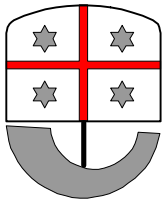


REGIONE LIGURIA

I CONCETTI FONDAMENTALI DI INTERNET

di Beppe Pavoletti (Servizio Programmi e Strutture Culturali)

*L'autore di questo testo è raggiungibile per commenti e suggerimenti all'indirizzo email giuseppe.pavoletti**@regione.liguria.it**.*

V 7.3 Aggiornamento: 26 dicembre 2000
Versione precedente: 7.2 6.10.2000

ATTENZIONE: chi legge questo documento utilizzando Winword deve attivare la visualizzazione layout di pagina per poter vedere lo schema contenuto al punto 3.2.6.5

SOMMARIO

1. ELABORAZIONE DEI DATI

L'uso di Internet è una attività che comporta l'uso del computer. Bisogna quindi cercare di comprendere (pena il non capire nulla di tutto il resto, o capirlo a metà, o capirlo al rovescio) i concetti fondamentali su cui si basa il funzionamento dei computer. La parte più teorica della trattazione, riguardante informazione e computabilità, si trova in appendice al punto 6.1: si raccomanda comunque di leggerla con attenzione perché contiene concetti di grande importanza.

1.1 **Calcolatori**

Un calcolatore è un dispositivo fisico atto ad eseguire dei calcoli. In questa sede, in cui non vogliamo addentrarci nella teoria della computabilità, diciamo che i calcoli sono manipolazioni di simboli che avvengono secondo certe regole e devono avere una precisa conclusione, ossia produrre un certo risultato. Poiché un calcolatore è un dispositivo fisico, i simboli che vengono manipolati sono eventi fisici, che possono essere considerati come portatori di informazione (cfr. 6.1). Si discute, tra i filosofi, se un sistema fisico possa essere intrinsecamente in uno stato computazionale, oppure se questo si possa dire solo in base ad una interpretazione effettuata da una intenzionalità esterna: questo problema però esula dagli scopi di questo documento.

Non c'è un motivo intrinseco per cui un calcolatore debba essere elettronico: in linea di principio potrebbe essere realizzato attraverso un insieme di rubinetti aperti o chiusi, o di persone che fanno certi segnali piuttosto che altri. Ci sono però dei motivi pratici: l'uso dell'elettronica consente di realizzare calcolatori molto più veloci e potenti di qualunque altra tecnica finora nota. Quindi un calcolatore elettronico è un sistema fisico che esegue calcoli attraverso eventi di carattere elettromagnetico. I calcolatori oggi utilizzati utilizzano un insieme di "componenti" (nel senso astratto di elementi costitutivi) ognuno dei quali può trovarsi in due stati (in genere tensione più alta o tensione più bassa), e quindi codifica 1 bit di informazione.

Andando ora decisamente nel concreto, diciamo che un calcolatore elettronico consta di diverse parti:

1. una unità di elaborazione, che è quella che esegue i calcoli veri e propri, e che legge delle informazioni e ne produce (ne scrive) delle altre; questo componente è detto CPU = central processing unit
2. una memoria in cui si trovano informazioni lette o scritte dall'unità di elaborazione; la memoria poi in pratica di distingue in
   - 2.1 memoria volatile, più veloce ma che non mantiene le informazioni mentre l'apparecchio non è alimentato (comunemente detta RAM = random access memory)
   - 2.2 memoria di massa, più lenta, ma che mantiene le informazioni in modo permanente (hard disk, floppy disk, nastri, CD-ROM, DVD ...)
3. una o più periferiche, cioè apparecchiature che servono a comunicare con l'esterno, per ricevere o fornire informazioni: le periferiche più tipiche sono video e stampante, comunque sono periferiche anche mouse, modem, plotter, strumenti di misura, macchine a controllo numerico ecc. (in realtà, volendo, anche le memorie di massa potrebbero essere considerate periferiche e quindi classificate qui)

1.2 **Tipi di calcolatori**

Al mondo ci sono tanti computer, che si possono suddividere in alcune grandi categorie.

Una distinzione utile è quella tra **workstation** (o stazione di lavoro, cioè computer destinato all'uso di un singolo utente, eventualmente collegato ad una rete) e **server**, cioè computer che, attraverso una rete, fornisce servizi ad altri computer (ad esempio contiene dati che vengono letti da tutte le stazioni di lavoro collegate in rete).

Ci sono poi famiglie di computer che si distinguono per la potenza e il tipo di clientela cui sono destinate. Una classificazione potrebbe essere:

- **Home computer**. Categoria ormai scomparsa, comprende computer storici come i Commodore 16, 64 e 128, Apple II, Sinclair ZX Spectrum e altri, che hanno segnato la nascita dell'informatica individuale
- **Personal computer**. Sarebbero computer destinati all'uso personale, ma in realtà ormai comprendono poderose workstation e server in grado di supportare centinaia di utenti. Sono rappresentati principalmente dai compatibili IBM (prima solo con CPU Intel, ora anche con una delle numerose CPU compatibili) e, in misura minore dagli Apple Macintosh (altri prodotti prima di nicchia e ora scomparsi sono il Commodore Amiga che, come i Macintosh dell'epoca era basato sulle CPU Motorola serie 68000 e l'Acorn, popolare quasi esclusivamente in Gran Bretagna). I sistemi operativi usati sono DOS, Windows nelle sue varie versioni, OS/2, SCO Unix, Solaris, Linux, FreeBSD, BeOS, QNX e - sui Macintosh - MacOs e ancora Unix (specialmente Linux).
- **Computer di fascia media**. Comprende workstation per uso professionale, soprattutto nel CAD e nella grafica, e server ad altissime prestazioni, in grado di gestire sistemi informatici di grande potenza ed affidabilità, utilizzati in particolare in aziende e organizzazioni medie e grandi. Sono rappresentati soprattutto da prodotti IBM, Digital, Hewlett-Packard, Sun, Bull, Unisys, Silicon Graphics, Olivetti. Il sistema operativo più utilizzato è Unix, ma si trova anche Windows NT e qualche altro prodotto, come Open VMS[1]. Il tipico server che si incontra su Internet appartiene a questa categoria. In passato questi calcolatori venivano definiti *minicomputer*, ma ora il termine sembra piuttosto in disuso.
- **Mainframe**. Sono grandi server, dal costo di diversi miliardi, utilizzati in applicazioni strategiche, con grandissime esigenze di affidabilità, e necessità di gestire enormi quantità di dati e di utenti. Si trovano, ad esempio, in ambiente bancario, ma non solo (l'indice centrale del Servizio Bibliotecario Nazionale è su mainframe). Tipici rappresentanti della categoria sono i mainframe IBM con sistema operativo MVS. Ci sono anche i mainframe compatibili IBM, ma di solito ogni marca (Unisys, Bull, Data General ecc.) utilizza il proprio sistema operativo.
- **Supercomputer.** Computer di altissimo costo, utilizzati quando è necessaria una enorme potenza di calcolo, soprattutto per applicazioni scientifiche e tecniche (simulazioni, modelli matematici). I nomi più noti sono SGI (che ha incorporato la Cray) e Convex, sui quali vengono adottati sistemi operativi particolari, specializzati soprattutto nella gestione del calcolo parallelo, che è la vera ragion d'essere di questi computer.

