DMIF, Università di Udine

Tecnologie Digitali per il Cibo e la Ristorazione

Reti e Internet

Andrea Brunello
andrea.brunello@uniud.it

A.A. 2021–2022

Introduzione

- La necessità di condividere risorse e scambiare informazioni fra calcolatori ha portato alla nascita di sistemi per la loro l'interconnessione, detti **reti**
- In questa parte del corso parleremo di come i calcolatori sono connessi fra loro, come vengono scambiate le informazioni, le applicazioni delle reti (fra le quali il World Wide Web), e i relativi problemi di sicurezza

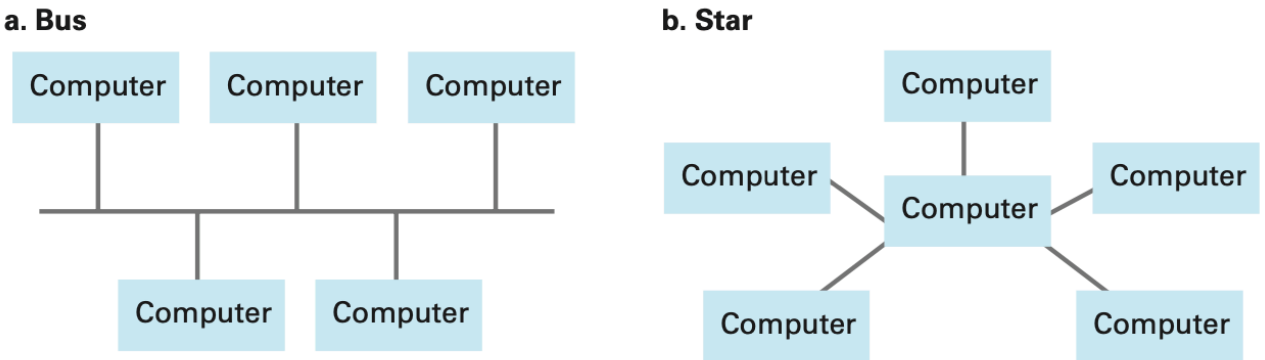
Classificazione per dimensione

- **PAN, Personal Area Network**: utilizzata per comunicazioni a breve distanza, ad esempio fra smartphone e cuffie bluetooth
- **LAN, Local Area Network**: composta da un insieme di calcolatori e altri dispositivi connessi all'interno di un singolo edificio, o un insieme di edifici (es., campus universitario)
- **MAN, Metropolitan Area Network**: rete di dimensioni intermedie, che copre un'intera comunità
- **WAN, Wide Area Network**: connette dispositivi posti a grande distanza, in diverse città, stati, o continenti

Classificazione per topologia

Le connessioni fra dispositivi possono essere realizzate in diversi modi. Le strategie più diffuse sono:

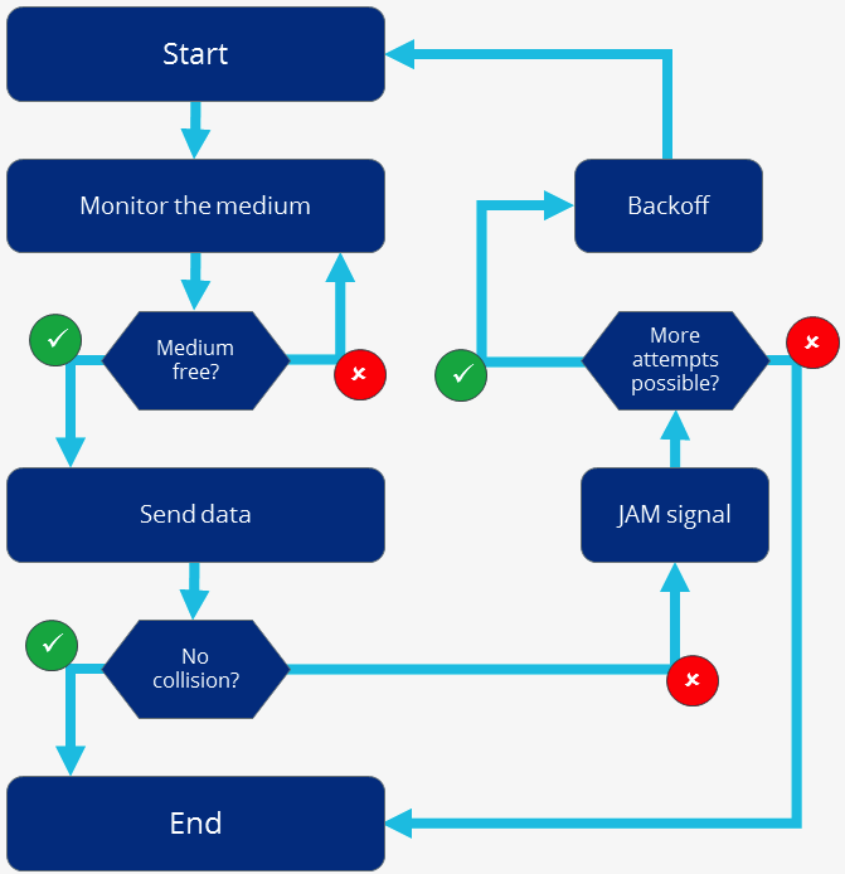
- **Bus**: dispositivi direttamente connessi fra loro attraverso una linea di comunicazione detta *bus* (es., *Ethernet*)
- **Stella**: ciascun dispositivo dialoga con gli altri per mezzo di un dispositivo centrale, che smista e inoltra i messaggi (es., access point Wi-Fi)



Protocolli di comunicazione

- Il ruolo dei protocolli è quello di disciplinare le comunicazioni attraverso una rete
- Senza protocolli, più dispositivi potrebbero cercare di comunicare nello stesso momento, provocando interferenze, o non rimanendo in ascolto dei messaggi provenienti dagli altri

- *Carrier Sense, Multiple Access With Collision Detection*
- Protocollo per reti bus, come Ethernet
- Ogni messaggio viene inviato a tutte le macchine connesse al bus, ma viene considerato solo dai destinatari interessati
- Per inviare un messaggio, una macchina attende fino a che il bus appare silenzioso
- A questo punto, inizia la trasmissione, continuando però ad ascoltare il bus
- Se un'altra macchina inizia a trasmettere nello stesso tempo, entrambe rilevano la *collisione*
- Come risposta entrambe interrompono la trasmissione e rimangono in attesa per un periodo di tempo casuale prima di ricominciare la trasmissione


Start
Monitor the medium
Medium free?
Send data
No collision?
JAM signal
More attempts possible?
Backoff
End

