Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma

Fondamenti di Informatica

Laurea in
Ingegneria Civile e Ingegneria per l'ambiente e il territorio

Principi di Reti di Calcolatori e Problematiche di Internet

Stefano Cagnoni e Monica Mordonini

Cenni su Reti di Calcolatori

- Cosa è una rete?
- Punto di vista logico: sistema di dati ed utenti *distribuito*
- Punto di vista fisico: insieme di *hardware, collegamenti, e protocolli* che permettono la comunicazione tra macchine remote

Reti di Calcolatori

- Una rete di calcolatore offre alcuni vantaggi rispetto all'uso di un calcolatore isolato:
  - Condivisione dell'Informazione
  - Condivisione delle Risorse
  - Accesso a Risorse Remote
  - Alta Affidabilità
  - Convenienza Economica
  - Crescita Graduale

Trasmissione dati

- Mezzi di trasmissione (bps=bit x sec) si utilizzano metodi per trasmissione telefonica
  - Doppino telefonico (vel. 2400-9600 bps)
  - Cavo coassiale (vel. 10^4-10^6 bps)
  - Fibra ottica (bit=assenza/presenza segnale luminoso) (vel. 10^9 bps)
  - Onde elettromagnetiche (es via satellite)

Digitale-Analogico

- Informazione=***digitale*** (seq. Bit)
- Segnali=***analogico*** (continuo)
- Il ***Modem*** (Modulatore-Demodulatore) si preoccupa di trasformare bit in segnali e viceversa
  - Modulazione in frequenza=il modem altera in frequenza una sequenza portante

    0=freq più bassa; 1=freq più alta
  - ...modulazione in ampiezza, fase,
- Es PC collegati ‘da casa’ attraverso un modem

Reti di Calcolatori

- Le reti sono classificate in base alla loro dimensione:
  - Rete locale (LAN)
  - Rete Metropolitana (MAN)
  - Rete Geografica (WAN)

Reti di Calcolatori

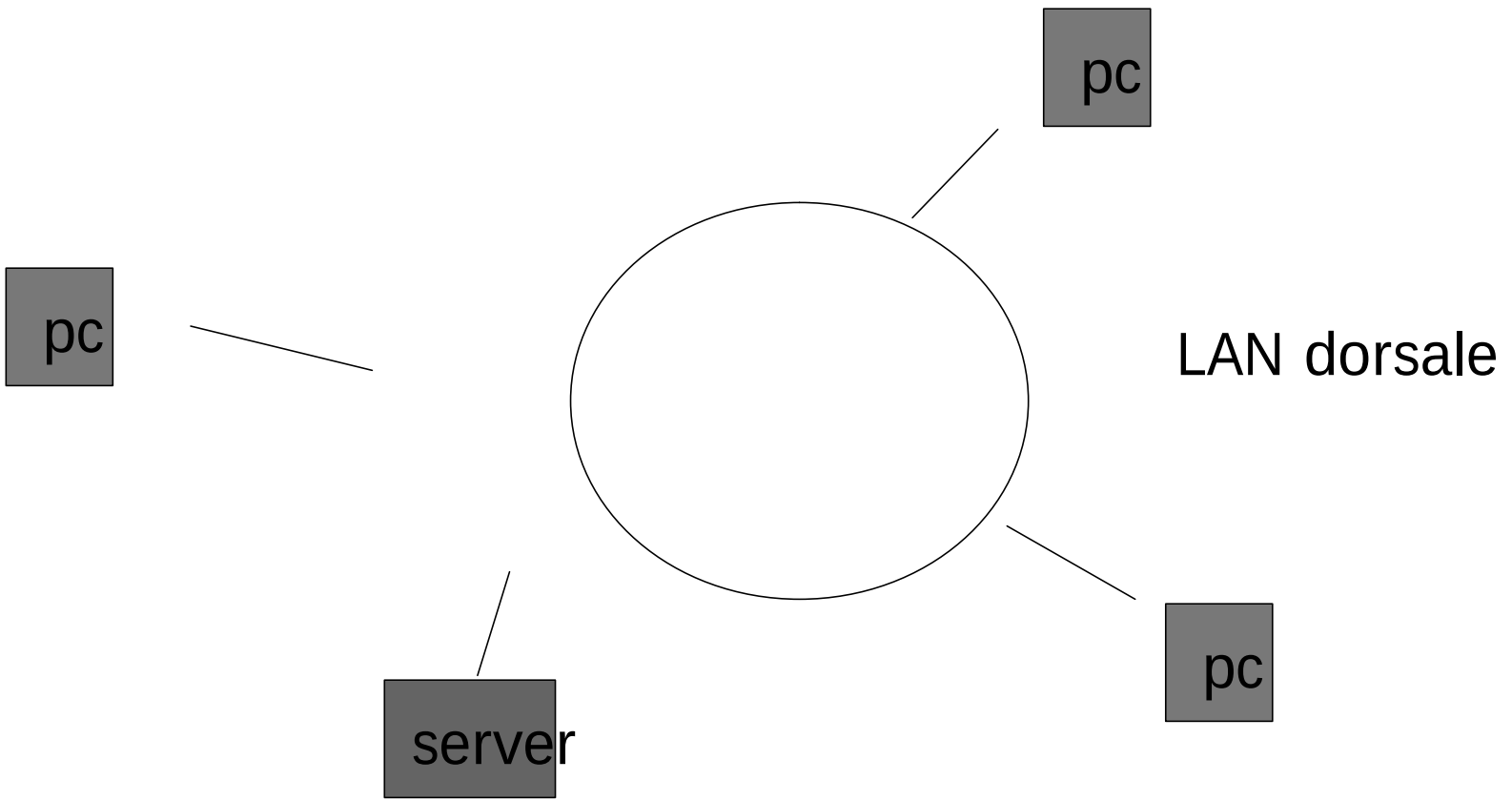
- Le prestazioni di una rete è indicata con la larghezza di banda.
  - La larghezza di banda indica la quantità di informazione che la rete è in grado di trasmettere.
  - La larghezza di banda viene misurata in bit al secondo:
    - un collegamento telefonico via modem (fino a 56 Kbps).
    - un collegamento telefonico dedicato ISDN o ADSL (fino a 6.4Mbps).
    - un collegamento di una rete locale (fino a 1Gbps).

Reti Locali

- Terminali nella stessa stanza/edificio possiamo utilizzare collegamenti diretti (senza passare per reti pubbliche)
- Rete locale più diffuse:
  - ***Ethernet e Fast Ethernet***

    insieme di componenti hardware e software particolari che gestiscono la trasmissione dati in una rete locale

Rete Locale



Reti Metropolitane e Geografiche

- Nodi distribuiti su medio-lunga distanza
- Possiamo usare
  - la rete di comunicazione pubblica utilizzando modem o affittando linee di trasmissione (PSTN)
  - Oppure reti digitali di trasmissione dati (ISDN)
- Trasmissione dati: attraverso messaggi

Topologia di una Rete

- A ***stella***=tutti nodi collegati ad un elaboratore centrale (che smista messaggi)
- Ad ***anello***=treno di messaggi
- A ***bus***=nodi disposti lungo un unico canale
- ***Irregolare***=attraverso host e nodi di trasmissione

Messaggi?

- Sequenze di bit (come al solito...)
  - Mittente
  - Destinatario
  - Caratteri di controllo (per consistenza)
  - Contenuto messaggi
- Nodo di trasmissione: riceve e ritrasmette un messaggio al destinatario o ad un nodo vicino se il destinatario non è collegato
- ***Routing*=**pecorso del messaggio nella rete può essere statico o dinamico

Protocolli di Comunicazione

- Utilizzati dai calcolatori per dialogare
- Come nel caso della codifica dei dati occorre utilizzare degli *standard* internazionali per problemi di compatibilità!
- Esempi
  - Modello OSI/ISO (Open System Interconnection)
  - Modello TCP/IP (standard de facto)

