

---

---

# **Introduzione ad Internet I protocolli e le applicazioni della pila TCP/IP**

Stefano Salsano

DIE - Università di Roma "Tor Vergata"

stefano.salsano@uniroma2.it

**Mi risulta difficile riconoscere tutti i contributi che ho utilizzato nella preparazione di questo materiale, ringrazio comunque in modo particolare Pietro Di Genio e Vincenzo Della Mea.**

Stefano Salsano

Marzo, 2002

## Outline

- L'architettura TCP/IP : storia
- La pila dei protocolli
- Il livello IP
- Il livello di trasporto
- Protocolli applicativi e servizi

Stefano Salsano

Marzo, 2002

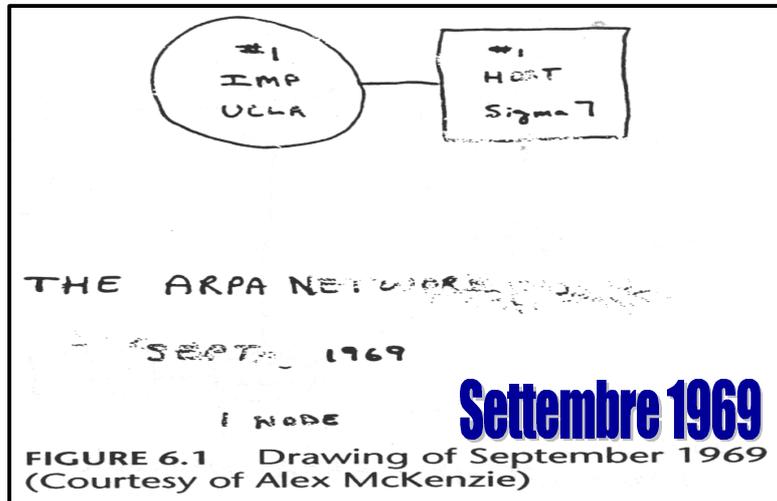
## Architettura TCP/IP: storia

- Le origini di Internet si possono far risalire al progetto DARPA (Defense Advanced Research Project Agency) del DOD Americano (inizi anni 70')
- Necessità di interconnettere reti dei centri militari, universitari e di ricerca: definizione della rete ARPANET
- 1973 viene commissionato all'Università di Stanford il progetto di una suite di protocolli standard che garantissero connettività a livello di rete
- Verso la fine degli anni '70, tale sforzo portò al completamento dell'**Internet Protocol Suite**, di cui i due protocolli più noti sono il TCP e l'IP
- Il nome più appropriato per l'architettura di rete rimane quello di Internet Protocol Suite, anche se comunemente si fa riferimento ad essa con la sigla TCP/IP
- I protocolli appartenenti a questa architettura sono specificati tramite standard denominati **RFC** (Request For Comments)

Stefano Salsano

Marzo, 2002

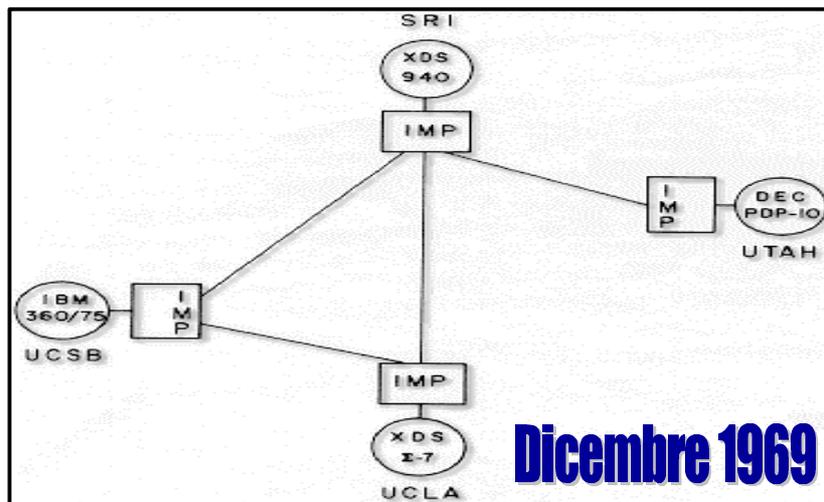
## ARPANET: il primo nodo



Stefano Salsano

Marzo, 2002

## ARPANET: quattro nodi!

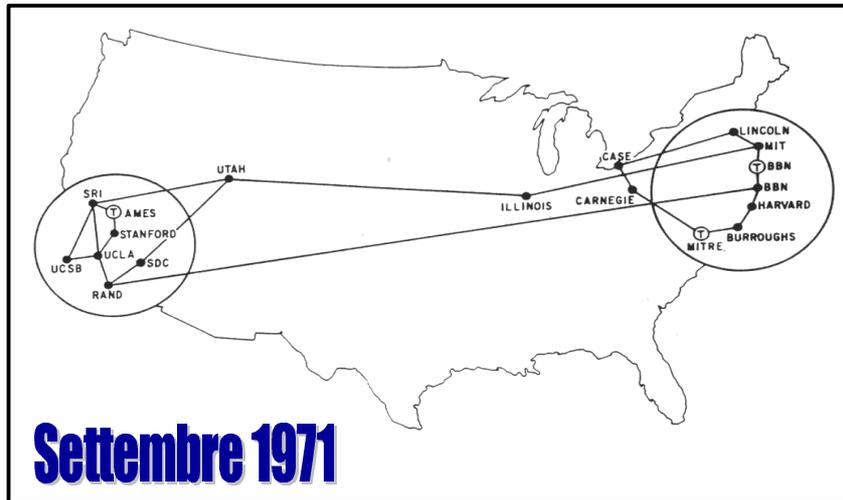


**Dicembre 1969**

Stefano Salsano

Marzo, 2002

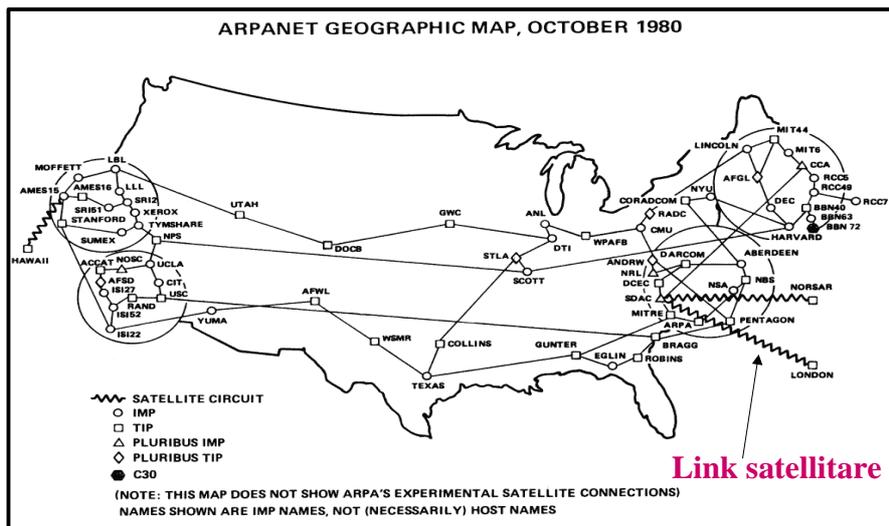
## ARPANET: inizia a crescere



Stefano Salsano

Marzo, 2002

## ARPANET: 1980



Stefano Salsano

Marzo, 2002

## Architettura TCP/IP

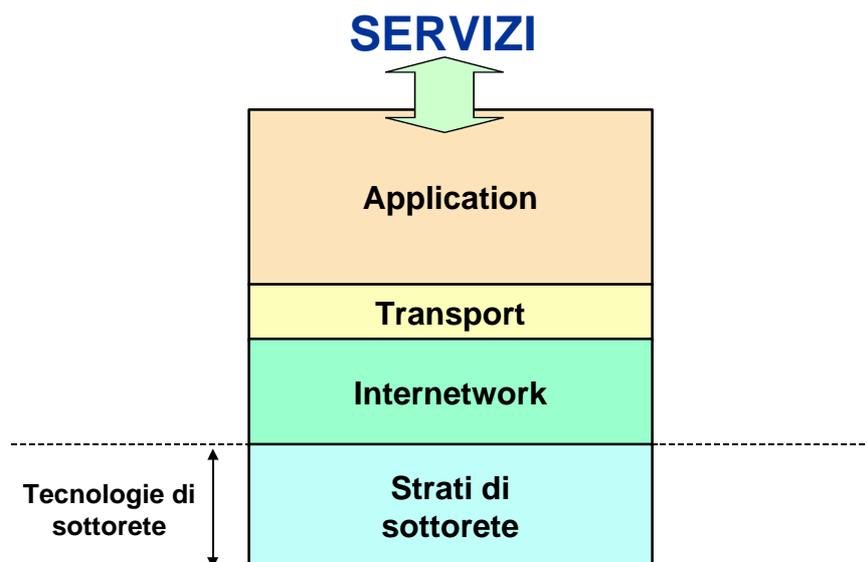
- Un'architettura Internet consiste di un insieme di reti interconnesse che possono essere considerate come un'unica entità
- Le reti sono collegate da particolari apparati detti Router, che hanno come funzione primaria quella di instradare i dati sulla base dell'indirizzo della rete destinataria