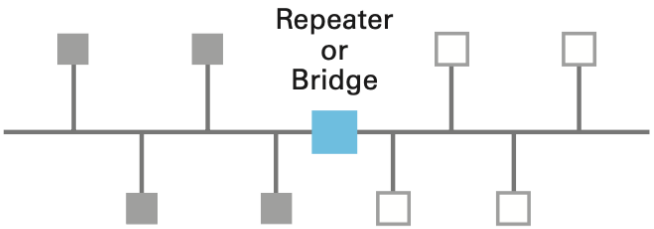
Protocolli di comunicazione

- Si osservi che il protocollo CSMA/CD non funzionerebbe su una rete a stella, in quanto una macchina potrebbe non essere in grado di rilevare che le sue trasmissioni stanno collidendo con quelle di un'altra
- Nel caso delle reti wireless, viene utilizzato il protocollo CSMA/CA (Carrier Sense, Multiple Access with Collision Avoidance)
- La formalizzazione di tale protocollo, assieme ad altri, all'interno dello standard IEEE 802.11, definisce il funzionamento del WiFi

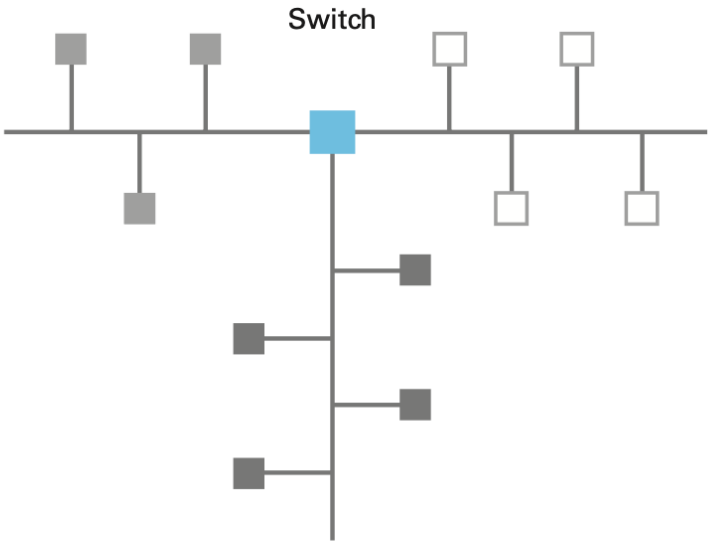
A volte è necessario connettere fra loro più reti al fine di creare un sistema di comunicazione più esteso. Ciò può essere fatto sfruttando diversi dispositivi:

- un **ripetitore** inoltra i messaggi fra due diversi bus, eventualmente amplificando i segnali
- un **bridge** connette due diversi bus, come un ripetitore. Tuttavia, inoltra da un bus all'altro solo i messaggi scambiati fra macchine che si trovano sulle due reti opposte. In questo modo, le comunicazioni che avvengono fra le macchine di una stessa rete non interferiscono con quelle fra le macchine dell'altra rete
- uno **switch** è essenzialmente un bridge che connette fra loro più di due bus

Combinazione di reti
Bridge VS Switch



a. A repeater or bridge connecting two buses



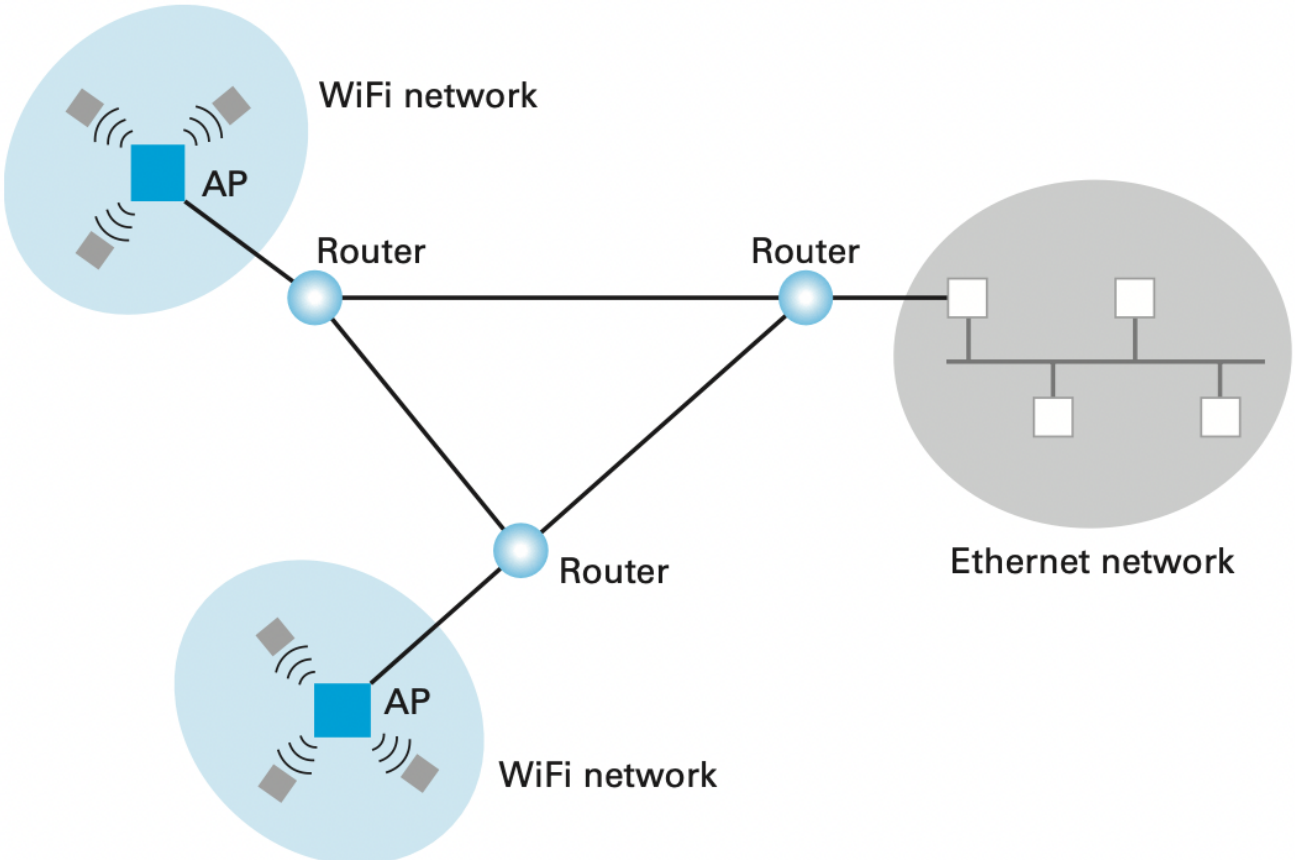
b. A switch connecting multiple buses

Combinazione di reti
Reti eterogenee

- Spesso, le reti da connettere hanno caratteristiche incompatibili (es., diverse topologie, diversi protocolli di comunicazione)
- In tal caso, si rendono necessarie strategie di connessione più complesse
- Si viene a creare una rete di reti, in cui le reti originali mantengono la loro individualità
- Tale rete di reti prende il nome di **internet** (da non confondere con Internet)

