

Introduzione a Internet e World Wide Web

Sommario

- Breve storia di Internet
- Commutazione di pacchetto e TCP/IP
- Il Web
 - HTTP
 - HTML
 - CGI
 - ...
- Connessione tra basi di dati e Web

Internetworking

(collegamento fra reti e sistemi eterogenei)

- Nasce da un progetto militare (DARPA)
→ ARPAnet (1975-80)
- Con l'aggregazione di laboratori di ricerca e università USA diventa Internet
- Sviluppo:
 - reti scientifiche nazionali e continentali
 - aggregazione reti private / commerciali
- Salto di qualità (anni '90):
Tecnologia **World Wide Web**

Internet: una definizione

Sistema globale di informazione che:

- è collegato da un unico spazio di indirizzi basato sul protocollo **IP**
- consente di comunicare mediante protocollo **TCP/IP** o sue estensioni
- rende accessibili, pubblicamente o privatamente, servizi basati sulla infrastruttura descritta

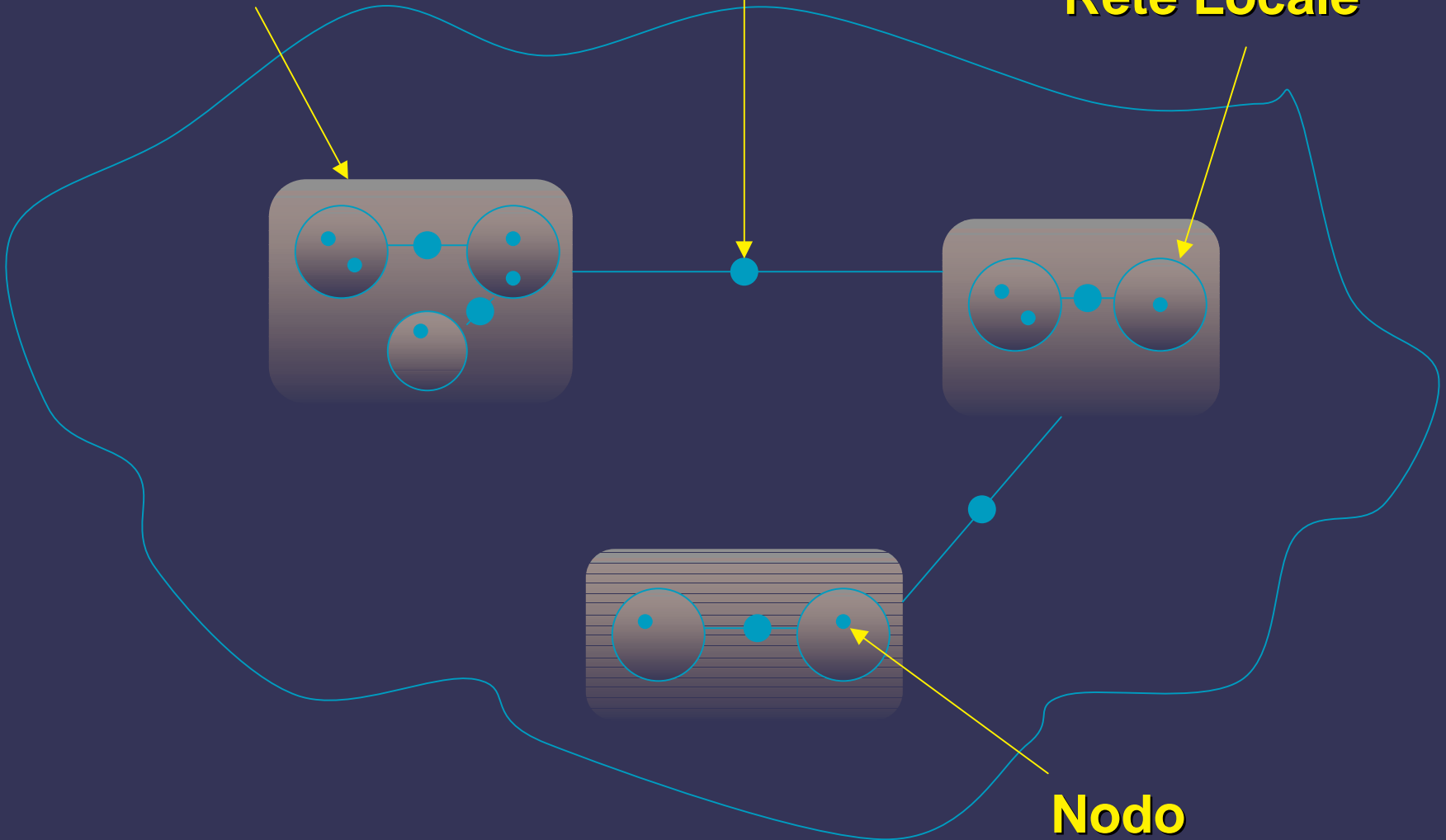
*Federal Networking Council,
24/10/1995*

