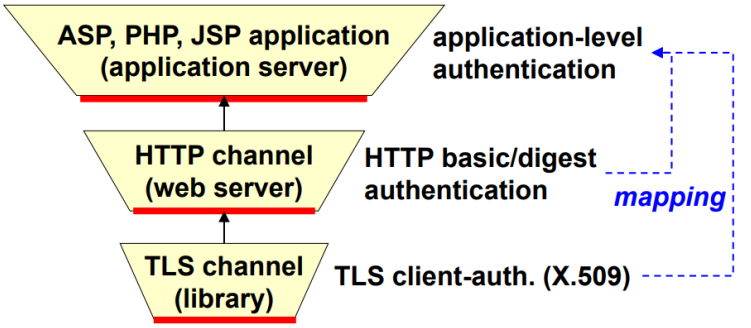
TSL AuthN at the Application Layer

Via client authentication it's possible to identify the user that opened the channel (without asking for his username and password).
The identification process of a client can be done at the application level, but it is not a standard feature of TLS.

AuthN in Web applications

The **earlier** access control is performed, the **smaller** is the attack surface.

Moreover, there is no need to repeat authentication across different parts of the application if the user's identity is securely propagated throughout the application. Some web servers support a (semi-)automatic mapping between the credential extracted from the X.509 certificate and the users of the HTTP service and/or the OS.



Web Form Security

Technically speaking, the security of the page containing the form is not important (e.g. http://www.ecomm.it/login.html).
The actual security depends on the URI of the method used to send the data to the server (e.g <form . . . action="https://www...>).
Yet, as a form of psychological help for the users, we should also look to implement security in the page containing the form

5.11 HTTP Strict Transport Security (HSTS)

The **HSTS extension** is used by HTTP server to declare that its interaction with **User Agent** (**UA** → e.g. browser, etc...) **must only be via HTTPS**, effectively preventing Protocol Downgrade and Cookie Hijacking attacks.
HSTS is applied **only** upon receiving a **valid HTTPS response**. Its validity is renewed at every access and it may:

- Include subdomains (recommended).
- Be pre-loaded into the servers.

Syntax

The syntax is:

```
Strict-Transport-Security:
   max-age = <expire-time-in-seconds>
   [ ; includeSubDomains ]
   [ ; preload ]
```

Example:

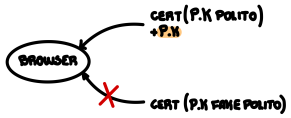
```
$ curl -s -D- https://www.paypal.com/ | fgrep -i strict
strict-transport-security: max-age=63072000
$ curl -s -D- https://accounts.google.com/ | grep -i
strict
strict-transport-security: max-age=31536000;
includeSubDomains
```

Upon receiving an answer from the server that will contain the **HSTS Header**, inside there will be specified:

- The expiration data in seconds.
- (Optionally) the subdomains.
- (Optionally) the option of being added to the client's pre-load list.

5.11.1 HTTP Public Key Pinning (HPKT)

The **HPKP extension** is used by HTTP servers to declare that it is using a certain **PK** by specifying the Digest of the PK and/or one or more CAs in its chain (except the root CA). The User Agent (UA) caches this key and will refuse to connect to a site presenting a different key.



- HPKP is a **TOFU** (**Trust On First USe**) technique.
- HPKP is dangerous if we lose control of the key.
- HPKP can cause problems with key updates → always include a backup key.

Syntax

The syntax is:

```
Public-Key-Pins:
   pin-sha256 = " <base64-sha256-of-public-key> ";
   max-age = <expireTime-in-seconds>
   [; includeSubDomains]
   [; report-uri = " <reportURI> "]
```

```
Public-Key-Pins-Report-Only:
   pin-sha256 = " <base64-sha256-of-public-key> ";
   max-age = <expireTime-in-seconds>
   [; includeSubDomains]
   [; report-uri = " <reportURI> "]
```