Stefano Salsano

Marzo, 2002

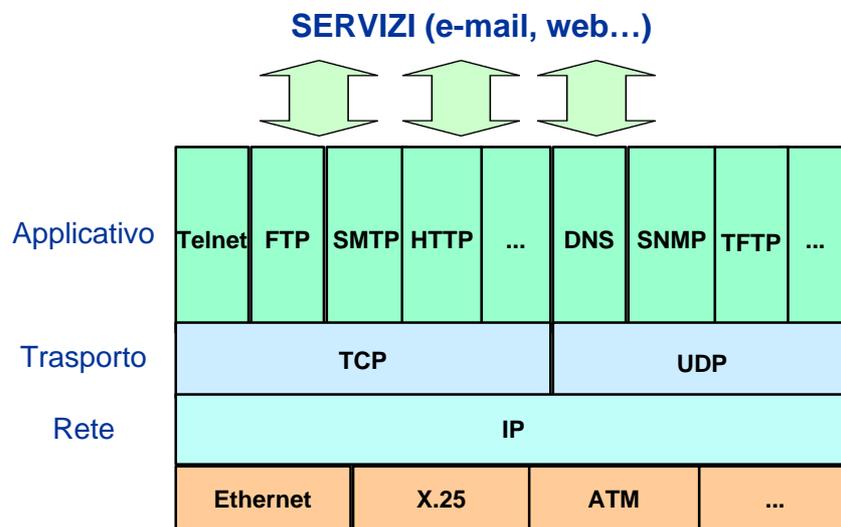
## Architettura stratificata



Stefano Salsano

Marzo, 2002

## Pila dei protocolli TCP/IP

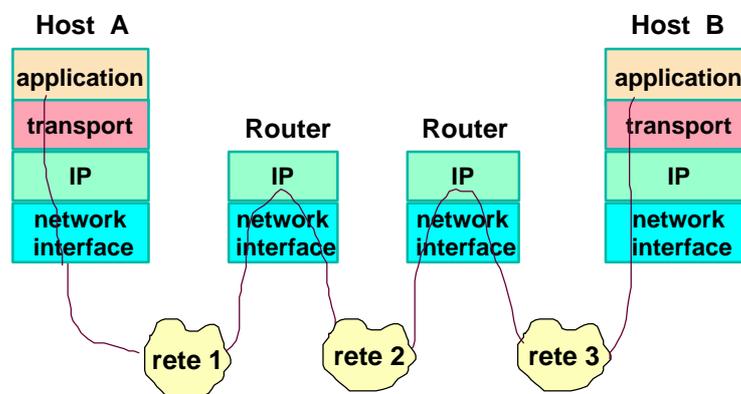


Stefano Salsano

Marzo, 2002

## Architettura TCP/IP

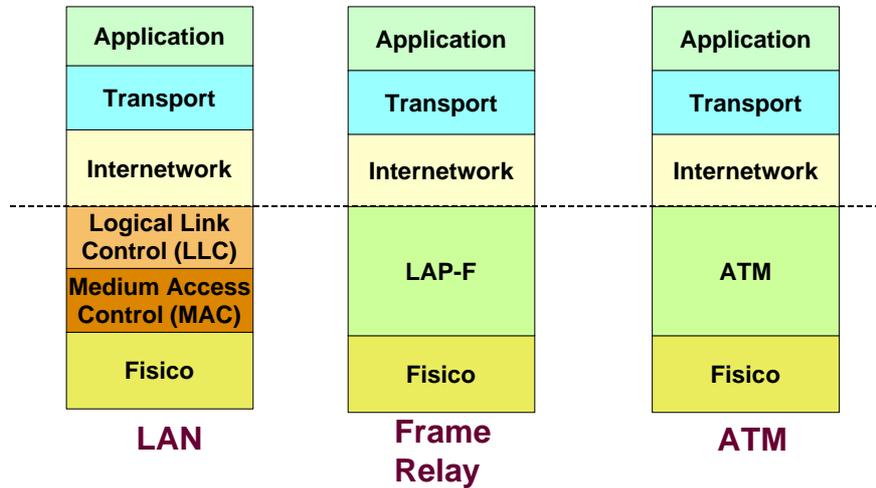
- Il percorso software di un messaggio che attraversa più reti interconnesse tramite un'architettura Internet può essere schematizzato come segue:



Stefano Salsano

Marzo, 2002

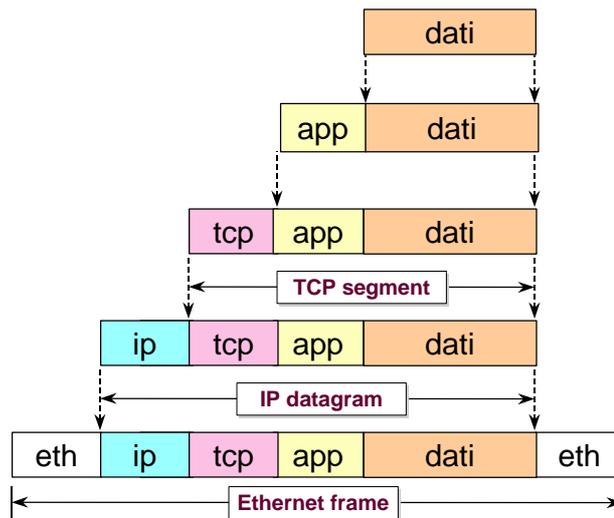
## Compatibilità con (tutte!) le tecnologie di sottorete



Stefano Salsano

Marzo, 2002

## Relazioni tra i livelli protocollari



Stefano Salsano

Marzo, 2002

## Strato di Rete: protocollo IP

- **Strato di internetworking-INTERNET PROTOCOL (IP- RFC 791)**
  - » Il protocollo definisce un meccanismo di consegna dati di tipo "non affidabile" e "non orientato alla connessione (connectionless)"
- **header IP**
  - » indirizzo mittente
  - » indirizzo destinatario
  - » protocollo di trasporto,....

Stefano Salsano

Marzo, 2002

## Internet Protocol (IP)

- **Il protocollo fornisce i servizi:**
  - » trasmissione di un datagram host-to-host (indirizzamento)
  - » funzioni di routing
  - » frammentazione e riassetaggio dei datagram
- **Il protocollo non fornisce:**
  - » controllo di flusso
  - » controllo d'errore
  - » controllo di sequenza
- **Il protocollo Internet tratta ogni datagram come un'entità indipendente di trasmissione scorrelata dalle altre.**

Stefano Salsano

Marzo, 2002

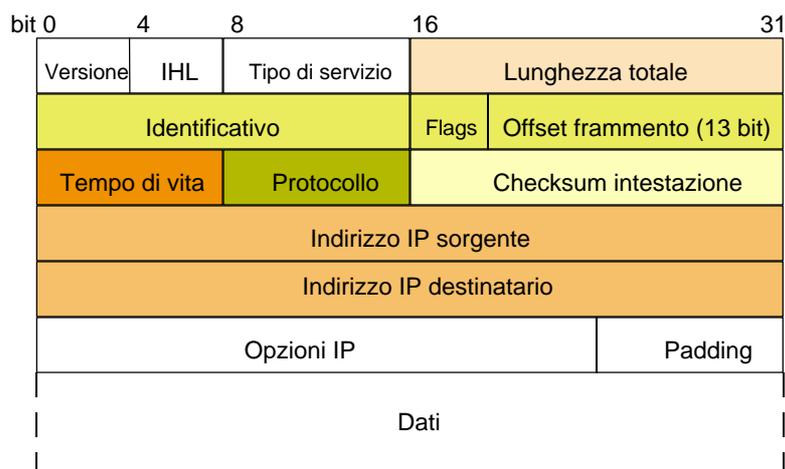
## Internet Protocol (IP)

- Le funzioni principali del protocollo IP sono:
- in trasmissione
  - » incapsula in datagram i dati provenienti dallo strato di trasporto
  - » prepone l'opportuna intestazione
  - » applica algoritmo di routing
  - » invia i dati verso l'opportuna interfaccia di rete
- in ricezione
  - » verifica la validità dei datagram in arrivo
  - » esamina l'intestazione
  - » verifica se sono dati da reinstradare
  - » se sono dati locali, sceglie l'opportuno protocollo dello strato di trasporto e provvede ad inviarglieli nel formato opportuno.

