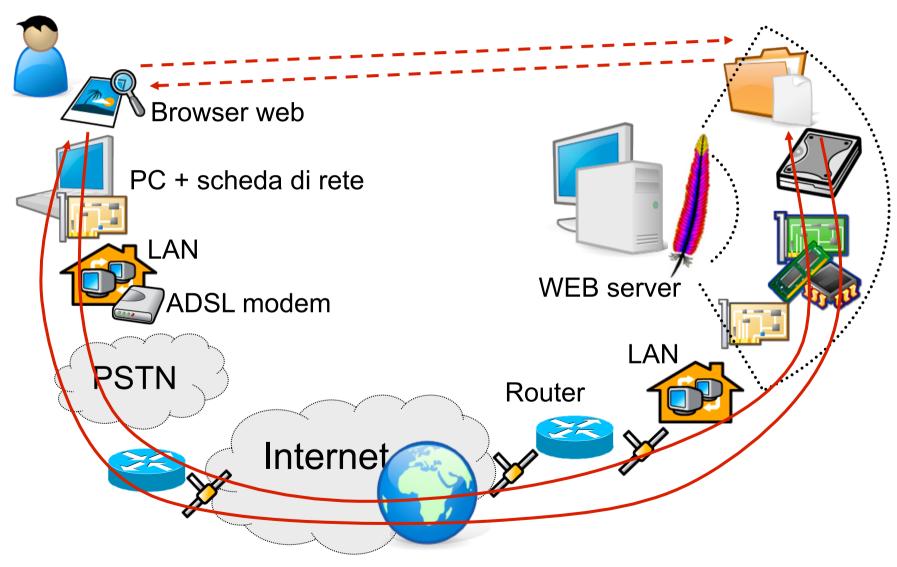


# Uso di Internet: introduzione

### Consultazione di una pagina WEB

- Un utente (U) è interessato a reperire una certa risorsa (R) in Internet
  - Risorsa = pagina di testo, immagine, video, documento multimediale, software
- Apre il proprio Browser Web (B)
  - Esistono numerosi "software" che svolgono la funzione di Browser (Es: Explorer, Firefox, Opera, ...)
- B per la rete è una Applicazione
  - B è utente di un terminale di rete (il calcolatore)
  - B deve avere una procedura di comunicazione per cercare e reperire R
- La risorsa web risiede su un server web (S)
  - S è anch'esso un'applicazione
  - S deve avere una procedura per rispondere alle richieste di B e consegnare R
- Ottenuta R il browser B la mostra all'utente finale U

### Chi è coinvolto?



#### Indirizzamento

- La risorsa R è univocamente identificata da un indirizzo
- Uniform Resource Locator (URL)
  - URL è un indirizzo complesso che riflette l'organizzazione a livelli della rete

Protocollo di dialogo con l'applicazione

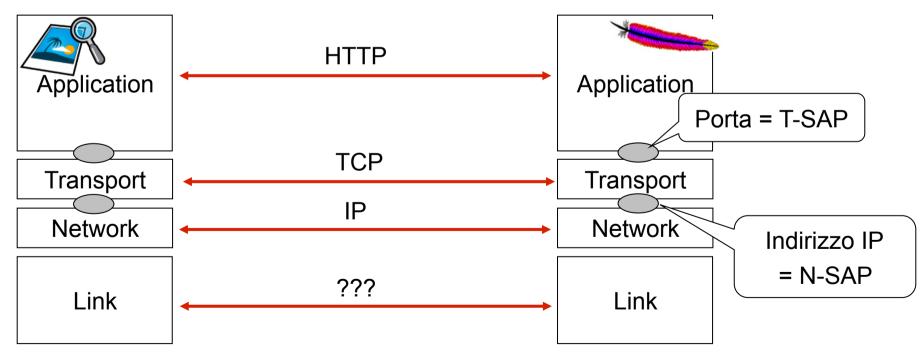
Nome simbolico del server web:
Indirizzo di rete (N-SAP)

Protocollo di dialogo con l'applicazione

Nome simbolico del server web:
Indirizzo di rete (N-SAP)

#### Protocolli ed interfacce

- Le applicazioni sono locali al calcolatore (terminale)
  - Alcune parti dell'URL hanno validità locale
    - Numero di porta
    - · Percorso nel filesystem
- Il calcolatore va identificato univocamente su Internet
  - Almeno una parte dell'indirizzo deve avere significato unico e universal
    - Indirizzo di rete (numero IP)



