

INTRODUZIONE A INTERNET

Cos'è una rete (1)

- Una rete è una “collezione di calcolatori *autonomi* fra loro collegati”
- Vantaggi:
 - un nuovo mezzo di comunicazione
(*email, videoconferenze, ...*)
 - condivisione di risorse
(*stampanti condivise, editing collaborativo, ...*)
 - scalabilità e maggiore affidabilità: ad esempio se una stampante non è disponibile o non funziona, può essere possibile usarne un'altra

Cos'è una rete (2)

- Le reti si classificano in funzione della loro estensione:
 - LAN (Local Area Network): reti con dimensione su scala di ufficio, edificio o campus
 - MAN (Metropolitan Area Network): reti con dimensione su scala di città
 - WAN (Wide Area Network): reti su dimensione geografica
- Ogni rete ha una determinata banda trasmissiva, ovvero capacità di trasferimento di dati nell'unità di tempo: la banda si misura in numero di Byte (o bit) trasferiti al secondo.

Ripasso: 1 bit (b) è una cifra digitale (0 o 1) e rappresenta la minima quantità di informazione rappresentabile. 1 Byte (B) è costituito da 8 bit. Il prefisso K (Kilo) davanti a Byte (o bit) indica $2^{10}=1024$ Byte (o bit). Analogamente il prefisso M (Mega) indica $2^{20}=1048576$ Byte, G (Giga) indica $2^{30}=1073741824$ Bytes, T (Tera) indica 2^{40} Bytes, ecc. Di solito si dice arrotondando che K indica 1000, M indica un milione e G indica un miliardo, ma tali cifre non sono esatte come si vede sopra. Per fare un esempio del significato associato alla quantità di banda trasmissiva disponibile in una rete consideriamo il trasferimento di 1 MB di dati lungo la rete. Se la rete avesse una banda di 1 KB/s (1 KiloByte al secondo) allora il trasferimento potrebbe avvenire in non meno di $(1\text{MB}/(1\text{KB/s}))=(1\text{MB}/1\text{KB})\cdot\text{s}=1024$ secondi. Se la banda fosse stata di un MB/s allora, ovviamente sarebbe stato sufficiente 1 secondo. Se la banda fosse stata di 1GB/s, invece sarebbe bastato $(1\text{MB}/(1\text{GB/s}))=(1\text{MB}/1\text{GB})\cdot\text{s}=1/1024$ di secondo. Attenzione! 1MB/s significa che la rete trasferisce 2^{20} Byte al secondo, mentre 1Mb/s significa che la rete trasferisce 2^{20} bit al secondo, quindi 1/8 della banda precedente (di solito la banda si esprime in b/s). Dovrebbe essere chiaro quindi come la banda, tanto più è grande, tanto più è efficace nel trasferimento di informazione. Purtroppo le reti a larga banda costano di più, e molto spesso, quando il traffico deve attraversare più porzioni di rete (come su Internet) la banda risultante tra due host A e B risulta essere pari alla banda più stretta incontrata durante il trasferimento tra tutte le porzioni di rete tra A e B. Tale

restringimento della banda risultante si dice “collo di bottiglia”, e può essere causato in maniera non controllabile da varie cause (congestione, guasti, errori nel routing).

- **I calcolatori collegati in rete comunicano seguendo il modello “client-server”:** nel modello client-server la comunicazione avviene sempre tra una coppia di elementi detti appunto Client e Server: con Client si intende di solito l’elemento della coppia che attiva o richiede un servizio di comunicazione, mentre il Server rappresenta il fornitore del servizio richiesto: ad esempio un browser (web client) e un Web server sono la coppia di elementi attiva in una comunicazione Web

Un po’ di storia (1)