Reti internet

- La connessione fra reti viene gestita da dispositivi detti **router**
- Compito dei router è gestire l'inoltro di messaggi fra le reti
- Si consideri il caso in cui un dispositivo *A* posto in una rete WiFi voglia inviare un messaggio ad un dispositivo *B* posto in una rete Ethernet, connesse in una rete internet:
  1. *A* invia il messaggio al proprio AP WiFi
  2. L'AP invia il messaggio al router associato
  3. Il router inoltra il messaggio ad un dispositivo router sulla rete Ethernet di destinazione
  4. Tale dispositivo diffonde il messaggio sul bus
  5. *B* riceve il messaggio dal bus
- Si osservi che, spesso, AP e router coincidono



Il "punto" in cui una rete è connessa ad una rete internet viene detto **gateway**.

- L'inoltro dei messaggi si basa su un sistema di indirizzamento globale
- Ad ogni dispositivo connesso alla rete internet viene assegnato un indirizzo univoco, che lo contraddistingue
- Dunque, ciascun dispositivo avrà due indirizzi associati: uno "locale", utilizzato per le comunicazioni all'interno della rete di appartenenza, e uno "globale", utilizzato per le comunicazioni fra le reti
- I router fanno uso di opportune **tabelle di instradamento** per determinare la direzione verso cui inviare i vari messaggi

Internet

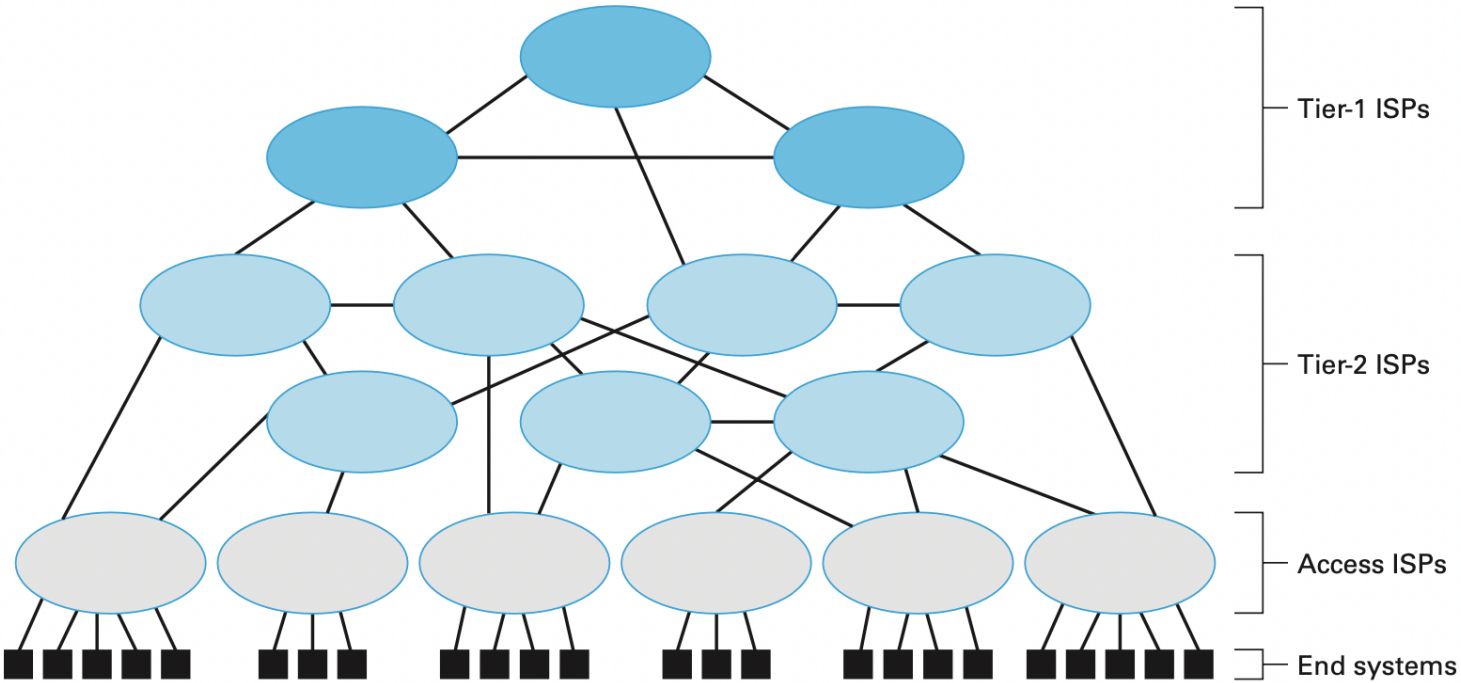
- **Internet** è il più famoso esempio di rete internet
- Rete globale che utilizza il protocollo TCP/IP per la gestione delle comunicazioni fra i dispositivi
- Le radici si hanno negli anni '60, in diversi progetti DARPA (U.S. Defense Advanced Research Projects Agency)
- Precursore di Internet è ARPANET, utilizzato negli anni '70 per connettere organizzazioni militari e accademiche
- *"ARPANET came out of our frustration that there were only a limited number of large, powerful research computers in the country, and that many research investigators, who should have access to them, were geographically separated from them"*
  Charles Herzfeld, ARPA Director (1965–1967)

ARPANET nel 1974

Architettura di Internet

- Le reti che compongono Internet sono gestite da organizzazioni dette **Internet Service Provider, ISP** (ISP può indicare anche le reti gestite dagli ISP stessi)
- Le reti gestite dagli ISP possono essere classificate in una gerarchia, a seconda del ruolo che giocano nella struttura generale di Internet:
  - Al vertice della gerarchia si trovano le **tier-1 ISP**, costituite da WAN ad elevata capacità e velocità, che si estendono su scala internazionale/intercontinentale
  - Alle tier-1 ISP si connettono le **tier-2 ISP**, che sono su scala regionale e hanno capacità inferiori
  - L'accesso alle ISP dei primi due livelli, che costituiscono il cuore di Internet, viene fornito da un intermediario detto **tier-3 ISP**. Quest'ultimo gestisce una propria rete indipendente, fornendo l'accesso ai clienti