1.3 **Programmi**

Un programma è quindi un insieme di istruzioni che vengono eseguite dal calcolatore, e corrispondono a stati fisici della CPU e della memoria. I dati sono informazioni sulle quali vengono eseguite le istruzioni, nonché il risultato della esecuzione delle istruzioni: per esempio, se l'istruzione è quella di eseguire una addizione, i dati sono gli addendi e il risultato. Anche i dati corrispondono a stati fisici della CPU e della memoria.

Normalmente però non si parla del funzionamento del calcolatore in termini di stati fisici, ma in termini astratti, per cui si dice semplicemente che la CPU sta eseguendo un programma e legge le istruzioni dalla memoria RAM e scrive dei dati nella RAM stessa o su una memoria di massa

E' possibile scrivere programmi in diversi modi, ad esempio scrivendo le istruzioni direttamente nella forma in cui vengono eseguite dal computer, ossia come sequenze di numeri binari. Scrivere programmi in questo modo è però estremamente difficile: per questo sono stati creati dei linguaggi simbolici, detti linguaggi di programmazione di alto livello, in cui le istruzioni vengono date in forma più comprensibile agli esseri umani, ad esempio il C, il Pascal o il Basic. C'è poi un programma specializzato (interprete o compilatore) che traduce queste istruzioni in forma eseguibile. Il programma scritto nel linguaggio di alto livello viene detto **codice** o **programma sorgente** (o semplicemente *sorgente*), quello in forma eseguibile viene detto appunto **codice** o **programma eseguibile** (o semplicemente *eseguibile*)

[1] Rientra in questa categoria la gamma AS/400 della IBM, con il sistema operativo OS/400. Si tratta di macchine utilizzate soprattutto nell'ambito gestionale, il cui uso come server Internet non è molto frequente, anche se ora supportano il TCP/IP

1.4 **Sistemi operativi**

Scendiamo ora più nel concreto, e diciamo che tutte le attività del calcolatore procedono in base all'esecuzione di qualche programma. Ogni programma serve per qualche scopo particolare: uno per scrivere, uno per disegnare ecc., e quindi ognuno è diverso dall'altro. Tuttavia ai primi degli anni '50, poco dopo l'avvento degli elaboratori elettronici, ci si rese conto che vi erano alcune cose che tutti i programmi, proprio per poter funzionare, dovevano fare: innanzitutto partire, poi utilizzare la RAM e le memorie di massa per leggere e scrivere, utilizzare la stampante ecc. Nacque così l'idea di scrivere un programma specializzato per effettuare queste operazioni, al quale tutti gli altri facessero ricorso. Questo programma è proprio il **sistema operativo**. Invece i programmi che svolgono le specifiche operazioni (il foglio elettronico, il word processor, il gioco ...) sono detti programmi applicativi. Si noti che in altri contesti la nozione di programma applicativo può essere più ristretta, ad esempio quanto si vuole porre in evidenza la differenza tra un particolare programma ed un ambiente di sviluppo come un compilatore o un data base.

È molto importante capire questa differenza, per sapere se, quando si effettua una certa operazione con il computer, si ha a che fare con il sistema operativo o con i programmi applicativi.

In sintesi, i principali compiti del sistema operativo sono:
- avviare i programmi
- gestire la RAM
- gestire le memorie di massa (file system)
- gestire le periferiche (stampante, video, mouse)
- permettere a più programmi di essere attivi contemporaneamente (multitasking)
- permettere a più utenti di utilizzare uno stesso computer o a più computer di comunicare tra loro (multiutenza o rete)

A proposito del multitasking, elemento fondamentale dei sistemi operativi moderni, osserviamo che esso può essere di tipo **cooperativo** se il sistema operativo può togliere un controllo al programma solo quando il programma stesso lo consente, oppure del più sofisticato tipo **preemptivo** quando il sistema operativo può comunque togliere il controllo a un programma. I sistemi operativi più potenti e moderni, da Windows 95 a Windows NT a Unix, adottano il multitasking preemptivo.

Un sistema operativo è composto da diversi elementi, ed in particolare i seguenti:
- il **kernel** è il cuore del s.o., perché è quell'insieme di procedure che si occupa della gestione dell'hardware, della sincronizzazione dei programmi e simili
- l'**interprete dei comandi** o **shell**, che può essere grafico o a riga di comando, permette all'utente di interagire con il kernel, e soprattutto di eseguire i programmi
- le **utilities di sistema** vengono fornite insieme al s.o. ma sono a tutti gli effetti programmi utente (come del resto, a rigor di termini, la shell); tali programmi di solito però sono progettati per svolgere funzioni strettamente collegate a quelle del sistema operativo, ad esempio la manutenzione dei file system

Si noti che la parte veramente essenziale è il kernel, che è quella parte che l'utente non può mai vedere direttamente. In genere i sistemi operativi permettono di utilizzare diverse shell, e a maggior ragione diverse utilities di sistema. Quello che l'utente vede del sistema operativo è in realtà la shell: in DOS la shell standard è il file command.com, in Windows a 16 bit è il Program Manager, in Windows 95/98/NT è l'Explorer (da non confondere con l'Internet Explorer), in OS/2 è il Presentation Manager, in Unix vengono usate a scelta dell'utente un gran numero di shell, come la C Shell, la Korn Shell, la Bourne-Again Shell e altre ancora.

Ci sono molti sistemi operativi: DOS, Windows 95 e 98, Windows NT, Windows ME, Windows 2000, OS/2, MacOS, BeOS, Unix (che indica una famiglia di s.o., ognuno con il suo nome: Linux, FreeBSD, SCO Unix, Solaris, AIX, HP-UX, Irix ecc.), VMS, Open VMS, OS/400, MVS, VM, OS/390, Unicos Max, QNX ecc.

