

Informatica Grafica
Corso di Laurea in Ingegneria Edile – Architettura

Internet

Michele Lombardi (su materiale originario di Paolo Torroni)

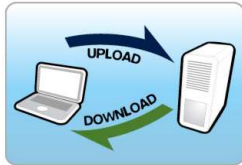
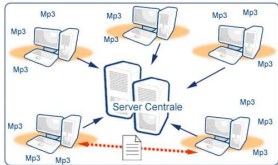
Dipartimento di Elettronica, Informatica e Sistemistica (DEIS)
Università degli Studi di Bologna

Anno Accademico 2010/2011

Un po' di dati personali:

- ▶ Michele Lombardi
- ▶ `michele.lombardi2@unibo.it`
- ▶ Ufficio: via Nosadella 51/A, SWIMM – tel. 051 20 93322
- ▶ Gruppo di Intelligenza Artificiale a UNIBO: `ai.unibo.it/`
- ▶ Home page: `ai.unibo.it/people/MicheleLombardi,`
`www.lompa.it/michelelombardi/`
- ▶ Ricevimento: previo appuntamento su email (firmata con nome e cognome)

Internet



► Internet

- I. Cos'è Internet? Com'è organizzata? Come funziona?
- II. Storia di Internet, dai primi anni ai giorni nostri. Quanto è grande il Web? Internet 2.
- III. World Wide Web: browser e ipertesti, HTTP(S), cookie (🖨️ Firefox)
- IV. Altri servizi di Internet: e-mail, collegamento remoto, newsgroup, chat, P2P.
- V. Aspetti di sicurezza su Internet

Parte I

Struttura di Internet

Reti di calcolatori

- ▶ Una delle più grandi innovazioni dell'informatica è stata la possibilità di collegare tra loro più calcolatori
 - ▶ **calcolatori singoli** collegati a risorse (es: stampanti)
 - ▶ **rete di calcolatori**: più calcolatori collegati tra loro mediante apposite infrastrutture di comunicazione (cavi, radio, ecc.)
- ▶ Esistono moltissime tipologie di reti:
 - ▶ **dimensioni**: LAN, MAN, WAN
 - ▶ **supporto di telecomunicazione**: Ethernet, coassiale, doppino, fibra, wireless, etc
 - ▶ ...
- ▶ Una rete coinvolge **sistemi eterogenei** \Rightarrow necessaria una convenzione (**protocollo**)
- ▶ Reti e protocolli permettono la realizzazione di **servizi**
 - ▶ browsing di pagine web, email, chat...

Internet: presentazioni

- ▶ **“Internet”** = rete di reti.
 - ▶ **internet** (per antonomasia) è la rete planetaria di tutte le reti collegate tra loro e che comunicano attraverso la coppia di protocolli TCP/IP.
 - ▶ Internet rende possibile la trasmissione e condivisione globale di informazione.
- ▶ **World Wide Web** = la principale *architettura informativa* basata su Internet
 - ▶ Un insieme di **ipertesti collegati tra loro** e che risiedono su nodi fisicamente diversi e distanti tra loro.

Trasmissione su rete

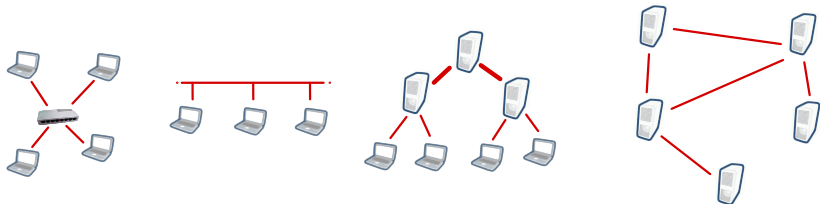
(Ovvia) Premessa: Perché due dispositivi comunichino, deve esistere un **collegamento fisico**

Con due dispositivi?

- ▶ Cavo ethernet, wifi, connessione bluetooth, fibra ottica ...
- ▶ Tutto dipende dal *supporto*

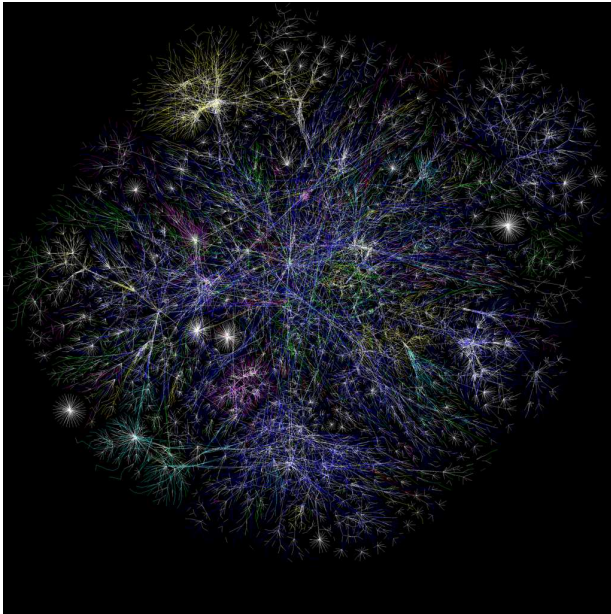
Con più dispositivi? È necessaria una **struttura** (topologia):

- ▶ stella, bus, albero, mesh...



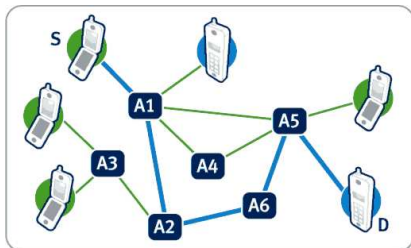
Sono equivalenti? Riuscite a vedere vantaggi e svantaggi?

Internet: quale topologia?