### Esempi di URL

- Interrogazione del server deisnet.deis.unibo.it utilizzando porte e percorsi ai documenti standard e non standard
- 1 porta 80 (default), file index.php (default)

http://deisnet.deis.unibo.it/Didattica/CorsiCE/RetiLA/testpage/

• 2 - porta 80 (default), file non di default

http://deisnet.deis.unibo.it/Didattica/CorsiCE/RetiLA/testpage/testpage.php

3 - porta 12345 (non default), file di default, URL diverso

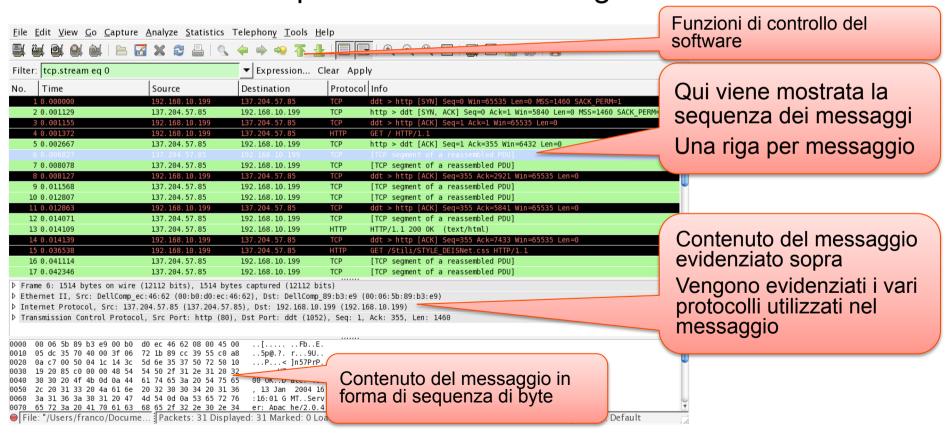
http://deisnet.deis.unibo.it:12345/testpage/

4 - porta 12345 (non default), file non di default

### Analisi di protocollo

 Esistono strumenti software per analizzare il traffico di rete

Wireshark http://www.wireshark.org/



### Protocollo applicativo

- Le applicazioni browser e server sono solitamente implementate per essere capaci di utilizzare diversi protocolli
  - HTTP (HyperText Transfer Protocol)
  - HTTPS (HyperText Transfer Protocol over Secure Socket Layer)
  - FTP (File Transfer Protocol)
  - **—** ...
- L'URL indica al browser quale protocollo utilizzare fra quelli a lui disponibili
  - HTTP è la scelta di default ed anche quella più utilizzata

#### HTTP

- Il messaggio HTTP è un puro messaggio di testo costruito come segue
  - Linea iniziale
  - Una o più linee di intestazione (header)
  - Una linea vuota (un carattere CRLF)
  - Un corpo del messaggio di tipo opzionale

linea iniziale, dipende da request a response>

Header1: valore1

Header2: valore2

Header3: valore3

< corpo del messaggio, può contenere testo o dati binari >

### Dialogo fra Entità http

- HTTP funziona con un meccanismo di richiesta/risposta
  - Non viene mantenuto uno stato della comunicazione (connectionless)
- Tipicamente una richiesta del Client a cui segue una risposta del Server
- Nel caso della richiesta di una pagina web
  - Il Client manda una <u>request</u> di tipo <u>GET</u> che chiede la pagina corrispondente all'URL (e inviando intanto vari parametri di set up)
  - Il Server invia la <u>response</u> (ancora con vari parametri di set up) con il contenuto della pagina
    - Il testo viene inviata direttamente come sequenza di caratteri, seguendo le regole del linguaggio HTML (HyperText Mark-up Language)
    - Per le immagini e gli eventuali altri oggetti il server invia l'URL dei files in cui sono contenuti
    - Il Client invia una GET per ottenere le altre componenti della pagina

#### Richiesta

Esempio di richiesta:

```
GET /Didattica/CorsiCE/RetiLB/
    index.html HTTP/1.1 ←
                                    metodo, file, versione
Accept: */* ←
                                    contenuto accettato
                                    preferenza linguistica
Accept-Language: en-us ←
                                    ultima versione nella cache
If-Modified-Since:
  Wed, 16 Jan 2002 16:37:40 GMT
User-agent:Mozilla/4.0 ←
                                  __ tipo di browser
                                   — host
Host: deisnet.deis.unibo.it ←
Connection: Keep-Alive ←
                                  ___ connessione permanente
```