Gerarchia di ISP



Indirizzi IP

- Ciascun dispositivo connesso alla rete Internet è identificato da un indirizzo univoco, detto **IP address** (Internet Protocol address)
- Originariamente, un indirizzo IP era costituito da un pattern di 32 bit (IPv4), es. 255.110.145.0
- Ad oggi è in corso il passaggio verso indirizzi IP a 128 bit (IPv6), per consentire uno spazio di indirizzamento più ampio
- Blocchi contigui di indirizzi IP sono assegnati agli ISP dall'ICANN (Internet Corporation for Assigned Names and Numbers, una compagnia non-profit nata per coordinare l'uso di Internet

Indirizzi mnemonici

- Utilizzare direttamente gli indirizzi IP sarebbe scomodo per gli esseri umani
- Internet dispone di un sistema di indirizzamento alternativo in cui i dispositivi sono identificati da nomi mnemonici, es. *uniud.it*
- Tale sistema di indirizzamento è basato sul concetto di **dominio**, che può essere considerato come una "regione" di Internet gestita da una singola autorità, quale un'università, un'azienda, o un'agenzia governativa
- I domini vengono registrati presso l'ICANN da appositi enti detti **registrar**
- All'atto della registrazione, ad un dominio viene assegnato un nome, univoco all'interno di Internet

Indirizzi mnemonici (2)

- Una volta registrato un nome di dominio, la relativa organizzazione può estenderlo liberamente per ottenere ulteriori identificatori univoci al suo interno, che si riferiscono a macchine specifiche, ad esempio *datasciencelab.dimi.uniud.it*
- Tali estensioni consentono anche l'organizzazione delle risorse facenti parte del dominio in *sottodomini*
- Ad esempio, a partire da *uniud.it*, è possibile specificare il sottodominio *dimi.uniud.it*

Passaggio da IP a indirizzo mnemonico

- Nonostante gli indirizzi mnemonici siano comodi per gli esseri umani, i dispositivi connessi a Internet dialogano fra loro sfruttando gli indirizzi IP
- È quindi necessario un sistema per operare la conversione fra i due sistemi di indirizzamento
- Tale conversione è effettuata attraverso l'uso di diversi server, detti **name server**, che possono essere intuitivamente considerati come degli elenchi telefonici
- L'insieme di tali server prende il nome di **Domain Name System** (DNS)
- Il processo di utilizzare DNS per effettuare la traduzione fra i due sistemi di indirizzamento viene detto **DNS lookup**

Il World Wide Web

- Il World Wide Web è un'applicazione software che consente lo scambio di documenti e altre risorse attraverso la rete Internet
- Trae origine dal lavoro del ricercatore Tim Berners-Lee al CERN di Ginevra, che intuì il potenziale rappresentato dagli **ipertesti**, vale a dire, documenti di testo collegati fra loro da **link**
- La prima implementazione ufficiale del World Wide Web risale al Dicembre 1990, e conteneva la definizione di:
  - un formato per la specifica di ipertesti (HTML)
  - un protocollo per trasferire ipertesti (HTTP), che segue il paradigma client-server
- In seguito, si ha l'estensione degli ipertesti agli ipermedia, i.e., documenti contenenti anche audio, video, immagini

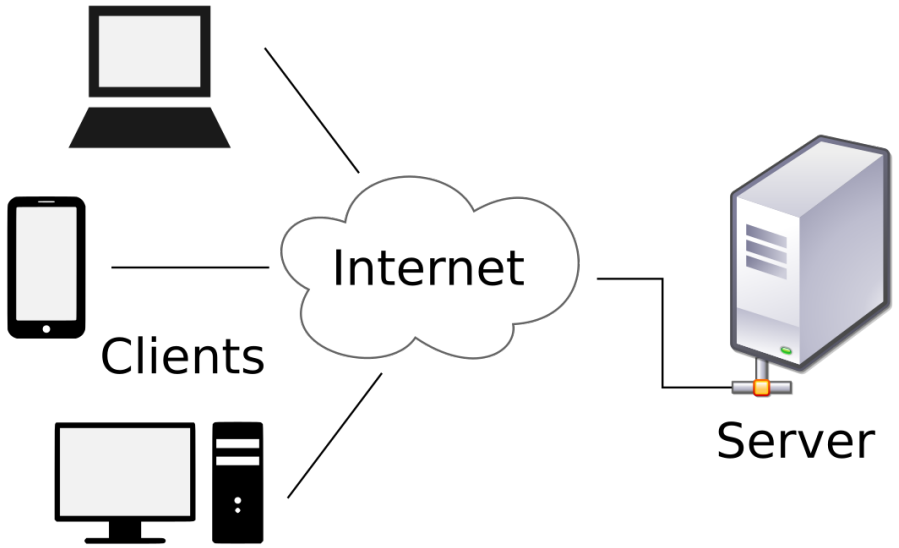
Implementazione del Web

I pacchetti software che consentono agli utenti di accedere agli ipertesti disponibili su Internet ricadono in due categorie principali:

- **browser**: risiede sulla macchina dell'utente e si occupa di recuperare il materiale da esso richiesto, presentandolo in maniera ordinata e fruibile (es., Firefox, Chrome)
- **webserver**: risiede su una macchina contenente ipertesti. Il suo compito è fornire l'accesso a tali dati agli utenti che ne fanno richiesta

