Master in Informatica del Testo ed Edizione Elettronica

Master professionale e di ricerca

Università degli Studi di Siena

Filologia e Critica delle Letterature Antiche e Moderne – Ingegneria dell'Informazione - Scienze Matematiche

ELEMENTI DI INFORMATICA DI BASE

Anno 2022

Dario Madeo

dario.madeo@unisi.it

https://www3.diism.unisi.it/~madeo/home.html

PROTOCOLLI E WEB

27 Gennaio 2022

Rivisitazione delle lezioni del prof. Stefano Melacci (2015)

Testo di riferimento

Tanenbaum Andrew S., **"Reti di calcolatori"**, Pearson Education Italia

Laura Lemay, Rafe Colburn, **"HTML & CSS"**, McGraw Hill

Outline

- Protocolli di trasmissione
- Web e HTML

Protocolli di trasmissione

Modello ISO-OSI

- OSI – Open System Interconnection

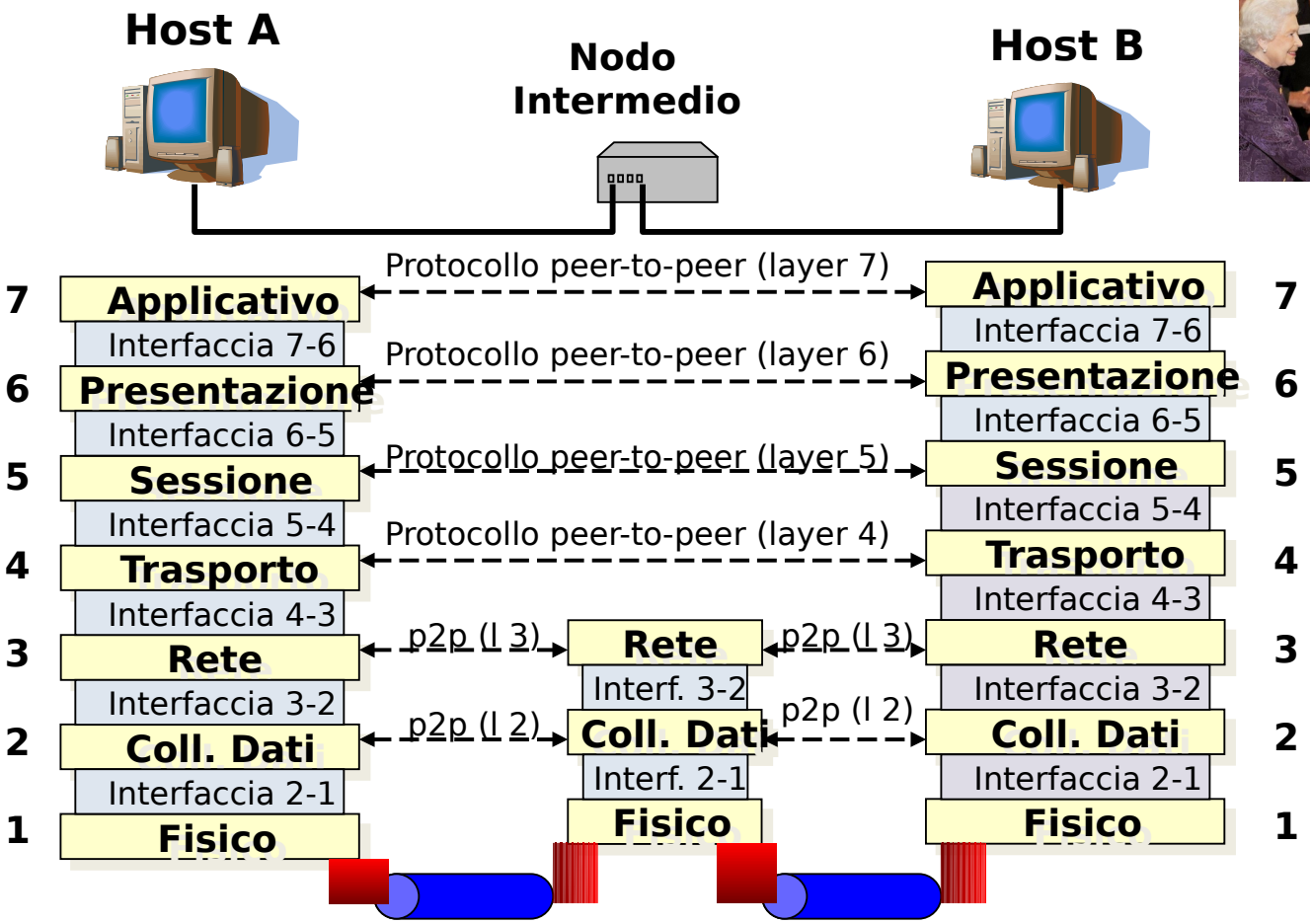
  E’ un modello per la progettazione di un’architettura di rete flessibile, robusta e inter-operativa

  Utilizza un riferimento a strati (layer) che descrivono la comunicazioni a livelli di astrazione successivi (stack OSI)