Commutazione di circuito

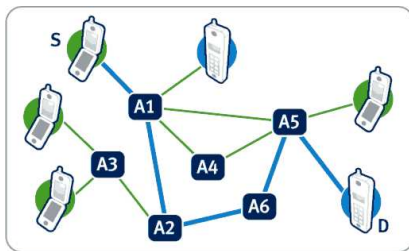
Come possono dialogare due dispositivi S e D in una rete multinodo?



Comunicazione di circuito: alcuni canali vengono *dedicati* alla comunicazione:

- ▶ modalità di connessione della normale telefonia vocale:
- ▶ le risorse (canali di comunicazione, interruttori, ripetitori, ecc.) che si trovano sul percorso $S, A1, A2, \dots, An, D$ sono **assegnate alla connessione** tra S e D
- ▶ non sono disponibili per altri, fino a quando S e D non le rilasciano (“è occupato”)

Problemi della commutazione di circuito



Inadatta alla comunicazione tra due calcolatori. Motivi:

1. *tempo per realizzare la connessione* tra S e D
2. utilizzo delle risorse
3. impatto di eventuali guasti al circuito

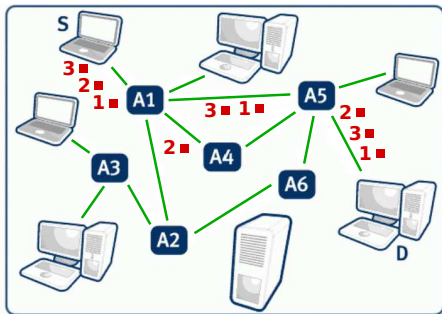
Comunicazione di pacchetto

Una soluzione alternativa: commutazione di pacchetto

1. Il nodo sorgente (S) suddivide il messaggio in **pacchetti**
2. S marca ogni pacchetto con:
 - ▶ la propria **firma**,
 - ▶ il **numero d'ordine** all'interno del messaggio,
 - ▶ l'**indirizzo** del destinatario (D) sulla rete.
3. S invia ciascuno dei pacchetti a uno dei calcolatori a cui è direttamente collegato.
4. Se questi non è il destinatario finale, quando riceve il pacchetto lo **inoltra** ad uno dei suoi vicini (*routing*);
 - ▶ il comportamento si ripete finché i vari pacchetti non raggiungono D .

Comunicazione tipicamente **best effort**

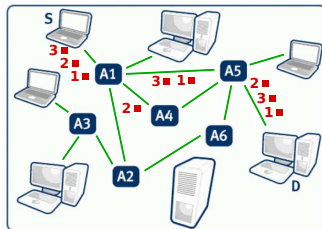
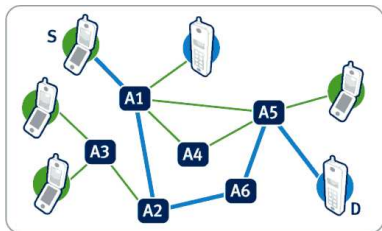
Una soluzione alternativa (commutazione di pacchetto)



Quali vantaggi?

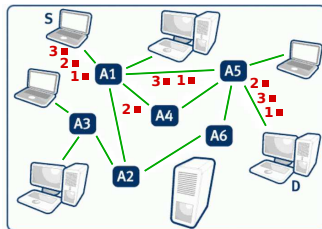
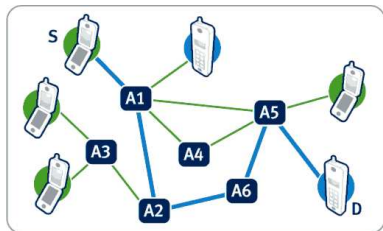
1. tempo per realizzare la connessione tra *S* e *D*: la trasmissione inizia immediatamente
2. utilizzo delle risorse: ogni pacchetto utilizza **tutta la banda di comunicazione disponibile** in quel momento (a **solo** in quel momento)
3. impatto di eventuali guasti al circuito: uso di percorsi alternativi, la comunicazione non si interrompe

Alcune caratteristiche della commutazione di pacchetto



- ▶ ogni pacchetto che arriva su un nodo viene memorizzato prima di essere ritrasmesso (meccanismo **store-and-forward**);
- ▶ molti pacchetti su una specifica connessione possono causare **congestione**: accodamento in attesa di usare la connessione;
- ▶ i pacchetti di uno stesso messaggio possono
 - ▶ seguire **percorsi diversi**;
 - ▶ arrivare a destinazione in **ordine arbitrario**.
- ▶ Via via che *D* riceve i pacchetti,
 - ▶ li **rimette in ordine**,
 - ▶ se necessario, richiede la **ritrasmissione** dei pacchetti persi.
 - ▶ scarta eventuali **duplicati**

Alcune domande sulla commutazione di pacchetto



- ▶ come decide un nodo a quale vicino trasmettere un pacchetto?
- ▶ come viene indicato l'indirizzo che identifica D in modo univoco?
- ▶ come conosce S questo indirizzo?
- ▶ come vengono gestiti gli errori durante la trasmissione?
- ▶ come viene gestita la congestione?
- ▶ chi si occupa di tutti questi dettagli?

Protocolli di comunicazione

Come risolvere i problemi indicati viene specificato da un **protocollo**.