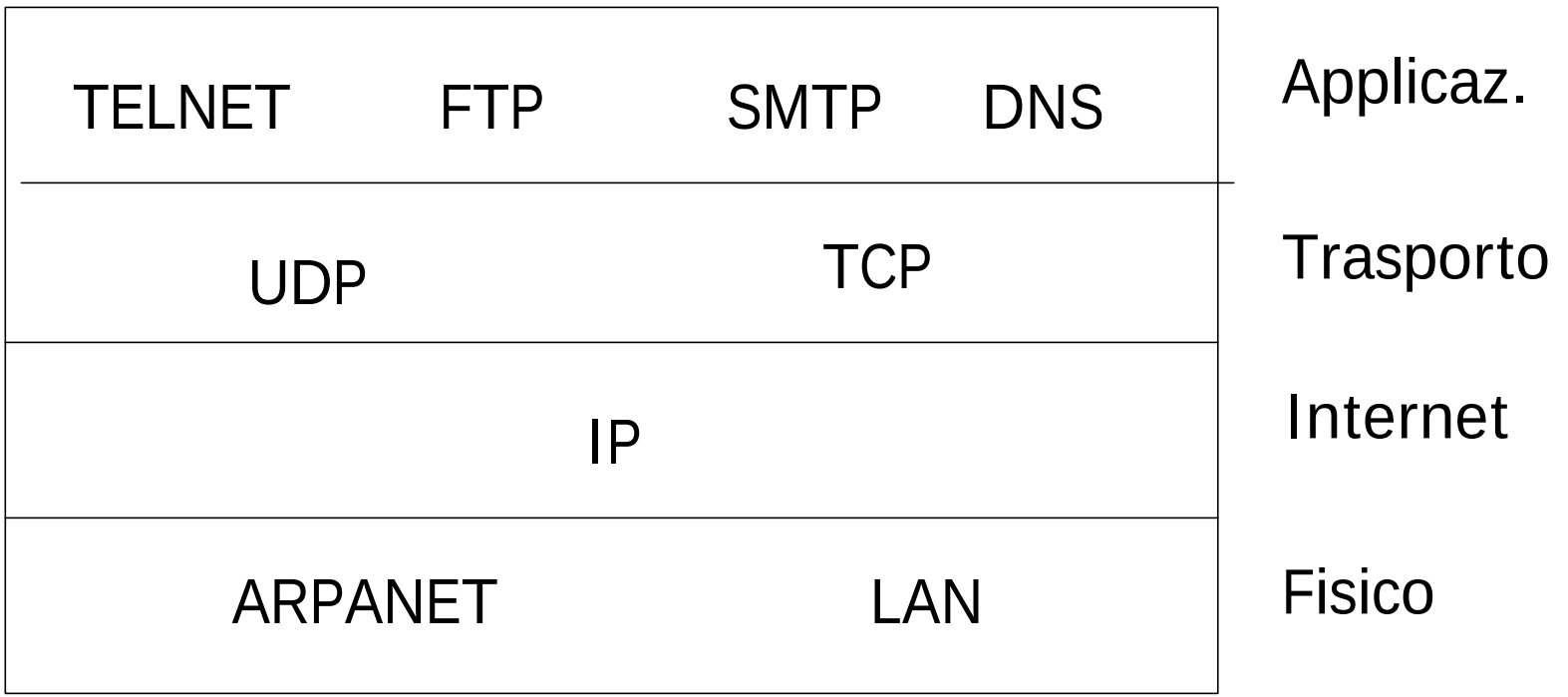
Internet

- Inter-rete (cioè che collega molte sottoreti tra loro) nata dalla fusione di diverse reti di agenzie governative americane (ARPANET) e reti di università
- Internet è una rete di calcolatori che permette potenzialmetne la comunicazione tra tutti i calcolatori del mondo:
  - Un indirizzo diverso per ogni calcolatore (indirizzo IP).
  - Un protocollo di comunicazione comune (TCP/IP) per lo scambio di messaggi tra i calcolatori.

Modello TCP / IP

- Suddiviso in vari livelli
  - Applicazione: software applicativo
  - Trasporto: trasforma dati in messaggi usando i protocolli TCP (trasmissione sicura) e UDP (trasmissione veloce)
  - Internet: protocollo IP di spedizione dei messaggi sulla rete
    - Indirizzo IP=indirizzo degli host in rete
  - Al di sotto aspetti legati al tipo di rete

Modello TCP/IP



Indirizzi IP

- Un indirizzo IP è composto da una sequenza di quattro numeri compresi tra 0 e 255.

  160.78.28.83

- Esiste un sistema detto Domain Name Server (DNS) che permette di associare dei nomi simbolici agli indirizzi IP.

  foresto.ce.unipr.it

  www.unipr.it

  WWW.UniPR.IT

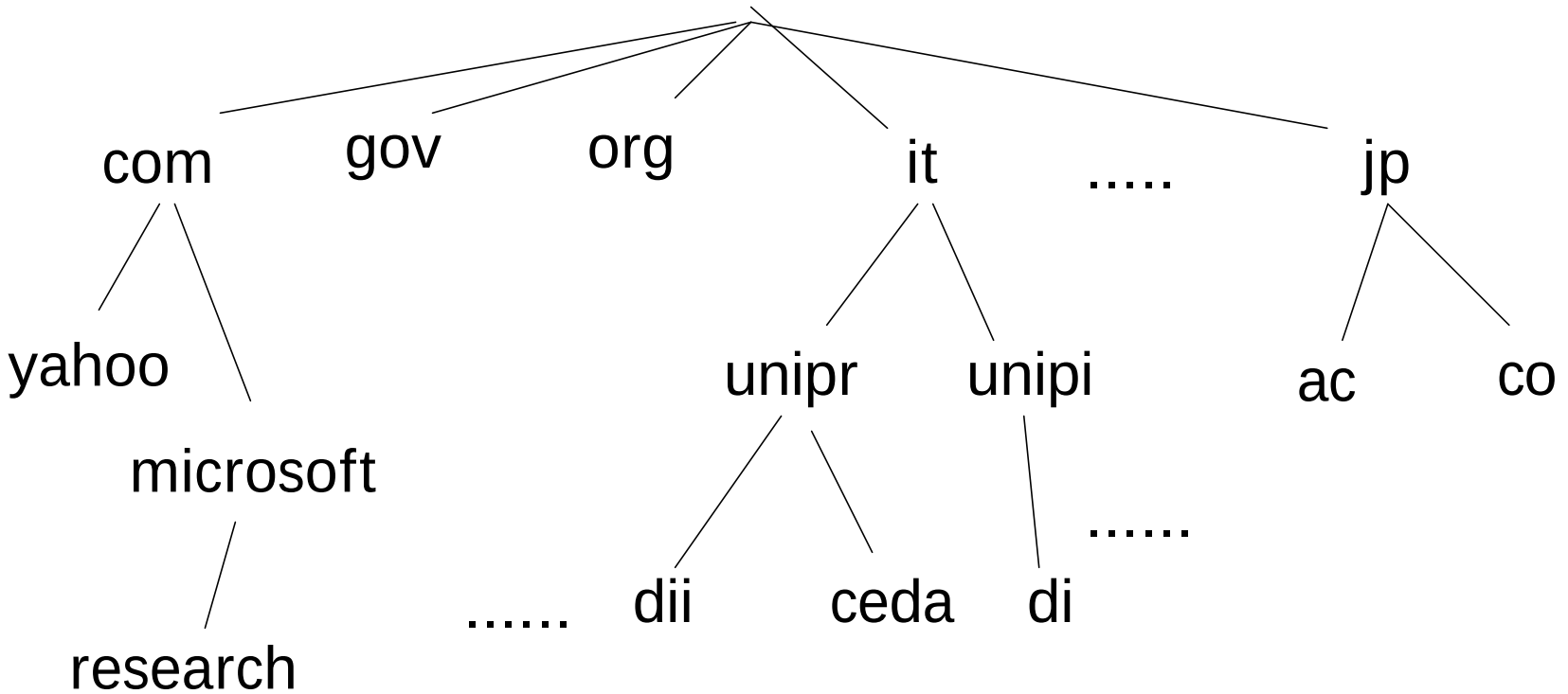
Indirizzi IP

- I nomi simbolici associati agli indirizzi IP non sono liberi, ma assegnati da uffici appositi.
- Il simbolo terminale è assegnato a livello internazionale e può essere di due tipi:
  - Indicante il tipo di organizzazione
    - com
    - edu
    - gov
    - net
    - mil
    - org
  - Indicante la nazione (it, uk, fr, …)

Dominio

- Domini=suddivisione logica di Internet per facilitare la gestione dei nomi delle risorse
- Internet è suddivisa in una multitudine di domini radice
  - Domini nazionali: *uk it de*
  - Domini generici: *com edu*
- Un dominio radice include una collezione di *host* e può essere suddiviso a sua volta in sottodomini e così via
  - Sottodominio del DII: ce.unipr.it

Spazio dei nomi dei domini



Name Server

- Lo spazio dei nomi è diviso in zone gestiti da un server principale e server secondari che mantengono la lista degli host inclusi nel dominio (database dei nomi)
- Se un server non trova un nome nel suo database manda una richiesta al server del dominio antenato o successore e così via (interrogazioni ricorsive)
- Si usa una memoria cache per mantenere gli indirizzi recuperati tramite altri server