  A ciascun livello devono corrispondere a funzioni definite diverse da quelle degli altri strati

  Ciascun livello utilizza i servizi dello strato inferiore

Livelli del modello ISO-OSI



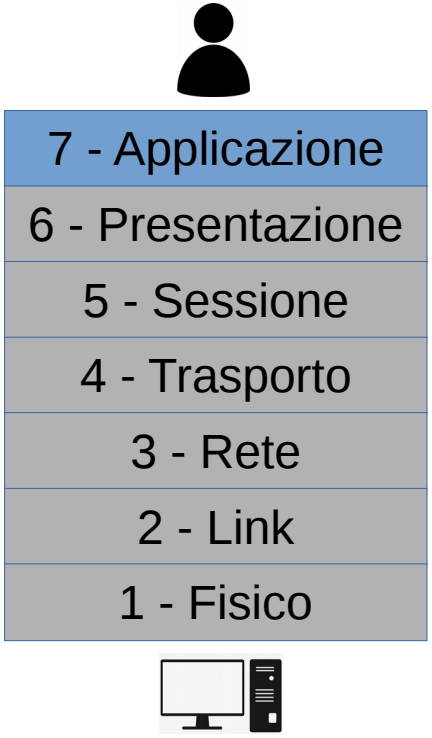
Livello 7 – Applicazione

Interfaccia l’utente con la macchina

Forrnisce un insieme di protocolli che operano a stretto contatto con le applicazioni.

Trasferimento file (FTP)
Trasferimento pagine internet (HTTP)
Posta elettronica (SMTP, POP)
Secure Shell (SSH)
Internet Relay Chat (IRC)

Protocolli diversi per applicazioni diverse!



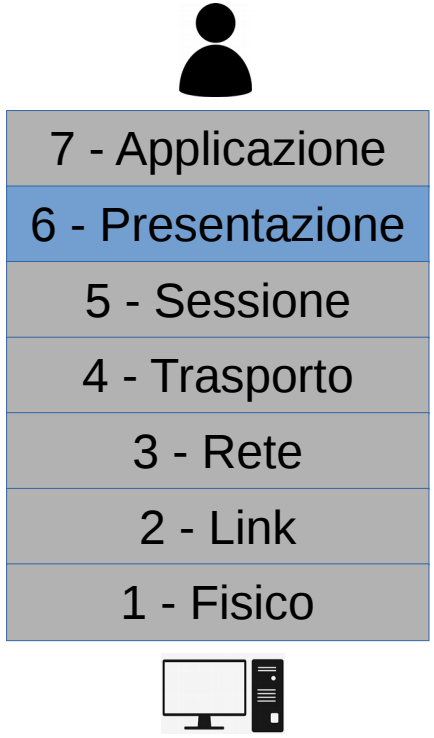
Livello 6 – Presentazione

Trasforma i dati forniti dalle applicazioni in un formato standardizzato

Consente di gestire la sintassi dell'informazione da trasferire.

American Standard Code for Information Interchange (ASCII)
Multipurpose Internet Mail Extensions (MIME)

$$(01100001)_2 = (97)_{10} = a$$



Livello 5 – Sessione

Controlla la comunicazione tra applicazioni.

Instaura e gestisce connessioni (sessioni) tra applicazioni cooperanti.
Si occupa anche della sincronia di invio/ricezione messaggi.

Es.

- sono connesso a Google, a Skype, … (client)
- sono un server Google ed ho 120 connessioni aperte con diversi utenti… (server)

A chi vanno i dati?

SOCKS e NetBIOS → permetono di effettuare connessioni **TCP**.

Livello 4 – Trasporto

Gestisce il trasferimento di dati in maniera trasparente ed affidabile tra due host.

Trasparente = gli indirizzi dei due host sono noti
Affidabile = implementa un controllo degli errori e delle perdite di dati *

La sua unità dati fondamentale è il segmento, user datagram o pacchetto**.**

Transmission Control Protocol (TCP *)

User Datagram Protocol (UDP **)

** Meno affidabile del TCP ma più veloce, ammette la perdita di dati, es. streaming

7 - Applicazione
6 - Presentazione
5 - Sessione
4 - Trasporto
3 - Rete
2 - Link
1 - Fisico

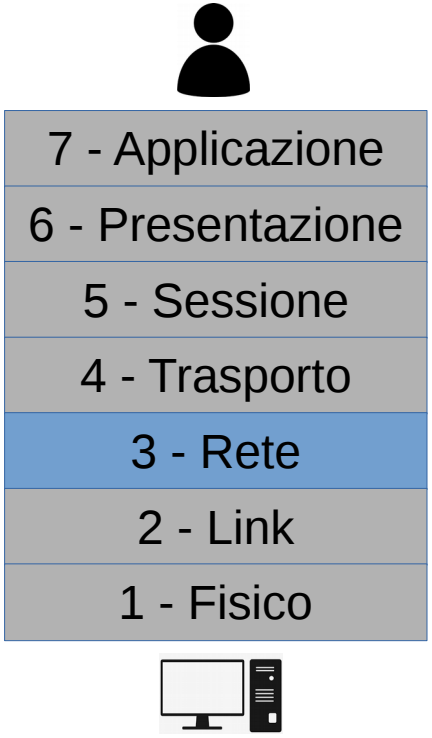
Livello 3 – Rete

Rende i livelli superiori indipendenti dai meccanismi e dalle tecnologie di trasmissione usate per la connessione.