### Risposta

#### Esempio di risposta positiva:

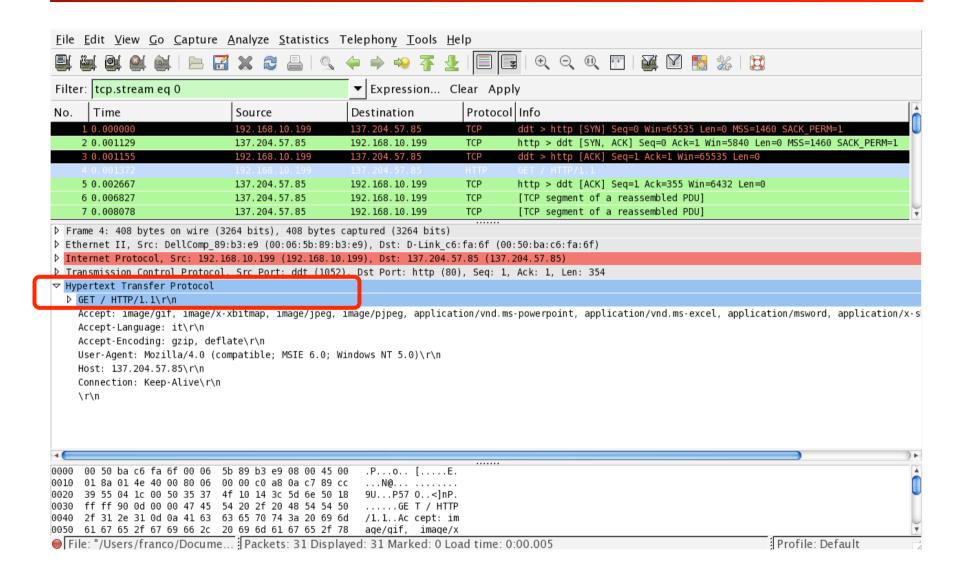
```
HTTP/1.1 200 OK
Date: Wed, 03 Mar 2004 17:37:44 GMT
Content-Length: 19692
Content-Type: text/html
Server: Apache/2.0.40 (Red Hat Linux)
Last-Modified: Mon, 01 Mar 2004 16:02:27 GMT
<html>
qui c'è il testo HTML della pagina richiesta (19692 byte)
</html>
```

### Risposta

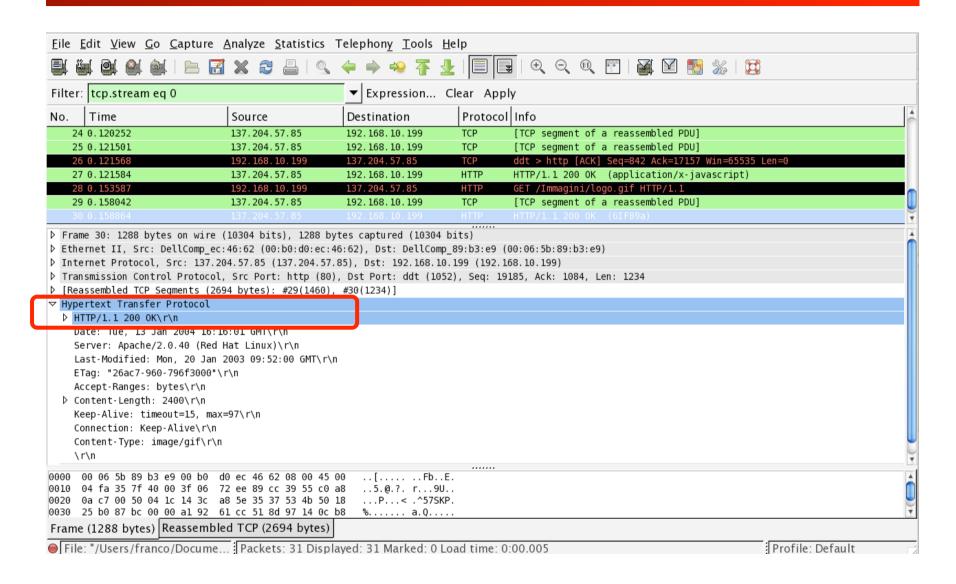
Esempio di risposta negativa (il file richiesto non esiste):

```
HTTP/1.1 404 Not Found
Date: Wed, 03 Mar 2004 17:38:37 GMT
Content-Length: 1067
Content-Type: text/html
Server: Apache/2.0.40 (Red Hat Linux)
<html>
Object not found
Error 404
</html>
```