Domain Name Server

- Lo spazio dei nomi è memorizzato quindi sotto forma di database distribuito (DNS)
- Ogni rete locale ha un proprio server DNS che mappa nomi logici (indirizzi DNS) in indirizzi fisici (indirizzi IP)
- Ricordate che
  - Indirizzi IP=codice binario utilizzato dal protocollo di invio dati del modello TCP/IP (livello Internet)
  - Es. 121.34.16.19

Servizi di Internet

- La rete internet fornisce quattro servizi principali:
  - FTP (File Transfer Protocol)
  - SMTP (Simple Mail Transfer Protocol)
  - TELNET
  - HTTP (HyperText Transport Protocol)

World Wide Web

- Assieme alla posta elettronica, World Wide Web (WWW o Web) è il modo più diffuso di utilizzare la rete Internet.
- Il Web permette agli utenti di internet di mettere a disposizione e di accedere a documenti via HTTP.
- Il Web si basa su due programmi:
  - Il Web server
  - Il Web client (browser)

L’ipertesto globale

- 1990 : l’idea della ragnatela a protocollo unico universale (CERN di Ginevra)
- 1993 : sviluppo della piattaforma ad interfaccia grafica per l’accesso ai siti (200 server web)
- 1998 : varie decine di milioni di server web
- 2000: 500 milioni di server web accessibili

Il successo del web

- Distribuzione planetaria : si serve del canale di distribuzione più vasto e ramificato del mondo (linee telefoniche)
- Facilità di utilizzo
- L’organizzazione ipertestuale
- Possibilità di trasmettere / ricevere informazioni multimediali
- Semplicità di gestione per i fornitori di informazioni (tutti gli utenti)

I concetti base del www

- Ipertesto : informazione organizzata in modo non sequenziale ma reticolare
- Esempio di informazione sequenziale : libro in cui le pagine sono lette in sequenza
- Multimedia: più mezzi (e linguaggi) in una stessa unità di messaggio comunicativo
- Ipertesto è costituito da unità informative (nodi), e collegamenti che permettono di passare da un nodo ad un altro
- Se i nodi sono costituiti da documenti multimediali , l'ipertesto si definisce ipermedia

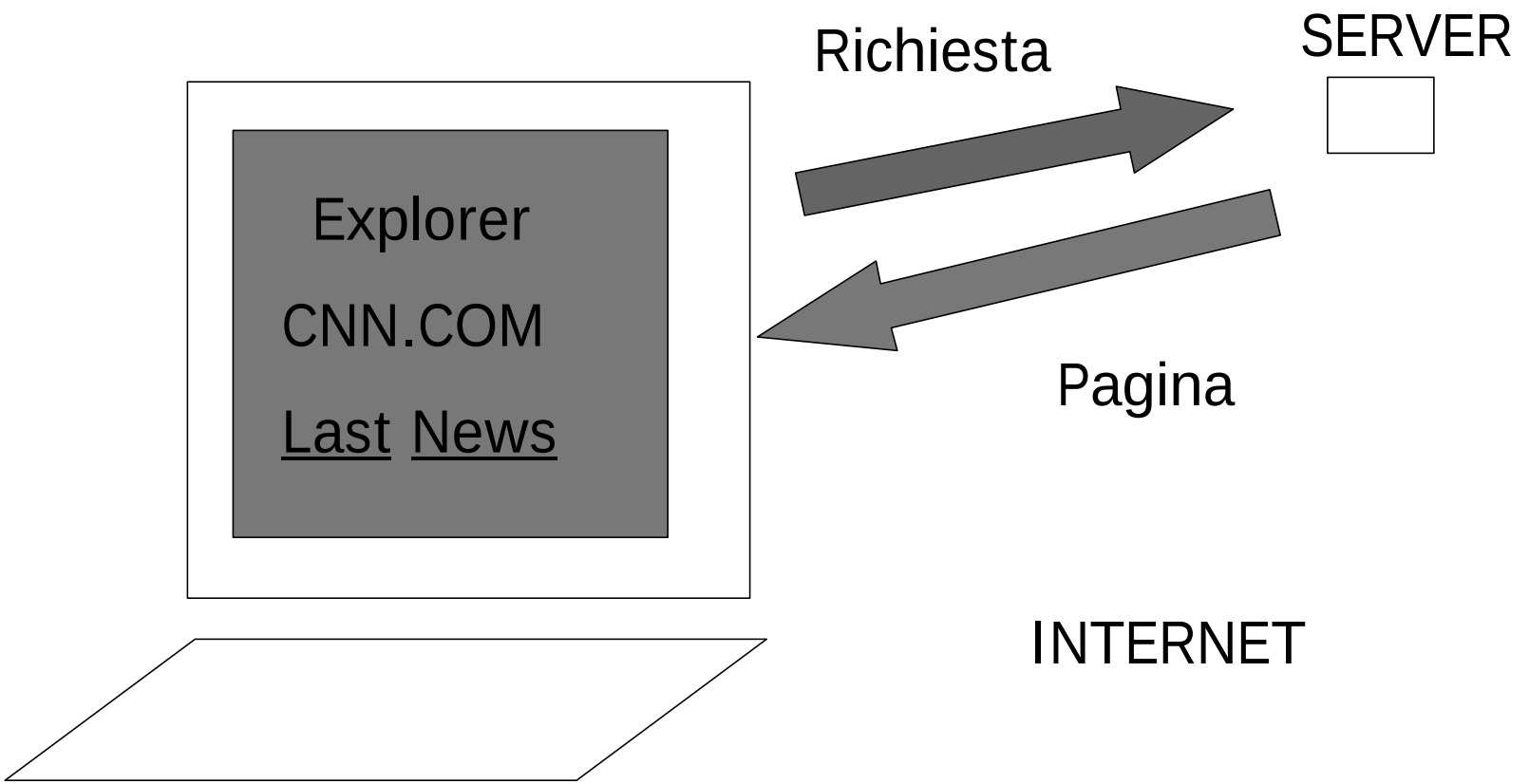
World Wide Web (WWW)

- Architettura software per gestire dati distribuiti geograficamente basata sulla nozione di ipertesto
- ***Pagine web***: ipertesti che possono contenere testo, immagini, suoni, programmi eseguibili
  - un utente legge le pagine, se seleziona un link la pagina viene sostituita con quella richiesta (scaricata dal sito remoto)
- Si appoggia a TCP/IP e quindi è compatibile con ogni tipo di macchina collegata ad Internet

Struttura del Web

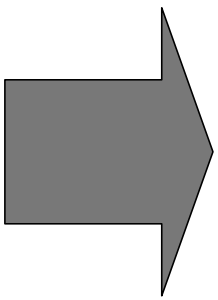
- Architettura ***Client-server***
- ***Client*** (ad esempio explorer) permette la navigazione nel web
  - trasmette le richieste di pagine/dati remoti, riceve le informazioni e le visualizza sul client
  - A volte utilizza programmi esterni (plug-in) per gestire i dati ricevuti
- Il ***Server*** è un ***processo*** sempre attivo che aspetta e serve le richieste dei client
  - Restituisce la pagina richiesta oppure un messaggio di errore

Client-Server



Le regole del web

- Formato universale dei documenti (HTML)
- Protocollo (linguaggio) di comunicazione “standard” tra l’utente (client) ed il server: HTTP



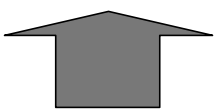
- Strumento essenziale : il browser: programma che riceve i comandi-utente, li trasmette al server, riceve da questo le informazioni (documenti), ne interpreta il formato e ne effettua la presentazione sul client

Le regole del web

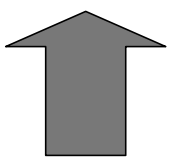
- Attraverso appositi programmi (Common Gateway Interface) il server-web esegue le richieste del client
- Il linguaggio HTML (Hyper Text Markup Language)
  - Le istruzioni contengono dei marcatori, detti tag (caratteri ASCII) che servono a descrivere la struttura, la composizione e l'impaginazione del documento testuale
  - Le immagini vengono gestite dal browser attraverso appositi programmi

La tecnica di indirizzamento

- URL /Uniform Resource Location : indirizzo unico della rete
- Indirizzamento alla risorsa : file, documento, pagina web, computer….
- Esempio di indirizzo:
  - http:// www.liberliber.it/index.htm



Nome computer



Nome del file

URL: indirizzi nel WEB

- Per accedere a un documento su Web bisogna conoscere il suo indirizzo.
- L'indirizzo è detto URL (Uniform Resource Locator) ed è composta da quattro parti:
  - Il protocollo (http in questo caso).
  - L'indirizzo del calcolatore su cui è in esecuzione il Web server.
  - Il numero di porta (opzionale: default 80).
  - Il percorso per accedere il file.