Stefano Salsano

Marzo, 2002

## Header IP



Stefano Salsano

Marzo, 2002

## Header IP

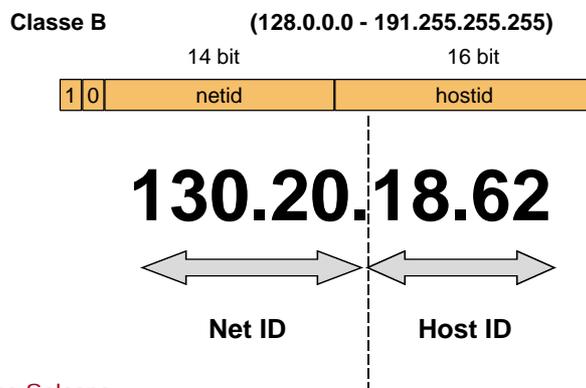
- Il datagram IP ha una lunghezza variabile:
  - » header: 20 - 64byte
  - » dati + header  $\leq$  MTU (Maximum Transmission Unit), altrimenti si ha la frammentazione
- Source address: indirizzo IP mittente
- Destination address: indirizzo IP destinatario
- Protocol: protocollo di trasporto
- Fragment offset: offset del frammento nell'ambito del totale dei dati da trasmettere
- Identification: identificativo del datagram
- Flag: indica se il datagram rappresenta un frammento
- Time-To-Live (TTL): numero di router che un datagram puo' attraversare
- Header checksum: controllo errore sull'header
- IHL: lunghezza dell'header
- Total length: lunghezza totale

Stefano Salsano

Marzo, 2002

## Indirizzi IP

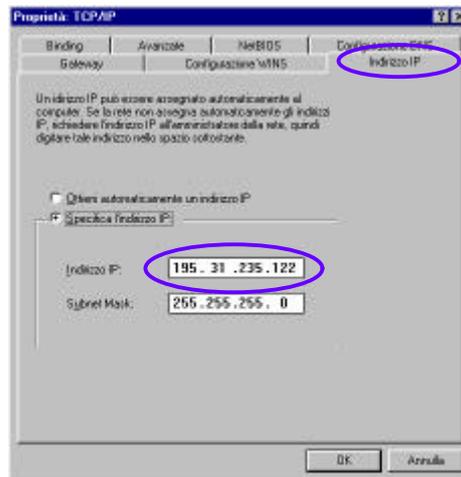
- Ogni host IP ha un indirizzo diviso in due parti:
  - » Host-ID identifica l'host all'interno della rete IP
  - » Network-ID identifica la rete IP su Internet
- Esempio:



Stefano Salsano

Marzo, 2002

## Configurazione Windows 95: indirizzo IP



Stefano Salsano

Marzo, 2002

## Router IP e routing dei pacchetti

- I router IP ricevono datagrammi IP da un'interfaccia e li inoltrano su un'altra
- Si distinguono dagli Host perchè:
  - > hanno in genere più di un'interfaccia
  - > utilizzano "protocolli di routing" che consentono di conoscere la topologia della rete e quindi il cammino che i pacchetti devono fare verso la destinazione
- I router IP hanno (normalmente) un indirizzo IP per ogni interfaccia
- L'inoltro dei pacchetti avviene guardando la "net-id" dell'indirizzo di destinazione del pacchetto

Stefano Salsano

Marzo, 2002

## Strato di Trasporto

- Lo scopo di tale strato è quello di fornire una comunicazione end-to-end a livello di processi
- Il software di questo strato è responsabile del meccanismo che permette di distinguere, all'interno di uno stesso host, il processo applicativo destinatario (o sorgente) dei dati
- Ogni elaboratore contiene un insieme di punti logici di accesso "ports"
- Ad ogni servizio è associato un port che consente di indirizzare il processo che realizza il servizio (server)
- Nell'architettura Internet esistono due standard principali di protocolli di trasporto:
  - » User Datagram Protocol (UDP - RFC 768)
  - » Transmission Control Protocol (TCP - RFC 793)

Stefano Salsano

Marzo, 2002

## Strato di Trasporto

- La divisione dei compiti fra lo strato di trasporto (UDP, TCP) e IP è la seguente:
  - » lo strato IP si occupa del trasferimento dei dati fra elaboratori collegati alle reti interconnesse; quindi l'intestazione IP identifica gli host sorgente e destinazione
  - » lo strato UDP (TCP) si occupa dello smistamento dei dati fra sorgenti o destinazioni multiple all'interno dello stesso host tramite il port number
- Per richiedere un servizio, fornito da un processo residente su un host remoto, il client deve conoscere il port number associato al servizio stesso

Stefano Salsano

Marzo, 2002

## Porte TCP e UDP

- Sono il mezzo con cui un programma client indirizza un programma server
  - » un client HTTP per connettersi ad un server HTTP indica:
    - » l'indirizzo IP dell'elaboratore remoto
    - » il numero della porta associata al server HTTP
- **Caratteristiche**
  - » identificate da un numero naturale su 16 bit
  - » 0...1023 = porte privilegiate
  - » 1024...65535 = porte utente

Stefano Salsano

Marzo, 2002

## Well Known Port

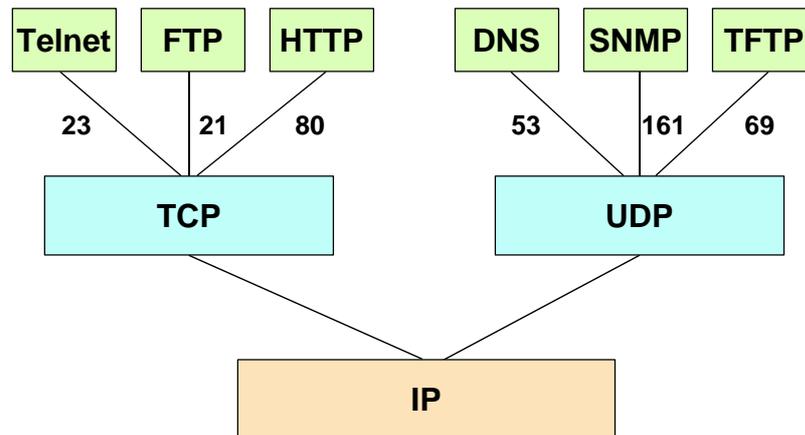
- Sono associate agli applicativi principali

Servizio	Porta	TCP	UDP
FTP	21	✓	
Telnet	23	✓	
SMTP	25	✓	
TFTP	69		✓
DNS	53	✓	✓
HTTP	80	✓	
SNMP	161		✓
POP3	110	✓	

Stefano Salsano

Marzo, 2002

## Well Known Port



Stefano Salsano

Marzo, 2002

## Transmission Control Protocol (TCP)

- **Trasferisce un flusso informativo bi-direzionale tra due host**
- **E' un protocollo con connessione**
- **Funzioni eseguite**
  - » controllo e recupero di errore
  - » controllo di flusso
  - » ri-ordinamento delle unità informative
  - » indirizzamento di uno specifico utente all'interno di un host

Stefano Salsano

Marzo, 2002

## User Datagram Protocol (UDP)

- UDP (User Datagram Protocol)
- Consente alle applicazioni di scambiare messaggi singoli
- Fornisce un livello di servizio minimo:
  - » E' un protocollo senza connessione
  - » Non supporta meccanismi di riscontro e di controllo d'errore
  - » Supporta funzionalità di multiplazione e demultiplazione attraverso l'introduzione delle porte (in modo analogo al TCP)
  - » Supporta il multicast (a differenza di TCP)

Stefano Salsano

Marzo, 2002

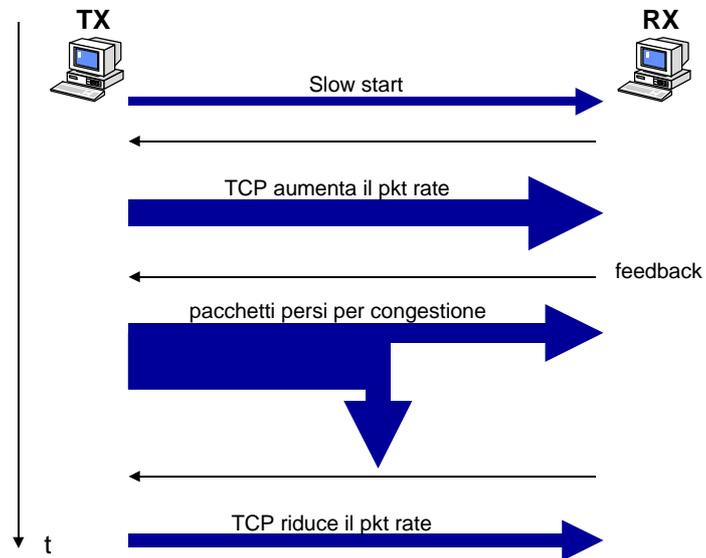
## Confronto fra TCP ed UDP

- TCP è affidabile (Connection Oriented), mentre UDP non lo è (Connectionless)
- TCP consente esclusivamente connessioni punto-punto, mentre UDP supporta broadcast e multicast
- TCP è stream-oriented, UDP message-oriented
- TCP recupera eventuali errori trasmissivi ed effettua controllo di flusso e di congestione, UDP no
- TCP introduce maggior overhead rispetto ad UDP

Stefano Salsano

Marzo, 2002

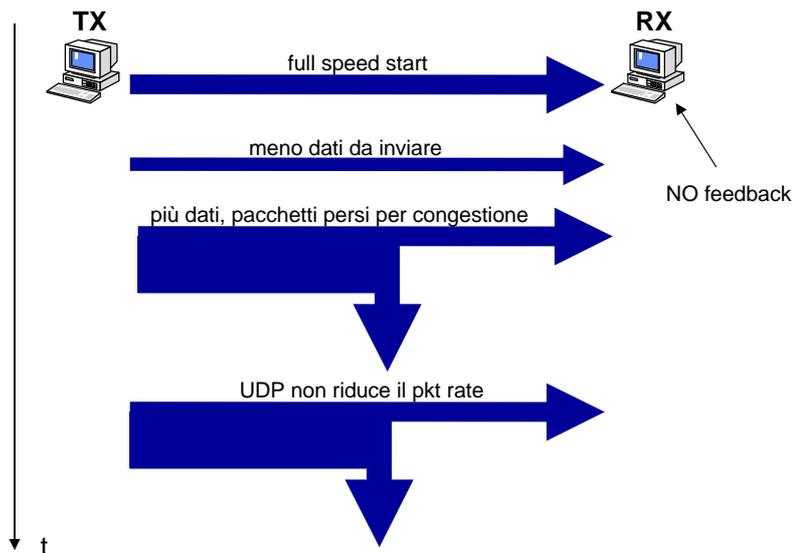
## TCP: comportamento in caso di congestione