2. RETI DI COMPUTER

La trattazione delle reti comprende molti aspetti strettamente interconnessi tra loro, che però ai fini dell'esposizione è necessario separare e trattare in punti diversi. Una piena comprensione della trattazione pertanto si potrà avere solo al termine della lettura di questo intero capitolo.

Una rete di computer è un insieme di dispositivi hardware (cioè di apparati materiali) e software (cioè programmi) che permettono a più computer di comunicare tra loro: ad esempio, da un computer collegato ad una rete si possono scrivere dati su un altro computer e prelevarne altri da un terzo, oppure eseguire un programma che si trova su un altro computer. Un concetto fondamentale, a proposito delle reti, è che esse, per poter funzionare, implicano diversi tipi di attività, che si riferiscono a diversi aspetti dei compiti delle reti: ognuno di questi aspetti viene regolamentato in modo diverso, attraverso standard differenti, e di solito un determinato standard può interfacciarsi con numerosi altri standard tra quelli che regolamentano **gli altri** aspetti del funzionamento della rete (v. anche quello che verrà detto di seguito a proposito del modello OSI).

Perché più computer possano essere collegati in rete non è necessario che siano simili: infatti sono proprio l'hardware e il software di rete che provvedono a quanto necessario affinché computer che internamente funzionano in modo diverso (ad esempio con sistemi operativi diversi) possano comunicare tra loro.

Una rete propriamente detta interconnette computer autonomi dotati di propria capacità di elaborazione, e dotati ciascuno di un identificatore univoco che lo distingue da ogni altro (indirizzo di rete). Non tutte le interconnessioni tra dispositivi informatici sono reti. **Alcuni esempi di cose che non sono reti**, anche se a prima vista possono sembrarlo:

- terminali collegati a un server centrale attraverso la porta seriale; in questo caso, i terminali (che possono anche essere computer opportunamente configurati) funzionano come semplici periferiche del server, analogamente al monitor e alla tastiera locale[2]
- dispositivi per la condivisione di una stampante tra più computer, che consistono in un interruttore che mette in comunicazione la stampante di volta in volta con uno specifico computer: anche qui si tratta solo del rapporto tra un computer e le sue periferiche, e il commutatore si limita a cambiare il collegamento fisico tra il computer e le periferiche
- dispositivi per la condivisione di monitor e tastiere, che permettono di utilizzare un monitor e una tastiera per controllare più computer; i più sofisticati tra questi dispositivi permettono di avere un certo numero di coppie monitor/tastiera ognuna delle quali può essere collegata a uno tra numerosi computer a scelta dell'utente; tali apparecchiature inoltre possono anche fornire funzioni di autenticazione, in modo da permettere l'uso dei computer solo agli utenti abilitati e non a chiunque; si tratta di soluzioni adottate quando una persona deve collegarsi in momenti diversi a numerosi computer a ognuno dei quali però nessuno deve essere continuamente collegato (si pensi ai tecnici che devono amministrare numerosi server lavorando un po' su uno e un po' sull'altro), e quindi permettono di risparmiare costi e ingombro evitando di comprare un monitor e una tastiera per ciascun computer; permettono anche di collocare monitor e tastiera lontani dal computer; questi apparecchi sono concettualmente del tutto analoghi a quelli per la condivisione delle stampanti: si tratta in ultima analisi di commutatori di collegamenti elettrici, anche se nei modelli sofisticati la commutazione può essere fatta via software

Queste cose che non sono reti anche se lo sembrano sono tutte tecniche per la condivisione di periferiche e/o per il loro uso a distanza, mentre le reti permettono l'interazione tra computer autonomi e non tra un computer e le sue periferiche. **Si faccia attenzione alla terminologia:** il termine *commutatore* (*switch*) viene usato anche nel campo delle reti vere e proprie, ma con significato diverso, che vedremo a suo tempo[3].

[2] Questa architettura è stata largamente usata per molti anni per i collegamenti ai mainframe e poi anche ai server Unix, ma ora è stata quasi ovunque abbandonata

[3] Del resto in inglese *switch* può essere anche il comune interruttore della luce, e quindi non è un termine particolarmente legato all'informatica (così anche *to switch on* vuol dire *accendere* detto di un apparecchio)

Bisogna ora introdurre alcune distinzioni molto importanti.

Una prima distinzione è quella tra **trasmissioni sincrone** e **trasmissioni asincrone**. Le trasmissioni **sincrone** sono quelle in cui i bit di dati vengono inviati a una velocità fissa sempre uguale e il trasmettitore e il ricevitore devono essere sincronizzati, in modo che il ricevitore sia pronto ad elaborare i dati man mano che questi arrivano. Le **trasmissioni asincrone** invece non prevedono alcuna sincronizzazione tra trasmettitore e ricevitore e possono avvenire a velocità variabile, ma richiedono l'introduzione, nei dati, di marcatori per individuare l'inizio e la fine delle unità di dati trasmesse. In altri termini, una trasmissione asincrona non consiste in un unico flusso di bit, ma in gruppi di bit delimitati da appositi marcatori, in modo che il ricevitore sia in grado di riconoscerli indipendentemente dalla velocità con cui arrivano. Quasi tutte le tecnologie di rete che esporremo nel seguito si basano sulle **trasmissioni asincrone**. L'unica eccezione significativa sono le reti SONET, reti ad alta velocità in fibra ottica usate sulle dorsali che raccolgono grandi quantità di trasmissioni, che sono sincrone come dice il nome stesso: la sigla SONET sta infatti per Synchronous Optical NETwork[4].

Una seconda distinzione importante è quella tra **LAN** (Local Area Network o rete locale), rete che si estende su di un'area limitata, per esempio un edificio o pochi edifici adiacenti, come avviene per le sedi regionali di Via Fieschi, Via Ravasco e Via D'Annunzio) e **WAN** (Wide Area Network o rete geografica), rete che può estendersi a qualunque distanza.