URL: specifiche standard

- Specifica:
  - *Come* si vuole accedere alla risorsa (metodo)
  - *Dove* si trova la risorsa (indirizzo server DNS)
  - *Nome* della risorsa (nome)
- Formato:
  - Metodo://host/nome

    http://www.ce.unipr.it/index.html

    http://www.ce.unipr.it/

    http://www.ce.unipr.it/people/

Protocolli (“metodi”)

- **http**: protocollo gestione ipertesti
- **ftp**: trasferimento file
- **news**: gruppi di discussione
- **telnet**: accedere a macchine remote
- **file**: accedere a documenti locali

Nome

- Nome (mnemonico) di dominio del *server DNS* al quale si vuole chiedere la risorsa
- Esempio:
  - Server Web DII-PR: www.ce.unipr.it
  - Server FTP DII-PR: ftp.ce.unipr.it

Nome risorsa

- Path name (cammino) che porta al file contente la risorsa (es pagina, foto, ecc) nello *spazio di dati* gestito dal server del sito che abbiamo contattato
  - Solitamente ogni sito ha una pagina di ingresso denominata index.html
  - Riepilogando:
  - http://www.ce.unipr.it/
  - ftp://ftp.ce.unipr.it/

Le funzionalità del browser

- La navigazione ipertestuale dalla pagina HTML
- Il ritorno alla pagina precedente/ successiva
- Cronologia delle pagine visitate
- Memorizzazione, stampa della pagina/documento
- Scrolling delle righe
- Ritorno alla pagina iniziale
- Attivazione d una nuova ricerca/indirizzo
- Uscita

“come” funziona il browser

- Uso della memoria cache per la ottimizzazione dei tempi di presentazione
- Uso del “proxy server” (server più vicino, appartenente alla stessa sottorete del client)
- Attivazione di tutti quei programmi che consentono la corretta visualizzazione delle pagine (il cui nome è memorizzato nella testata di trasmissione tra Client e Server; es. winword / applet Java…)

Funzioni avanzate del web

- **Plug –in:** modulo software – in grado di integrarsi con i browser che può attivarlo ed eseguirlo online (Acrobat Reader, shockwave, quicktime……)

- **Java** : linguaggio di programmazione (object oriented), multiforma , interoperabile, compatibile con HTML (applet)
  - Utile per rappresentare:
    - Suoni,video,animazioni,grafica

Linguaggi del Web

Il linguaggio HTML

- Il linguaggio HTML (*HyperText Markup Language*) utilizza annotazioni per descrivere come verrà visualizzato il documento sul browser di un cliente
  - Es: La prossima parola è in neretto
- Il browser interpreta le annotazioni traducendole in effetti grafici
  - Es: La prossima parola è in **neretto**
- Alcuni tool forniscono direttamente l'effetto desiderato senza dover usare HTML

Il linguaggio HTML

- Un documento HTML contiene:
  - Testo.
  - Comandi HTML (tag).
  - Collegamenti ad altri documenti.

Il linguaggio HTML

- I comandi HTML hanno in genere la forma:

  <tag> … testo … </tag>

- Un documento HTML ha in genere la forma:

```
<html>
…
<head>
…
</head>
<body>
…
</body>
</html>
```

Tag HTML

- I tag HTML possono essere divisi in cinque gruppi:
  - Tag di intestazione
  - Tag di formattazione fisica
  - Tag di strutturazione logica
  - Tag di collegamento ipertestuale
  - Tag di inclusione di immagini e programmi

Tag di intestazione e formattazione fisica

- I tag di intestazione vengono utilizzati nella parte di intestazione di un documento HTML.

  <meta>

  <meta name="author"content="M. Mordonini">

  <title>

  <title>List of recomended books</title>

Tag di intestazione e formattazione fisica

- I tag di formattazione fisica permettono di impaginare il documento.
  - font arial
  - , <i>, <ul> Grassetto
  - <hr>,

Tag di strutturazione logica

- I tag di strutturazione logica permettono di organizzare la struttura del documento.
  - <h1>, …, <h6> <h2>informazioni utili</h2>
  - , corsivo
  - <address>, <blockquote>, <cite>, <p>

<address>
Monica Mordonini
Università di Parma
Parco Area delle Scienze 181A
43100 Parma
</address>

Tag di strutturazione logica

- <table>, <th>, <tr>, <td>

```
<table border=1>
<tr><th>Nome</th><th>Cognome</th><th>Città</th></tr>
<tr><td>Mario</td> <td>Rossi</td> <td>Parma</td></tr>
<tr><td>Paola</td> <td>Bianchi</td> <td>Piacenza</td></tr>
</table>
```

Nome	Cognome	Città
Mario	Rossi	Parma
Paola	Bianchi	Piacenza

Tag di strutturazione logica

- , <ul>,

```
<ol>
<li>Uno</li>
<li>Due</li>
</ol>
```

1. Uno
2. Due

```
<ul>
<li>Uno</li>
<li>Due</li>
</ul>
```

- Uno
- Due

Tag di collegamento ipertestuale

- I tag di collegamento ipertestuale permettono di accedere al contenuto di altri documenti.

DII-Parma

Tag di inclusione di immagini e programmi

- I tag di inclusione di immagine permettono di inserire delle immagini in un documento.

  <img src="http://www.ce.unipr.it/image/pippo.gif">

- I tag di inclusione di programmi permettono di inserire dei programmi in un documento.
  - <script>, <applet>

Sorgente pagina web

<HTML>
<BODY>
Marco Rossi
PhD Student
Universit&agrave di Parma
<IMG SRC="marco.gif"><hr>
Per scaricare la mia tesi premi qui sotto

TESI
</BODY>
</HTML>

Pagina visualizzata su browser

Marco Rossi
PhD Student
Università di Parma

Per scaricare la mia tesi premi qui sotto

TESI

Form e interazione con cliente

- Si possono creare pagine che permettono all'utente di immettere dati attraverso FORM (moduli da compilare)
- I dati vengono gestiti poi da programmi residenti sul server (es CGI)

Form esempio

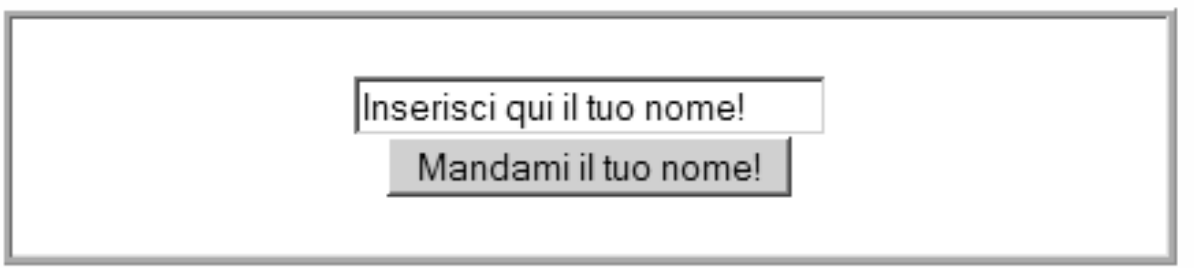
- Un tipico esempio di form potrebbe essere un motore di ricerca.

- Inviando questo form, ecco cosa succede:
- Le parole di ricerca vengono mandate a un programma sul server.
- Il programma cerca un database per i riscontri.
- Il programma crea una pagina web con i risultati.
- La pagina web con i risultati viene rimandata al visitatore.

Form: esempio

```
<html>
<head>
<title>La Mia Pagina</title>
</head>
<body>
<form name="myform"
action="http://www.mydomain.com/myformhandler.cgi"
method="POST">
<div align="center">
<br><br>
<input type="text" size="25" value="Inserisci qui il tuo nome!">
<br> <input type="submit" value="Mandami il tuo nome!"><br>
</div>
</form>
</body>
</html>
```

Ecco il risultato::



Xml: cos'è?

- Xml: eXtensible Markup Language
- Linguaggio di markup: permette di evidenziare il contenuto di un documento, all'interno di marcatori descrittivi, che ne definiscono gli attributi, ad es. linguaggio html.
- Estensibilità: in realtà Xml è un metalinguaggio di markup, cioè viene utilizzato per creare linguaggi di markup personalizzati.
- NON è un linguaggio di programmazione.