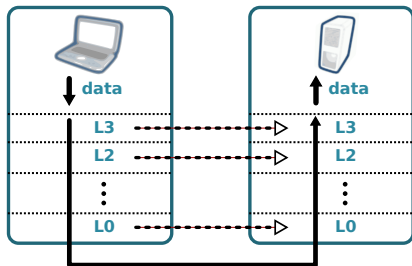
Un protocollo:

- ▶ è un accordo tra due parti sulle **modalità** con cui la loro **comunicazione** deve procedere:
- ▶ insieme di **regole** che definiscono
 - ▶ il **formato** e l'**ordine** delle **comunicazioni** spedite tra le due parti,
 - ▶ le **azioni da compiere** al momento della **trasmissione** e del **ricevimento** della comunicazione

Protocolli e livelli

Si utilizzano più protocolli, organizzati in una **pila (stack)**:

- ▶ ogni livello i della pila affronta un insieme specifico di problemi
- ▶ sfruttando le funzionalità offerte dal livello $i - 1$
- ▶ livelli non adiacenti **non comunicano**



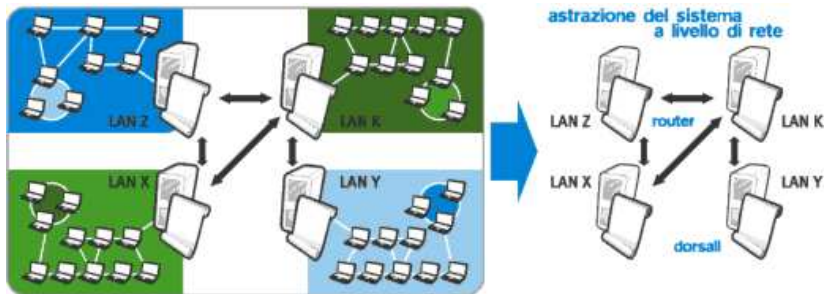
- ▶ La comunicazione avviene (concettualmente) allo stesso livello
- ▶ Ma in realtà attraversa tutti i livelli
- ▶ In ogni nodo sono presenti tutti i livelli

I livelli dei protocolli di Internet

Internet è progettata su cinque livelli

- ▶ **Livello applicazione** Richiesta, scambio e ricezione dati.
 - ▶ **HTTP** (scambio di servizi sul WWW),
 - ▶ **SMTP** (posta elettronica), ecc.
- ▶ **Livello trasporto** Impacchettamento dati del livello applicazione e loro invio sulla rete; ricostituzione del messaggio, richiesta di rinvio pacchetti perduti.
 - ▶ **TCP (Transmission Control Protocol)**.
- ▶ **Livello rete** Instradamento (*routing*) dei messaggi (a quale nodo inoltrare un pacchetto?)
 - ▶ **IP (Internet Protocol)**.
- ▶ **Livello data-link** (MAC-LLC) Accesso al mezzo di comunicazione, usando un apposita codifica dei dati digitali.
- ▶ **Livello fisico** Voltaggio e durata dei segnali, ecc.

Una rete di reti eterogenee



Il meccanismo dei livelli, consente di collegare reti eterogenee:

- ▶ eleggere in ciascuna rete un nodo rappresentante (**router**),
- ▶ collegare i router con un canale di comunicazione,
- ▶ far passare dai router tutte le comunicazioni tra le due reti.

Lo stesso meccanismo per isolare una **rete TCP/IP privata (intranet)**:

- ▶ un software specifico (**firewall**) sul router filtra il traffico in ingresso ed in uscita

Una rete di reti eterogenee

- ▶ La comunicazione tra router segue un protocollo comune.
- ▶ I router gestiscono *ai livelli inferiori* le differenze tra reti:
 - ▶ velocità di trasmissione,
 - ▶ dimensione dei pacchetti,
 - ▶ condizioni d'errore, ecc.
- ▶ Ciascun router vede un'altra rete come costituita soltanto dal suo router.
- ▶ Non occorre sincronizzare le reti (commutazione di pacchetto)
- ▶ È una soluzione **scalabile**:
 - ▶ per aggiungere una nuova rete basta poter indirizzare un pacchetto verso un nodo della nuova rete
 - ▶ il destinatario (non la rete) ha la responsabilità di rimettere insieme il messaggio e scartare i pacchetti duplicati

Come identificare un nodo su Internet?

- ▶ I nodi delle reti sono individuati da *identifier univoci* (numeri), detti **indirizzi**, assegnati come segue:
 - ▶ l'universo Internet è suddiviso in **reti fisiche**;
 - ▶ ad ogni rete fisica è assegnato **in modo centralizzato** un certo indirizzo (es. 137);
 - ▶ se questa è divisa in **sottoreti**, gli indirizzi ai suoi nodi vengono assegnati **in modo gerarchico**, mediante concatenazione
 - ▶ Ad es 137::204;
 - ▶ infine, l'indirizzo di ciascun nodo è la concatenazione dell'indirizzo della sua sottorete con un numero che lo individua in modo **univoco** nella sottorete
 - ▶ Ad es: 137::204::6132
- ▶ I numeri ai nodi delle reti fisiche, o alle sottoreti, sono **assegnati dai gestori** delle reti stesse.
- ▶ La gestione degli indirizzi è di pertinenza del **livello di rete**, quindi del **protocollo IP**

Indirizzi IP

- ▶ Un indirizzo IP è una sequenza di 4 numeri decimali $\in [0..255]$, separati da un punto. Es: 137.204.24.12
 - ▶ Indirizzo della rete fisica: 137.204
 - ▶ Identificativo della sottorete: 24
 - ▶ Identificativo del nodo: 12
 - ▶ (... a chi appartiene questo indirizzo? ;-))
- ▶ I numeri alle **reti fisiche** vengono assegnati dalla **Internet Assigned Number Authority (IANA)**, www.iana.org
- ▶ IANA delega analoghi **organismi regionali** all'assegnamento dei numeri IP all'interno delle relative zone geografiche
- ▶ Per l'Europa: RIPE NCC (Réseaux IP Européens Network Coordination Centre), www.ripe.net
- ▶ I gestori delle singole reti fisiche (es: CeSIA) assegnano i numeri ai loro nodi.