Un'altra distinzione molto importante è quella tra reti **broadcast**, nelle quali ogni messaggio viene inviato a tutti i computer della rete, anche se viene accettato solo dal computer al quale è diretto e rifiutato da tutti gli altri, e reti **commutate**, nelle quali i messaggi vengono instradati nell'una o nell'altra parte della rete a seconda del computer al quale sono diretti (come esempio di rete commutata possiamo pensare anche alla rete telefonica, nella quale la voce di un utente che parla non viene inviata a tutti coloro che sono collegati alla rete, ma solo all'utente di cui è stato composto il numero). Le reti commutate sono composte da un insieme di collegamenti **punto-punto**, che cioè collegano due punti specifici e solo quelli, a differenza delle reti broadcast, in cui tutti sono collegati con tutti gli altri.

Le LAN sono tradizionalmente reti broadcast: tuttavia questo sistema è poco efficiente, perché comporta la trasmissione, attraverso tutta la rete, di dati che vengono in realtà scartati da tutti quelli che li ricevono meno uno, cioè il vero destinatario. Per questo si vanno diffondendo le LAN commutate (switched LANs), che sono divise in sezioni e sono dotate di appositi dispositivi che instradano i dati solo alla sezione dove in realtà si trova il destinatario. Le reti geografiche invece sono tutte commutate[5].

La commutazione a sua volta può essere di due tipi:

- commutazione **di circuito** quando viene creato un circuito permanente tra due entità che dialogano sulla rete, come avviene quando due persone si telefonano
- commutazione **di pacchetto** quando i dati scambiati vengono suddivisi in unità di struttura predefinita (pacchetti), ognuna delle quali viene instradata indipendentemente (e magari anche lungo percorsi diversi) verso il destinatario. In una rete a commutazione di pacchetto è però possibile creare dei circuiti virtuali, cioè permettere a due utenti di comunicare come se tra di loro ci fosse un circuito permanente. La maggior parte delle tecnologie di rete che tratteremo di seguito sono a commutazione di pacchetto (o di cella - v. seguito), ma con alcune eccezioni significative, come frame relay e ISDN.

[4] Anche l'architettura di rete SNA, che è una tecnologia proprietaria IBM, impiegata nei mainframe e negli AS400, fa uso di tramissioni sincrone ma, essendo una architettura alternativa a TCP/IP, non può essere impiegata su Internet direttamente ma solo attraverso dei gateway. Si noti che comunque ora anche i mainframe e gli AS400 della IBM supportano in modo nativo TCP/IP e i relativi protocolli applicativi.

[5] L'unica eccezione è probabilmente un segmento di rete per la TV via cavo quando viene utilizzato per trasmissione dati via cable modem; v. anche nel seguito

Più precisamente, il termine **pacchetto** designa di solito un blocco di informazioni di lunghezza variabile, per lo più dell'ordine di alcuni Kb[6]. Il fatto che i pacchetti abbiano lunghezza variabile comporta la necessità di dotarli di appositi campi atti ad identificare l'inizio e/o la fine del pacchetto, il che rende la trasmissione meno efficiente: infatti questi dati aggiuntivi occupano spazio, ed inoltre devono essere elaborati dalle varie apparecchiature di rete. Si parla invece di **cella** quando i dati sono suddivisi in unità di lunghezza fissa e di piccole dimensioni (53 byte nelle reti ATM), atti ad essere elaborati a grande velocità anche da programmi implementati in hardware, e quindi più veloci. Sia nei pacchetti che nelle celle si distingue un **payload** (carico utile) cioè i dati veri e propri che devono essere trasportati e un **header** (intestazione)che contiene le informazioni utili ai vari protocolli per svolgere i loro compiti. Il rapporto tra header e payload varia molto a seconda del protocollo.

Prima di proseguire, è opportuno tornare alla commutazione per osservare che in diverse architetture, e segnatamente in ATM, si parla di **circuiti virtuali**, che non sono circuiti fisici, risultati da un insieme di commutatori elettrici, ma circuiti logici implementati via software, in modo tale che i pacchetti non vengano gestiti ciascuno in modo indipendente, ma instradati in un circuito nel senso definito sopra. Si deve poi distinguere tra **PVC** (Permanent Virtual Circuit = Circuito Virtuale Permanente), che è un percorso statico definito manualmente, e **SVC** (Switched Virtual Circuit = Circuito Virtuale Commutato) che è un percorso stabilito dinamicamente su richiesta dell'utente e poi disattivato quando non più necessario. Gli SVC sono assai più complessi da gestire rispetto ai PVC, perché richiedono sofisticati protocolli di segnalazione che "attraversano" la rete per verificare la presenza delle risorse necessarie, riservarle e gestire poi il circuito fino al momento di chiuderlo.

Per meglio comprendere il funzionamento delle reti è utile fare riferimento al **modello OSI**, che rappresenta una rete come un insieme di strati:

7 Applicazione
6 Presentazione
5 Sessione
4 Trasporto
3 Rete
2 Data Link LLC MAC
1 Fisico

Ogni strato o livello (*layer)* fornisce i suoi servizi a quelli superiore e usufruisce dei servizi di quelli inferiori. Quando uno strato riceve le informazioni da quelli superiori, vi aggiunge altre informazioni sue proprie, necessarie per lo svolgimento dei suoi compiti. Quando invece passa le informazioni a quelli superiori, elimina queste informazioni proprie del livello, e mantiene solo quelle utili al livello superiore. Risulta quindi evidente che ciascuno strato non ha bisogno di conoscere quello che avviene nell'ambito degli strati inferiori e superiori: è sufficiente che sia in grado di ricevere (dagli strati inferiori) e fornire (agli strati superiori) informazioni conformi a quanto previsto dal protocollo che viene utilizzato.

Analogamente, perché i computer interessati possano comunicare tra loro, è sufficiente che possano produrre e riconoscere dati nel formato previsto per i vari strati dell'architettura di rete, mentre non ha importanza come svolgono le loro operazioni interne (ad esempio come avviano un programma o come scrivono sul disco).

Tra i livelli sono di particolare importanza: il 7 (applicazione), dove si trovano le applicazioni dell'utente, per esempio quella per inviare posta elettronica; il 4 (trasporto), che gestisce la connessione logica tra due computer (si dice anche connessione end-to-end perché le entità che si prendono in considerazione sono solo i

[6] Questa è una definizione generale di pacchetto; spesso si parla di *pacchetti* al livello OSI 3 e di *frames* al livello OSI 2; cfr. nel seguito la descrizione dei livelli OSI

due computer che comunicano, e non tutti i dispositivi che si trovano tra di loro, e che sono di pertinenza del livello 3); il 3 (rete), che gestisce l'interconnessione di reti e l'indirizzamento, cioè provvede a fare in modo che i messaggi vengano effettivamente indirizzati al computer cui sono destinati; il 2 (data link), che gestisce l'accesso al mezzo fisico; l'1 (fisico), che è rappresentato dal vero e proprio mezzo fisico per la trasmissione, per esempio i cavi e i connettori.