Non esistono documenti scritti in XML, ma documenti scritti in un linguaggio di markup basato su XML

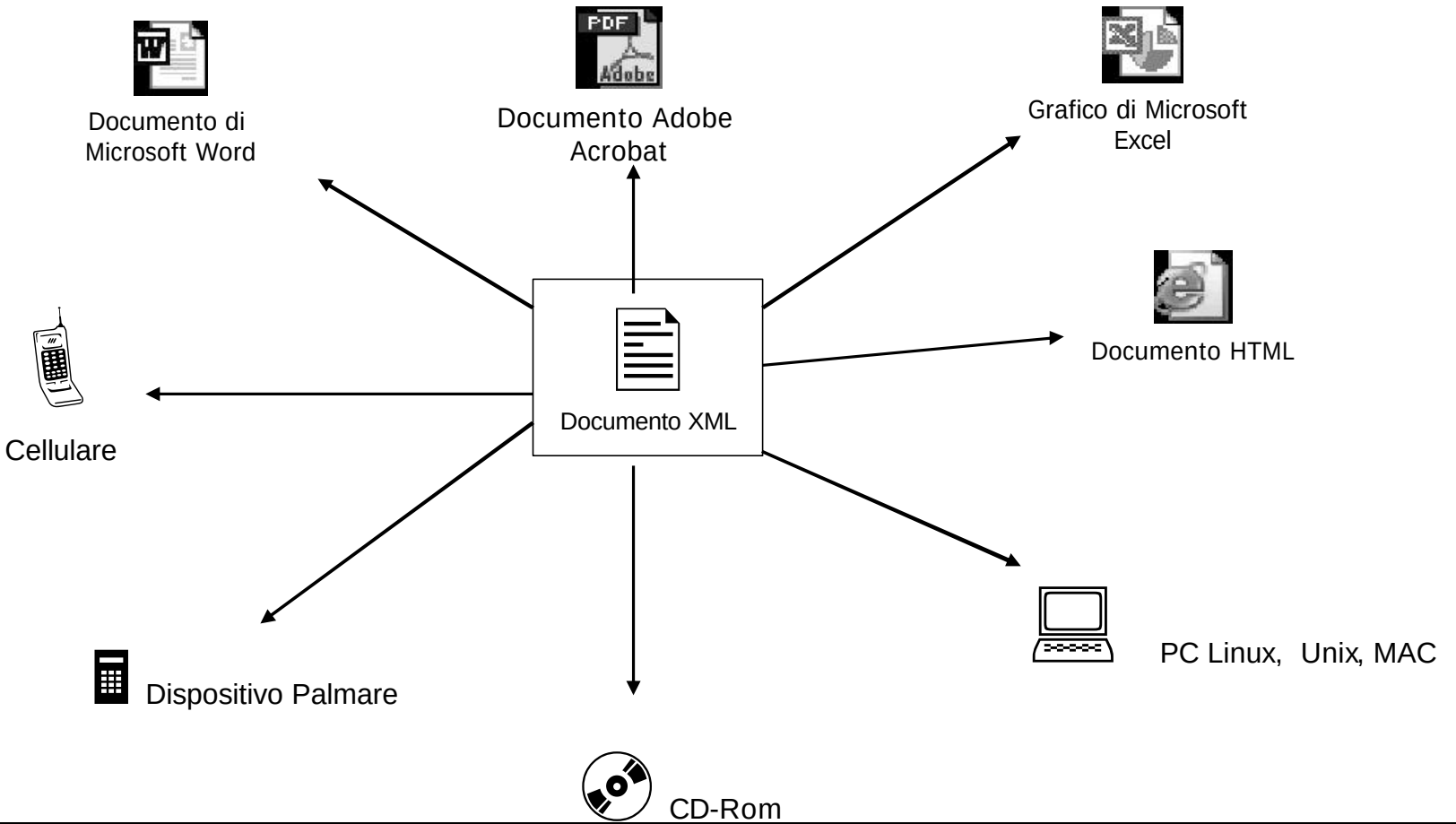
Alcune caratteristiche di XML

- Permette di organizzare a proprio piacimento le informazioni, consentendo di inviarle a chiunque in modo libero e gratuito.
- Standard aperto, no royalty, no brevetti, no copyright o segreti industriali.
- Salvataggio dei dati: se avete dei file in un formato proprietario quale Visicalc, Lotus Jazz, magari su floppy da 5”¼, la possibilità di recuperarli è piuttosto bassa. Il formato di solo testo resiste abbastanza bene alle alterazioni magnetiche.
- Permette di condividere sul Web documenti con strutture anche molto complicate che contengono dati particolarmente complessi.

Alcune caratteristiche di XML

- L'importanza di uno standard di comunicazione tra applicazioni, per evitare che i dati possano essere letti e "capiti" solo da pochissimi programmi.
- La gran parte dei flussi di informazioni viaggia attraverso sistemi proprietari e costosi, ad es. Oracle, SQL, DB2.
- Ciò che invece, possiamo ottenere con XML è questo:

XML


Documento di Microsoft Word
Documento Adobe Acrobat
Grafico di Microsoft Excel
Documento HTML
Documento XML
Cellulare
Dispositivo Palmare
PC Linux, Unix, MAC
CD-Rom

XML Alcuni punti chiave considerati nella progettazione

- Supporto di una vastissima gamma di applicazioni (non solo Web).
- I documenti XML devono rimanere leggibili da parte dell'uomo, e ragionevolmente chiari.
- Minimizzare le caratteristiche opzionali.
- Facile da redarre, ossia creabile anche con semplici editor di testo.
- Formale e conciso: piccoli insiemi di regole ma assolutamente precise.

Struttura XML

```
<?xml version="1.0"  ?>
<Media>
    <CDs>
        <CD quantity=“10”>
            <Title>Glory and consequence</Title>
            <Artist>Ben Harper</Artists>
            <Album>The wheel to live</Album>
        </CD>
        <CD quantity=“6”>
            <Title>Bigmouth strikes again</Title>
            <Artist>The Smiths</Artists>
            <Album>Meat is murder</Album>
        </CD>
    </CDs>
    <Books>
            <Book quantity=“3”>
                <Title>Caves bird</Title>
                <Author>Ted Hughes</Author>
            </Book>
    </Books>
</Media>
```

I tag XML vengono chiamati nodi o elementi.

Il tag che contiene tutti gli altri viene detto elemento radice.

Annidamento, nodi figli.

Nodo di testo, Attributi

Regole di sintassi e strutturazione

- Quando un documento XML rispetta le regole strutturali definite dal W3C, si dice che è un documento "*ben formato".* Gli interpreti XML rifiutano e ignorano i documenti mal formati. Ciò invece non accade in Html.
- Tutti i documenti XML devono avere una tag radice.
- Tutti i tag devono necessariamente essere chiusi.
- L' annidamento dei tag deve rispettare la gerarchia di apertura. <tag1><tag2>Getafix</tag1></tag2> NON e' consentito in XML (in Html si!)
- Il valore degli attributi DEVE sempre essere racchiuso tra virgolette.
- I nomi degli elementi e degli attributi devono iniziare con una lettera, un underscore _ o un segno di due punti (:), poi possono seguire un numero qualsiasi di lettere, numeri, due punti, punti singoli, trattini e underscore. Attenzione perché XML è "case sensitive".

La ricerca delle informazioni

Internet come biblioteca virtuale

La funzione “principe” di internet: la ricerca

- La metafora della più grande biblioteca del mondo:
  - Disordinata
  - Non “certificata”
- condizioni necessarie:
  - conoscere la rete (attori che vi pubblicano le informazioni
  - saper utilizzare gli strumenti (come operano i “motori di ricerca”)

Caratteristiche dell'informazione in Internet

- Informazione
  - NON strutturata ed eterogenea (testi, immagini, video, ….)
  - Ad alto dinamismo e caotica : modifiche continue, non controllabili
  - Non "classificabile": banche dati specialistiche ed a contenuto di alto valore professionale sono disponibili " allo stesso livello" di contenuti a fini divulgativi e meramente commerciali

➢ La ricerca può avvenire solo per "parole" e combinazioni di "parole" nel testo (pagina)

Gli accorgimenti per una “buona ricerca”

- Per l’utilizzatore:
  - Strategie di ricerca
  - L’organizzazione dei risultati ottenuti
- Per il fornitore di informazione:
  - Costruire **buona** informazione e **buoni** ipertesti (“collegamenti funzionali allo scopo del sito”

Ma attenzione ai conflitti di interesse:

- Il fornitore di informazione vuole ottenere il maggior numero di visitatori possibile e per il maggior tempo possibile (meglio se acquista i prodotti pubblicizzati)
- Il fruitore naviga per:
  - Studio
  - Ricerca di informazioni particolari e/o specialistiche
  - Vuole “confrontare” più informazioni tra loro

Gli strumenti

- I motori di ricerca
  - di gran lunga i più usati

- Altri strumenti per la ricerca di :
  - Programmi o file “particolari”
    - Gopher per archivi catalogati attraverso strutture ad albero (ambito accademico)
    - Selezione di newsgroup , per la ricerca di gruppi di interesse specifici (newsgroup filter)

Come funziona un motore

- Un particolare programma scorre sistematicamente e periodicamente TUTTO il web generando degli indici aggiornati con tutte le parole trovate, assegnando loro dei pesi per ogni pagina

- Al momento della ricerca il motore scorre solo questi indici e propone la lista dei siti con le parole trovate, evidenziando alcune informazioni di carattere generale (occupazione di memoria della pagina, data dell'ultimo aggiornamento….)