Indirizzi simbolici di dominio (naming)

- ▶ Un indirizzo IP identifica univocamente un dispositivo fisico in Internet, ai fini del protocollo IP.
- ▶ Per l'uso umano, è tutt'altro che pratico
- ▶ Si usano degli **indirizzi simbolici**, o **nomi logici**: caratteri invece di numeri. Es: `www.myspace.com`, `mail.unibo.it`
- ▶ *L'insieme* e la *struttura* di questi nomi costituiscono il **Domain Name System**, o **DNS**, di Internet.

Corrispondenza tra nomi logici ed indirizzi:

- ▶ Ad un nome logico corrisponde **un unico indirizzo IP** (non esattamente vero...)
- ▶ Ad un indirizzo IP possono corrispondere **più nomi logici**
- ▶ **Attenzione**: la struttura dei nomi logici non ha **nulla a che vedere** con la gerarchia degli indirizzi IP
 - ▶ `www.miur.it` → `193.206.6.24`
 - ▶ `pop.cs.unibo.it` → `130.136.1.110`
 - ▶ `ing.unibo.it` → `137.204.24.12`

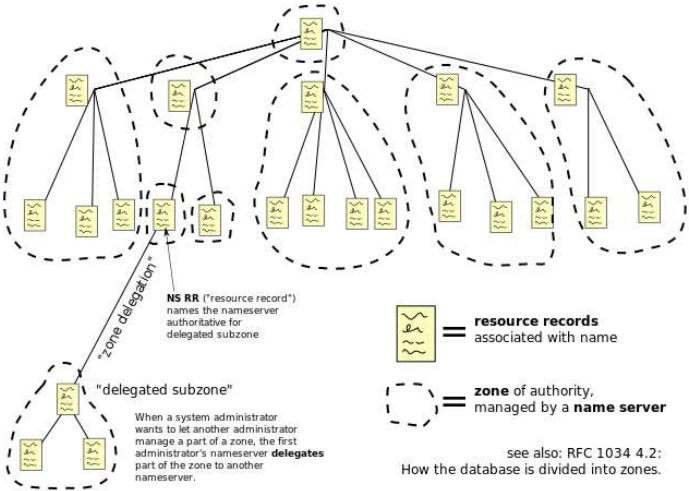
Risoluzione di nomi logici e indirizzi IP

Domain Name *Space*

- ▶ L'ultima parte del nome logico identifica il **Top Level Domain** (dominio di primo livello): es. `it`, `com`
- ▶ La penultima identifica il dominio di secondo livello, etc.
- ▶ Non esistono limiti al numero di livelli di un nome logico (es. `www.lia.deis.unibo.it`)
- ▶ Alcuni nodi della rete (**DNS**, Domain Name Server) si occupano di tradurre i nomi logici in indirizzi IP
 - ▶ Quindi per ottenere l'indirizzo IP di `www.myspace.com` basta sapere l'indirizzo IP di un DNS che lo conosce
- ▶ Se un DNS non conosce l'indirizzo IP di un determinato nodo, inoltra la richiesta a un altro DNS, via via fino a uno dei **DNS root server**

Risoluzione di nomi logici

Domain Name Space



DNS root server



- ▶ Ci sono molti DNS root server sparsi un po' in tutto il mondo (130 posizioni in 53 paesi a settembre 2007)
- ▶ Si tratta di un servizio fornito da organizzazioni, selezionate dallo IANA, che si assumono l'obbligo di fornirlo

Corrispondenza tra indirizzi IP e nomi logici

- ▶ Ad un indirizzo IP possono corrispondere **più nomi logici**
- ▶ Non è sempre possibile ottenere informazioni significative da un semplice indirizzo IP
- ▶ Esistono però dei servizi di *lookup* (**WHOIS**)
 - ▶ RIPE, <http://www.db.ripe.net/whois>
 - ▶ APNIC, www.apnic.net/apnic-bin/whois.pl
 - ▶ ARIN, www.arin.net/whois/
 - ▶ Network Solutions,
www.networksolutions.com/en_US/whois
- ▶ **Reverse DNS lookup**
 - ▶ <http://remote.12dt.com/>
- ▶ **Localizzazione geografica**
 - ▶ <http://www.infosniper.net/>

Chi assegna i nomi logici di dominio?

- ▶ Domini di primo livello: esiste un organismo internazionale indipendente che sovrintende alla loro ripartizione e definizione
 - ▶ **Internet Corporation for Assigned Names and Numbers**
`www.icann.org`
 - ▶ Primi domini di primo livello:
 - ▶ `edu` (educational: università e scuole), `com` (commerciale), `mil` (militare), `gov` (governativo), `int` (internazionale), `net` (fornitori di connettività).
 - ▶ Successivamente:
 - ▶ Domini nazionali e regionali, `it`, `fr`, `jp`, `eu`, ...
 - ▶ Altri domini: `biz`, `coop`, `museum`, `name`, `org`, ...
- ▶ Ogni dominio di primo livello ha un organismo di gestione
 - ▶ **Registration Authority** per il dominio `it`: `www.nic.it`
- ▶ I domini di terzo livello sono assegnati dal titolare del dominio di secondo livello, etc.

Dynamic Host Configuration Protocol (DHCP)

- ▶ In sostanza per connettersi ad internet occorre:
 1. un **collegamento fisico** (rete ethernet, wireless...)
 2. un **indirizzo IP**
 3. l'**indirizzo IP del router** a cui inviare i pacchetti
 4. l'**indirizzo IP di un server DNS**
- ▶ Ma nessuno di voi (probabilmente) immette a mano tutti questi dati!
- ▶ Dynamic Host Configuration Protocol (DHCP):
 - ▶ quando un dispositivo viene connesso fisicamente alla rete
 - ▶ cerca nelle immediate vicinanze un **server DHCP**
 - ▶ se ne trova, richiede **automaticamente** al server i dati necessari alla sua configurazione

Riassumendo. . .