Bisogna a questo punto fare accenno ancora ad alcuni concetti importanti: innanzitutto quello di **sottorete** (subnet); la sottorete è l'insieme di tutta l'infrastruttura di comunicazione, cioè di tutte le linee e gli apparti diversi dai sistemi finali, ossia dalle stazioni di lavoro e dai server a disposizione degli utenti: essa quindi comprende in particolare gli apparati di switching e di routing, che saranno illustrati nel paragrafo 2.3; non bisogna confondere questa accezione generale di sottorete con il significato che il termine assume nell'ambito dell'indirizzamento IP, e che sarà esposto in seguito; altra nozione importante è quella di **sistema autonomo** (autonomous system o AS): un sistema autonomo è un insieme di una o più reti che si trovano sotto una gestione comune (quindi la LAN di Via Fieschi e la rete geografica costituiscono un sistema autonomo, perché sono entrambe sotto giurisdizione tecnica ed amministrativa della Regione Liguria).

2.1 **Tipi di reti**

Tutto quanto pertiene ai livelli 1 e 2 tende ad essere un po' trascurato, ma in realtà è assolutamente fondamentale, e anzi spesso quando si fa riferimento a reti di questo o di quel tipo (es. Ethernet, Token Ring) ci si riferisce proprio a diverse architetture di livello fisico e data link. Citiamo quindi almeno le principali di queste architetture di rete:

- **Ethernet** e **Token Ring** (o più esattamente **IEEE 802.3** e **IEEE 802.5**) sono le due più diffuse tipologie di reti locali (Ethernet molto più di Token Ring), che funzionano a velocità di 4, 16 o 100 Mb/s per Token Ring, e 10 o 100 Mb/s oppure 1 Gb/s per Ethernet; Ethernet è una rete con topologia a bus o a stella basata sul riconoscimento della collisione, mentre Token Ring è una rete ad accesso arbitrato con topologia ad anello; più in particolare distinguiamo:
    - Ethernet a 100 Mb al secondo o più precisamente **100Base-T** (IEEE 802.3u) e **Gigabit Ethernet** (IEEE 802.3z) che si basano sugli algoritmi di riconoscimento della collisione
    - **isoEthernet** (isochronous Ethernet - IEEE 802.9a), poco diffusa, che grazie ad una codifica più efficiente aggiunge ai 10 Mb/s della Ethernet tradizionale un canale isocrono a 6 Mb/s, cioè in grado di trasmettere pacchetti ad intervalli di tempo predeterminati e garantiti, molto adatto per la trasmissione di audio e video
    - **Token Ring su ATM**: è possibile velocizzare le reti Token Ring, tradizionalmente limitate a 16 Mb/s trasportando pacchetti Token Ring su reti ATM ad alta velocità
    - **HSTR** (High Speed Token Ring - IEEE 802.5r) è un nuovo standard che permette di realizzare reti Token Ring a 100 Mb/s
    - **Gigabit Token Ring** (IEEE 802.5v) è uno standard ancora più nuovo, non ancora definitivo, per realizzare reti Token Ring a 1 Gb/s; con questi nuovi standard le reti Token Ring potrebbero nuovamente diventare competitive rispetto a Ethernet, facendo valere, insieme alla velocità, le loro qualità di migliore sfruttamento della banda in presenza di traffico elevato, ma è ancora presto per dire se riusciranno effettivamente ad affermarsi[7]
- **IEEE 802.11b** è uno standard per reti wireless, cioè senza cavi, che collega gli apparecchi tramite trasmissioni in radiofrequenza; offre velocità fino a 11 Mb/s (quindi paragonabili alla tradizionale Ethernet a 10 Mb/s) su distanze di 100-150 m.; per implementare questa rete vengono usati degli apparecchi, detti *access point*, funzionalmente paragonabili agli hub ma che gestiscono trasmissioni radio anziché via cavo e a loro volta possono essere collegati con un cavo ad una rete tradizionale
- **Bluetooth** è una tecnologia wireless introdotta inizialmente dalla Ericsson e ora supportata dal consorzio SIG (*Special Interest Group*), cui aderiscono oltre 2000 aziende (si noti quindi che questa tecnologia non proviene dagli enti di standardizzazione, ma da una iniziativa privata) ed ha lo scopo di collegare, appunto

[7] Le reti Token Ring ad alta velocità sono promosse dalla HSTRA (High Speed Token Ring Alliance)

senza l'uso di cavi, apparecchiature elettroniche di qualsiasi genere per mezzo di dispositivi economici e a basso assorbimento di energia (requisito indispensabile nei dispositivi portatili, anche se desiderabile ovunque); permette una velocità massima di 1 Mb/s (divisi tra le due direzioni) e una portata che normalmente è di 10 m, ma può arrivare fino a 100 con trasmettitori più potenti; può essere utilizzata per piccole reti locali, ma per portata e prestazioni non può competere con gli standard espressamente concepiti per le reti