Stefano Salsano

Marzo, 2002

## UDP: comportamento in caso di congestione



Stefano Salsano

Marzo, 2002

# I protocolli applicativi

Stefano Salsano

Marzo, 2002

## Il paradigma client/server



- » si connette al server (calcolatore) sulla porta specifica ad esso associata
- » invia dei messaggi composti secondo il protocollo di comunicazione
- » aspetta i risultati

- » è sempre attivo (daemon)
- » "ascolta" i messaggi in arrivo su una porta
- » li interpreta (grazie al protocollo) ed effettua il servizio richiesto
- » rispedisce indietro i risultati

Stefano Salsano

Marzo, 2002

## I protocolli di Internet

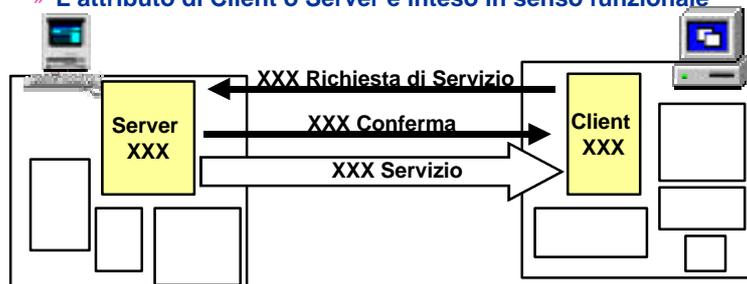
- Come accennato, sono “stratificati”
- I protocolli applicativi rispettano tutti lo stesso principio client/server
- Definiscono:
  - » il formato dei messaggi scambiati
  - » il significato di alcune parti del messaggio (es. richiesta servizi)
- Sono descritti nelle "RFC" (Request for Comments), liberamente disponibili su Internet

Stefano Salsano

Marzo, 2002

## Modello Client / Server

- Le applicazioni Internet e alcuni protocolli della famiglia TCP/IP (DNS, ARP/RARP, ...) sono basate sul modello Client / Server
  - » L'attributo di Client o Server è inteso in senso funzionale



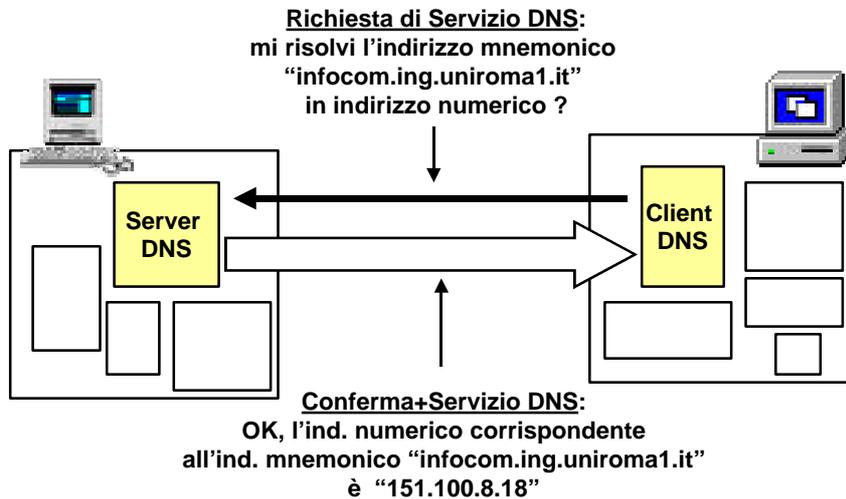
- » Il verso della freccia “XXX Servizio” identifica il verso logico di fornitura del servizio, e può non coincidere con il verso fisico percorso dai dati

Stefano Salsano

Marzo, 2002

## Modello Client / Server

- Esempio 1 (DNS)

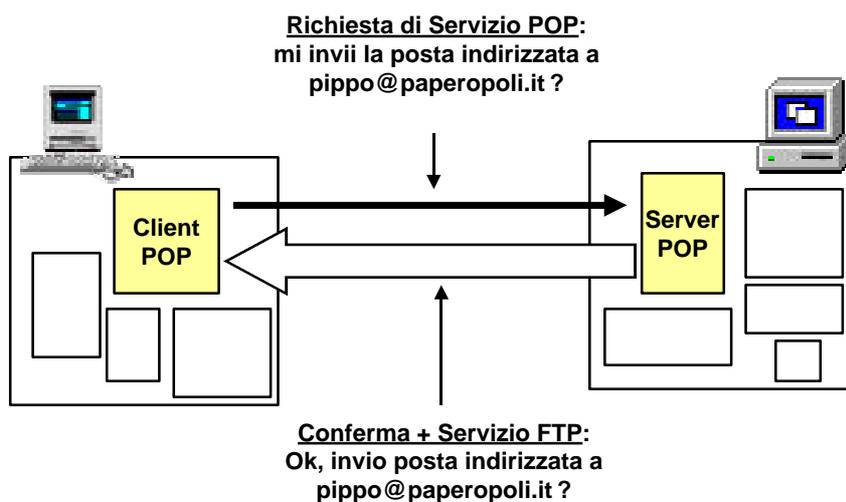


Stefano Salsano

Marzo, 2002

## Modello Client / Server

- Esempio 2 (POP)

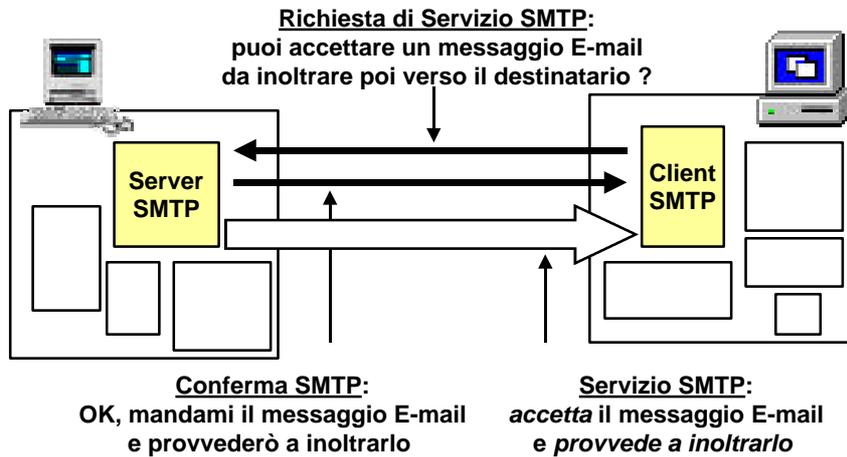


Stefano Salsano

Marzo, 2002

## Modello Client / Server

- Esempio 3 (SMTP)