Internet come federazione di reti

Rete Componente

Router

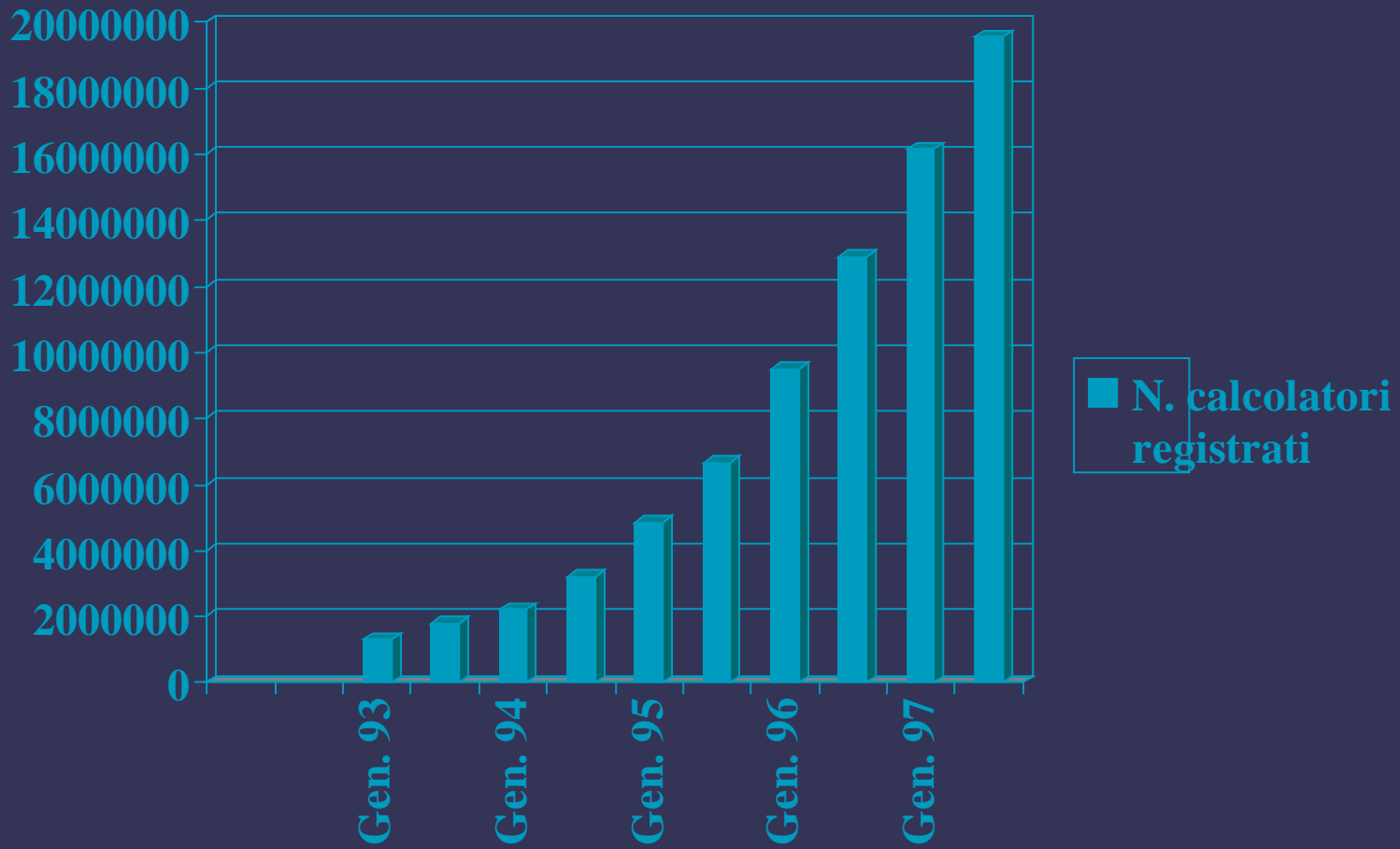
Rete Locale



Internet come federazione di reti

- Ogni nodo ha un indirizzo numerico (**Indirizzo IP**)
es. **137.204.57.150**
- Ad ogni nodo è associato un **nome simbolico** utilizzato nelle applicazioni per riferirsi ad esso
es. **www-db.deis.unibo.it**
- La conversione da nome simbolico ad indirizzo IP è gestita dai **Domain Name Server** (**deis.unibo.it** è un nome di **dominio**)

Lo sviluppo di Internet



World Wide Web: una definizione