Example

```
$ $ curl -s -D - https://scotthelme.co.uk/
    | fgrep -i public-key
public-key-pins:
pin-sha256="9dNiZZueNZmyaf3pTkXxDgOzLkjKvI+Nza0ACF5IDwg=";
pin-sha256="X3pGTSOuJeEVw989IJ/cEtXUEmy52zs1TZQrU06KUKg=";
pin-sha256="V+J+71HvE6X0pqGKVqLtxuvk+0f+xowyr3obtq8tbSw=";
pin-sha256="9lBW+k9EF6yyG9413/fPiHhQy5Ok4UI5sBpBTuOaa/U=";
pin-sha256="ipMu2Xu72A086/35thucbjLfrPaSjuw4HIjSWsxqkb8=";
pin-sha256="+5JdLySIa9rS6xJM+2KHN9CatGKln78GjnDpf4WmI3g=";
pin-sha256="MWfCxyqG2b5RBmYFQuLllhQvYZ3mjZghXTRn9BL9q10=";
includeSubDomains; max-age=2592000;
report-uri="https://scotthelme.report-uri.com/r/d/hpkp/
   enforce"
public-key-pins-report-only:
pin-sha256="X3pGTSOuJeEVw989IJ/cEtXUEmy52zs1TZQrU06KUKg=";
pin-sha256="Vjs8r4z+80wjNcr1YKepWQboSIRi63WsWXhIMN+eWys=";
pin-sha256="IQBnNBEiFuhj+8x6X8XLgh01V9Ic5/V3IRQLNFFc7v4=";
max-age=2592000;
report-uri="https://scotthelme.report-uri.com/r/d/hpkp/
   reportOnly"
```

The **HPKP Header** includes:

- The digest of the PK computed with SHA-256.
- The expiration date in seconds.
- (Optionally) the subdomains.
- (Optionally) the report URI that is used to report violations.

N.B. HPKP can work in **enforcing** or **report-only mode**.

5.12 E-Payment System

What happen in the past:

- Failure of digital cash due to technical and legal reasons.
- Failure of a dedicated payment protocol. (SET-Secure Electronic Transaction) caused by technical and organizational challenges.

Beware

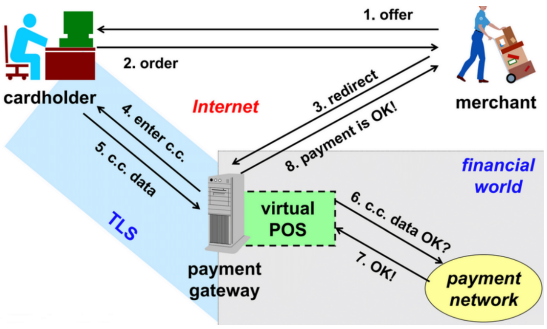
Providing users with specialized software can compromise the security of the system.

Currently, the most widely used approach is transmitting credit card numbers over a TLS channel. However, this does not guarantee protection against fraud: in the past, VISA Europe reported that internet transactions accounted for about 50% of fraud attempts, despite representing only 2% of total transactions.

5.12.1 Web-Based Payment Systems

The steps are:

1. The **merchant M** presents the products or services on their website for the **carholder C** to browser.
2. **C** places an order through the merchant's website, initiating the payment process.
3. **M** redirects **C** to a **payment gateway PG**.
4. **PG** will create a **Secure TLS channel** between the **PG** itself and **C**.
5. **C** provide the credit card data.
6. Inside the **PG**, a **Virtual POS (Point Of Sale)** will handle the data and will ask to the payment network (of the specific credit card) if they are valid.
7. If the provided credit card data are correct, the payment network will return a positive answer.
8. Finally **M** is informed of the confirmation of the transaction and it will provide the goods to **C**.



Different **PG** can have different **deduction fees** (e.g 5%,10%,etc...) which cumulates with the price proposed by **M**.
In order for all of this work, **C must have TLS enabled** in its browser.

Consequences

- The effective security depends upon the configuration of both the server and the client.
- The **PG** has all the information (payment + goods) while **M** knows only info about the goods.

5.12.2 PCI DSS (Payment Card Industry Data Security Standard)

The **PCI DSS** standard is required by **all credit issuers** or internet-based transactions. It contains very detailed technical prescriptions (compared to other security standards):

- **Design, Build and Operate a Protected Network:**
  - **R1**:install and maintain a configuration with a firewall to protect access to cardholders' data.
  - **R2**: do not use pre-defined system passwords or other security parameters set by the manufacturer.
- **Protect the Cardholders' Data:**
  - **R3**: protect the stored cardholders' data.
  - **R4**: encrypt the cardholders' data hen transmitted across an open public network.
- **Establish and Follow a Program For Vulnerability Management:**
  - **R5**: use an antivirus and regularly update it.
  - **R6**: develop and maintain protected applications and systems.
- **Implement Strong Access Control:**
  - **R7**: limit the access to the cardholders' data only to those needed for a specific risk.
  - **R8**: assign a single unique ID to each user.
  - **R9**: limit physical access to the cardholders' data.
- **Regularly Monitor and Test the Networks:**
  - **R10**: monitor and track all accesses to the network resources and cardholders' data.
  - **R11**: periodically test the protection systems and procedures.
- **Adopt a Security Policy:**
  - **R12**: adopt a security policy.