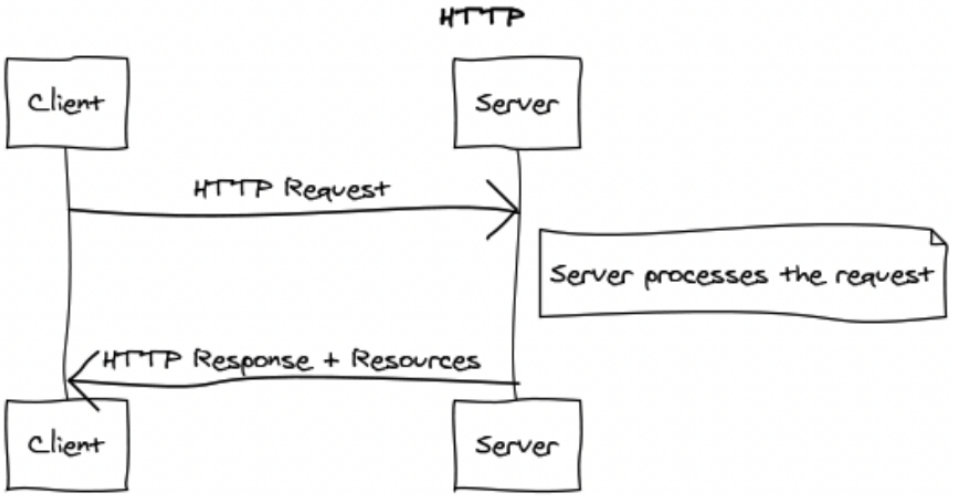
Paradigma client-server

Browser e webserver aderiscono al cosiddetto paradigma client-server



Un computer client o terminale si connette ad un server per la fruizione di un determinato servizio, quale ad esempio la condivisione di una risorsa hardware/software con altri client, appoggiandosi alla sottostante architettura protocollare

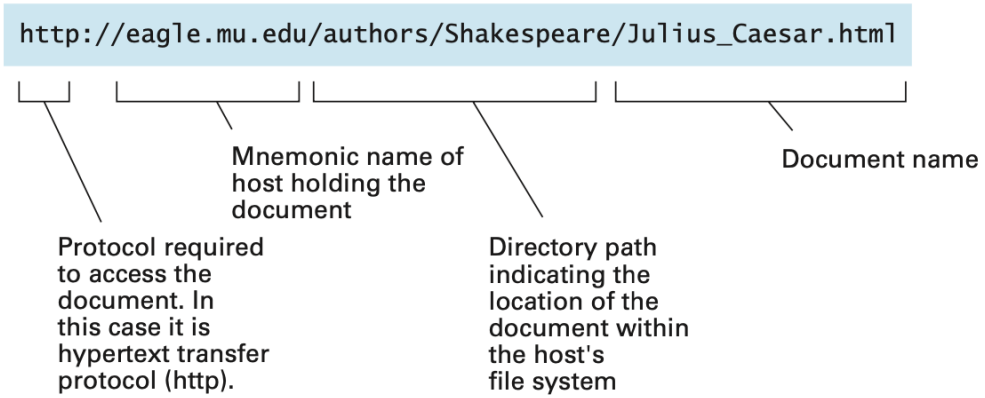
Paradigma client-server (2)



1. *Client*: attiva la connessione e richiede il servizio
2. *Server*: accetta la connessione, identifica il client, risponde alla richiesta e chiude la connessione
3. *Client*: ottiene la risposta e chiude la connessione

URL

- Al fine di poter localizzare e recuperare documenti sul Web, a ciascuno di essi viene assegnato un indirizzo univoco, detto *Uniform Resource Locator* (URL)
- Un URL contiene tutta l'informazione di cui un browser necessita per contattare il server e richiedere il documento desiderato



"Contatta il webserver presente sulla macchina nota come eagle.mu.edu utilizzando il protocollo HTTP e recupera il documento Julius_Caesar.html che si trova al percorso /authors/Shakespeare/"

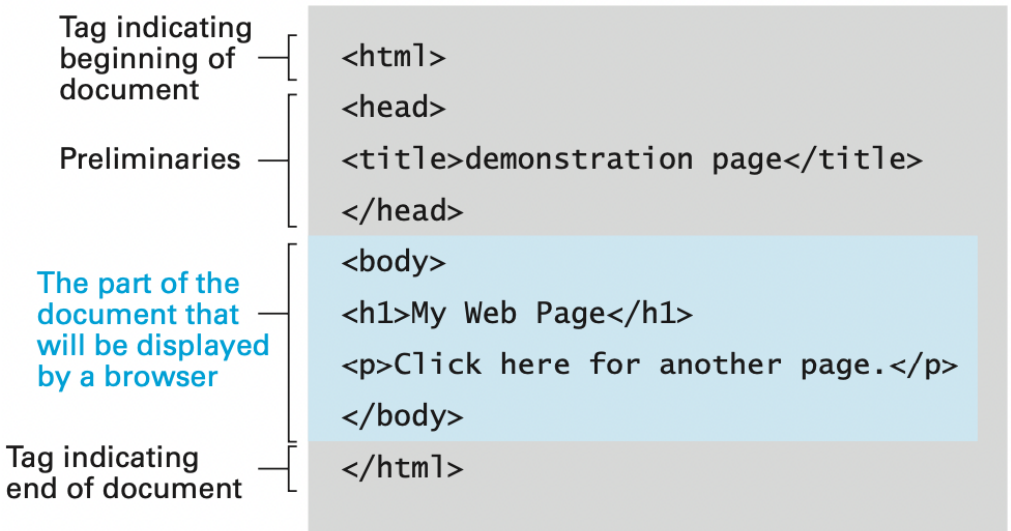
URL (2)

- A volte, un URL consiste solamente della parte relativa al protocollo e al nome mnemonico di una macchina, senza quelle relative al percorso e al file da recuperare
- Ad esempio, *http://www.google.com*
- In tal caso, il webserver installato sulla macchina restituisce una pagina (documento) di default, tipicamente chiamata *home page*
- Infine, anche la parte relativa al protocollo può essere tipicamente ignorata. In tal caso, il browser assume "http://". Ad esempio, *www.google.com*

Hypertext Markup Language (HTML)

- Un documento ipertestuale è simile ad un normale file di testo, in quanto contiene stringhe codificate utilizzando ASCII o Unicode
- Oltre al testo, un documento ipertestuale contiene anche dei simboli speciali, detti **tag**, che descrivono come esso dovrebbe essere visualizzato a schermo, quali risorse multimediali dovrebbero accompagnare il testo, e quali parti del contenuto sono collegamenti ad altri documenti
- Intuitivamente, è un processo simile al fornire ad un tipografo un testo e delle immagini, annotati con le indicazioni grafiche su come stamparli

- Un documento HTML si compone di due sezioni:
  - **head**: contiene metadati (dati sui dati) riguardanti il documento, quali titolo, set di caratteri utilizzato, ecc
  - **body**: contiene gli elementi principali di un documento, come titoli, paragrafi, immagini, collegamenti, tabelle, liste, ecc



HTML, esempio

Esempio interattivo:

- `https://www.w3schools.com/html/tryit.asp?filename=tryhtml_basic_document`
- cambiare il colore del testo:
  `style="color:red"`
- specificare un collegamento:
  `<a href="https://www.uniud.it/">Link a Uniud</a>`
- inserire un'immagine:
  `<img src="pic_trulli.jpg">`