Collezione distribuita di documenti ipertestuali e multimediali disponibili su Internet.

Alla base di WWW c'è:

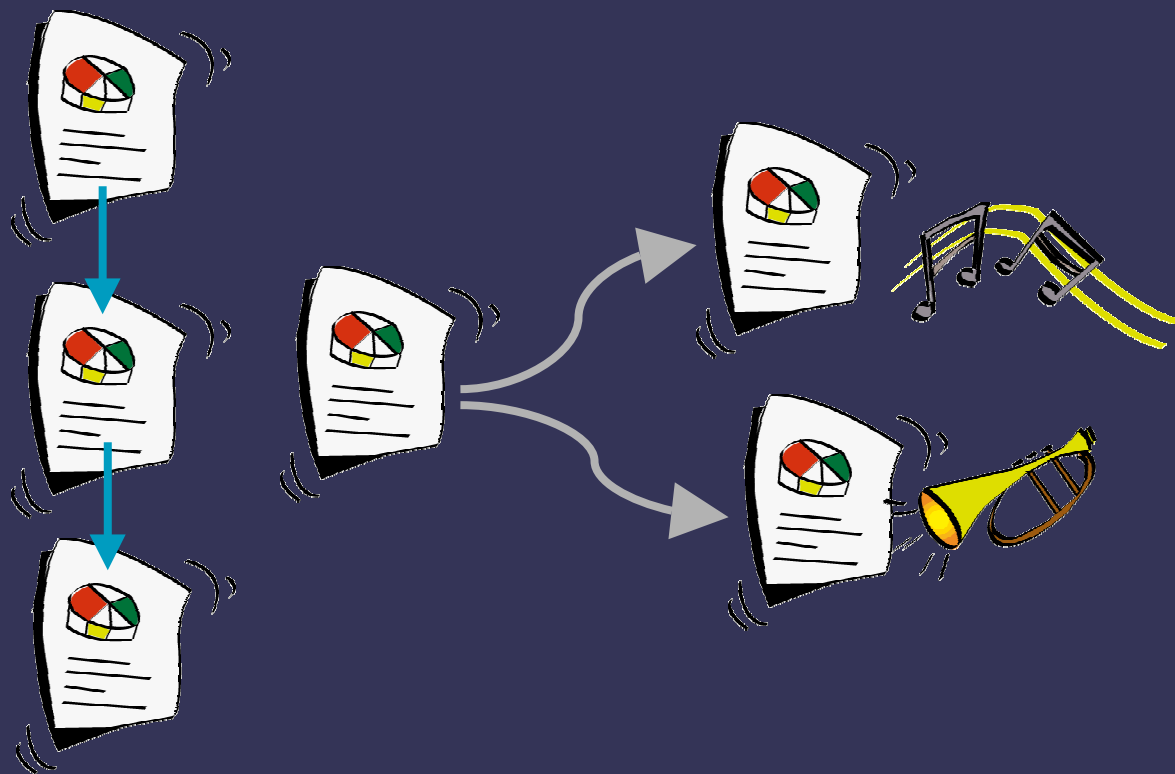
- Uno standard per identificare risorse di rete e poterle specificare all'interno di documenti ipertestuali (**URL** - **U**niform **R**esource **L**ocator)
- Uno o più linguaggi per la realizzazione di documenti ipertestuali (**HTML** - **H**yper**T**ext **M**arkup **L**anguage, XML...)
- Un protocollo di comunicazione client-server per lo scambio di documenti ipertestuali (**HTTP** - **H**yper**T**ext **T**ransfer **P**rotocol)