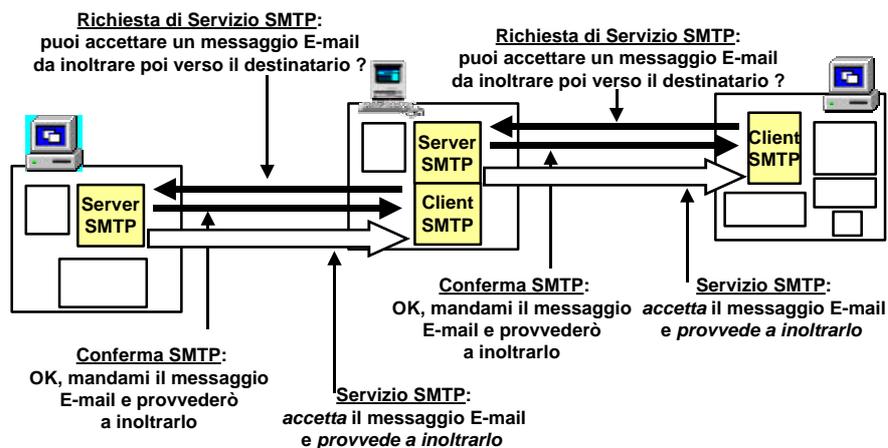


Stefano Salsano

Marzo, 2002

## Modello Client / Server

- La strutturazione Client / Server può essere concatenata

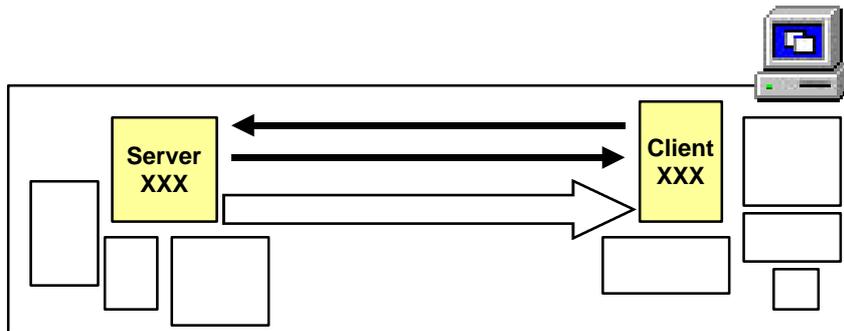


Stefano Salsano

Marzo, 2002

## Modello Client / Server

- Le componenti Client / Server possono risiedere nello stesso sistema



Stefano Salsano

Marzo, 2002

## Domain Name System (DNS)

Stefano Salsano

Marzo, 2002

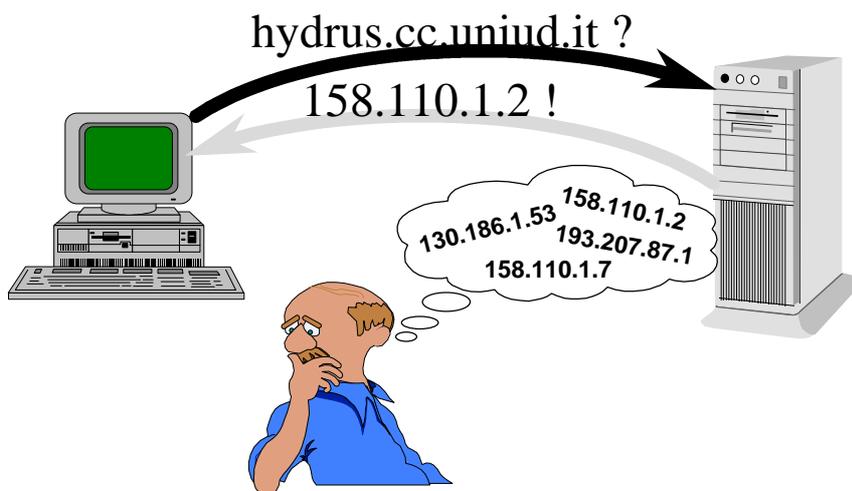
## Domain Name System

- Oltre alla notazione dotted viene spesso utilizzata anche un'altra forma di notazione (mnemonica):  
"151.100.8.18"="infocom.ing.uniroma1.it"
- E' necessaria la funzionalità di traduzione di nomi mnemonici in indirizzi e viceversa
- Questa traduzione è attuata da un protocollo di alto livello implementato in un meccanismo noto come Domain Name System (DNS)

Stefano Salsano

Marzo, 2002

## Domain Name System



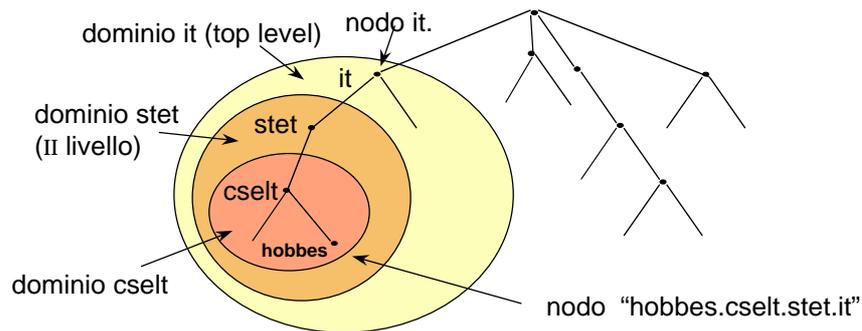
Stefano Salsano

Marzo, 2002

Slide 44

## Domain Name System

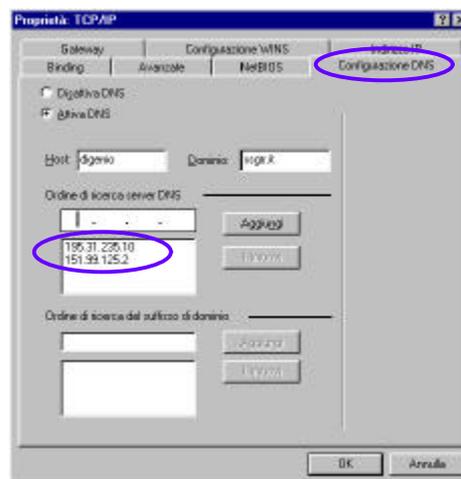
- In Internet i nomi sono organizzati gerarchicamente in Domini
  - » I nomi sono costituiti da stringhe separate da “.”
  - » La parte più significativa è a destra



Stefano Salsano

Marzo, 2002

## Configurazione Windows 95: DNS



Stefano Salsano

Marzo, 2002

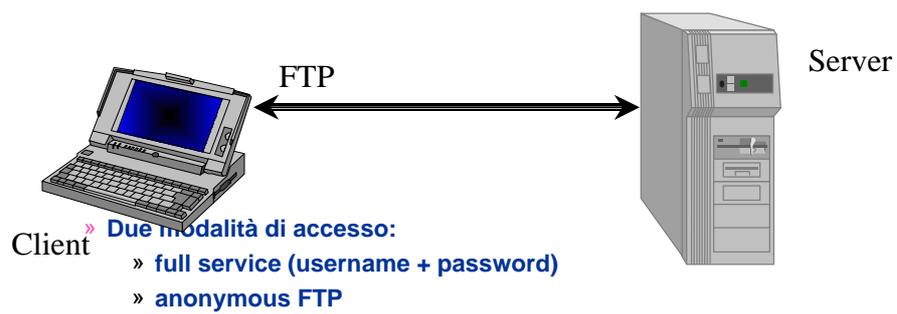
# File Transfer Protocol (FTP)

Stefano Salsano

Marzo, 2002

# File Transfer Protocol (FTP)

- protocollo/software per il trasferimento di file



Stefano Salsano

Marzo, 2002

## **File Transfer Protocol (FTP)**

- protocollo per il trasferimento di files tra sistemi remoti
- è uno dei protocolli applicativi più vecchi
- permette la gestione locale del file system del sistema remoto tramite una serie di comandi per la navigazione nel file system, ed il trasferimento dei files in entrambe le direzioni
- si basa su convenzioni tipiche di Unix (file system, comandi e loro output)
- rispetta i privilegi di accesso ai files del sistema remoto
- è un protocollo a connessione persistente

Stefano Salsano

Marzo, 2002

## **Telnet**

Stefano Salsano

Marzo, 2002

## **TELNET**

- Fornisce un “terminale virtuale” su una rete TCP/IP
- E' un PROTOCOLLO che fornisce una “comunicazione ad 8 bit bidirezionale”
- E' anche una APPLICAZIONE che supporta il protocollo TELNET
- Molte applicazioni utilizzano il protocollo TELNET

Stefano Salsano

Marzo, 2002

## **La posta elettronica (SMTP, POP)**

Stefano Salsano

Marzo, 2002

## La posta elettronica

- **Caratteristiche:**
  - » velocità
  - » versatilità
  - » economicità
  - » Indipendenza dal tempo e dallo spazio



Stefano Salsano

Marzo, 2002

## La posta elettronica

- **Gli strumenti necessari:**
  - » Mailbox (casella postale)
  - » indirizzo posta elettronica
  - » PC connesso ad Internet
  - » programma "client" sul PC



Stefano Salsano

Marzo, 2002

## La mailbox

- Normalmente risiede su un server permanentemente connesso alla rete
- Ha associato un indirizzo di posta elettronica



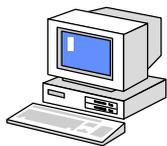
E' il contenitore elettronico dei messaggi ricevuti

Stefano Salsano

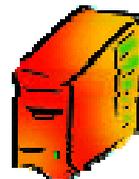
Marzo, 2002

## E-mail protocols

**Mail User Agent  
(Client)**



**Mail Transfer Agent  
(Server)**



*Formato messaggi  
RFC822/MIME*

SMTP

POP3/IMAP

- » ha un'interfaccia utente per l'inserimento dei messaggi
- » "conosce" il protocollo per spedire messaggi (SMTP)
- » ed il protocollo per riceverli (POP3 o IMAP4)
- » "conosce" come comporre i messaggi (RFC822/MIME)

- » un server SMTP (porta 25): gestisce la spedizione e ricezione dei messaggi da/verso altri server SMTP
- » un server POP3 (porta 110) oppure IMAP4: gestisce la spedizione dei messaggi al client

Stefano Salsano

Marzo, 2002

## Mail User Agent

- **Funzioni del MUA:**
  - » Composizione dei messaggi
  - » Trasferimento del messaggio da mittente a destinatario
  - » Rapporto al mittente sullo stato della spedizione
  - » Visualizzazione dei messaggi (facile se sono solo testuali, più complessa se sono multimediali)
  - » Gestione dei messaggi spediti e ricevuti: eliminazione, memorizzazione in mailbox, etc.