Altri applicativi Internet

Oltre al Web, su Internet sono state costruite molte altre applicazioni, ad esempio:

- Posta elettronica: protocolli SMTP, POP3, IMAP
- Comunicazioni telefoniche: VoIP (Voice over IP)
- Streaming multimediali: RTSP
- Trasferimento di file: FTP

Aspetti relativi alla sicurezza

- Si stima che, una volta connesso a Internet, un dispositivo sarà soggetto ad almeno un attacco nei primi 20 minuti
- La maggior parte degli attacchi si concretizza nel tentativo di installare codice malevolo (**malware**) sui dispositivi
- Il software malevolo può essere di diversi tipi: virus, worm, trojan, spyware, …

- Un virus è un software che infetta un dispositivo inserendosi all'interno di programmi che risiedono nella macchina bersaglio
- Quando l'utente avvia il programma, attiva inconsapevolmente anche il virus, che è in grado di svolgere operazioni più o meno distruttive sull'hardware o sul software del sistema
- Un virus necessita quindi dell'intervento umano per la sua attivazione e diffusione, che avviene tipicamente distribuendo copie del codice infetto ad utenti inconsapevoli
- *Take home message*: non aprite mai allegati e-mail provenienti da mittenti sconosciuti

- I worm, a differenza dei virus, non necessitano di un programma ospitante per operare: sono a sé stanti
- Una volta che un worm è entrato nel sistema operativo, in genere attraverso una connessione di rete o un file scaricato, è in grado di copiarsi più volte e diffondersi attraverso la connessione di rete
- Poiché ogni copia successiva di un worm può a sua volta auto-replicarsi, le infezioni si possono diffondere molto rapidamente tra le reti
- Il risultato è un degrado delle prestazioni delle macchine infettate, nonché un sovraccarico dell'intera rete
- *Take home message*: non collegatevi a siti non attendibili

- L'idea di un programma auto-replicante fu teorizzata per la prima volta da John von Neumann nel 1949
- Il primo codice in grado di auto-replicarsi e diffondersi si ebbe solo nel 1971 con *Creeper*, considerato il primo worm della storia, scritto per diffondersi su Arpanet
- *Morris* viene considerato il primo worm dell'epoca moderna (1988), distribuito via Internet
  - Creato da uno studente della Cornell University, USA, e poi diffuso dai laboratori del MIT
  - Inconsapevolmente, scatena un attacco di tipo denial of service sull'intera rete Internet, causando danni per milioni di dollari
  - Robert Tappan Morris diventa la prima persona ad essere condannata negli USA per pirateria informatica
  - Nel 2006, diventa professore al MIT

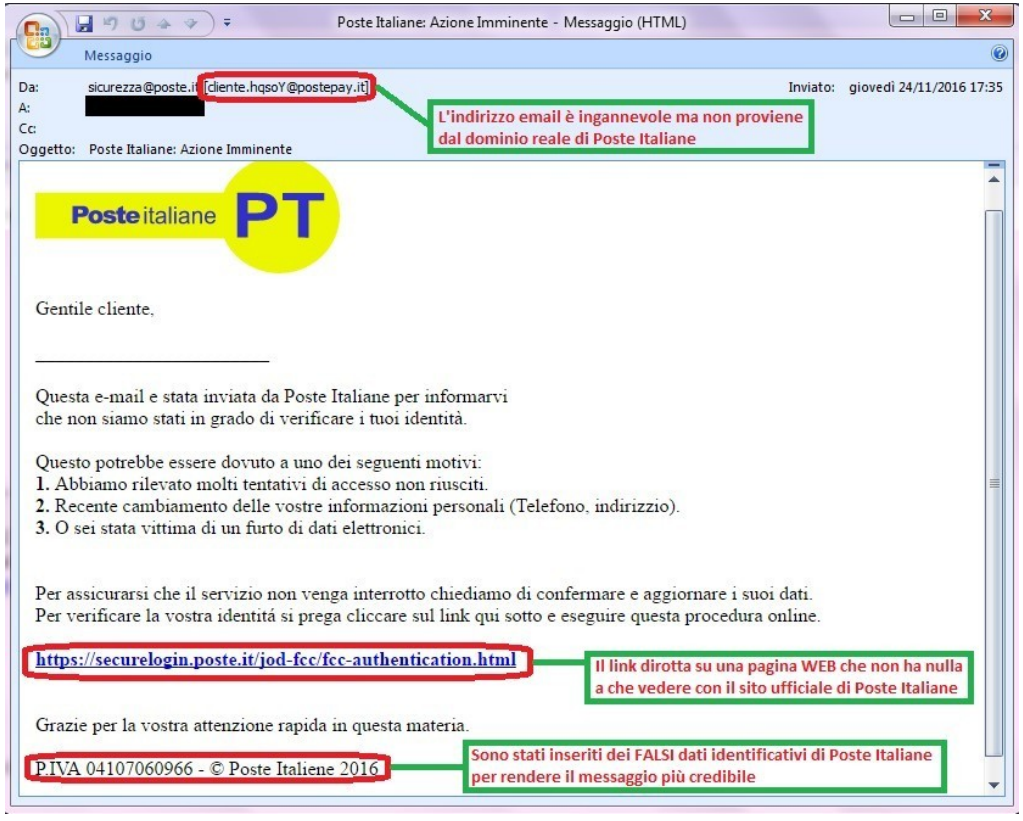
- Un trojan è un programma nocivo che penetra in un sistema camuffandosi da programma desiderato dall'utente, come un videogame
- Una volta nel sistema, il trojan può rimanere dormiente fino al verificarsi di una condizione di attivazione, ad esempio basata su una determinata data e ora
- Una volta attivato, il trojan inizia a svolgere attività dannose, come aprire le aprire le porte ad altre infezioni
- A differenza dei virus e dei worm, i trojan non sono in grado di replicarsi
- *Take home message*: il sistema potrebbe essere infetto anche in assenza di sintomi; eseguite regolari scansioni antivirus