Iper testo (multimediale)

Documento con una struttura non sequenziale, costituito da varie porzioni fra loro collegate, al fine di favorirne la consultazione a seconda delle esigenze.

Può contenere:

- testo
- immagini
- video
- suoni
- ecc.



Protocolli e linguaggi di Internet

TCP / IP, HTTP, HTML

Il protocollo TCP/IP

- **T**ransfer **C**ontrol **P**rotocol / **I**nternet **P**rotocol
- Insieme di protocolli (**suite**) strutturati in maniera gerarchica adibiti alla comunicazione (in Internet)
- Ogni livello della gerarchia (detta **PILA**) usa i servizi forniti dal livello inferiore e offre servizi al livello superiore
- Fra livelli contigui è definita una interfaccia di comunicazione

La pila TCP/IP

application layer offre servizi utilizzabili dalle applicazioni o dall'utente (es. telnet, ftp, smtp, http)

transport layer fornisce meccanismi di comunicazione tra applicazioni (TCP, UDP)

internet layer fornisce il meccanismo di trasmissione dati attraverso la rete Internet (IP)

network interface layer nasconde i dettagli di accesso alla rete fisica ai livelli superiori

TCP e IP

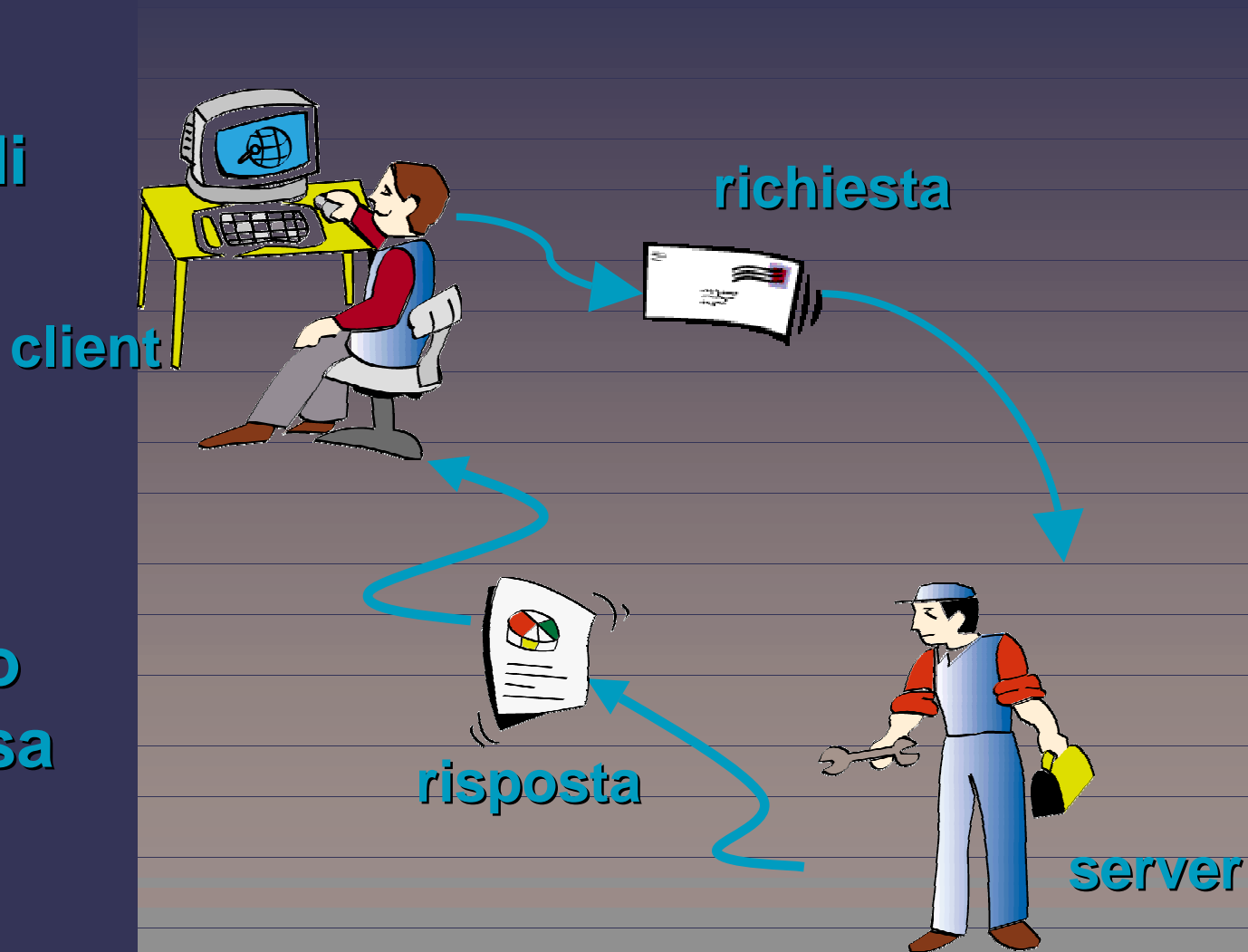
- **I**nternet **P**rotocol
 - si occupa della spedizione in rete di unità di informazione (**datagram**)
 - definisce il meccanismo di indirizzamento dei nodi di Internet (**IP address**)
- **T**ransfer **C**ontrol **P**rotocol
 - converte il flusso dati proveniente dalla applicazione in pacchetti (**segment**) trasmissibili in rete
 - verifica integrità e correttezza dei dati trasmessi (recupero/ritrasmissione pacchetti persi, rimozione duplicati ...)