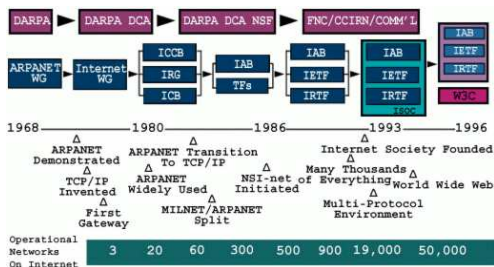
- ▶ Con “*Internet*” si indica il sistema informativo globale che:
 - (i) è logicamente connesso mediante **un unico spazio globale di indirizzi** basato sul protocollo IP o sulle sue estensioni;
 - (ii) permette di **supportare le comunicazioni** utilizzando la coppia di protocolli **TCP/IP** o le sue estensioni e/o altri protocolli compatibili con IP;
 - (iii) fornisce, utilizza o rende accessibili, in modo pubblico o privato, **servizi ad alto livello** sfruttando i livelli di comunicazione e le infrastrutture che sono stati descritti ai punti precedenti.

Federal Networking Council, 24 ottobre 1995

Parte II

Un po' di storia

Breve storia di Internet: i primi anni



1968 DARPA (Defense Advanced Research Projects Agency) incarica BBN (Bolt, Beranek & Newman) di creare ARPAnet

1970 Primi 5 nodi: UCLA, Stanford, UC Santa Barbara, Utah University, e BBN

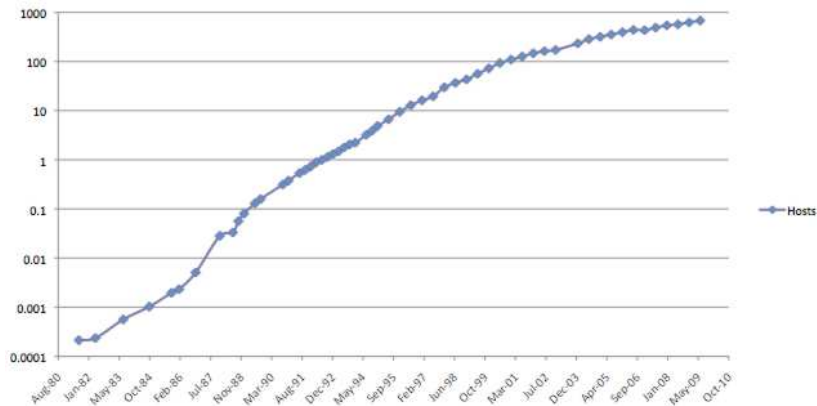
1974 Specifiche del protocollo TCP (Vint Cerf)

1984 Internet (1000 nodi) adotta TCP/IP in massa

1995 Apertura di Internet a scopi commerciali

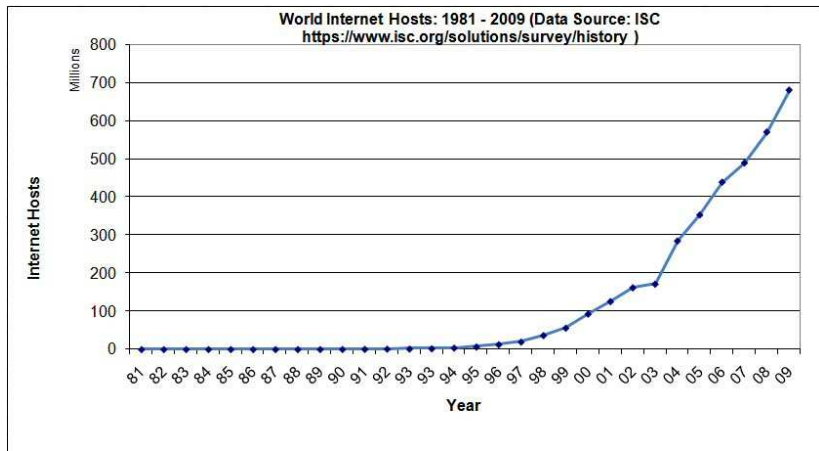
► Per saperne di più: <http://www.isoc.org/internet/history/>

Quanto è grande Internet



- ▶ Grafico in **scala logaritmica** (milioni di nodi)
- ▶ Fonte: Internet Systems Consortium, Internet Host Count History <https://www.isc.org/solutions/survey/history>

Quanto è grande Internet



- ▶ Grafico in **scala lineare** (milioni di nodi)
- ▶ Fonte: Internet Systems Consortium, Internet Host Count History <https://www.isc.org/solutions/survey/history>

Le scelte cruciali

- ▶ Idee chiave nel progetto di TCP/IP (Robert Kahn):
 1. **nessuna modifica interna** deve essere richiesta per collegarsi a Internet;
 2. **best effort**: se un pacchetto non raggiunge la destinazione, sta al mittente originario (non ad altri nodi intermedi) ritrasmetterlo;
 3. **semplicità dei router**, che non devono mantenere alcuna informazione sul flusso dei singoli pacchetti attraverso di essi
 4. **nessun controllo globale della rete** al livello della sua operatività.

Un bilancio storico

- ▶ Internet ha sempre coinvolto insieme ricerca pubblica e iniziativa privata.
- ▶ Allo stesso tempo, è sempre stata una **rete non proprietaria**, le cui **specifiche e caratteristiche** erano e sono **pienamente disponibili e utilizzabili da chiunque** intenda investirvi.
- ▶ Aspetti chiave nella storia di Internet:
 1. **tecnologia**: commutazione di pacchetto mediante protocolli in continuo miglioramento;
 2. **gestione operativa e strategica** di una infrastruttura globale complessa;
 3. **aspetti sociali**: vasta comunità di persone che sviluppano la tecnologia anche grazie alla rete stessa;
 4. **aspetto economico**: transizione da struttura di ricerca ad infrastruttura informativa di vasta scala.