Alcuni motori di ricerca

- Altavista
- Lycos
- Excite
- Google
- Copernic
- Excite
- Virgilio

i motori all'interno dei grandi portali fanno uso di motori più noti ed efficienti

I motori di ricerca

- Ricerca attraverso una o più parole
- La “prevalenza” è stabilita dall’ordine delle parole ma la presentazione dei risultati avviene secondo regole (pesi) interne al motore stesso (e diverse da motore a motore)
- Esempi: numero di volte che la parola compare nella pagina, “dove” compare…)

L’ordine “logico” dei risultati può essere stravolto da accordi commerciali tra il fornitore di informazione e l’ente che governa il motore di ricerca

Consigli per la ricerca

- Formulare “a priori” uno schema concettuale di ciò che si desidera trovare
- visionare sempre la “home-page” del sito di interesse
- Controllare la validità dei “link” messi a disposizione dal sito

Gli elementi di forza del navigatore

- Scelta della “rotta di navigazione”:
  - Quale motore
  - Quali pagine aprire
  - Come navigare sui siti scelti
  - Come raccogliere le informazioni e quali
  - Come organizzarle (stampa, archiviazione..)

Informazione organizzata in rete

- ***Portale***

- Prodotto editoriale on-line che svolge funzione di punto privilegiato di accesso al web per gli utenti e che fornisce loro le risorse informative, i servizi di comunicazione personale e gli strumenti con cui ricercare contenuti ed altri servizi sul web

Portale: informazione organizzata in rete

- Prodotto editoriale su web
  - Insieme di pagine web organicamente collegate e coerenti secondo uno specifico punta di vista (di argomento,di riferimento istituzionale, pubblicitario, di impresa…….)
- Il problema della fidelizzazione
  - Nomadismo del navigatore (*a differenza del lettore di giornale)*
  - Chi e che cosa possono attrarre l'utente che si accinge a navigare nella rete?

I portali si diversificano

- Portali orizzontali
  - Generalisti (ampio spettro tematico)
  - Offrono servizi “general-purpose” (su web)
  - Utenza indifferenziata

- Portali verticali
  - Domini tematici particolari (finanza, sport, cinema,informatica ecc.)
  - Gruppi sociali e comunità caratterizzanti interessi comuni (affinity portal)

I requisiti per mettere informazione in rete

- Saper realizzare pagine web
  - Conoscere il linguaggio HTML: linguaggio di marcatura , formale e non di programmazione, logica semplice che permette di indirizzare le parti di una normale pagina
  - In alternativa : utilizzare prodotti “end-user” che permettono la costruzione “visiva” di pagine web, senza conoscere l’HTML (ES. Netscape Composer, Frontpage…..

I requisiti per mettere informazione in rete

- Acquistare spazio web, in generale presso il fornitore di connettività (ISP); verificarne le caratteristiche di sicurezza (molte linee, molto veloci, anche in relazione al numero di utenti)
  - Alcuni Internet Provider offrono spazio web (limitato) anche a titolo gratuito (siti amatoriali)
  - Dotarsi di un server proprio e stipulare un contratto con un ISP

Le frodi informatiche

Le frodi informatiche

- Accessi non autorizzati
  - Sicurezza informatica (tecnologia)
- Violazione della privacy
  - Normativa vigente estesa alla rete
- Violazione dei diritti del copyright : occorre una rivisitazione della normativa con un respiro planetario (già in essere)
- Frodi finanziarie : tecnologia di protezione

➢ Tentativi “censori” o “limitativi” ad hoc non sortiscono effetti validi per la natura stessa della rete

Problemi di sicurezza in rete

- La parte di File System del server accessibile al client è controllata dal server (i nomi delle risorse sono relativi a tale parte di file system!)
- Il client può scaricare dalla rete programmi (es Java) che vengono poi ***automaticamente*** eseguiti dal browser (ad es animazioni); tali programmi hanno permessi molto limitati per evitare intrusioni nel sistema del client

Problemi di sicurezza in internet

- **intranet:** rete privata (aziendale) utilizzata in modo da sfruttare al suo interno i servizi offerti dalle reti geografiche
- **firewall**: macchina dedicata utilizzata da una rete privata per filtrare il traffico tra la rete privata e la rete esterna (internet).

Problemi di sicurezza: crittografia delle informazioni

- I messaggi trasmessi sulla rete possono essere **ascoltati** da un **"intruso"** che può anche **modificare** il messaggio o inviare un messaggio **apocrifo**.
- **Crittografia**
- I dati possono essere **crittati** (**cifrati**) per impedire che vengano letti o alterati mentre vengono spediti sulla rete. Verranno poi **decrittati** (**decifrati**) dal ricevente.
- Avremo così:
  - **testo in chiaro**
  - **testo cifrato**
- Le tecniche di crittografia usano una **chiave crittografica** conosciuta solo dai due partner della comunicazione.

Tecniche elementari di crittografia

- **Sostituzione**
  - Alfabeto: a b c d e
    …
  - Alfabeto cifrato: n z q a h
    …
- Problemi: tecniche statistiche (conoscenza della frequenza delle lettere, delle coppie etc. permettono una facile decrittazione.

Tecniche elementari di crittografia

Trasposizione

Si altera l`ordine delle lettere del testo secondo una chiave.

Testo in chiaro: *this is a lovely day*

3	2	4	1	**chiave crittografica**
T	H	I	S	
_	I	S	_	
A	_	L	O	
V	E	L	Y	
_	D	A	Y	

Testo crittato: S_OYY HI_ED T_AV_ ISLLA

Crittografia a Prodotto

Usa contemporaneamente sostituzione e tras-posizione.

Algoritmi di crittografia

- **Algoritmi a chiave privata (simmetrici):**
  - Si usa la stessa chiave per la codifica e per la decodifica. La chiava deve essere privata e conosciuta dalle due persone che comunicano.
- **Algoritmi a chiave pubblica (asimmetrici):**
  - Si usano due differenti chiavi. Una **pubblica**, nota da tutti per **crittare**, ed una **privata**, nota solo al destinatario e da mantenere segreta, per **decrittare**.

Algoritmi simmetrici

- Un algoritmo di questo tipo generalmente utilizzato al momento è il **DES** (Data Encription Standard), nelle sue versioni *con chiavi a 56 e 112 bit.*
- Questi algoritmi **non** si presta bene a garantire la riservatezza nella comunicazione continuativa fra n soggetti indipendenti:
  - **Occorre una chiave privata per ogni coppia di soggetti**
  - **Ogni soggetto è costretto a possedere n-1 chiavi**, a mantenerle segrete ed a ricordare quale sia la chiave da utilizzare per comunicare con ciascuno degli altri soggetti.
  - Nel caso in cui la chiave sia generata autonomamente dal soggetto che avvia la comunicazione, è necessario che venga trasmessa al destinatario affinché questo possa decifrare i messaggi che riceve. E **durante il trasferimento la chiave potrebbe essere intercettata**.

Algoritmi asimmetrici

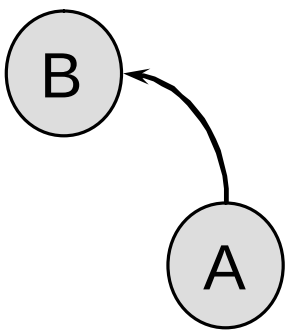
- Gli algoritmi **asimmetrici** sono di concezione recente (1976) ed utilizzano due chiavi distinte per cifrare e decifrare, con alcune proprietà fondamentali:
  - un documento cifrato con una chiave può essere decifrato con l'altra **e viceversa**.
  - le chiavi vengono generate in coppia da uno speciale algoritmo ed è di fatto impossibile ottenere una chiave a partire dall'altra.
  - Una qualsiasi delle due chiavi viene detta **pubblica**, è può essere distribuita. L'altra, detta privata, deve essere mantenuta segreta.

Algoritmi asimmetrici

- L'algoritmo **RSA**, proposto da Rivest, Shamir e Adleman nel 1978, è attualmente considerato come standard per la crittografia a chiave pubblica.
  - RSA basa la sua robustezza sulla complessità algoritmica della scomposizione in fattori primi, operazione per la quale non è attualmente noto un algoritmo efficiente.
  - Esistono varie implementazioni di RSA, che variano in funzione della dimensione in bit delle chiavi, e quindi del grado di sicurezza offerto. Chiavi di 512 bit sono un buon compromesso fra sicurezza e prestazioni.