Si fa carico della consegna dei pacchetti.

Internet Protocol (IP)

Spesso si parla di TCP/IP poichè è la coppia più usata.



TCP

Dato

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura
ché la diritta via era smarrita.
Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!
Tant'è amara che poco è più morte;
ma per trattar del ben ch'i' vi trovai,
dirò de l'altre cose ch'i' v'ho scorte.

Pacchetto 1
Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura
ché la diritta via era smarrita.

Pacchetto 2
Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!

Pacchetto 3
Tant'è amara che poco è più morte;
ma per trattar del ben ch'i' vi trovai,
dirò de l'altre cose ch'i' v'ho scorte.

TCP

Offset	Ottetto	0								1								2								3							
Ottetto	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Source port																Destination port															
4	32	Sequence number																															
8	32	Acknowledgment number (se ACK è impostato)																															
12	96	Data offset				Reserved 0 0 0 0				CWR	ECE	URG	ACK	PSH	RST	SYN	FIN	Window Size															
16	128	Checksum																Urgent pointer (se URG è impostato)															
20	160	Options (facoltativo)																															
20/60 ...	160/480 ...	Data																															

Campi “tecnici”

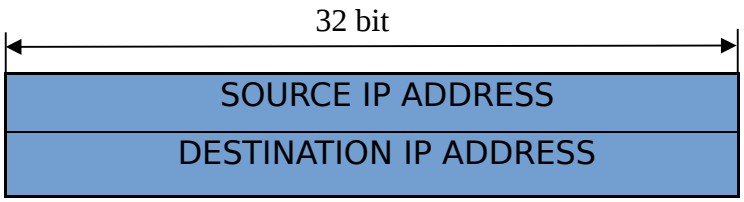
- Porta del mittente
- Porta del destinatario
- Numero del pacchetto
- Indicatore di avvenuta ricezione / Numero del pacchetto atteso
- Dati

IP

IP (Internet Protocol)

- È un protocollo per la trasmissione di singoli pacchetti (datagram)
- I datagram possono seguire percorsi diversi e raggiungere la destinazione in un ordine diverso a quello di spedizione
- Il protocollo IP non si occupa di riordinare i pacchetti dopo il loro arrivo

Indirizzamento IP



- L'indirizzo IP a 32 bit identifica univocamente un dispositivo (host o router) sulla rete
- Gli indirizzi IP pubblici sulla rete Internet sono assegnati dal NIC (Network Information Center)
- Due dispositivi connessi alla rete Internet non possono avere lo stesso IP
- Un indirizzo IP identifica una connessione alle rete
  - Un host che ha più interfacce di rete ha più indirizzi IP

Struttura degli indirizzi IP

- Un indirizzo IP è composto di 4 byte
  - Si rappresenta con 4 numeri decimali (0-255) separati da punti (.)

Indirizzo IP binario (32 bit)	Indirizzo numerico
10000001 00110100 00000110 00000000	129.52.6.0
11000000 00000101 00110000 00000011	192.5.48.3
00001010 00000010 00000000 00100101	10.2.0.37
10000000 00001010 00000010 00000011	128.10.2.3
10000000 10000000 11111111 00000000	128.128.255.0

- L'indirizzo può essere scomposto in due parti
  - Indirizzo di rete
  - Indirizzo di host nella rete
- La struttura a due livelli (rete/host) permette una gestione più efficiente dell'instradamento

Funzionalità del TCP

Trasmissione

- Riceve un flusso di dati dall'applicazione
- Li organizza in unità lunghe al massimo 64Kb
  - Eventualmente bufferizza i dati prima di spedire il pacchetto (es. input da tastiera)
- Spedisce le unità di dati **come datagram IP**

Ricezione

- Riceve i **datagram IP**
- Ricostruisce il flusso di byte originale nella sequenza corretta

Ritrasmissione dei datagram non ricevuti, riordinamento dei datagram arrivati in ordine sbagliato

Le porte

Per comunicare è necessario conoscere l'indirizzo IP del destinatario ma anche la "porta" su cui richiedere il servizio

Le porte attive definiscono i servizi TCP disponibili

Per connettersi ad un servizio specifico su un server si deve conoscere il numero di porta su cui il processo server accetta le connessioni

- Le porte inferiori alla 1024 sono dette “well-known ports” e corrispondono a servizi standard
- Ad esempio
  - -la porta 80 di TCP corrisponde al servizio HTTP (Hypertext Transfer Protocol) ovvero al server Web

Domain Name System (DNS)