Le prospettive: Internet2

- ▶ Internet progettato per collegare grandi calcolatori e fornire semplici servizi a migliaia di utenti.
- ▶ Oggi: cambiati i numeri e la tipologia di utenti (wireless) e il tipo e qualità dei servizi offerti.
- ▶ Problemi:
 1. infrastruttura di comunicazione lenta e troppo sfruttata;
 2. non gestisce problematiche di sicurezza, identità, autorizzazione e autenticazione degli utenti;
 3. protocolli hanno raggiunto i loro limiti fisici (indirizzi IP disponibili... 4,228,250,625?!?);
 4. applicazioni limitate dalla tecnologia di rete invece che ruolo guida per lo sviluppo di tecnologie innovative.
- ▶ Consorzio Internet2 (<http://www.internet2.edu>), IPv6 (Internet Engineering Task Force, <http://www.ietf.org/>)

Parte III

Servizi di Internet

- ▶ **Protocollo** insieme di regole e di convenzioni da rispettare perché due calcolatori possano comunicare.
- ▶ **Servizio**: prestazione fornita da una macchina servente (server) a una macchina cliente (client) che ne fa richiesta.
- ▶ **Applicazione**: programma.
- ▶ **Attenzione**: a volte uno stesso termine può essere usato per indicare un protocollo, un servizio, un'applicazione
 - ▶ Esempio: FTP
- ▶ Tipicamente, un servizio si basa su un *protocollo di livello applicativo* e richiede l'utilizzo di uno o più *programmi*
 - ▶ un browser web è un programma
 - ▶ un server web è un *programma* (di fatto il termine si usa anche per indicare la macchina su cui esegue)
 - ▶ Skype è un programma

Modelli di interazione

Esistono due modelli di interazione/comunicazione tra calcolatori su Internet che permettono di fornire un servizio.

- ▶ **Client/server**: quando un computer (client) ha bisogno di un servizio:
 - ▶ il client *chiede* ad un altro computer (server) di fornire il servizio in questione
 - ▶ il server *esegue* l'elaborazione necessaria per fornire il servizio ed *invia* la risposta
 - ▶ asimmetria tra client e server
 - ▶ Esempio: navigazione Web.
- ▶ **Peer-to-Peer**: i nodi della rete che comunicano/collaborano giocano ruoli interscambiabili e svolgono le stesse funzioni
 - ▶ comunicazione diretta tra nodi, senza intermediari
 - ▶ i nodi sono simmetrici
 - ▶ Esempio: Napster, BitTorrent

Parte IV

Un servizio: World Wide Web

Ipertesti: la storia

1945 Vannevar Bush, MIT: idea di organizzare la conoscenza su base reticolare e associativa, invece che sequenziale

1965 Ted H. Nelson usa per la prima volta la parola “ipertesto”

un corpus di materiali scritti o grafici interconnessi in un modo così complesso da non poter essere ragionevolmente presentato o rappresentato su carta. Può contenere sommari, o schemi dei suoi stessi contenuti e delle loro relazioni reciproche; può contenere annotazioni, aggiunte e note [...] Un tale sistema potrebbe crescere senza limiti, inglobando gradualmente una parte sempre più ampia della conoscenza scritta esistente al mondo.

1962 Doug Engelbart, SRI, inventa il mouse.

- ▶ Ancora nessuno aveva pensato di sfruttare Internet come supporto.

Un ipertesto ;-)

Home - WEBROMANZO BINACA - Google Chrome

Home - WEBROMANZ... x

www.webromanzo.it

PRIMAPAGINA E-MAIL

MAPPA DEL SITO

PERSONAGGI	ALTRI INTERPRETI	ALTRI ANCORA
Prima pagina	Il Dottore	Dervis
Io	Il garbino	Muratori Ivan
Binaca	La bici	Il mio nonno
Il figlio di Papini	La figa	Australia
Il mio babbo	La mezzapunta	Pensione Cinzia
I genitori del figlio di Papini	Le scandinave	La morte della povera Ida
L'altro	Olimpiadi	La moglie tedesca
Il cugino del figlio di Papini	Mamma	Giornata tipo
La povera Ida	CCCP	Le due gemelline
L'Italia		La Signora
		Gli scafati



WEBROMANZO

by PIER PAOLO CEVOLI

BINACA, Di Paolo Cevoli <http://www.webromanzo.it/>

Dagli ipertesti al Web

- 1989 Tim Berners-Lee, CERN, specifica un sistema di documenti ipertestuali distribuiti, per la collaborazione di gruppi di ricerca in tutto il mondo
- ▶ Idea nuova: un documento possa essere indicato in modo univoco attraverso un *Universal Document Identifier* (oggi, URL).
 - ▶ Progetto di HTML per specificare documenti
 - ▶ Progetto del protocollo HTTP per recuperare ipertesti in Internet.
- 1993 Marc Andreessen, U. Illinois, scrive Mosaic, il primo browser multimediale
- 1994 Mosaic ha un bacino di milioni utenti nel mondo. Andreessen fonda Netscape.
- oggi 1 miliardo di persone usano il Web attraverso browser multimediali.

Ipertesti e WWW

- ▶ Iper testo: *documento che contiene al suo interno collegamenti ad altri documenti o a sezioni dello stesso documento.*
- ▶ Informazione organizzata in modo reticolare
- ▶ **World Wide Web (WWW, o Web)**: un gigantesco ipertesto multimediale distribuito dotato di un'interfaccia di facile uso
 - ▶ tecnologia informatica per realizzare link facili da usare
 - ▶ documenti composti di testo, immagini, video, audio, ecc.
 - ▶ diverse parti di un ipertesto possono risiedere su calcolatori diversi e distanti tra loro (in modo trasparente all'utente)
 - ▶ **browser**, visualizza in modo uniforme i dati multimediali e permette il raggiungimento dei dati remoti.