#### Richiesta

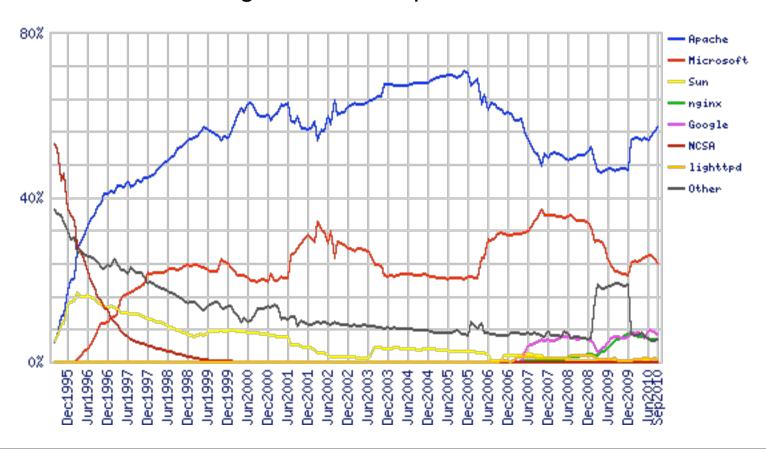


### Risposta



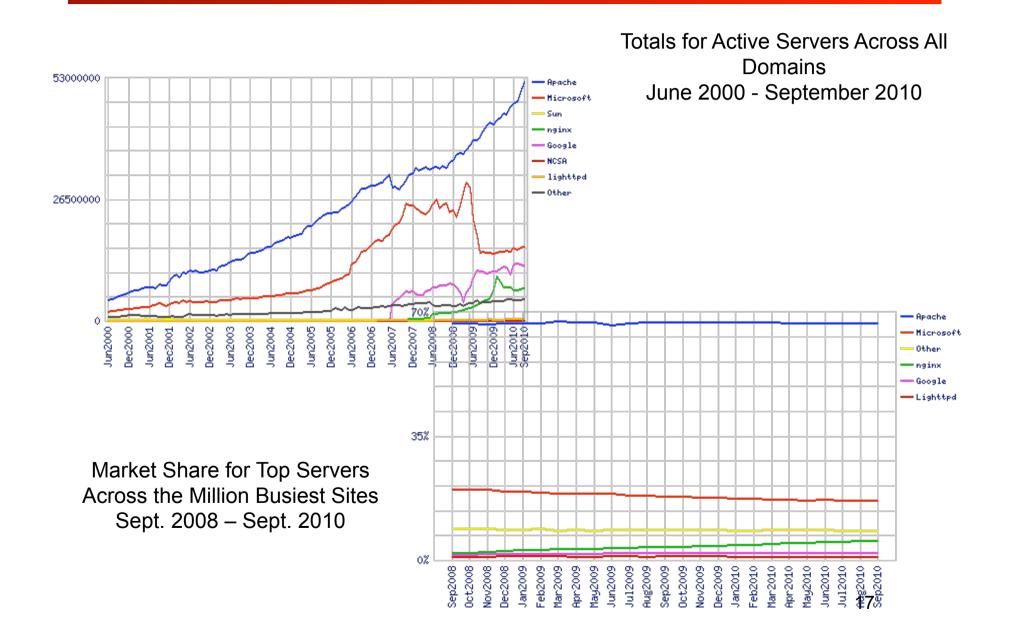
#### Diffusione Server HTTP

#### Market Share for Top Servers Across All Domains August 1995 - September 2010



Source: http://news.netcraft.com/archives/web\_server\_survey.html

#### Diffusione Server HTTP

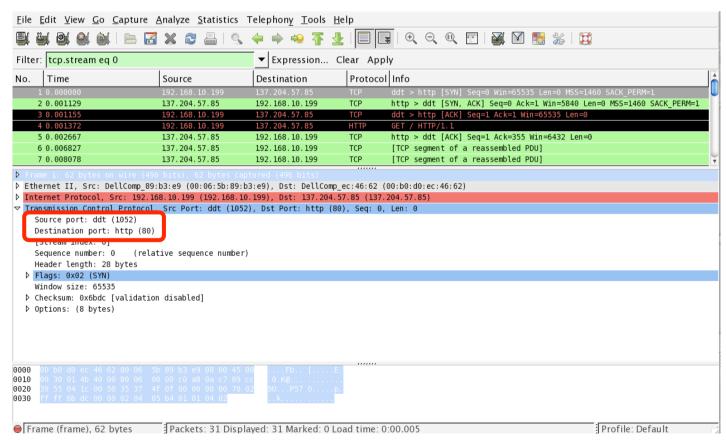


### Protocollo di trasporto

- Il protocollo di trasporto si occupa del trasporto dei dati end-to-end
- Può trasportare i dati pertinenti ad una qualunque applicazione
- I flussi dati di diverse applicazioni sono distinguibili sulla base del numero di porta
- Esistono diversi protocolli di trasporto
  - UDP User Datagram Protocol
  - TCP Transmission Control Protocol
  - RTP Real Time Transmission Protocol
    - Il protocollo di trasporto per una determinata istanza applicativa viene scelto in funzione delle caratteristiche che il trasporto dei dati deve avere

### Numero di porta

- Indirizzo di 16 bit
  - Valori decimali da 0 a 65535
- Locale al singolo calcolatore, ripetute su tutti i calcolatori
- Condiviso fra tutti i protocolli di trasporto



### Classificazione dei numeri di porta

- Regole d'uso
  - Da 1 a1023 (in origine da1 a 255): Riservati
    - possono essere usati solo dai server
  - da 1024 a 49151: Registrati
    - Sono usati da alcuni servizi ma anche da client
  - Da 49151 a 65535: ad uso dei client
- Una parte dei numeri di porta sono riservati (Well Known Ports)