E’ un database distribuito che permette di convertire i nomi simbolici degli host negli indirizzi IP numerici

E’ più facile ricordare (e dà più informazioni)

www.ing.unisi.it

che

193.205.7.2

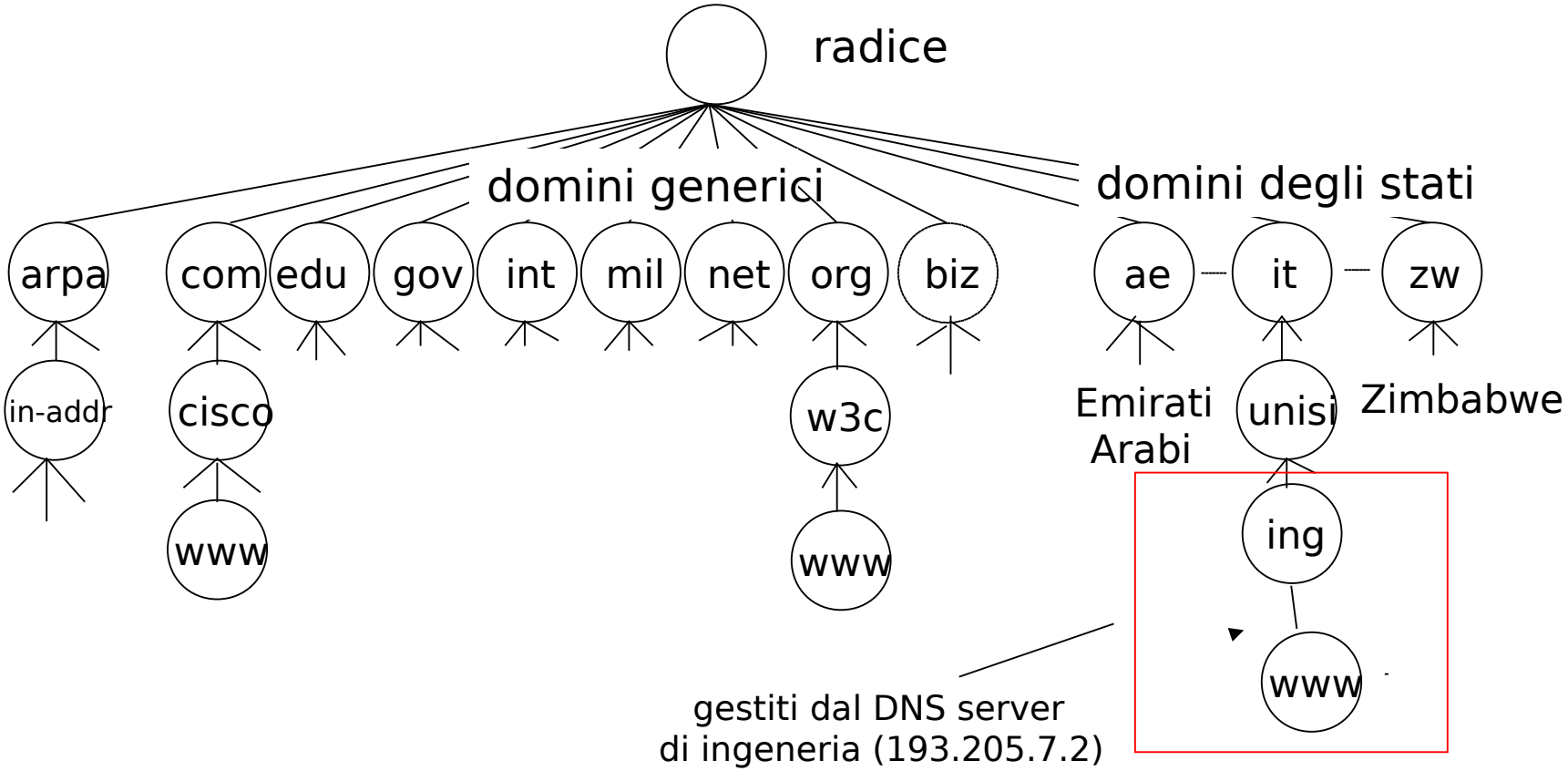
Per comunicare su rete serve l’indirizzo IP!

Il DNS è distribuito: ogni sito mantiene il suo database relativo agli host locali eseguendo su una macchina il server DNS che può essere interrogato da altri sistemi su Internet

Può essere usato dalle autorità nazionali per limitare la visibilità di alcuni siti
- Pro: utile per limitare accessi a siti “criminali”
- Contro: possibile censura da parte degli stati

Lo spazio dei nomi

- I nomi in Internet sono organizzati in modo gerarchico



I domini

- I nomi dei componenti sono stringhe (case-insensitive) fino a 63 caratteri mentre il nome completo di un cammino non può superare 255 caratteri
- Ciascun dominio controlla l’allocazione dei suoi sottodomini
- I domini delle nazioni sono indicati con i codici previsti dallo standard ISO 3166

Esempio

Server con Apache

Dominio registrato: pinkopallo.com

Sul server:

C:\www\home.html → www.pinkopallo.com/home.html

C:\www\info\contacts.html → www.pinkopallo.com/info/contacts.html

C:\private\doc\human-resources.html → private.pinkopallo.com/doc/human-resources.html

www, www\info, private\doc sono cartelle sul computer “server”.
Per renderle visibili all’esterno, bisogna oppurtanemente configuare Apache, ovvero dichiarare quali sono le cartelle sul disco che contengono il sito.