- **La “madre” di INTERNET è ARPANET (Advanced Research Projects Agency NET):** una rete sviluppata tra calcolatori universitari per scopi legati alla ricerca, e non con le finalità che abbiamo in mente oggi
- **Il primo prototipo venne terminato nel 1969 e collegava 4 calcolatori universitari (SRI, UCLA, UCSB, Utah):** in seguito si sono aggiunti altri host (calcolatori con funzioni di comunicazione in rete) da parte di agenzie governative, militari, e solo in seguito commerciali. La crescita e lo sviluppo di Internet sono quindi avvenute sotto la spinta dei contributi distribuiti di un numero molto grande di costituenti.
- **I calcolatori collegati a quella che viene chiamata oggi INTERNET sono già più di 20.000.000**

Un po’ di storia (2)

- **Nei suoi primi anni la rete era un esperimento informatico gestito ed in uso a soli ricercatori**
- **Negli anni ‘90 la rete assume una valenza di infrastruttura globale con implicazioni economiche e sociali su larga scala**

INTERNET: la rete delle reti (1)

- **INTERNET è la “rete delle reti”:** il termine Internet infatti viene usato quando si intende fare riferimento alla collezione di reti la cui infrastruttura globale di collegamento forma una connessione tra host su livello planetario
- **Non esiste un giorno di inaugurazione per le ragioni legate al suo sviluppo**
- **Nasce assieme alla consapevolezza che tutte le reti cresciute da ARPANET comunicavano fra loro adottando un unico protocollo di comunicazione chiamato TCP/IP:** questo protocollo, o meglio questa coppia di protocolli è ancora oggi alla base della comunicazione su Internet

INTERNET: la rete delle reti (2)

- **INTERNET è costituita da due elementi base:**
 - **Il collegamento fisico:**
 - **linee dedicate (dorsali) per la comunicazione di grandi quantità di dati nell’unità di tempo**

- router (instradatori) per lo smistamento dei dati durante il cammino fisico dei dati lungo le linee di comunicazione. Ogni router ha lo scopo primario di inviare una porzione (pacchetto) di dati proveniente da un host destinatario (source) verso l'host destinatario (destination)
- Il protocollo di comunicazione:
 - indirizzamento IP (Internet Protocol): è una tecnica di assegnamento di nomi (indirizzi) univoci ad ogni host connesso a Internet. Se almeno due host diversi avessero lo stesso indirizzo IP potrebbero nascere problemi di ambiguità nella consegna dei dati ad essi indirizzati. Gli indirizzi IP sono di solito indicati attraverso 4 valori numerici interi tra 0 e 255 separati da un punto (es. 136.124.234.5). Esiste anche uno spazio dei nomi mnemonici per gli host connessi alla rete (ad esempio www.dsc.unibo.it) che risulta utile per ricordare meglio un indirizzo, ma che viene sempre mappato sul corrispondente indirizzo IP (es. 136.124.234.5) per far funzionare la rete secondo il protocollo IP. I nomi mnemonici dei domini Internet vengono assegnati in maniera centralizzata (per evitare conflitti di nome) dal NIC. Al fine di tradurre i nomi mnemonici negli indirizzi IP assegnati dal NIC, esistono degli host specificamente demandati a tale servizio (traduttori dei nomi in indirizzi IP), detti Domain Name Server (DNS). Ogni Web Server o ISP deve quindi avere configurato un DNS attivo (ovvero deve conoscerne l'indirizzo IP), affinché sia possibile tradurre i nomi dei riferimenti (link o URL) via via definiti.
 - spedizione a pacchetti: la trasmissione dei dati in rete non avviene con un flusso continuo di bit lungo il canale (la rete) ma avviene suddividendo opportunamente l'insieme di Byte da trasmettere in porzioni di dimensione variabile, detti pacchetti. Questa scelta è motivata da vari fattori:
 - una volta spedito un pacchetto tra due host, il canale può essere usato per la spedizione di altri pacchetti di altri host, in modo che ogni coppia abbia l'illusione di avere una linea di trasmissione dedicata e attiva. Naturalmente la BANDA (ovvero il numero di Bytes che è possibile trasmettere in un secondo) disponibile per ogni comunicazione (coppia di host source e destination) non è pari alla banda totale della rete usata, ma solo a una porzione di essa.