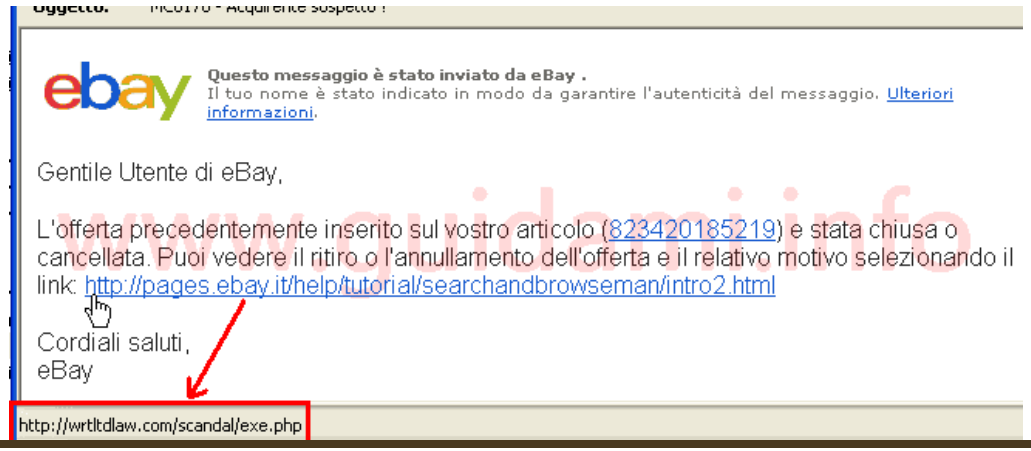
- Un'altra famiglia di software nocivi è data dagli spyware
- Obiettivo degli spyware è collezionare informazioni riguardanti i sistemi infettati
- Ad esempio, possono cercare di recuperare dati relativi all'utente del sistema con il fine di delinearne un profilo di interesse commerciale
- In altri casi, l'attacco è più subdolo, come la registrazione dei pulsanti digitati dall'utente sulla tastiera, con l'obiettivo di carpire password e numeri di carte di credito
- Infine, le informazioni raccolte vengono inviate all'attaccante

Aspetti relativi alla sicurezza
Phishing

- Il termine phishing non si riferisce ad un software malevolo, ma ad una tecnica utilizzata per ingannare l'utente affinché esso comunichi spontaneamente informazioni riservate all'attaccante
- Un esempio è dato da email accuratamente camuffate in modo da sembrare a prima vista lecite, ed in grado di superare i controlli anti-spam
- Ad esempio, un utente potrebbe ricevere una falsa email da parte di un servizio di streaming al quale è abbonato, in cui viene invitato ad accedere al proprio account per aggiornare determinate impostazioni. Cliccando su un link, l'utente viene invece indirizzato su un sito malevolo
- *Take home message*: verificate accuratamente il mittente delle e-mail; se in dubbio, esaminate gli indirizzi dei collegamenti riportati in esse prima di cliccarli

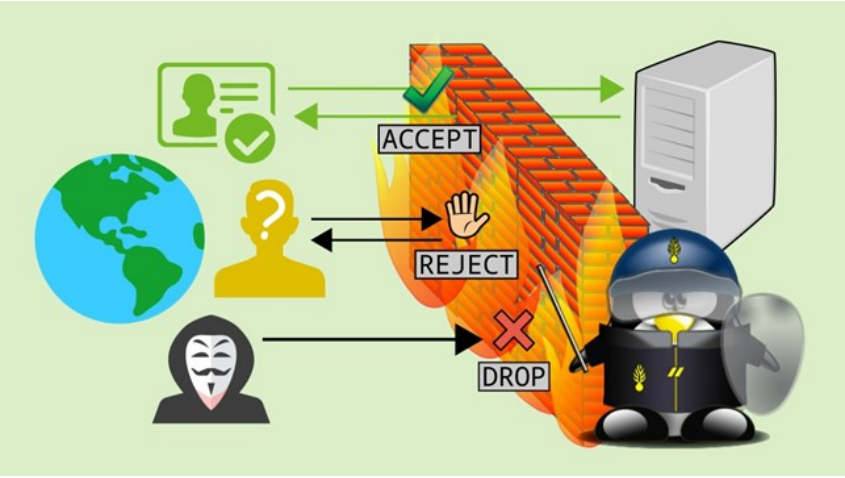
Phishing
Esempi





- Un attacco DoS mira a sovraccaricare una macchina o un'intera rete tramite l'invio di un elevato volume di messaggi non desiderati
- In numero sufficiente, tali messaggi possono portare ad un rallentamento o ad un totale blocco dei servizi forniti dalla macchina o rete oggetto dell'attacco
- La fase di attacco DoS è in genere preceduta dall'infezione di diversi sistemi con software dormienti i quali, risvegliati nello stesso istante, iniziano a sommergere di messaggi l'obiettivo
- L'insieme di dispositivi (possibilmente ignari) che prendono parte all'attacco viene detto **botnet**

- Una tecnica primaria di prevenzione degli attacchi è filtrare il traffico che attraversa un determinato punto della rete, utilizzando un software detto **firewall**
- Ad es., un firewall potrebbe essere installato all'interno di un gateway che connette una rete aziendale ad Internet
- Compito del firewall è bloccare i messaggi uscenti destinati a specifici indirizzi, o i messaggi entranti provenienti da origini non affidabili



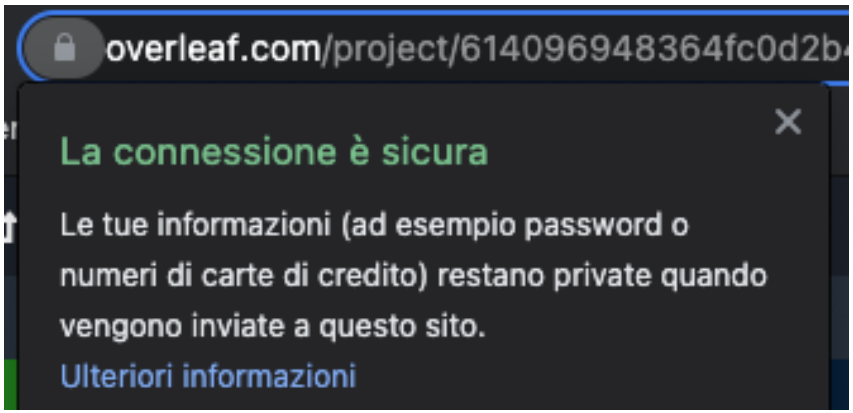
- I filtri antispam possono essere considerati come dei firewall specializzati, il cui compito è distinguere le e-mail desiderate/lecite da quelle indesiderate/fraudolente

- Le tecniche utilizzate dai filtri antispam possono essere molto sofisticate e basate, ad esempio, sull'analisi del testo e su strumenti di apprendimento automatico (machine learning)