I domini generici

com	organizzazioni commerciali
edu	istituzioni accademiche (università USA)
gov	organizzazioni governative USA
int	organizzazioni internazionali
mil	rete militare USA
net	fornitori di servizi di rete
org	organizzazioni no profit

- In genere non sono riferiti ad una posizione geografica
- Sono attivati nuovi domini generici (es. biz)
- Alcuni paesi inseriscono sottodomini equivalenti a quelli generici (es. .ac.uk e .co.uk - istituzioni accademiche e commerciali del Regno Unito)

Livelli 2 ed 1 – Link e Fisico

Livello 2

Permette il trasferimento affidabile di dati attraverso il livello fisico.

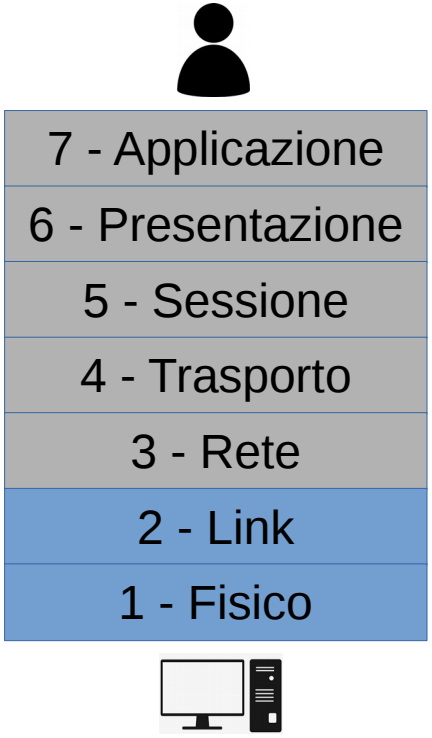
Invia i dati con la necessaria sincronizzazione ed effettua un controllo degli errori e delle perdite di segnale.
Tutto ciò consente di far apparire, al livello superiore, il mezzo fisico come una linea di trasmissione esente da errori di trasmissione.

Ethernet
WiFi

Livello 1

Trasmette un flusso di dati non strutturati attraverso un collegamento fisico.

ADSL
Fibra ottica



Web e HTML

Il Web

- Nato nel 1989 al CERN di Ginevra come mezzo per scambiare informazioni
- w3c - Consorzio World Wide Web (1994)

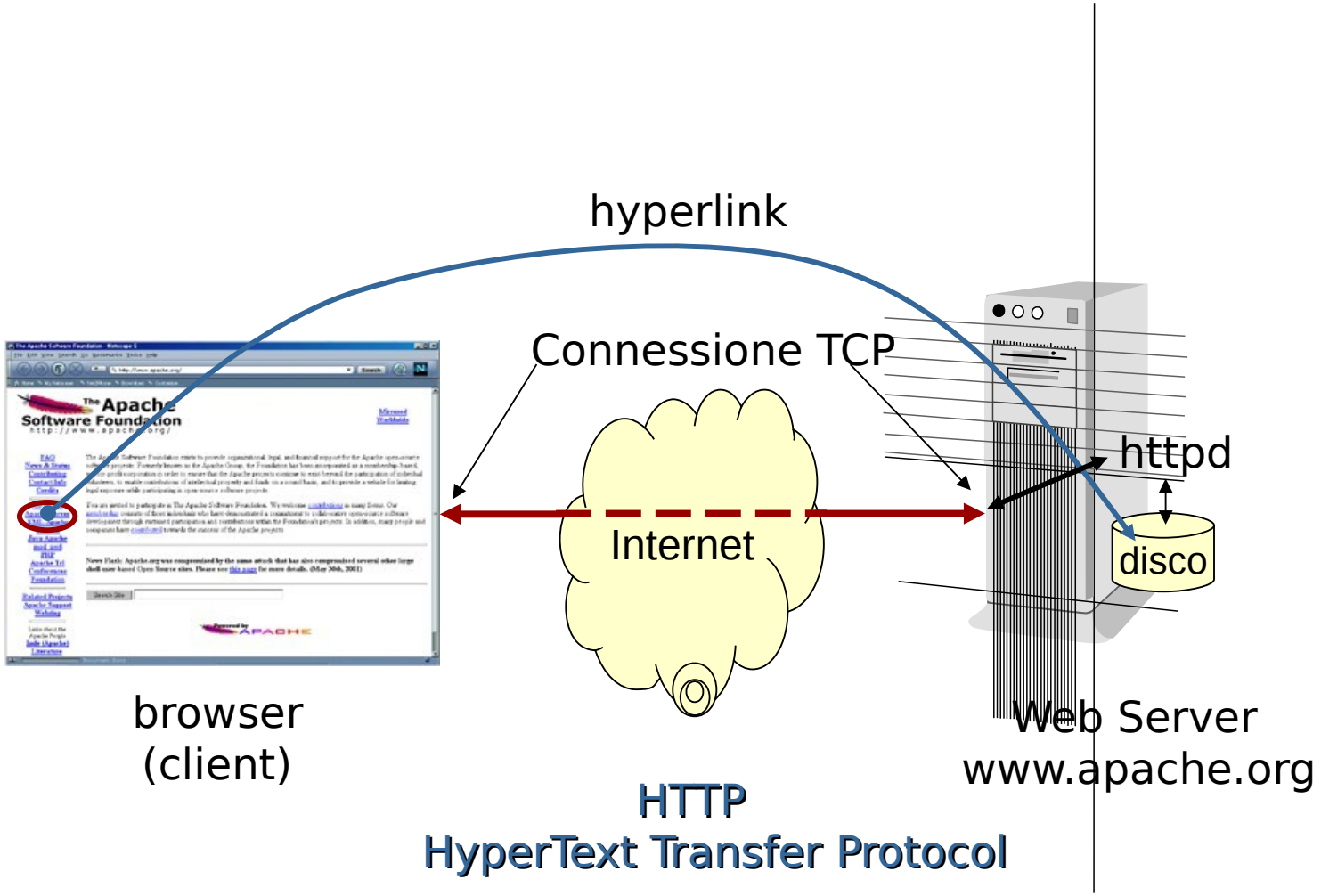
  - Standardizzazione dei protocolli su Web
  - Interoperabilità fra i siti
  → http://www.w3c.org