Accedervi con il proprio P.C.

- Per accedere a INTERNET con il proprio P.C. occorre:
 - un abbonamento ad un Provider: ISP (Internet Service Provider) è un fornitore di servizio di accesso a Internet. L'ISP mette a disposizione dei suoi utenti un pool di modem ai quali ogni utente può collegarsi da casa, attraverso la normale linea telefonica. La linea telefonica diventa quindi un'estensione

della rete Internet. Il pool di modem è connesso a un router in grado di smistare i pacchetti trasmessi e ricevuti sulla rete. Ogni utente collegato, al momento della connessione via modem, riceve un indirizzo IP temporaneo (cioè valido solo per il tempo della connessione), in modo che tutta la comunicazione in rete possa avvenire seguendo i protocolli (es. TCP/IP) standard. L'ISP mette anche a disposizione altri servizi, quali Web Server per la pubblicazione di pagine Web dei propri utenti, Mail server per la gestione dei messaggi di e-mail, ecc. Oltre alla tariffa di servizio da pagare all'ISP (oggi molti ISP offrono un servizio gratuito) rimane da pagare la tariffa per il collegamento telefonico da casa all'ISP. Questo è il motivo per cui conviene avere un ISP con il pool di modem accessibile entro l'area di tariffa urbana.

– un modem

- Un provider fornisce un accesso a INTERNET tramite modem: i dati da comunicare sono definiti da sequenze di valori binari (bit). I segnali trasmissibili sulla linea telefonica sono invece segnali analogici (cioè onde sonore).un modem è un dispositivo che trasmette i valori digitali (bit) che rappresentano l'informazione da comunicare trasformandoli in onde sonore, cioè “cantandoli” al telefono. Il problema di questo tipo di trasmissione è dato dalla banda limitata disponibile, ovvero dal numero limitato di bit/sec che è possibile cantare al telefono. I modem più performanti oggi sono in grado di trasmettere fino a 56Kb/s. Oggi esiste la possibilità di usare linee telefoniche digitali ISDN per le quali non è necessario disporre di un modem. Il vantaggio delle linee ISDN è dato dalla maggiore banda disponibile (128 Kb/s) contro i 64 Kb/s della linea telefonica analogica.
- Il modem permette di collegarsi (via linee telefoniche) al provider (ISP)

I servizi di INTERNET

- INTERNET fornisce una infrastruttura di comunicazione a livello planetario: come utilizzarla?
- Storicamente si sono via via stabiliti dei servizi di base che sono divenuti degli standard
- Potenzialmente chiunque potrebbe definire propri servizi personalizzati, a condizione che chi li utilizza condivida le scelte effettuate.

Colleg. Remoto / FTP (1)

- TELNET o RLOGIN sono servizi di collegamento a calcolatore remoto:
 - permettono di realizzare sessioni di lavoro su calcolatori diversi da quello in uso
- FTP (File Transfer Protocol) è un servizio di collegamento a file system remoto e permette di trasferire file tra file system su calcolatori diversi: un file system è un meccanismo per il mantenimento dell'informazione strutturata all'interno di un insieme di supporti di memorizzazione del calcolatore. Per fare un esempio,

basta pensare all'organizzazione dei file e cartelle sul disco fisso di un calcolatore: tale struttura è mantenuta dal file system. FTP è un servizio e allo stesso tempo un protocollo nato per consentire il trasferimento di file tra due host diversi. I file controllati e definiti dal file system sorgente possono essere trasferiti e copiati sull'host destinazione, dove saranno controllati e gestiti tramite il file system dell'host stesso.