- All'interno di un singolo computer, è possibile garantire la sicurezza delle informazioni ad esempio limitando l'accesso al dispositivo tramite l'uso di password
- Tale sistema ha un utilità limitata nel caso in cui i file vengano scambiati attraverso la rete
- La **crittografia** consente lo scambio sicuro di messaggi attraverso la rete
- Anche se intercettati, i messaggi crittografati non possono essere letti dall'attaccante
- Ad oggi, diverse applicazioni Internet tradizionali hanno incorporato tecniche di crittografia, dando vita alle loro varianti cosiddette "sicure"

Aspetti relativi alla sicurezza
HTTPS

- Un esempio di primaria importanza è la versione sicura di HTTP, nota come HTTPS
- HTTPS viene utilizzato, ad es., nel caso dell'online banking
- L'uso di HTTPS viene tipicamente indicato da un piccolo lucchetto sulla barra degli URL
- Si confrontino, ad esempio:
  - `http://www.ardiss.fvg.it/`
  - `https://www.uniud.it/it`



- HTTPS si basa sul protocollo denominato **TLS** (Transport Layer Security), preceduto da SSL (Secure Sockets Layer)
- Si tratta di un protocollo basato sul cosiddetto sistema di *crittografia a chiave pubblica*
- La crittografia a chiave pubblica prevede l'utilizzo di due chiavi:
  - *chiave pubblica*, utilizzata per crittografare i messaggi
  - *chive privata*, necessaria per decifrare i messaggi
- Ciascun soggetto coinvolto nella comunicazione genera una tale coppia di chiavi (es., tramite algoritmo RSA); condivide poi la chiave pubblica con gli altri soggetti, mentre conserva in un luogo sicuro la chiave privata

Bob & Alice

Comunicazione tramite chiavi

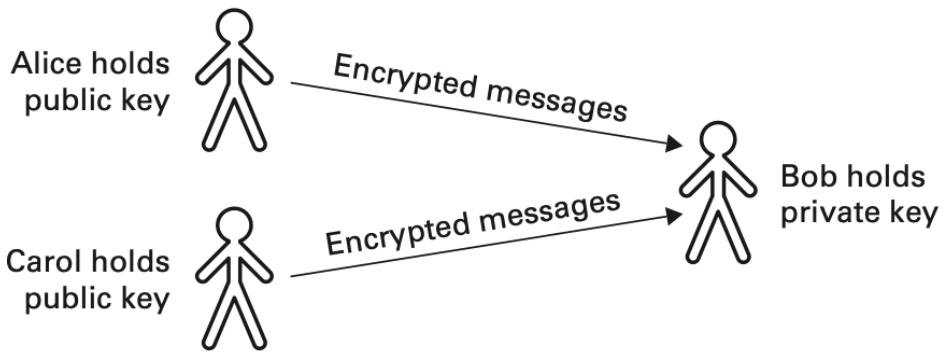
Bob desidera comunicare con il sito web della sua banca in modo riservato:

1. Il server che ospita il sito web della banca comunica la propria identità a Bob inviando il proprio *certificato* di chiave pubblica garantito da un'autorità di certificazione
2. Il browser di Bob verifica che l'URL al quale è collegato corrisponda all'identità contenuta nel certificato
3. Il browser, utilizzando la chiave pubblica del server, invia in modo sicuro la chiave pubblica di Bob al server insieme, ad es., ai suoi username e password per accedere al conto corrente
4. I successivi scambi di messaggi possono avvenire privatamente utilizzando le due chiavi

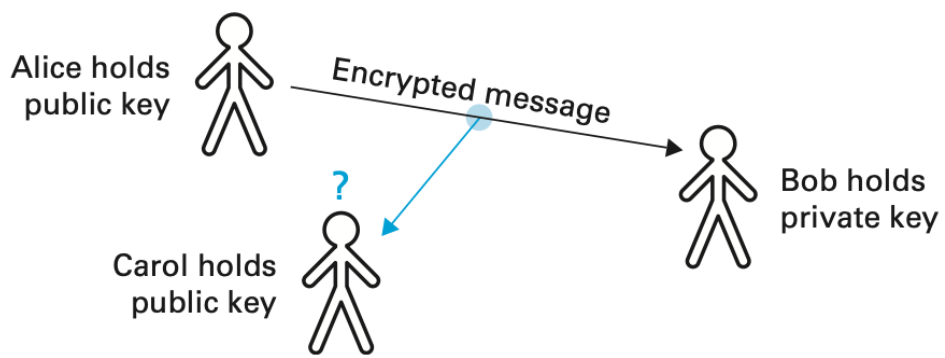
Non condividere mai la chiave privata!

- Il concetto fondamentale è che un potenziale attaccante non è in grado di decifrare i messaggi scambiati dalle due controparti, senza possedere le relative chiavi private
- Questo è vero anche nel caso in cui l'attaccante conosca le chiavi pubbliche e l'algoritmo utilizzato per cifrare i messaggi
- Tuttavia, è fondamentale che le chiavi private non cadano in mano a terzi

Comunicazione tramite chiavi, recap



Both Alice and Carol can send encrypted messages to Bob.



Carol cannot decrypt Alice's message even though she knows how Alice encrypted it.

Autenticazione dei messaggi

- Gli enti di certificazione giocano un ruolo fondamentale, in quanto assicurano che la chiave pubblica ricevuta da Bob sia effettivamente della banca (e non di un malintenzionato che si camuffa come tale)
- Si osservi che, grazie agli enti di certificazione, il sistema di crittografia a chiave pubblica può essere utilizzato anche ai fini di autenticazione
- L'idea è quella di scambiare il ruolo delle due chiavi, i.e., crittografare con la chiave privata, e decifrare con la chiave pubblica

Comunicazione fra Bob e Alice

- Supponiamo che Bob voglia inviare un messaggio ad Alice, in forma privata, e che sia già in possesso della chiave pubblica di Alice
- Bob cifra il messaggio tramite la propria chiave privata
- In seguito, Bob invia ad Alice il messaggio cifrato, assieme al proprio certificato contenente la sua chiave pubblica, cifrando a sua volta il tutto con la chiave pubblica di Alice
- Alice riceve il messaggio cifrato e lo decifra con la propria chiave privata
- Alice decifra poi il messaggio in esso contenuto con la chiave pubblica di Bob, fornita assieme al certificato

Osservazioni

- Così facendo, Alice è certo dell'identità di Bob e dell'autenticità del documento da esso inviato
- Infatti, dal momento che unicamente Bob ha accesso alla propria chiave privata, solamente lui potrebbe essere stato in grado di cifrare il messaggio inviato
- Il tutto si basa sulla presenza dell'ente certificatore, che garantisce che la chiave pubblica in possesso di Alice sia effettivamente quella di Bob
- Meccanismo alla base della **firma digitale**