Il protocollo HTTP

- **HyperText Transfer Protocol**
- Protocollo a livello di applicazione per lo scambio di ipertesti multimediali
- Prescrive il formato di
 - nomi delle risorse (**URI, URL**)
 - domande
 - risposte
- Versioni: HTTP/0.9, 1.0, 1.1
- Riferimento: Tim Berners Lee, Request for Comment 1945, HTTP/1.0

Gli attori di HTTP

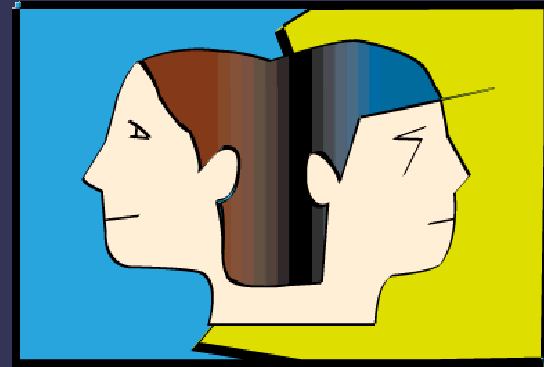
- client
(browser):
genera
richieste di
risorse



- server:
depositario
della risorsa
richiesta

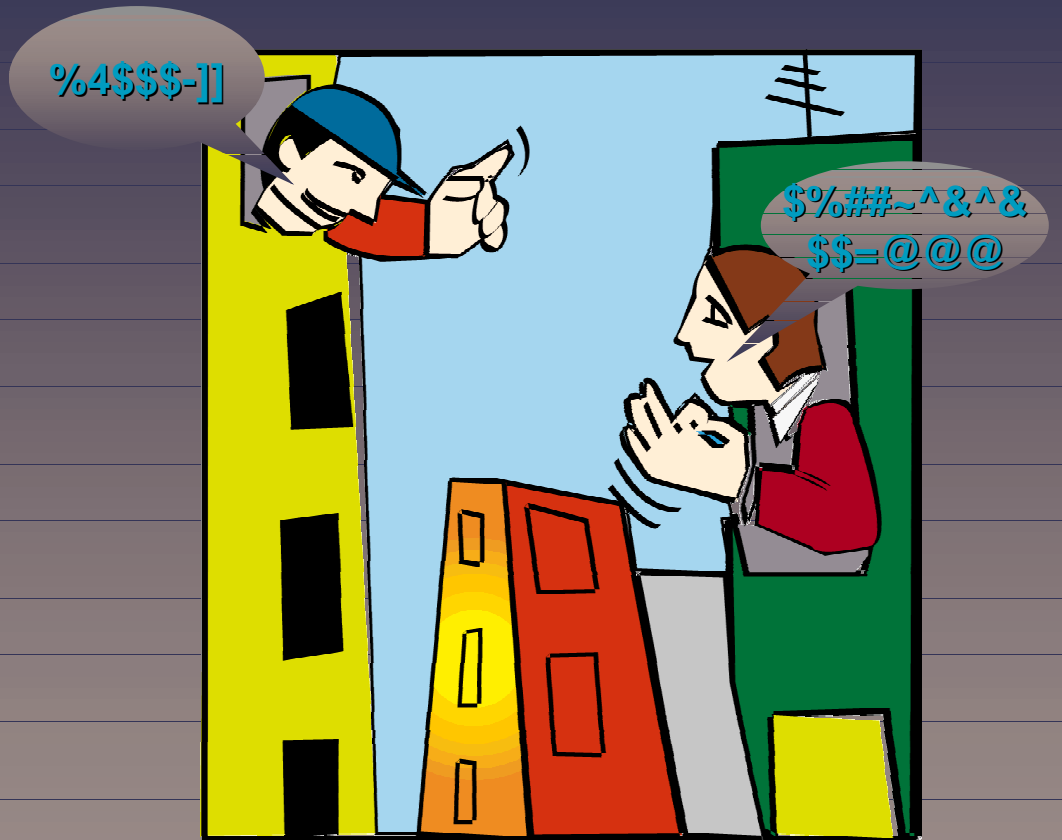
Gli attori di HTTP

- **proxy:**
nodo
intermedio
che può
fungere sia
da client che
da server



Gli attori di HTTP

- gateway:
nodo
intermedio
che può
agire solo
da server,
per tradurre
richieste
in altri
protocolli
o a fini di
protezione



Uniform Resource Locator (URL)

- E' una stringa strutturata, ad esempio:
`http://www-db.deis.unibo.it/~fgrandi/index.html`
- Protocollo: http, ma anche ftp e gopher
- Indirizzo della macchina:
 - simbolico: `www-db.deis.unibo.it`
 - numerico (IP): `137.204.57.150`
 - può includere il numero di porta (es. `:8080`)
- Path: sequenza di direttori
- Nome risorsa: identificativo di un file
 - se la risorsa e' un file html, puo' includere un indirizzo interno, es. `index.html#curriculum`

Formato delle risposte

- **full-response** :- **status-line**
 *(general-header |
 request-header |
 entity-header)
CRLF [entity-body]
- **status-line** :-
 version SP status SP message CRLF
- **Codici di stato:**
1XX (informativo), 2XX (successo),
3XX (redirezione), 4XX(errore client),
5XX (errore server)

Formato dei documenti

- I documenti risiedono sul server come file e vengono inviati al client usando il campo `entity-body` della risposta
- Es: Iper testo = file ASCII scritto secondo la sintassi HyperText Markup Language (HTML)
- Estensioni tipiche del file: `*.html` `*.htm`
- Il documento può contenere inserti multimediali (immagini, audio, video)
 - il programma client carica e visualizza il testo
 - poi richiede al server i file corrispondenti agli inserti multimediali

Il Browser (client HTTP)

- Applicazione in grado di:
 - accedere alla rete secondo il protocollo HTTP
 - richiedere risorse identificate da un URL a un server
 - interpretare e rendere a video la risposta del server
 - gestire la navigazione (click su link)