Stefano Salsano

Marzo, 2002

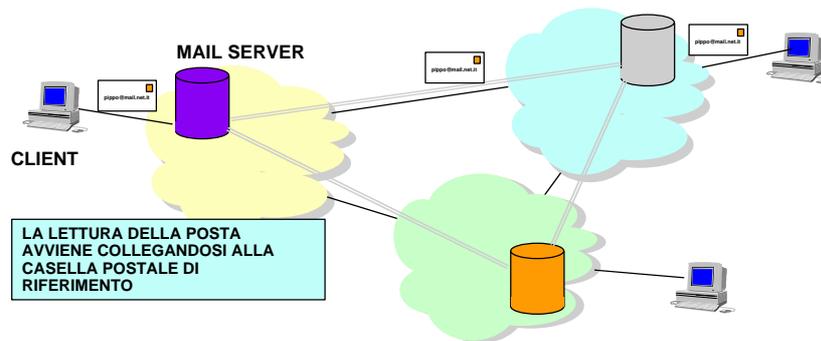
## Messaggio e-mail

- Il formato base è descritto nella RFC 822
- E' costituito da un testo ASCII organizzato genericamente in righe separate da coppie di caratteri CR LF
- E' logicamente suddiviso in:
  - » Intestazione (header)
    - » mittente (From:)
    - » destinatario (To:)
    - » oggetto (Subject:)
    - » ...
  - » Corpo del messaggio
    - » testo qualsiasi, separato dall'intestazione con una linea vuota, e terminato da un punto singolo

Stefano Salsano

Marzo, 2002

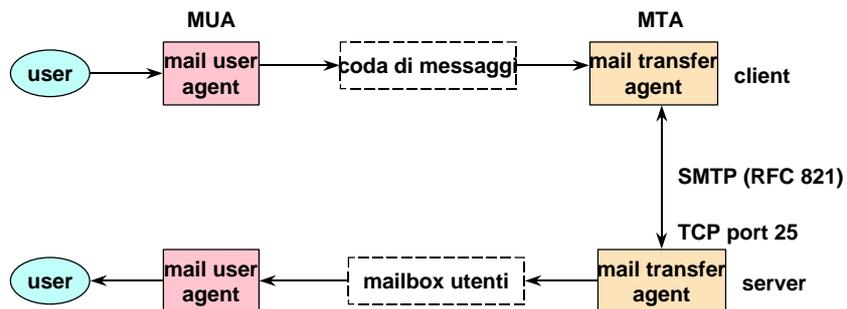
## SMTP (Simple Mail Transfer Protocol)



Stefano Salsano

Marzo, 2002

## SMTP (Simple Mail Transfer Protocol)



Stefano Salsano

Marzo, 2002

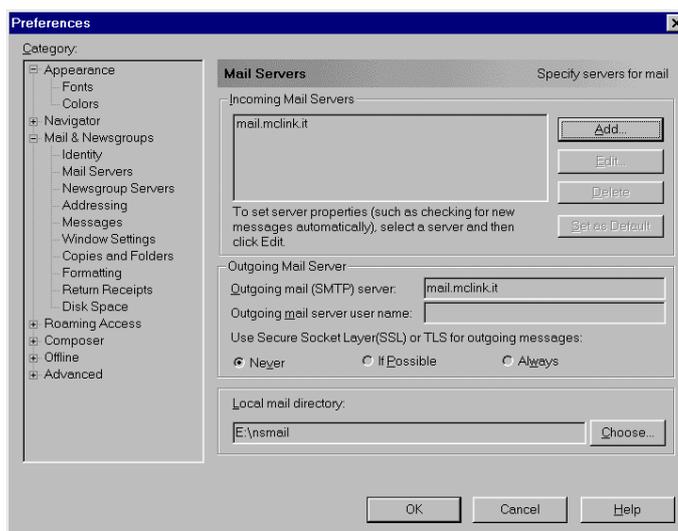
## Come ricevere la posta?

- Ogni utente che ha accesso ad un sistema di solito ha una sua mailbox (casella di posta elettronica) che corrisponderà ad uno spazio nel file system; in quello spazio il server SMTP locale salverà i messaggi ricevuti per lo specifico utente
- Se l'utente accede direttamente al sistema (es. tramite shell di Unix), il programma di email leggerà direttamente dal file system i messaggi;
- Altrimenti, via rete, sarà necessario un protocollo apposito per la lettura dei messaggi:
  - » POP3: permette di scaricare i messaggi come tali, senza funzionalità di gestione
  - » IMAP: accesso alla mailbox, i messaggi rimangono sul server e sono quindi accessibili da più sistemi

Stefano Salsano

Marzo, 2002

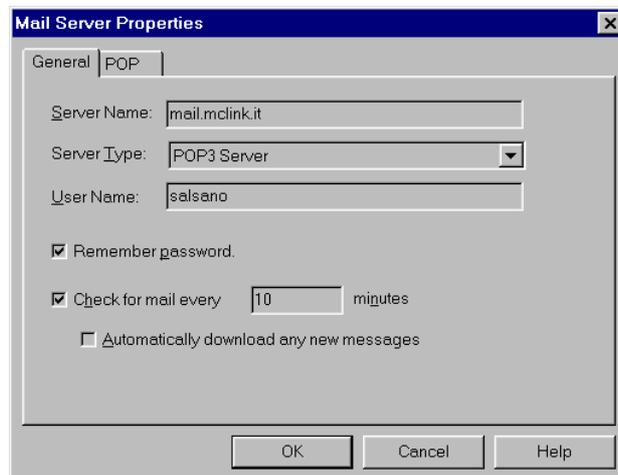
## Esempio: netscape client



Stefano Salsano

Marzo, 2002

## Esempio: netscape client



Stefano Salsano

Marzo, 2002

## La posta elettronica multimediale

- **Nuovi computer multimediali -> necessità di spedire messaggi non limitati al solo testo**
- **MIME: Multipurpose Internet Mail Extensions**
- **Esigenze progettuali:**
  - » rappresentazione di immagini, audio, video, etc
  - » presenza in rete di infrastrutture (=server SMTP) difficilmente e costosamente sostituibili
  - » e quindi vincolo dei 7 bit, e modifiche al solo corpo del messaggio
- **Realizzazione**
  - » campi aggiuntivi nell'intestazione
  - » definizione di tipi di contenuto per interpretare il corpo
  - » metodi canonici di codifica di file binari
- **E' più generale di una soluzione multimediale**

Stefano Salsano

Marzo, 2002

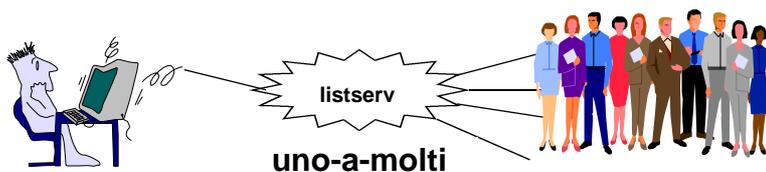
## Tipi di contenuto

- Valori che assume il campo Content-Type, nella forma tipo/sottotipo
- Sette tipi predefiniti:
  - » text, per messaggi testuali; text/plain è il formato testuale RFC822. Altro sottotipo è text/enriched.
  - » image, per rappresentare immagini, che a loro volta possono essere di sottotipo jpeg o gif.
  - » audio, per dati sonori, con sottotipo audio/basic.
  - » video, per la rappresentazione di filmati, con sottotipo iniziale video/mpeg.
  - » application, per files binari di altro tipo: es., con il sottotipo octet-stream, files binari generici.
  - » multipart: composizione di messaggi secondo i seguenti sottotipi: mixed, alternative, parallel, digest.
  - » message: per incorporare ulteriori messaggi, per esempio RFC822

Stefano Salsano

Marzo, 2002

## La Mailing List

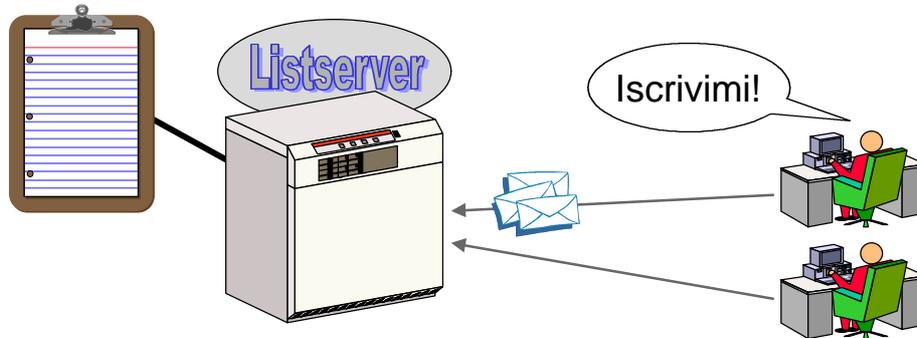


- » lista di indirizzi E-mail
- » iscrizione, rimozione: a richiesta
- » invio di un messaggio a tutti gli iscritti
- » liste tematiche

Stefano Salsano

Marzo, 2002

## La Mailing List: il Listserver

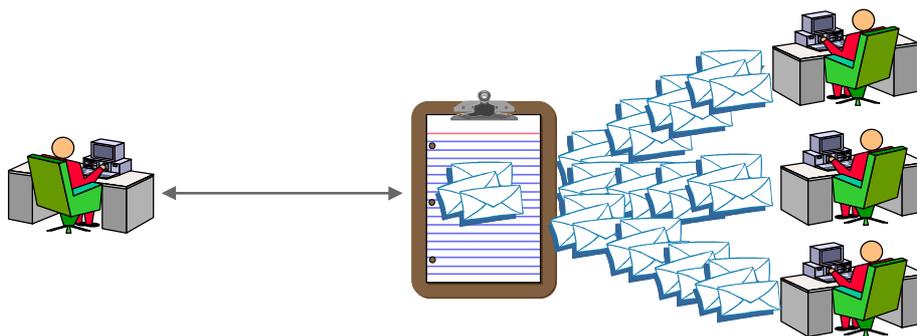


- Programma che gestisce la lista
- Indirizzo email specifico (listserver@host.domain)
- Richieste di iscrizione, rimozione: via mail al listserver

Stefano Salsano

Marzo, 2002

## La Mailing List: la lista vera e propria



- Indirizzo e-mail specifico (nomelista@host.domain)
- Da NON confondersi con quello del Listserver
- Spedire all'indirizzo della lista per spedire copia a tutti

Stefano Salsano

Marzo, 2002

## La Mailing List: riassumendo

Per ogni lista almeno DUE indirizzi E-mail:

- l'indirizzo della lista (per comunicare con gli iscritti)  
Esempio: `www-uniud@cc.uniud.it`
- l'indirizzo del LISTSERVER (iscrizione, rimozione)  
Esempio: `mailserv@cc.uniud.it`



Stefano Salsano

Marzo, 2002

## Hyper Text Transfer Protocol (HTTP) e il web

Stefano Salsano

Marzo, 2002

## WWW: cos'è? (per l'utente...)

- Collezione di documenti (testo, immagini, etc) che risiede su calcolatori sparsi in giro, e connessi tramite Internet
- I documenti vengono chiamati pagine
- Ogni pagina può contenere dei collegamenti (link) ad altre pagine correlate, situate ovunque
- L'utente può seguire questi collegamenti, cliccando sui link e muovendosi di pagina in pagina
- Le pagine sono viste grazie ad un programma detto browser, che recupera e mostra le pagine, interpreta le richieste dell'utente, etc.