- **100VG-AnyLAN** (IEEE 802.12) ha alcune somiglianze con la Ethernet, ma se ne distacca su un punto fondamentale: invece di usare un algoritmo per il riconoscimento della collisione usa un sistema centralizzato di arbitraggio per l'accesso al canale trasmissivo, basato su di hub centrale (v. seguito per la trattazione degli hub), per cui ha una struttura fortemente centralizzata
- **FDDI**: rete locale ad alta velocità su fibra ottica; usata soprattutto per l'interconnessione di LAN, ma sempre a livello locale; **CDDI** è una implementazione di FDDI su cavo in rame (a volte il termine CDDI non viene usato e si parla semplicemente di FDDI su cavo in rame)
- **ISDN** è una rete digitale a commutazione di circuito progettata per trasportare sia dati che comunicazioni telefoniche, usata per l'interconnesione di LAN o per la connessione di LAN o anche di singoli computer a reti geografiche (tipico il suo uso per l'accesso a Internet)
- **IEEE-1394** o **Firewire** è un protocollo usato normalmente per connettere ad un PC periferiche che devono trasmettere dati ad alta velocità (come macchine fotografiche digitali di fascia alta o hard disk molto veloci) ma può anche essere usato per collegamenti in rete a breve distanza (questo uso però non è frequente)
- **Fibre Channel** (ANSI X3T11) ha la particolarità di essere in grado di collegare a velocità fino a 1 Gb/s sia reti che altri apparecchi, come unità disco ecc.; funziona su cavo coassiale, doppino o fibra ottica a distanze fino a 10 km (a seconda del cablaggio), ma è scarsamente supportato dai router, per cui non è adatto per l'integrazione tra reti locali e reti geografiche
- **PPP** è un protocollo atto a trasportare traffico TCP/IP su linea seriale, come la linea telefonica: è la tipica connessione utilizzata per collegarsi a Internet via modem; in precedenza per lo stesso scopo era utilizzato il meno sofisticato protocollo **SLIP**; comunque PPP è un protocollo molto potente, il cui uso non è limitato alle connessioni casalinghe, ma si estende anche a collegamenti seriali ad alta velocità su linee dedicate; PPP può trasportare un gran numero di protocolli, non solo IP, ma anche IPX (reti Novell), NetBIOS, SNA (mainframe e AS400 IBM), OSI CLNP, DECnet, Banyan Vines e altri; un'altra interessante caratteristica di PPP è il suo supporto per l'autenticazione degli utenti[8]
- **xDSL** è una famiglia di tecnologie per comunicazioni ad altissima velocità su doppino in rame, e quindi accessibili agli utenti tramite il loro tradizionale collegamento telefonico: servono appunto per portare connessioni ad alta velocità fino all'utente finale sfruttando l'onnipresente cablaggio telefonico, e quindi evitando l'installazione di nuovi cavi; di per sé sono tecnologie di livello 1 piuttosto che di livello 2, per cui sopra xDSL possono essere usati altri protocolli di livello 2, come PPP, ATM o anche PPP su ATM; il raggiungimento di elevate velocità su un mezzo trasmissivo limitato come il doppino in rame è reso possibile da sistemi di codifica particolarmente efficaci a prezzo di un ingente carico computazionale, che ora però è possibile ottenere anche in apparecchiature di costo accessibile; i tipi più interessanti di collegamenti xDSL sono i seguenti:
    - **ADSL:** consente comunicazioni asimmetriche, cioè in traffico in entrata è molto più veloce di quello in uscita (caratteristica molto adatta nel collegamento a Internet, in cui di solito uno riceve molto più di quanto trasmetta); in Italia ha cominciato a diffondersi nella seconda metà del 1999, e sta destando grande interesse, per cui molti provider la offrono come soluzione per l'accesso a Internet, con velocità fino a 640 Kb/s in ricezione e 128 Kb/s in trasmissione[9] a costi fissi indipendenti dal traffico; l'adattatore ADSL si collega alla normale linea telefonica (è possibile telefonare anche mentre è in corso la trasmissione dati), mentre la connessione col computer avviene tramite una porta USB o una interfaccia Ethernet, il permette di collegare via ADSL una intera LAN (se l'interfaccia Ethernet è quella di un router)
    - **HDSL:** comunicazioni simmetriche a velocità fino a 2 Mb/s; viene ora offerta in Italia da diversi fornitori; può essere una interessante alternativa ai CDN

[8] Una trattazione di PPP particolarmente chiara ed esaustiva si trova in *[Advanced 1999]*

[9] ADSL in sé consentirebbe prestazioni anche superiori

    - **VDSL:** per ora è il top della gamma xDSL, non ancora offerta commercialmente in Italia; consente comunicazioni sia simmetriche che asimmetriche a velocità variabili a seconda della lunghezza del collegamento; ad esempio su di una lunghezza di 1,5 km, interessante per scopi come l'accesso a un provider Internet o a una rete privata, si possono avere 6,5 Mb/s simmetrici oppure 13 Mb/s in ricezione e 1,6 Mb/s in trasmissione
- **DQDB** è una architettura per reti metropolitane (MAN, una via di mezzo tra reti locali e reti geografiche)
- **X.25, SMDS, Frame Relay, HSSI** sono varie tecnologie utilizzate per reti metropolitane e/o geografiche ad alta velocità; in particolare:
    - X.25 è un vecchio protocollo a commutazione di pacchetto progettato per operare su reti scarsamente affidabili e quindi appesantito da un gran numero di funzionalità di controllo e correzione dei dati, utili sulle reti di un tempo ma largamente superflue su quelle assai più affidabili di oggi; è limitato ad una velocità di 56 Kb/s;
    - Frame Relay è concettualmente simile ad X.25 ma, potendo ormai operare su reti intrinsecamente molto affidabili, è notevolmente più semplice e quindi più efficiente, per cui può operare a velocità molto maggiori, fino a 2 Mb/s e oltre; si tratta di un protocollo a commutazione di circuito in cui l'utente può richiedere (e pagare) esattamente la capacità trasmissiva di cui ha bisogno; in particolare, frame relay permette di scegliere la CIR, cioè Committed Information Rate, che si potrebbe tradurre Capacità Trasmissiva Garantita, ossia la minima banda che viene garantita all'utente, e la CBR, Committed Burst Rate, cioè Capacità Garantita di Traffico di Picco, ossia una capacità trasmissiva superiore alla CIR che viene assicurata solo per brevissimi istanti per far fronte a picchi di traffico
    - SMDS fornisce un servizio di trasporto senza connessione di celle ATM
- **SONET** è una architettura sincrona utilizzata soprattutto nelle dorsali ad altissima velocità per trasportare sia dati che traffico telefonico digitalizzato
- **ATM** ha gli stessi scopi di SONET, ma si può utilizzare sia in rete locale che in rete geografica, per cui sta destando sempre maggiore interesse, e forse sostituirà molti altri protocolli attualmente usati; non è del tutto esatto dire che ATM è uno standard pertinente ai livelli OSI 1 e 2, perché è difficile trovare una assoluta corrispondenza tra ATM e OSI (e tra ATM e TCP/IP), per cui ATM ingloba compiti che sarebbero di pertinenza dei livelli 3 e addirittura 4; di fatto, spesso si trasporta IP su ATM, ma si potrebbe anche pensare di eliminare IP e trasportare direttamente TCP su ATM; ATM fornisce un servizio orientato alla connessione e basato su celle di lunghezza fissa di 53 byte; ATM è stata progettata per la QoS e solo secondariamente per i servizi best-effort (per questi concetti v. seguito), per cui è nativamente adatta per il traffico multimediale
- **HIPPI** e la sua evoluzione **Super HIPPI** sono tecnologie nate per le reti di supercomputer che ora si stanno estendendo ad altri impieghi, anche se lentamente, vista la sempre maggiore necessità di capacità trasmissiva; sono le reti più veloci in assoluto: fino a 6,4 Gb/s in rete geografica che dovrebbero diventare 12,8 nel 1999; in questo modo si possono avere su rete geografica velocità maggiori di quelle che con altre tecnologie si hanno solo su rete locale, per cui alcuni pensano che sia questa, e non ATM, la vera architettura di rete del futuro; peraltro fino ad oggi queste reti restano pochissimo diffuse e molto costose
- **cable modem**: si tratta di apparecchi che permettono di utilizzare per trasmissione dati (essenzialmente per l'accesso a Internet) la rete della TV via cavo, sconosciuta da noi ma molto diffusa in USA; il vantaggio di questa tecnica è di fornire una elevata capacità trasmissiva tramite una rete fisica già esistente; gli svantaggi derivano dal fatto che tale rete era stata progettata per trasmettere video e non dati, per cui la trasmissione avviene in broadcast su ogni segmento di rete, con possibili inconvenienti per la riservatezza dei dati e per le prestazioni, che tendono a degradarsi man mano che si aggiungono utenti, non essendo possibile garantire ad un utente un livello minimo di servizio
- **CDA:** non è un protocollo di rete, ma significa Circuito Diretto Analogico; si tratta di una normale linea telefonica, che però viene affittata in modo esclusivo ad un utente per collegare due punti tramite modem simili a quelli usati per l'accesso alla rete telefonica pubblica; anche le prestazioni sono simili, cioè limitate; attualmente viene ancora utilizzata soprattutto quando è necessario avere un linea sempre a disposizione, senza procedure di collegamento, composizioni di numeri ecc., ma sulla quale si deve far passare poco traffico, ad esempio in ambito scientifico o industriale per collegare sistemi di controllo o