Algoritmi asimmetrici

- Essendo asimmetrico, **RSA** risulta molto più versatile di DES, ed è alla base di tutti i servizi di sicurezza.
- Nella comunicazione fra n soggetti, ad esempio, RSA garantisce riservatezza con 2n chiavi, una coppia (pubblica,privata) per ciascun soggetto.

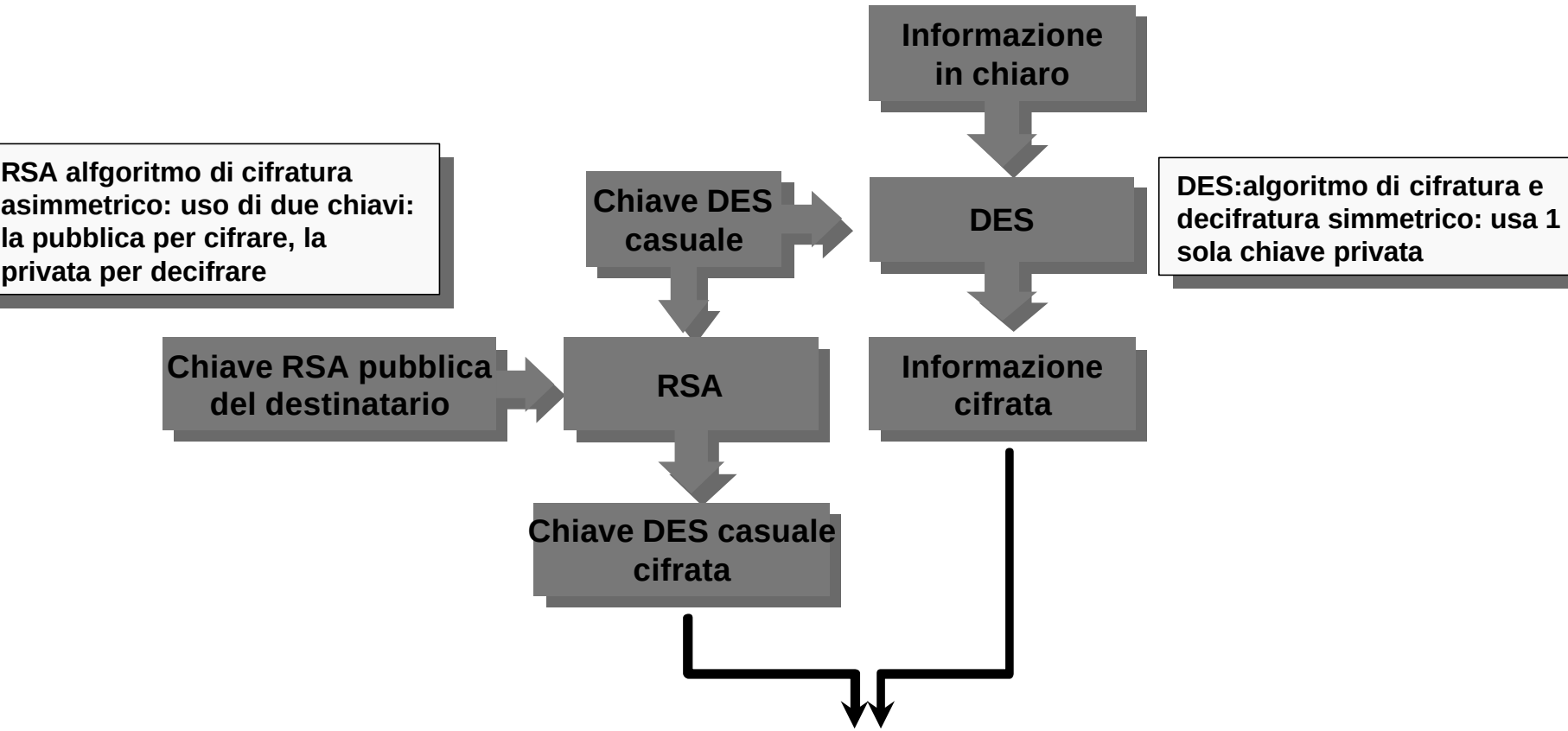


Se il soggetto A vuole inviare un messaggio riservato al soggetto B, ad esempio, cifra il messaggio con la chiave pubblica di B che, in quanto pubblica, è nota a tutti. In questo modo il messaggio sarà decifrabile soltanto con la chiave privata di B che, in quanto privata, solo B conosce.

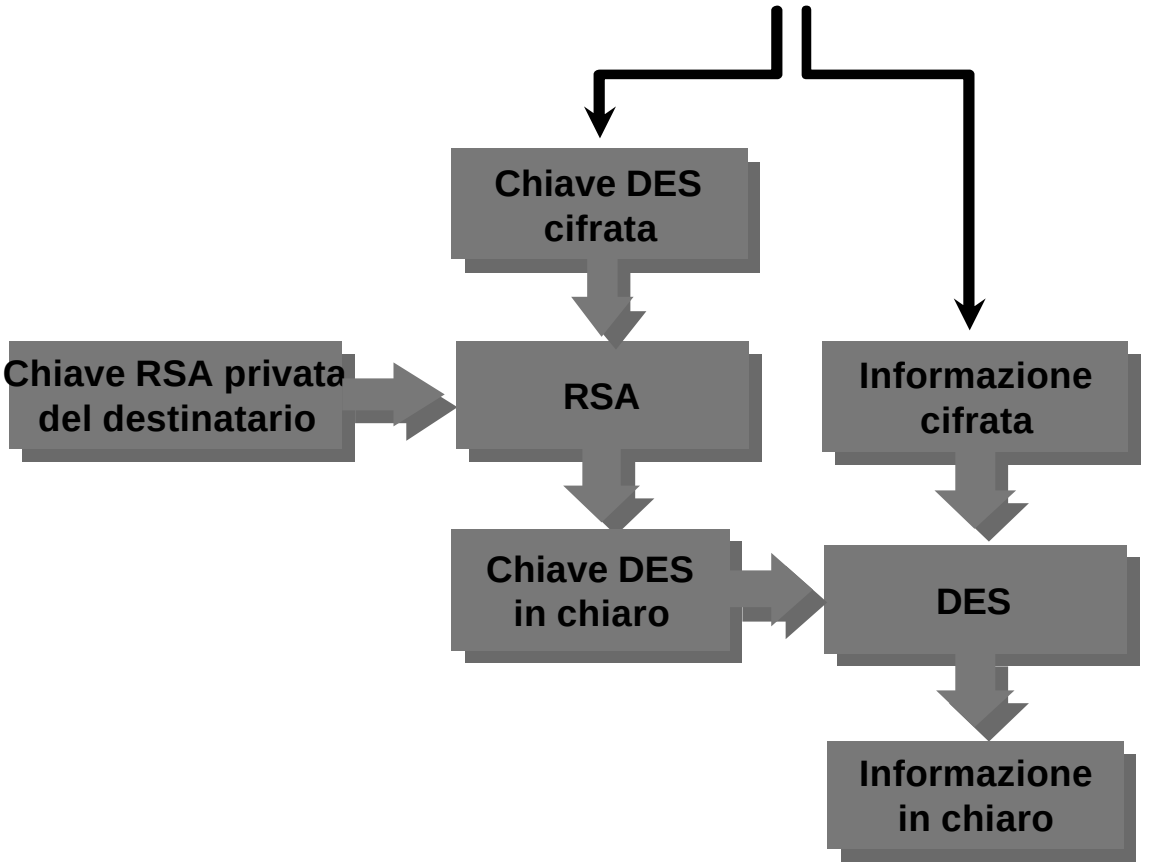
Algoritmi asimmetrici

- A causa della complessità algoritmica, le implementazioni di RSA sono generalmente troppo lente per cifrare direttamente i documenti
  - Per questo motivo RSA si utilizza spesso in congiunzione con altri algoritmi.
  - Per inviare un documento di grandi dimensioni in modo riservato, ad esempio, si genera una password casuale, si cifra il documento con DES (algoritmo simmetrico) utilizzando la password casuale, poi si cifra la password stessa (112 bit al massimo) con RSA, ed infine si invia il tutto (documento e password entrambe cifrati) al destinatario.

Introduzione di riservatezza



Rimozione di riservatezza



Politiche di Sicurezza

- L'insieme delle misure (di carattere organizzativo e tecnologico) tese ad assicurare a ciascun utente autorizzato (e a nessun altro) tutti e soli i servizi previsti per quell'utente, nei tempi e nelle modalità previste.
- Più formalmente, l'insieme delle misure atte a *proteggere* i requisiti che si desidera il sistema soddisfi, in termini di
  - **disponibilità**
  - **integrità**
  - **riservatezza**

Politiche di sicurezza: Requisiti

- Il sistema deve impedire la **alterazione** diretta o indiretta delle informazioni, sia da parte di utenti e processi non autorizzati, che a seguito di eventi accidentali.