Il web è una collezione di documenti ipertestuali distribuita su server collegati alla rete Internet

La ragnatela (web) è un grafo i cui nodi sono i singoli documenti collegati fra loro da puntatori (hyperlink)

La maggior parte dei documenti su web è in formato HTML (HyperText Markup Language) ma si trovano anche altri formati (Macromedia Flash, pdf, postscript, ecc..)

Web servers


hyperlink
Connessione TCP
The Apache Software Foundation
http://www.apache.org/
Internet
httpd
disco
browser
(client)
Web Server
www.apache.org
HTTP
HyperText Transfer Protocol

HTTP

- Ogni sito Web ha un processo server in ascolto su una porta TCP
- La porta standard è la numero 80
- HTTP definisce il **protocollo** di comunicazione fra il client e il server

  - E’ utilizzato per trasferire ogni tipo di risorsa (file) su WWW

  - Una risorsa è un’entità individuata da un URL (Uniform Resource Locator)

    - file di qualsiasi formato
    - prodotto dell’esecuzione di una interrogazione o di uno script

- Utilizza un modello client-server
- Il client HTTP apre la connessione e invia un messaggio di richiesta al server HTTP

- Il server invia una risposta che in genere contiene la risorsa richiesta e poi chiude la connessione

Richiesta HTTP

La linea iniziale di una richiesta è formata da 3 parti

- Metodo
- Percorso locale della risorsa richiesta
- Versione di HTTP usata

Esempio di richiesta

Linea Iniziale

header

```
GET /index.html HTTP/1.1
Host: univac.dii.unisi.it
User-Agent: Mozilla/5.0 (Windows; U; Win98; en-US; m18) Gecko/20010131
    Netscape6/6.01
Accept: */*
Accept-Language: en
Accept-Encoding: gzip,deflate,compress,identity
Keep-Alive: 300
Connection: keep-alive
```

Risposta HTTP

- La linea iniziale di una risposta costituisce una linea di stato

```
HTTP/1.1 200 OK
Date: Wed, 06 Jun 2001 22:44:40 GMT
Server: Apache/1.3.20 (Win32)
Last-Modified: Wed, 06 Jun 2001 22:32:26 GMT
ETag: "0-64-3b1eaf7a"
Accept-Ranges: bytes
Content-Length: 100
Connection: close
Content-Type: text/html
```

headers

risorsa

```
<HTML>
<HEAD>
<TITLE>Prova</TITLE>
</HEAD>
<BODY>
<H2>File HTML di Prova</H1>
</BODY>
```

Codici di stato

La linea di stato riporta un codice di stato e la sua “spiegazione”