Colleg. Remoto / FTP (2)

- **Entrambi i servizi usano**
 - **l'indirizzamento IP per definire i calcolatori a cui collegarsi: siccome il traffico viene trasmesso su Internet, è necessario appoggiarsi sui protocolli standard, per garantire la comunicazione a tutti i livelli intermedi (router intermedi).**
 - **un meccanismo di *user* e *password* per permettere il riconoscimento dell'utente: siccome il trasferimento di file può essere soggetto a restrizioni (informazioni riservate) è necessario prevedere un meccanismo di autenticazione degli utenti che intendono eseguire trasferimenti di file. Il modo più semplice e a volte più efficace per fare ciò è utilizzare un meccanismo di autenticazione basato su Login e Password, simile al metodo di accesso ai calcolatori del laboratorio di SdC.**
- **FTP permette la definizione di utenti anonimi che possono accedere solo a porzioni pubbliche di file system: se esistono file o dati che si intendono mettere a disposizione del pubblico, è possibile renderli disponibili tramite FTP, senza richiedere alcun tipo di autenticazione dell'utente (FTP anonimo)**

E-mail / News / Chat (1)

- **INTERNET permette la definizione di caselle di posta elettronica: si rende necessario un Mail server in grado di gestire tutto il traffico e la memorizzazione di messaggi per ogni dominio di posta. Tutti i messaggi destinati a utenti di un certo dominio Internet dovranno essere gestiti e smistati dal mail server del dominio stesso. Se il dominio è quello di un ISP, allora funzionerà il mail server dell'ISP, mentre se il dominio è quello di un dipartimento funzionerà il mail server del dipartimento.**
- **Ogni casella è identificata da un nome, es: *bononi@cs.unibo.it***
- **La parte di nome prima di @ identifica l'utente, e di solito (ma non necessariamente) corrisponde al nome di Login dell'utente presso quel dominio**
- **La parte che segue definisce il dominio presso cui esiste il mail server attivo per quell'utente.**

E-mail / News / Chat (2)

- **Le NEWS sono l'equivalente dei giornali su INTERNET**
- **Il vantaggio è che ognuno può pubblicare i propri articoli/messaggi**
- **Il risultato è la realizzazione di gruppi di discussione (news di carattere tematico)**

E-mail / News / Chat (3)

- **E-mail e News sono asincrone: la scrittura e la lettura avvengono in istanti differenti**
- **IRC (Internet Relay Chat) permette una comunicazione sincrona tra utenti (simile alla comunicazione telefonica) basata su comunicazione di testo scritto. Questa viene realizzata attraverso finestre condivise dove ogni utente inserisce i propri messaggi**

World Wide Web (1)

- **Il WEB è spesso considerato come un insieme di documenti di carattere ipertestuale che sono distribuiti sulla intera INTERNET. In generale lo si può ritenere una rete di risorse informative tra loro collegate.**
- **La codifica e la definizione di questo tipo di documenti è realizzata tramite un linguaggio chiamato HTML (HyperText Markup Language)**

World Wide Web (2)

- **Le singole parti dei documenti sono memorizzate in file memorizzati su file system remoti: ciò significa che ogni documento del Web risiede presso un Web server da qualche parte su Internet. Un web client (es. un browser) richiede le informazioni al web server specificando il nome del documento**
- **I singoli documenti hanno nomi (URL Uniform Resource Locator) del tipo:**
 - http://www.cs.unibo.it/index.html**
 - **Dove http specifica lo schema o protocollo di trasmissione usato: HyperText Transfer Protocol per le pagine Web**
 - **www.cs.unibo.it: rappresenta il dominio presso il quale esiste il Web Server da contattare (in pratica è il nome mnemonico al quale è associato l'indirizzo IP del Web server che gestisce il documento richiesto)**
 - **index.html: rappresenta il nome del file o documento richiesto**

• **What is the World Wide Web?**

- **The *World Wide Web (Web)* is a network of information resources. The Web relies on three mechanisms to make these resources readily available to the widest possible audience:**
- **A uniform naming scheme for locating resources on the Web (e.g., URIs).**
- **Protocols, for access to named resources over the Web (e.g., HTTP).**
- **Hypertext, for easy navigation among resources (e.g., HTML).**

World Wide Web (6)

- **Every resource available on the Web -- HTML document, image, video clip, program,**

etc. -- has an address that may be encoded by a *Universal Resource Identifier*, or "URI".