acquisizione dati con un centro di elaborazione; il CDA provvede un collegamento a livello fisico, sul quale possono essere usati protocolli come PPP o HDLC

- **CDN:** anche questo non è un protocollo, ma la sigla di Circuito Diretto Numerico, noto anche come ***linea dedicata***; è un collegamento punto punto-punto affittato ad uso esclusivo di un utente, ma un collegamento di tipo interamente digitale, che consente velocità anche elevatissime fino a centinaia di Mb/s, sempre che si abbiano i soldi per pagarle (l'effettiva offerta commerciale dipende dai fornitori dei servizi di telecomunicazioni, per cui non necessariamente ciò che è tecnicamente realizzabile viene sempre offerto ovunque sul mercato); il CDN fornisce un collegamento fisico sul quale possono girare protocolli come i soliti PPP o HDLC

È utile dire qualche parola anche sul **cablaggio**, cioè sui cavi da utilizzare per collegare i vari apparecchi. Si tratta di un aspetto che passa spesso inosservato, ma che evidentemente è indispensabile. Attualmente vengono usati principalmente i seguenti tipi di cavi (particolari tipi di reti possono utilizzare altri cavi, oppure utilizzare gli stessi cavi in modi diversi da quelli qui indicati; in ogni caso quanto descritto qui vale sicuramente per le reti Ethernet):

- **cavo coassiale a 50 Ω** (simile al cavo televisivo), detto anche BNC; questo cavo passa da un computer all'altro come in una sorta di catena, che alle due estremità è chiusa da due resistenze, poste sul cavo, dette *tappi*; questo tipo di cablaggio è semplice da realizzare e non richiede apparecchi aggiuntivi, ma è poco affidabile perché basta una interruzione in un qualunque punto del cavo o in un connettore per bloccare tutta la rete; inoltre non è facile individuare in quale punto esattamente è avvenuto il guasto; questo cablaggio consente velocità fino a 10 Mbit/s
- **doppino ritorto non schermato** (UTP = Unshielded Twisted Pair) **categoria 5**; si tratta di un cavo simile a quello telefonico, ma con caratteristiche elettriche superiori, che consente velocità anche oltre i 100 Mbit/s; con questo cavo si possono collegare direttamente due computer, ma se i computer sono in numero maggiore bisogna impiegare un apparecchio aggiuntivo detto **hub** o **concentratore**, che è dotato di numerosi connettori, ad ognuno dei quali è collegato un computer: i vari computer comunicano quindi non attraverso un cavo che passa da uno all'altro, ma attraverso l'hub; questo tipo di cablaggio viene detto a stella; nelle reti più grandi un hub può anche essere collegato ad altri hub, oltre che a singoli computer
- **doppino ritorto non schermato categoria 5e**, simile alla categoria 5 ma di caratteristiche elettriche superiori, particolarmente indicato per le reti a 100 Mb/s
- **doppino ritorno non schermato categoria 6 e 7**; di recentissima introduzione, è un cavo con caratteristiche ancora superiori alla categoria 5, che quindi consente velocità superiori, ed è quindi particolarmente indicato per la Gigabit Ethernet
- **fibre ottiche**; si tratta di cavi in vetro opportunamente modellato, che trasmettono onde luminose invece di segnali elettrici; alle due estremità del cavo vi sono dispositivi che convertono il segnale elettrico in impulsi luminosi e viceversa; gli impulsi luminosi vengono prodotti tramite LED o laser; le fibre ottiche, a prezzo di costi maggiori (circa 10 volte di più del doppino), offrono le migliori prestazioni; a seconda del tipo di rete si possono utilizzare con gli hub oppure ad anello, come nelle reti FDDI

Vi sono ancora altre apparecchiature pertinenti al cablaggio. A volte è necessario connettere un cavo ad un apparecchio che non ha connettori per quel tipo di cavo, ad esempio un coassiale ad un hub che ha solo connettori per il doppino: il problema si può risolvere con un apparecchio detto **media converter**, che fa appunto da adattatore tra i due tipi di cavo. Esistono in commercio molti tipi di media converter (non tutti facili d reperire) che permettono di effettuare molte combinazioni tra diversi cavi e connettori). Oltre ai media converter vi sono i ben più complessi e costosi **convertitori di interfaccia** che permettono la comunicazione tra apparati che impiegano protocolli diversi a livello fisico.

Alcuni apparecchi, per esempio molti router hanno un connettore per così dire "generico", detto AUI (Attachment Unit Interface) nella Ethernet a 10 Mb/s e MII (Media Independent Interface) nella Ethernet 100Base-T, al quale si collega un cavo molto breve che a sua volta va collegato ad un adattatore, detto **transceiver**, specifico per il tipo di cavo che si vuole impiegare (coassiale, doppino o fibra ottica); in questo

modo un apparecchio si può collegare a diversi tipi di cavo sostituendo solo il transceiver (peraltro abbastanza costoso).