#	Protocol		Servizio
21	FTP-CONTROL	File Transfer Protocol	Trasferiento file (control)
20	FTP-DATA	File Transfer Protocol	Trasferimento files (dati)
23	TELNET		Accesso via terminale
25	SMTP		Trasferimento di posta elettronica
53	DNS	Domain Name System	Accesso al DNS
80	HTTP		Web server
109	POP2	Post Office Protocol (Version 2)	Lettura posta elettronica
22	SSH	Secure Socket	Accesso via terminale cifrato
110	POP3	Post Office Protocol (version 3)	Lettura posta elettronica
137	NETBIOS Name Service.		Convisio di roto por applicazioni in
138	NETBIOS Datagram Service.		Servizio di rete per applicazioni in ambiente DOS (Windows)
139	NETBIOS Session Service.		
443	HTTPS	HTTP over SSL/TLS	WEB cifrato

### IANA (Internet Assigned Numbers Authority)

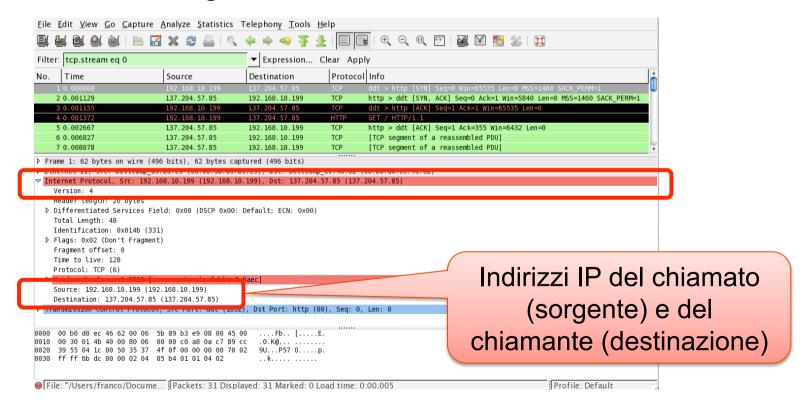
```
PORT NUMBERS
(last updated 2010-09-21)
The port numbers are divided into three ranges: the Well Known Ports,
the Registered Ports, and the Dynamic and/or Private Ports.
The Well Known Ports are those from 0 through 1023.
DCCP Well Known ports SHOULD NOT be used without IANA registration.
The registration procedure is defined in [RFC4340], Section 19.9.
The Registered Ports are those from 1024 through 49151
DCCP Registered ports SHOULD NOT be used without IANA registration.
The registration procedure is defined in [RFC4340], Section 19.9.
The Dynamic and/or Private Ports are those from 49152 through 65535
```