Stefano Salsano

Marzo, 2002

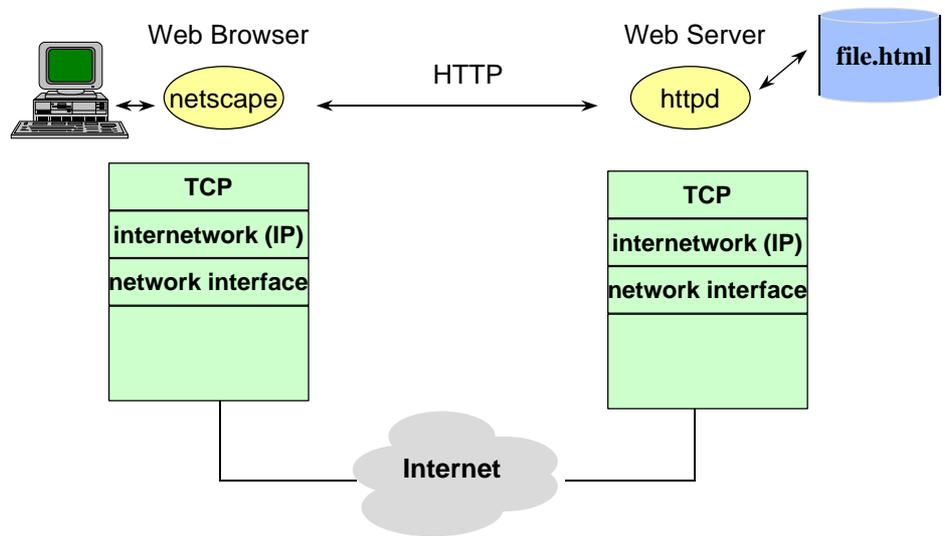
## World Wide Web (Web, WWW, W<sup>3</sup>)

- è un unico ipertesto distribuito sulla rete Internet
- si basa su diversi formati di memorizzazione delle informazioni, e qualsiasi tipo di informazione
  - » quello che dà le caratteristiche ipertestuali è HTML
- utilizza diversi protocolli applicativi di Internet (potenzialmente tutti)
  - » quello principalmente usato è HTTP
- **Storicamente:**
  - » nasce nel 1989 al CERN per la collaborazione su progetti di ricerca internazionali di fisica, con condivisione di documenti (di testo)
  - » 1991: prima demo pubblica, solo testo
  - » 1993: prima interfaccia grafica (NCSA Mosaic) -> successo
  - » 1994: nasce Netscape, CERN e MIT fondano il W3 Consortium
  - » 1995: Netscape, sebbene in rosso, raccoglie investimenti

Stefano Salsano

Marzo, 2002

## World Wide Web



Stefano Salsano

Marzo, 2002

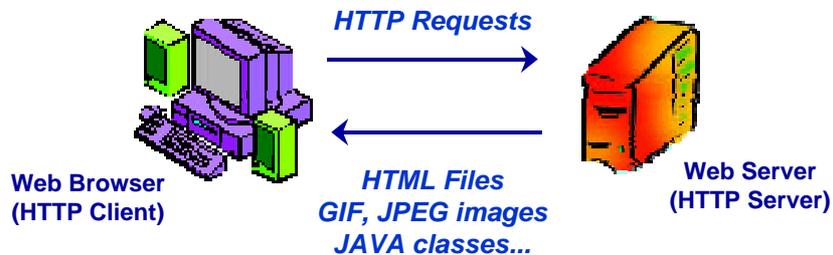
## Hyper Text Transfer Protocol (HTTP)

- **Principale protocollo utilizzato nel World Wide Web**
- **Utilizzato per la realizzazione di sistemi ipermediali distribuiti**
- **E' essenzialmente un protocollo per il trasferimento di file**
- **La porta standard è la 80**

Stefano Salsano

Marzo, 2002

## Il protocollo HTTP



- » le richieste sono costituite da un comando, una risorsa su cui applicare il comando, la versione del protocollo ed un eventuale messaggio MIME con parametri, informazioni sul client e contenuto
- » il server risponde con una linea di stato indicante almeno successo o fallimento del servizio, e un eventuale messaggio MIME con del contenuto
- » comandi: GET, HEAD, POST, PUT, DELETE

Stefano Salsano

Marzo, 2002

## Esempio di transazione HTTP

GET "www/index.html"

HTTP/1.0 404 Not Found

MIME-version: 1.0

Content-type: text/html

Server: QuidProQuo/2.1

<H3>The requested file was not found</H3> <HR>

<A HREF="http://www.socialeng.com/"> Quid Pro Quo  
2.1</A>

Stefano Salsano

Marzo, 2002

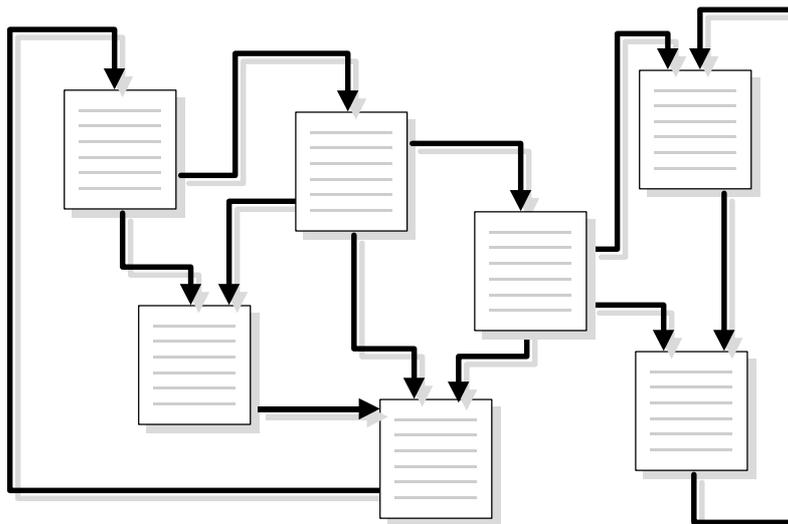
## Multimedia/Ipermedia

- I documenti multimediali comprendono informazioni provenienti da sorgenti diverse: testo, immagini, audio, video, ...
- I documenti ipermediali sono documenti multimediali con capacità ipertestuali, ovvero con possibilità di accesso non lineare
- Un nodo di un documento ipermediale sarà quindi costituito da testo, immagini, etc., ed un link potrà puntare a frammenti costituiti da questi particolari tipi di documenti

Stefano Salsano

Marzo, 2002

## Iper testo



Stefano Salsano

Marzo, 2002

## WWW: ipermedia distribuiti

- **WWW è un documento ipermediale distribuito:**
  - » distribuito perché lo spazio di memorizzazione fisica è costituito dai file system di computer siti ovunque e connessi ad Internet
  - » ipermediale perché i nodi possono essere costituiti da qualsiasi tipo di documento, il cui "tessuto" ipertestuale è dato da HTML
  - » i nodi ed il modo di accesso sono identificati da nomi appositi (URL)
- **Si basa sul principio client/server**
  - » il client si occupa della visualizzazione dei nodi, e dell'interpretazione delle richieste di navigazione dell'utente
  - » i dati sono invece mantenuti e distribuiti da appositi server in modalità diverse

Stefano Salsano

Marzo, 2002

## URL - Uniform Resource Locator

- **In generale, i nodi di un ipertesto devono essere identificati in qualche modo**
- **Se l'ipertesto è anche distribuito, allora i problemi sono:**
  - » dare un nome al nodo
  - » identificare dove il nodo è memorizzato
  - » indicare come accedere al nodo
- **L'URL serve a questo scopo, ed è fatto così:**
- **Schema://indirizzo.su.Internet/identificatore/locale/della/pagina**
  - » schema è il protocollo per l'accesso
  - » l'indirizzo identifica il calcolatore su Internet ove risiede il nodo (pagina)
  - » l'identificatore locale è usualmente la posizione del nodo nel file system locale, o sua abbreviazione