HTTP/1.1 200 OK

- I codici più comuni sono

200 OK
Richiesta con successo. La risorsa è nel corpo della risposta

404 Not Found
La risorsa richiesta non esiste

301 Moved Permanently

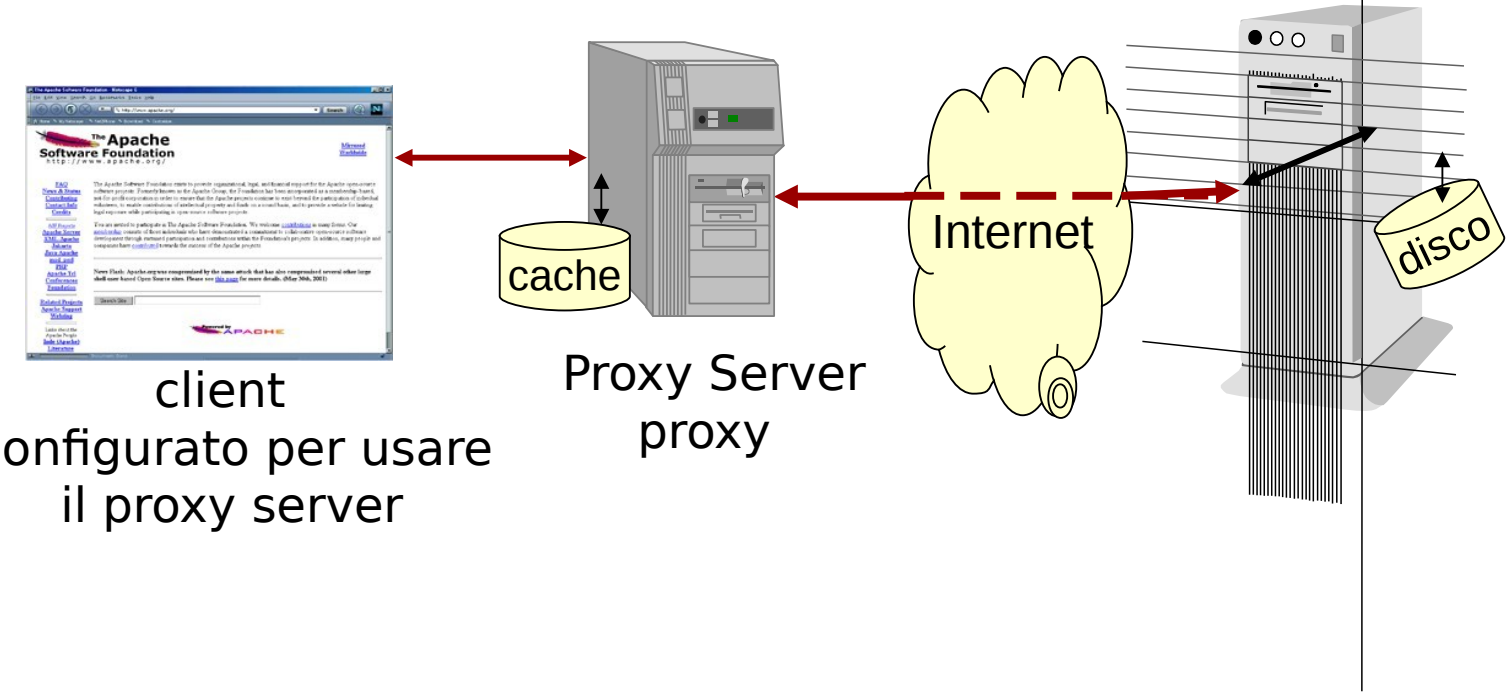
302 Moved Temporarily

303 See Other (in HTTP 1.1)
La risorsa è stata spostata ad un altro URL specificato nel campo Location: dell’intestazione. Il client dovrebbe saltare a tale locazione (redirect).

500 Server Error

Proxy HTTP

- Un proxy HTTP agisce da intermediario fra il client e il server
  - Riceve le richieste dal client
  - Propaga la richiesta al server corretto
  - Memorizza le pagine richiesta
  - Vantaggi: sicurezza e minor consumo di banda

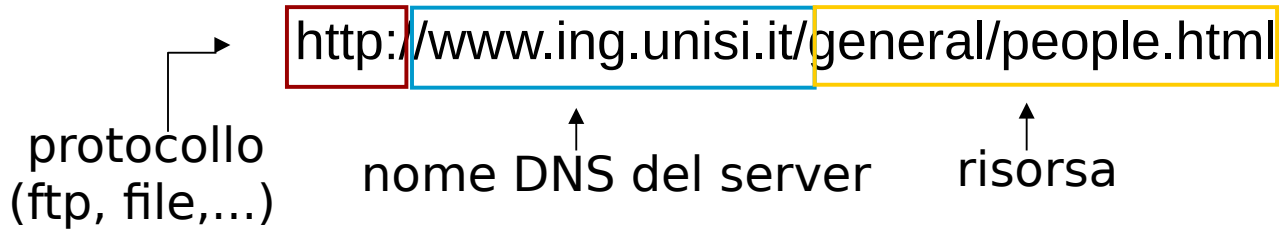


client
configurato per usare
il proxy server

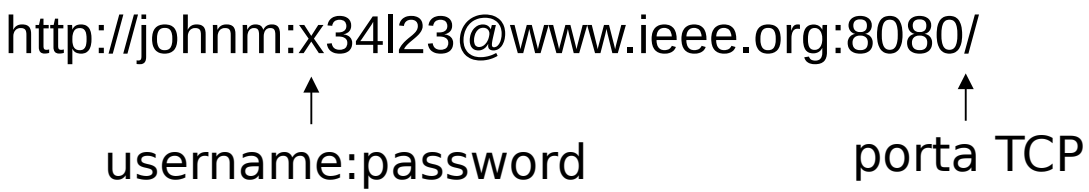
Proxy Server
proxy

URL

- Un URL (Uniform Resource Locator) individua univocamente una risorsa su web



- Al nome possono essere aggiunte
  - la porta TCP (altrimenti si usa quella di default del protocollo)
  - una coppia username/password per accedere a risorse che richiedono autenticazione



HTML

HyperText Markup Language (HTML)

Linguaggio di markup.