#### **Transmission Control Protocol**

- È un protocollo connection-oriented
- Connessione = associazione di 4 indirizzi
  - Numero di porta (T-SAP) e numero di rete (N-SAP) dei due host coinvolti nella comunicazione
- Come previsto dal modello connection-oriented
  - Esistono funzioni di controllo delle "connessioni"
  - Per ogni connessione sono previste le procedure per
    - l'instaurazione
    - il controllo del corretto andamento
    - la chiusura

### Protocollo di rete

- Garantisce il corretto indirizzamento ed instradamento dei dati
- Deve necessariamente essere unico in una rete globale
- Internetworking Protocol IP



### L' indirizzo IP

- Indirizzi di lunghezza fissa pari a 32 bit
- Scritti convenzionalmente come sequenza di 4 numeri decimali, con valori da 0 a 255, separati da punto (rappresentazione dotted decimal)

```
10001001.11001100.11010100.00000001
137.204.212.1
```

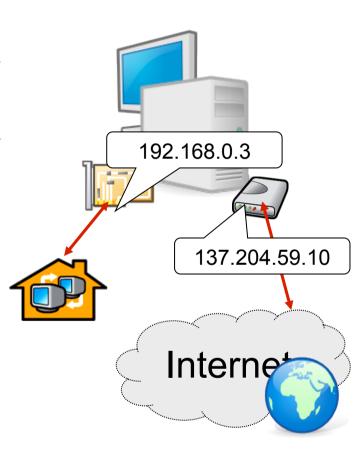
Numero teorico max. di indirizzi

$$2^{32} = 4.294.967.296$$

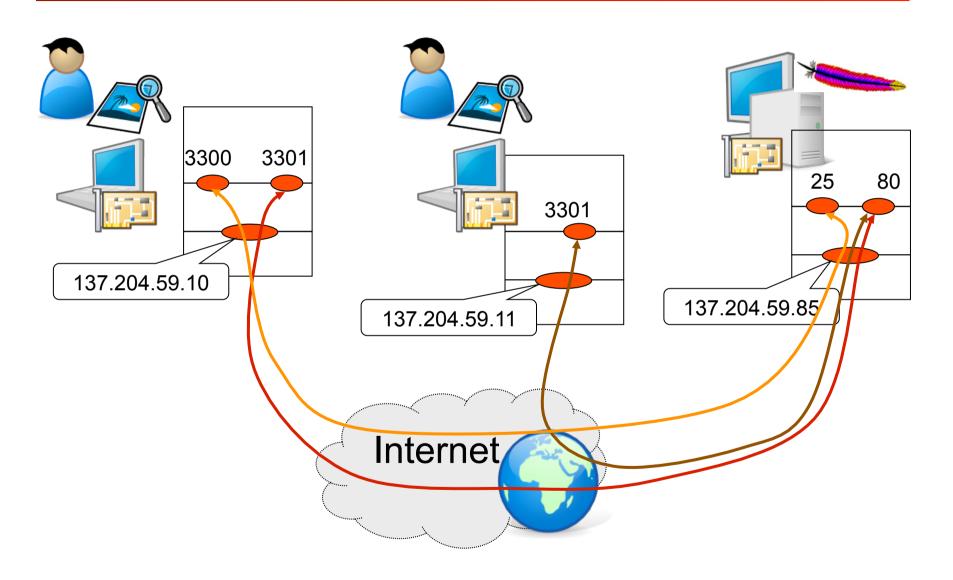
In realtà si riesce a sfruttare un numero molto inferiore

#### Indirizzi e interfacce di rete

- L'indirizzo identifica i punti di interconnessione di un host con la rete
  - Non identifica un host individuale, ma una delle sue interfacce di rete
- Multi-homed hosts
  - host con due o più interfacce di rete
- Esempio: un router che collega N reti ha
  - N interfacce di rete
  - N distinti indirizzi IP, uno per ogni interfaccia di rete



### Connessioni



#### L'interfaccia Socket

- Gli standard non specificano come gli applicativi debbano interagire con i protocolli
- L'interfaccia fra applicazione e TCP
  - non è standardizzata
  - dipende dall'implementazione del sistema operativo
  - viene comunemente chiamata Socket
- Si dice Indirizzo della Socket (Socket Address) un Numero di porta concatenato ad un indirizzo di rete