Stefano Salsano

Marzo, 2002

## URL e indirizzamento univoco

### URL: Uniform Resource Locator

**Sintassi:** *servizio://host.domain/nomeoggetto*  
*servizio:nomeoggetto*

**Esempi:** <http://www.uniud.it/www/welcome.html>  
<gopher://gopher.uniud.it/>  
<ftp://ftp.switch.ch/mirror/msdos/>  
<news:it.comp.aiuto>  
<mailto:brunetta@cc.uniud.it>

Stefano Salsano

Marzo, 2002 Slide 81

## La componente ipertestuale: HTML

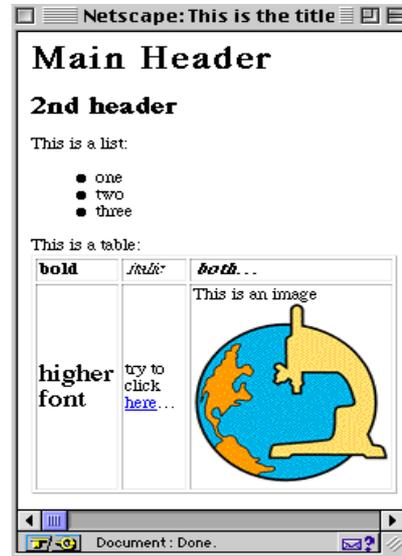
- HTML (HyperText Markup Language) è un linguaggio di markup, cioè i cui comandi sono inseriti esplicitamente all'interno del testo
- I comandi permettono la formattazione del testo similmente a quel che accade in un word processor, e però anche la creazione di link

Stefano Salsano

Marzo, 2002

## Un esempio

```
<HTML><HEAD>
<TITLE>This is the title</TITLE>
</HEAD>
<BODY BGCOLOR=White>
<H1>Main Header</H1>
<H2>2nd header</H2>
This is a list:
<UL><LI> one <LI> two
<LI> three</UL>
This is a table:
<TABLE BORDER>
<TR><TD><B>bold</B></TD>
<TD><I>italic</I></TD>
<TD><I><B>both...</B></I></TD>
<TD><FONT SIZE=5>higher
font</FONT></TD>
<TD>try to click <A
HREF="link.html">here</A>...</TD>
<TD>This is an image <BR>
<IMG SRC="img.gif"></TD>
</TABLE></BODY></HTML>
```



Stefano Salsano

Marzo, 2002

## Il browser

- Funge da interfaccia uniforme ad ogni servizio di rete
- Gestisce i protocolli indicati negli schemi degli URL, autonomamente o tramite programmi esterni detti helper
- Visualizza dinamicamente il testo ricevuto dai server secondo le istruzioni di formattazione specificate dai tag di HTML
- Visualizza almeno una serie di formati multimediali standard, autonomamente o tramite helper
- Interpreta le richieste dell'utente di selezione dei link, nascondendo ogni passo necessario alla selezione del protocollo, alla ricerca del server ed alla richiesta della risorsa
- Esempi: Explorer, Netscape, Opera, Lynx, Mosaic

Stefano Salsano

Marzo, 2002

## I server

- Ogni calcolatore che funge da server ha in genere in funzione un programma per ogni protocollo servito (http, ftp, ...)
- Le risorse distribuite da HTTP di solito consistono in
  - » files (HTML, JPEG, GIF, MPEG, ...) presenti sul file system locale in appositi sottoalberi
  - » a volte programmi di cui viene distribuito l'output (es. database, ed in generale CGI)
  - » a volte programmi che vanno ad essere eseguiti sul calcolatore client (classi JAVA)

Stefano Salsano

Marzo, 2002

## Nodi a contenuto attivo

- Sono stati sviluppati metodi per inserire contenuto attivo (cioè programmi) nelle pagine web
  - » CGI, ASF: programmi in funzione sul server, che emettono dell'output in corrispondenza di richieste dell'utente tramite form HTML
  - » Javascript: linguaggio di programmazione direttamente inserito nelle pagine HTML, ed interpretato dal browser
  - » Java: linguaggio di programmazione completo, gestito da un'apposita macchina virtuale richiamata automaticamente dal browser quando incontra un'applet

Stefano Salsano

Marzo, 2002

## Alcune caratteristiche di WWW

- **E' un grande ipertesto**
  - » costituito da decine di milioni di pagine
  - » non molto ordinato
  - » eterogeneo per contenuti e stile
- **(Internet) non ha un proprietario centrale**
  - » è possibile pubblicare/trasmettere tutto (netiquette)
  - » non viene garantita la qualità del servizio
  - » gli utenti si collegano da ovunque ed accedono ovunque
  - » pubblicazione/trasmissione sono a basso costo
  - » facile ed ampia diffusione di qualsiasi prodotto elettronico
- **Problemi**
  - » difficoltà di reperimento delle informazioni utili
  - » non tutti gli "utenti" sono in buona fede

Stefano Salsano

Marzo, 2002

## Ricerca delle informazioni

- **Buona parte delle tante informazioni presenti su Internet (via Web, ftp, ...) non è strutturata, ma testuale e libera**
- **La ricerca per parola chiave viene usualmente fornita da appositi siti:**
  - » Motori di ricerca libera
  - » Portali

Stefano Salsano

Marzo, 2002

## Motori di ricerca su Web

- **Permettono di trovare pagine Web che soddisfino particolari criteri sul contenuto (presenza di parole specifiche, connettori logici, ...)**
- **Gli indirizzi delle pagine vengono mostrati all'utente secondo un ordine di rilevanza**
- **Esempi di motori a ricerca libera**
  - » [www.altavista.com](http://www.altavista.com),
  - » [www.lycos.com](http://www.lycos.com)
  - » [www.virglio.it](http://www.virglio.it)
  - » [www.google.com](http://www.google.com)
  - » Metacrawler, Savvy Search: usano gli altri motori
- **Differenze: criteri di ricerca e visualizzazione**

Stefano Salsano

Marzo, 2002

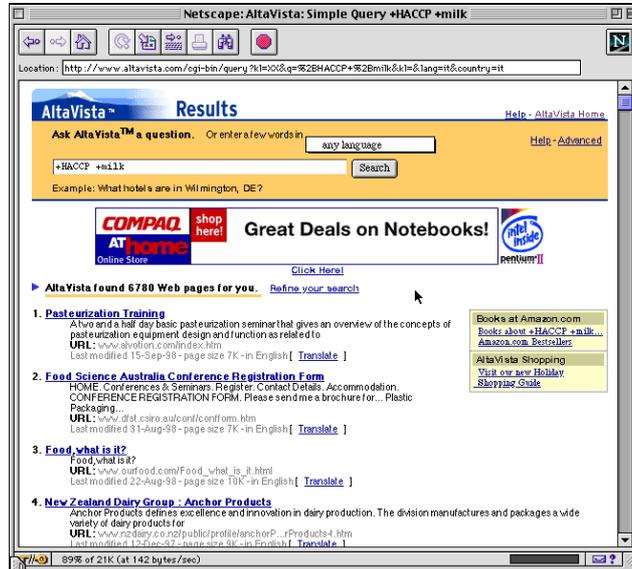
## Problemi dei motori di ricerca

- **Si trovano pagine che contengono delle parole, indipendentemente dal loro uso**
  - » quindi si possono trovare pagine che citano casualmente ciò che cerchiamo
- **Pur essendo aggiornato in continuazione, inevitabilmente il database**
  - » contiene riferimenti a pagine non più esistenti
  - » contiene riferimenti a pagine con contenuto cambiato
  - » non contiene riferimenti alle pagine più recenti
  - » non contiene riferimenti a pagine spostate

Stefano Salsano

Marzo, 2002

## Motori di ricerca su Web



Stefano Salsano

Marzo, 2002

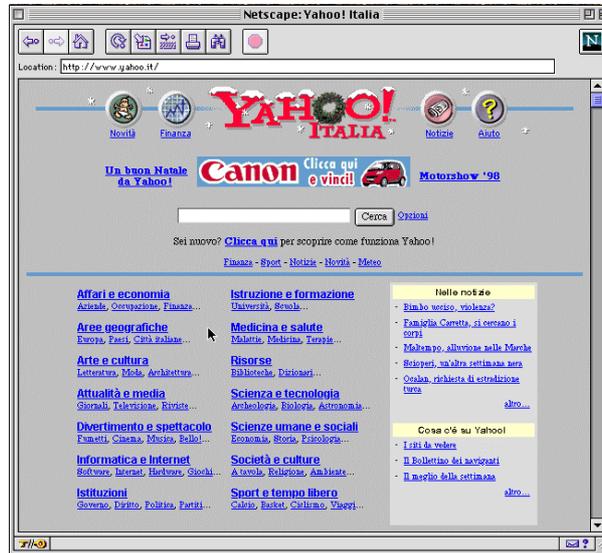
## Portali

- **Alcuni siti forniscono un servizio di catalogazione sistematica delle risorse, in cui ogni pagina viene valutata ed associata ad una particolare categoria (usualmente determinata gerarchicamente)**
  - » in pratica: sovrapposizione di struttura di link organizzativi
  - » in modo simile a quanto avviene nelle catalogazioni bibliografiche
- **Spesso si associano a motori di ricerca tradizionale**

Stefano Salsano

Marzo, 2002

# Portali



Stefano Salsano

Marzo, 2002