Fruizione del Web

- ▶ Basata su servizi client/server
 - ▶ **server**: mette a disposizione **le informazioni** (documenti ipertestuali), residenti sul server stesso
 - ▶ **client**: accede ai documenti attraverso un **indirizzo**
- ▶ Quando un client effettua una richiesta di un ipertesto a un server, il server invia al client le varie componenti che costituiscono il documento
- ▶ Principali client: Firefox, Safari, Opera, Chrome, Internet Explorer, etc.

Hyper Text Transfer Protocol (HTTP)

Protocollo di livello applicativo, permette la fruizione del WWW.

Client (es. *browser*)

- ▶ richiesta di una **risorsa**
- ▶ attesa della risorsa
- ▶ Ricezione della risposta e sua analisi
- ▶ Sono necessarie ulteriori risorse per riprodurla? es. immagini, codice javascript, file audio, etc.
- ▶ Nel caso, richiesta delle singole rimanenti risorse
- ▶ A mano a mano che i vari elementi arrivano, riproduzione della risorsa

Lato Server

- ▶ Il server HTTP sta in attesa di una comunicazione su una determinata *porta* di comunicazione
- ▶ una *porta* sta ad un IP come un *numero di interno* sta ad un indirizzo
- ▶ ogni volta riceve una richiesta da parte di un client (**URL**), la soddisfa inviando la risorsa
- ▶ abbandona la connessione e si rimette in ascolto

Come è descritto un ipertesto?

- ▶ **2 tipi di informazione:**
 - ▶ Informazione (testo)
 - ▶ la struttura (titolo, corpo del testo, link, etc.)
- ▶ **Hyper Text Markup Language (HTML):** Linguaggio per la descrizione di ipertesti.
- ▶ Informazioni sulla struttura del documento
 - ▶ Diversi browser possono visualizzare in modo diverso la struttura descritta da HTML
- ▶ Informazioni su come il documento deve essere visualizzato

Hyper Text Markup Language (HTML)

```
<HTML>
  <HEAD>
    <meta name=description content=Esemplificazione di HTML
    <title>Documento di prova</title>
  </HEAD>
  <BODY>
    <H1> Questo egrave; il titolo</H1>
    Questo documento describe <I>con un esempio</I> l'uso di
    HTML.
    <H2> Essenza di HTML </H2>
    Un documento HTML consiste di testo immerso in <b>tag</b>
    HTML. I comandi descrivono la struttura del
    documento. HTML permette:
    <UL>
      <LI> di descrivere la struttura del documento;</LI>
      <LI> inserire elementi non testuali, come questo
      stemma: <IMG src=./logoUnibo.gif ALT="logo UNIBO" >;</LI>
      <LI> inserire riferimenti ipertestuali:
      <A HREF=http://www.ing.unibo.it Facolta' di Ingegneria</A></LI>
    </UL>
  </BODY>
</HTML>
```


E questo è il risultato



Questo è il titolo

Questo documento descrive *con un esempio* l'uso di HTML.

Essenza di HTML

Un documento HTML consiste di testo "immerso" in **tag** HTML. I comandi descrivono la struttura del documento. HTML permette:

- di descrivere la struttura del documento;
- inserire elementi non testuali, come questo stemma;
- inserire riferimenti ipertestuali: [Facoltà di Ingegneria](#)

Identificazione di una risorsa sul WWW

- ▶ Ogni **risorsa** sul Web è univocamente individuata da una sequenza di caratteri che ne costituisce l'indirizzo
 - ▶ **Uniform Resource Locator (URL)**
 - ▶ `http://www.lia.deis.unibo.it/Courses/InfoGrafica/lezioni.html`
- ▶ Contenuto di un URL:
 - ▶ il **protocollo** da utilizzare per recuperare il documento
 - ▶ `http, ftp, https, news, ...`
 - ▶ l'**indirizzo** della risorsa sul Web; per `http`:
 - ▶ il **nome logico** del server su cui risiede (o il suo **indirizzo IP**)
 - ▶ opzionalmente, *username* (es. **user @nomeserver**)
 - ▶ opzionalmente, *numero di porta* (es. **nomeserver:porta**)
 - ▶ un **percorso locale** che individua uno specifico documento all'interno del server (es `/Ingegneria/default.html`)
 - ▶ Il dominio è *case insensitive*
 - ▶ Il percorso locale *può essere case sensitive* (dipende dal server)

!!! Cautela (phishing)

Contenuti non testuali

- ▶ Informazioni per il browser

```
<meta name=description
      content=Esemplificazione di HTML
<meta http-equiv=Content-Type
      content=text/html; charset=iso-8859-1
```

- ▶ Informazioni sul contenuto del documento

```
<meta name=keywords content=university
      teaching computers logics >
```

- ▶ Link a risorse multimediali e varie

- ▶ **immagini** (gif, jpg, tiff, bmp),
- ▶ **audio** (mov, wav, mp3),
- ▶ **video** (wmv, mpg, mov, rm, mpeg, avi),
- ▶ **documenti in formato proprietario** (doc, pdf, swf, etc.)