Esempio: 137.204.57.1:80

### Implementazioni dei servizi in Internet

- Comunicazioni fra calcolatori (Host) = scambio di messaggi fra processi applicativi (Applicazioni)
  - Un messaggio in arrivo ad un host è utilizzabile se è in esecuzione (running) un processo applicativo che legge il messaggio e sa cosa farsene

#### Client-server

- Nel modello classico gli host in rete sono classificabili in due tipologie:
  - Server: mettono a disposizione risorse di elaborazione e dati
  - *Client*: ospitano applicazioni che, al fine di svolgere le relative funzioni, si connettono ai server per ottenere risorse ed informazioni
- La variante Peer-to-peer (P2P)
  - Gli host in rete sono tutti equivalenti (peer, appunto) e fungono alternativamente sia da client che da server verso altri nodi
    - In una rete P2P qualsiasi nodo utilizza e mette a disposizione contemporaneamente risorse ed informazioni in rete

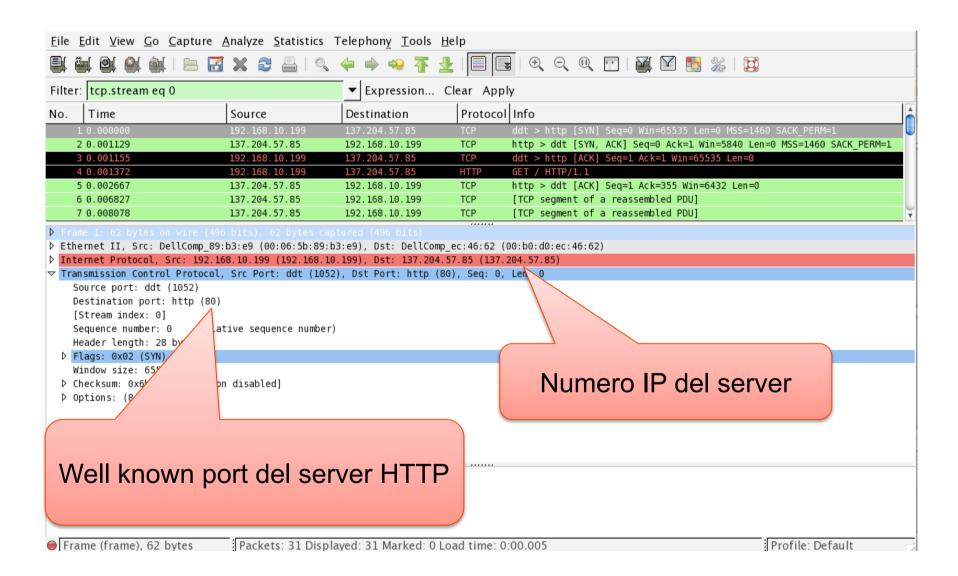
#### Client-server

- Il processo Server si predispone a ricevere una connessione eseguendo una apertura passiva
  - Crea una socket e si mette in ascolto in attesa dell'arrivo di una richiesta di connessione (questo processo nel mondo Unix è chiamato Demone)
- Il processo Client esegue una apertura attiva tentando di collegarsi al processo server di destinazione

#### La ricerca della destinazione

- Il client deve conoscere indirizzo IP e il Numero di porta del server di destinazione
  - Come fa a scoprirli?
- Nell'URL sono specificati
  - Protocollo applicativo
    - A cui corrisponde una well known port
  - Eventuale numero di porta non standard
  - II numero IP o il nome del server
    - Il nome deve tramutarsi in un numero IP
    - Il numero IP identifica in modo univoco il punto di accesso alla rete del server
- Come fa il nome a diventare un numero?

## Esempio



#### In conclusione

- L'utente finale interagisce con il software di applicativo
- L'applicazione dialoga con una o più applicazioni remote utilizzando i protocolli applicativi necessari
- I protocolli applicativi sfruttano il servizio di trasporto di uno dei protocolli di trasporto per raggiungere l'applicazione remota
- Il protocollo di trasporto utilizza le capacità di instradamento di IP per la consegna dei dati al calcolatore remoto dove risiede l'applicazione
- IP consegna i dati sfruttando l'infrastruttura di rete a cui gli host sono connessi tramite l'interfaccia di rete