- Il sistema deve impedire la **alterazione** diretta o indiretta delle informazioni, sia da parte di utenti e processi non autorizzati, che a seguito di eventi accidentali.

- Nessun utente deve poter ottenere o dedurre dal sistema informazioni che non è autorizzato a conoscere.

Politiche di sicurezza

- Occorre partire dal presupposto che, a dispetto delle misure attuate, un evento indesiderato possa comunque violare i requisiti di disponibilità, integrità e riservatezza, attraverso meccanismi che non avevamo previsto.

- Proteggere i requisiti di sicurezza di un sistema significa, in termini realistici,
  - Ridurre ad un valore accettabile la **probabilità** che vengano violati.
  - Individuare tempestivamente quando ed in quale parte del sistema questo accade.
  - Limitare i **danni** e ripristinare i requisiti violati nel minor tempo possibile.

Analisi del Sistema Informatico

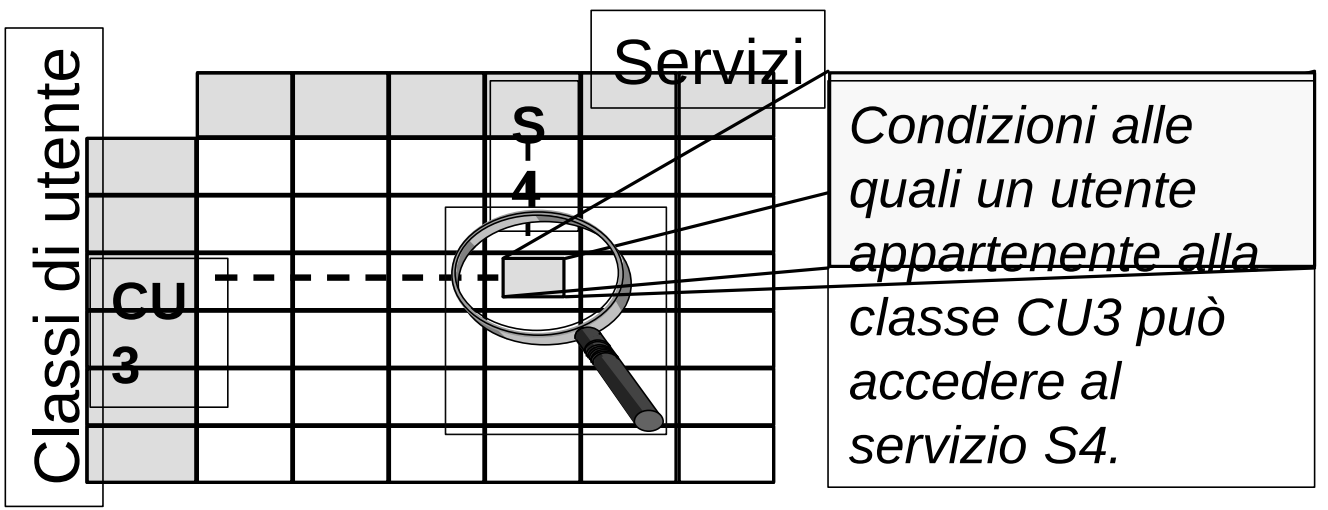
- Risorse **fisiche**
  - Il sistema visto come insieme di dispositivi che vanno protette da furti e danni materiali.
- Risorse **logiche**
  - Il sistema come insieme di informazioni, flussi e processi. Vanno classificate in base al loro valore per l'organizzazione, al contesto in cui si opera, al grado di riservatezza
- Analisi delle dipendenze fra risorse
  - Per ciascuna risorsa del sistema, fisica o logica, occorre individuare di quali altre risorse ha bisogno per funzionare correttamente.
  - Questa analisi tende ad evidenziare, almeno in prima battuta, le risorse **potenzialmente critiche** del sistema, cioè quelle da cui dipende il funzionamento di un numero rilevante di altre risorse

Analisi e Classificazione utenti

- Un presupposto essenziale per la sicurezza di un sistema, è che gli utenti possano controllarlo ed accedere alle informazioni **esclusivamente** attraverso i servizi da lui stesso messi a disposizione.
- È quindi fondamentale individuare con precisione tutti i servizi offerti dal sistema informatico, al fine di verificare poi, in maniera sistematica, che ogni servizio risponda pienamente a **tutte e sole** le specifiche di progetto (e non presenti, ad esempio, pericolosi *side-effects).*

Diritti di accesso: esempio di matrice utenti / servizi)

- **Per ogni coppia**
- (Classe di utente **CU,** Servizio **S**),
- si definiscono le condizioni che regolano ***se, come e quando*** un utente appartenente alla classe **CU** può accedere al servizio **S**.



La password

Assegnazione di un **profilo** il cui accesso è controllato mediante **password**; al profilo (e quindi all'utente) sono associate una serie di azioni, di permessi e di divieti.

Questi controlli vengono realizzati a diversi livelli:

- delle basi di dati,
- del sistema operativo
- di gruppi di utenti

Esempio: una ***password*** è richiesta per la connessione da casa al *provider Internet* o per entrare in un'area di servizi a cui l'utente è abilitato ad accedere gratuitamente(previa iscrizione con invio di dati identificativi) o dietro pagamento di un canone.

Password : criteri di scelta

- La *password* deve essere la più *lunga* possibile
- La *password non* deve essere in alcun modo *collegata* alla vita privata dell'utente (soprannomi, diminutivi, date di nascita ecc.)
- La *password non* deve essere una *parola comune* riportata in un vocabolario
- La *password non* deve venire *scritta* da nessuna parte
- La *password deve essere variata periodicamente*

Eventi indesiderati

- **Un qualsiasi accesso** (a servizio o informazione) **che non sia esplicitamente permesso** dalla rispettiva matrice dei diritti.
- L'insieme degli eventi indesiderati, **tuttavia**, è **più esteso** in quanto comprende eventi che non sono affatto degli accessi, dal guasto di un disco all'attacco di un virus.
- Occorre condurre una indagine **sistematica** al fine di individuare il maggior numero possibile di eventi indesiderati.
- A tal fine è possibile in generale distinguere
  - **Attacchi** intenzionali
  - **Eventi accidentali**

Attacchi

- Classifichiamo gli attacchi intenzionali in funzione di
  - **la risorsa**, fisica o logica, oggetto dell'attacco
  - **la tecnica** utilizzata per condurre l'attacco.
- E' naturale pensare che una risorsa può essere attaccata con un piu' tecniche contemporaneamente ed un sistema sara' attaccato su piu' risorse contemporaneamente
- Le tecniche di attacco possono essere classificate in funzione del livello al quale operano (logico o fisico)

È

Tecniche di attacco

- Gli attacchi a livello **fisico** sono principalmente tesi a sottrarre o danneggiare risorse critiche.
  - **Furto.** Prevedibile per nastri di backup, dischi o interi server. *È un attacco alla disponibilità ed alla riservatezza*.
  - **Danneggiamento.** Prevedibile per apparecchiature e cavi di rete, più difficilmente per calcolatori, dischi ed apparecchiature di supporto come trasformatori di corrente ed impianti di condizionamento. *È un attacco alla disponibilità ed alla integrità.*

Tecniche di attacco

- Gli attacchi a livello **logico** sono tesi a sottrarre informazione o degradare la operatività del sistema.
- Un attacco logico può essere classificato come di
  - **Intercettazione e deduzione** *(attacco alla riservatezza, incrocia informazioni tratte dall'osservazione del sistema con informazioni ottenute per altre vie. Es: informazione negata)*.
  - **Intrusione** *(attacco alla integrità ed alla riservatezza. Es:* Accesso con password di altro utente, introduzione di backdoor *)*.
  - **Disturbo** *(attacco alla disponibilità. Es: virus, worms, denial of service)*.

Individuazione e Integrazione delle contromisure

- Individuazione del **sotto-insieme** di **costo minimo** che al contempo rispetti alcuni **vincoli** essenziali:

  - **Completezza**. Il sotto-insieme delle contromisure scelte deve comunque far fronte a tutti gli eventi indesiderati individuati per il sistema in esame.
  - **Omogeneità**. Tali contromisure devono essere compatibili ed integrabili tra loro in modo da minimizzare il costo della loro attuazione congiunta.
  - **Ridondanza controllata**. La ridondanza delle contromisure ha un costo e deve quindi essere rilevata e vagliata accuratamente. Può accadere, ad esempio, che più contromisure siano inutilmente ridondanti, che ad esempio neutralizzino un medesimo evento valutato a basso rischio. D'altra parte, è anche possibile che un evento ad alto rischio, che potrebbe e dovrebbe essere neutralizzato da più di una contromisura, di fatto non lo sia.