→ insieme di regole che descrivono i meccanismi di rappresentazione (strutturali, semantici, presentazionali) o layout di un testo

Nato per la formattazione e impaginazione di documenti ipertestuali disponibili nel web.

Il termine markup (o marcatura) deriva dall'ambiente tipografico dove si usava contrassegnare con annotazioni le parti del testo che andavano evidenziate o corrette, allo scopo di segnalarle al compositore o al dattilografo.

HTML (HyperText Markup Language)

- Una pagina prevede un'intestazione e un corpo compresi fra il tag **<HTML>** e **</HTML>**

  <HEAD>...**</HEAD>** definiscono l'intestazione (titolo, stili, meta, ..)

  <BODY>...**</BODY>** definisce il corpo del documento

- E' un linguaggio che usa annotazioni (markup) per definire la formattazione del testo

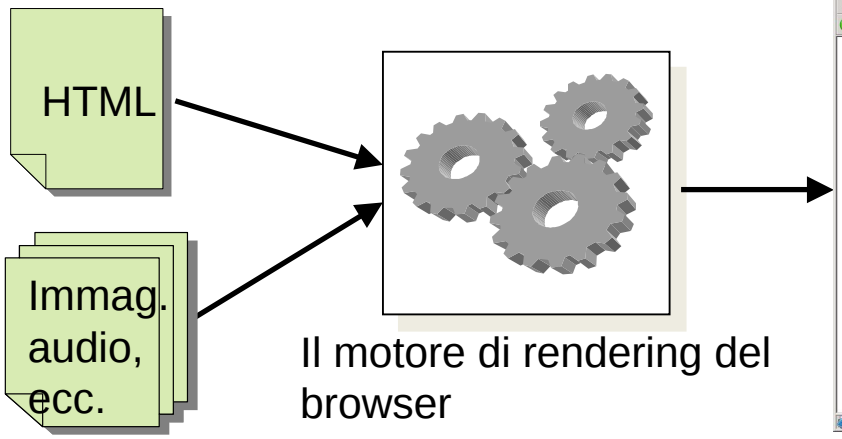
<B>scrittura in grassetto**</B>**

tag di apertura

tag di chiusura

HTML, tag, rendering

- I tag, inseriti nel testo, indicano al browser come deve impaginare e visualizzate la pagina Web nel suo insieme
  - Il browser interpreta i tag, formatta la pagina Web e ne genera la rappresentazione grafica (il rendering)



Il motore di rendering del browser

La finestra del browser

Sintassi dei Tag

- Quasi tutti i tag sono composti da un **tag di apertura** e da un **tag di chiusura**
  - **<nome_tag>**testo sul quale il tag agisce**</nome_tag>**
  - Es.: **<title>**`Questo è un esempio`**</title>**

- Alcuni tag non hanno la corrispondente chiusura:
  - **<nome_tag>**
  - Es.: **
**

- Per molti tag è possibile specificare degli attributi:
  - **<nome_tag nome_attributo="valore_attributo">**
  - Es.: **<body bgcolor="red">** … **</body>**

Esempio di pagina HTML

```
<html>
	<head>
		<title>Questo è un esempio</title>
	</head>
	<body>
		Ciao! Questo è
		un esempio di pagina HTML
	</body>
</html>
```

- La pagina è identificata dall'estensione **.html** o **.htm**

Esempio di pagina HTML

```
<html>
	<head>
		<title>Questo è un esempio</title>
	</head>
	<body>
		<h1>Titolo principale</h1>
		<h2>Titolo secondario</h2>
		<p>Testo all'interno di un
		<strong>paragrafo</strong></p>
		<p>Altro paragrafo</p>
		Testo al di fuori di un paragrafo
	</body>
</html>
```

Collegamenti Ipertestuali (Link)

- Servono per creare connessioni navigabili
  - Link interni alla stessa pagina
  - Link tra pagine

- Link:

  ****`qualsiasi cosa`****

  - Facendo click su `qualsiasi cosa`, il browser si sposta a `indirizzo`
  - Indirizzo può essere una **URL**

Immagini

- E’ possibile includere immagini:

  `<img src="indirizzo e nomefile immagine">`

- Alcuni attributi permettono di modificare le dimensioni dell’immagine:
  - Es.: **`<img src="a.gif" height="50" width="70">`**
  - Es.: **`<img src="a.gif" height="50%" width="5%">`**