- URIs typically consist of three pieces:
 - The naming scheme of the mechanism used to access the resource.
 - The name of the machine hosting the resource.
 - The name of the resource itself, given as a path.
- Consider the URI that designates the current HTML specification:
- <http://www.w3.org/TR/REC-html40/>

World Wide Web (7)

- This URI may be read as follows: There is a document available via the HTTP protocol (see [\[RFC2068\]](#)), residing on the machine www.w3.org, accessible via the path `"/TR/REC-html40/"`. Other schemes you may see in HTML documents include "mailto" for email and "ftp" for FTP.
- Here is another example of a URI. A user's mailbox:
 - *...this is text...*
 - For all comments, please send email to
 - `Joe Cool`.
 - *Note.* Most readers may be familiar with the term "URL" and not the term "URI". URLs form a subset of the more general URI naming scheme.

World Wide Web (8)

–Fragment identifiers

- Some URIs refer to a location within a resource. This kind of URI ends with "#" followed by an anchor identifier (called the *fragment identifier*). For instance, here is a URI pointing to an anchor named `section_2`:
- http://somesite.com/html/top.html#section_2

World Wide Web (9)

–Relative URIs

- A *relative URI* doesn't contain any naming scheme information. Its path generally refers to a resource on the same machine as the current document. Relative URIs may contain relative path components (e.g., "." means one level up in the hierarchy defined by the path), and may contain *Errore. Riferimento a collegamento ipertestuale non valido.*
- `Suppliers`

HTML (1)

- **What is HTML?**

- To publish information for global distribution, one needs a universally understood language, a kind of publishing mother tongue that all computers may potentially understand. The publishing language used by the World Wide Web is HTML (from HyperText Markup Language).

HTML (2)

- HTML gives authors the means to:
 - Publish online documents with headings, text, tables, lists, photos, etc.
 - Retrieve online information via hypertext links, at the click of a button.
 - Design forms for conducting transactions with remote services, for use in searching for information, making reservations, ordering products, etc.
 - Include spread-sheets, video clips, sound clips, and other applications directly in their documents.

HTML (3)

- HTML documents should work well across different browsers and platforms. (develop only one version of a document).
- risk that the Web will devolve into a proprietary world of incompatible formats
- Each version of HTML reflect greater consensus among industry players (documents will not become unreadable in a short period of time)
- HTML developed for all devices: PCs cellular telephones, hand held devices...

HTML (4)

- HTML: the history
- Tim Berners-Lee (CERN) 1988: first development
- first browser: Mosaic (NCSA)
- 1990: start explosive growth of the Web
- HTML estensions --> joint work on specifications
- November 1995: HTML 2.0 IETF (RFC1866)
- HTML+ and HTML 3.0: new features
- 1996: W3C HTML 3.2
- Today: HTML 4.0

HTML 4.0 (1)

- HTML 4.0 extends HTML with mechanisms for style sheets, scripting, frames, embedding objects, improved support for right to left and mixed direction text, richer tables, and enhancements to forms, offering improved accessibility for people with disabilities.

HTML 4.0 (2)

- Internationalization: documents may be written in every language and be transported easily around the world
- Accessibility: encouraging the use of style sheets, better forms, access keys
- well-designed HTML documents that distinguish structure and presentation will adapt more easily to new technologies

HTML 4.0 (3)

- Better Tables (shown incrementally)
- better Compound documents: objects for including images, video, sound, mathematics, specialized applications, and other objects in a document
- Style sheets: simplify HTML markup and alignment, font size, and text color
- Scripting: create dynamic Web pages
- better Printing: W3C's Resource Description Language (RDF)