- ▶ **Codice: javascript, applet.**

Contenuti non testuali

```
<HTML>
  <HEAD>
    <meta name=description content=Esemplificazione di HTML
    <title>Documento di prova</title>
  </HEAD>
  <BODY>
    <H1> Questo egrave; il titolo</H1>
    Questo documento descrive <I>con un esempio</I> l'uso di
    HTML.
    <H2> Essenza di HTML </H2>
    Un documento HTML consiste di testo immerso in <b>tag</b>
    HTML. I comandi descrivono la struttura del
    documento. HTML permette:
    <UL>
      <LI> di descrivere la struttura del documento;</LI>
      <LI> inserire elementi non testuali, come questo
      stemma: <IMG src=./logoUnibo.gif ALT="logo UNIBO" > ;</LI>
      <LI> inserire riferimenti ipertestuali:
      <A HREF=http://www.ing.unibo.it Facolta' di Ingegneria</A></LI>
    </UL>
  </BODY>
</HTML>
```

Helper e plug-in

- ▶ Per riprodurre documenti di altro formato (es. pdf), è possibile
 - ▶ ricorrere a programmi esterni al browser (**helper**)
 - ▶ estendere le funzionalità del browser con dei componenti esterni ma incorporati nel browser (**plug-in**)
- ▶ È possibile specificare che azioni intraprendere (quale helper o plugin) per ciascun tipo di risorsa
- ▶ I tipi delle risorse sono specificati in un formato chiamato **MIME type**
- ▶ Mentre un helper resta aperto dopo la chiusura del browser, un plug-in vive in funzione delle azioni di navigazione del browser

Chi coordina il WWW?

- ▶ Il Web nasce aperto, senza formati proprietari (copyright)
- ▶ Importante garantire l'esistenza unitaria del Web e il suo carattere aperto e libero
- ▶ 1994: **World Wide Web Consortium, (W3C)**, presso il MIT (<http://www.w3.org/>)
 - ▶ Sede europea: ERCIM, <http://www.ercim.org/>
- ▶ Scopo principale: sviluppare protocolli, specifiche, software e strumenti che garantiscano l'interoperabilità del Web.
- ▶ Attenzione: qui si parla di protocolli *di livello applicativo*

Ricerca delle informazioni sul WWW

- ▶ L'accesso a risorse su Web avviene tramite una URL
- ▶ Se non si conosce la URL, bisogna avere uno strumento per **cercare** informazioni (URL) sul Web
- ▶ Caratteristiche del Web da tenere in considerazione:
 - ▶ dimensioni della rete e numero delle risorse
 - ▶ dinamica della rete e delle risorse
 - ▶ eterogeneità nelle strutture dei documenti
 - ▶ varietà dei contenuti
- ▶ Non basta un semplice “elenco telefonico”!

Motori di ricerca e directory

Motori veri e propri	Directory
Automatici	Risultato di lavoro umano
Indicizzano URL (singole risorse)	Classificano interi siti
Mirano a indicizzare tutto il Web	Mirano a diventare guide alle migliori risorse Web
Si basano su algoritmi matematici	Si basano su giudizi di valore qualitativi ed umani
Interrogabili attraverso combinazioni di parole chiave	Percorribili per categorie e sotto-categorie

- ▶ Search engine: <http://www.google.com/>
- ▶ Directory: <http://dir.yahoo.com/>

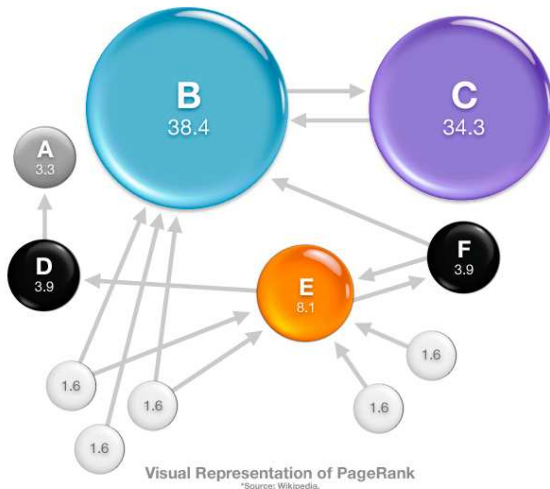
Com'è fatto un motore di ricerca?

- ▶ Un motore di ricerca si basa su diversi componenti:
- ▶ **Back end:**
 - ▶ Un programma che interroga periodicamente il Web per scaricare le pagine e catalogarle.
 - ▶ Usa un pool di **agenti autonomi** chiamati **spider (o crawler)**
 - ▶ Un **database** delle pagine catalogate.
- ▶ **Front end:**
 - ▶ Un'**interfaccia** di interrogazione (**query**).
 - ▶ Ricerche per **parole chiave**
 - ▶ Linguaggio di interrogazione: Or, And, Not, +, -, Near, "...", *

Google PageRank

- ▶ Il cuore della tecnologia di ricerca di Google è un algoritmo: **PageRank**
- ▶ PageRank assegna un voto (**ranking**) alle pagine Web
- ▶ Si basa sulla struttura ipertestuale del Web (link) come indicatore del valore di una data pagina
 - ▶ Link da A a B \Rightarrow A “vota” per B
- ▶ Considera il **numero dei voti** e il **contenuto delle pagine che esprimono questi voti**.
 - ▶ Voti che provengono da pagine più importanti hanno un peso maggiore
 - ▶ Siti Web di alta qualità ricevono un voto più alto di altri
- ▶ In dettaglio, Google stima la **probabilità di capitare su una data pagina** durante una visita casuale del Web (random surfing).

Rappresentazione grafica del PageRank



La dimensione dei nodi del grafo (URL) dipende dal numero dei link in ingresso e dalla loro provenienza.

Altri motori di ricerca

Motori di ricerca specializzati:

- ▶ Viaggi:
 - ▶ eDreams (<http://www.edreams.it/>)
 - ▶ Expedia (<http://www.expedia.it>)
- ▶ Mappe:
 - ▶ Google Maps (<http://maps.google.it>)
 - ▶ Bing Maps (<http://www.bing.com/maps/>)
- ▶ Riferimenti bibliografici:
 - ▶ Portale delle Biblioteche di UNIBO
(<http://www.cib.unibo.it/portale>)
 - ▶ Google Scholar (<http://scholar.google.com/>)

Motori embedded

- ▶ Ricerca in siti web (es: UniBo www.unibo.it, Il Corriere www.corriere.it Repubblica www.repubblica.it, etc.)