Si noti che invece dei cavi si possono utilizzare anche onde radio, microonde o connessioni satellitari. Abbiamo già parlato di alcuni tipi di reti wireless, per lo più alquanto limitate come capacità trasmissiva. Possiamo ancora ricordare che esistono apparecchiature per collegamenti wireless che impiegano un raggio laser consentendo velocità di 155 Mb/s e supportando tutti i tipi di protocollo per distanze di alcune centinaia di metri (è necessario che sia possibile il puntamento diretto tra le apparecchiature laser alle due estremità della connessione, per cui non possono esserci ostacoli frapposti). Si tratta di apparecchi di livello fisico, che devono poi essere collegati a switch, hub o router. Questa tecnica, molto costosa, implementa un collegamento punto-punto, per cui è adatta soprattutto per collegare due LAN o due segmenti di LAN ad alta velocità.

2.1.1 **A ognuno il suo modem (e anche di più)**

Nell'opinione comune, le telecomunicazioni e i collegamenti tra computer sono spesso associati all'immagine del **modem**: il modem però è solo un apparecchio operante a livello 1 che permette di collegare computer tramite la linea telefonica, operazione effettuata essenzialmente dalla piccola utenza, per cui dal punto di vista tecnico non è certo il principale tra gli apparecchi usati per il collegamento tra computer. Naturalmente il modem ha una notevole importanza commerciale, proprio perché viene utilizzato dalla massa degli utenti finali di servizi telematici, soprattutto - ora - per il collegamento a Internet.

Questi sono i modem più comunemente usati, che vediamo esposti in tutti i negozi di informatica e raffigurati in tutti i cataloghi. Il termine modem è però utilizzato anche per altri tipi di apparecchi, tutti operanti a livello fisico e aventi lo scopo di permettere ad una apparecchiatura l'accesso ad una linea di trasmissione dati:

- **cable modem** (è stato trattato nel paragrafo *Tipi di reti*)
- **modem ISDN**, detto più propriamente **Terminal Adapter** o **TA**, che è l'apparecchio che sta tra il dispositivo utente (come il PC, il router o il fax ISDN) e la linea ISDN
- **Short Haul Modem** (**SHM**) o **line driver**: servono per realizzare collegamenti punto-punto, sincroni o asincroni, su cavo in rame; si usano su linee private in varie topologie: ad esempio possono collegare tra loro due computer, oppure più computer ad un unico server; trattandosi di dispositivi punto-punto, si usano sempre in coppia, per cui in quest'ultimo caso il server deve disporre di tanti SHM quanti sono i computer che vi si devono collegare; esistono molti modelli di SHM, che forniscono varie combinazioni di velocità di trasmissione e distanza massima: ad esempio alcuni modelli vanno da 19,2 Kb/s su 1,6 Km a 2,4 Kb/s su 6,4 km, altri arrivano a 115,2 kb/s su 1,2 km, altri da 19,2 Kb/s su 3,2 km a 2,4 kb/s su 9,7 km, altri ancora da 256 b/s su 1 km a 2,4 kb/s su 6,1 km, infine altri sono ottimizzati per la distanza più che per la velocità e vanno da 64 kb/s su 8,5 km a 32 kb/s su 14,5 km; come si vede, gli SHM sono adatti quando si devono coprire distanze notevoli senza ricorrere alle reti pubbliche di trasmissione dati e senza necessità di velocità particolarmente elevate; ad esempio, vengono impiegati per collegare numerose postazioni ad un host per applicazioni a carattere, o per collegare a un host registratori di cassa, lettori di carte di credito o strumenti di misura[10]; possono anche essere collocati tra due router per collegare tra loro due LAN o una LAN con una rete geografica
- **baseband modem** (modem in banda base): concettualmente simili agli SHM, si usano anch'essi in coppia per collegamenti punto-punto, ma utilizzano tecniche più sofisticate di gestione del segnale, che permettono prestazioni molto superiori sia come velocità che come portata (i baseband modem sono peraltro molto più costosi degli SHM); tra questi dispositivi vi sono anche i modem xDSL, che non si limitano ai dispositivi per collegarsi ai servizi pubblici ADSL, a includono, ad esempio, i modem HDSL che si usano su linee private in rame[11] soprattutto per collegare tra loro reti locali o segmenti di rete (naturalmente il modem dovrà essere collegato a un hub o a un router); per quanto riguarda le prestazioni, vi sono numerose combinazioni velocità/distanza: ad esempio i baseband modem per linea a 2 fili vanno da 512 kb/s su 5 km a 19,2 kb/s su 13 km, i modem mDSL permettono velocità da 2,3 Mb/s su 4,9 km a 128 Kb/s su 8,9 km,

[10] L'host sarà ovviamente un computer che deve elaborare tali dati, tipicamente registrandoli in un database
[11] Si noti che, come abbiamo già detto, esistono anche servizi HDSL pubblici

mentre con HDSL si ottiene la velocità di 2,048 Mb/s su distanze da 4 a 12 km a seconda del diametro dei cavi

- **line driver a fibre ottiche**: sono analoghi agli SHM ma funzionano su cavi a fibre ottiche anziché in rame e consentono prestazioni superiori, ad esempio 76,8 kb/s su 4 km o addirittura 2,048 mb/s su 50 km (come si può ben immaginare, quest'ultima è una apparecchiatura estremamente costosa)[12]
- **modem eliminator** (eliminatore di modem): si tratta di apparecchiature per connessioni sincrone che si usano singolarmente e non in coppia, per cui sostituiscono vantaggiosamente una coppia di modem (di qui il nome) per il collegamento di apparecchiature sincrone; alcuni modem eliminator fanno anche da convertitore di interfaccia per collegare apparecchiature sincrone con apparecchiature asincrone; anche i modem eliminator forniscono varie combinazioni velocità distanza: le velocità arrivano fino a 2,048 Mb/s, le distanze fino a qualche centinaio di metri

Ma l'elenco delle apparecchiature pertinenti alla connessione a livello fisico può continuare. Molto importanti sono i **concentratori** e **multiplatori** (detti anche **multiplexer** o **MUX**), il cui scopo è di concentrare, appunto, diversi canali trasmissivi su un unico mezzo di maggiore capacità. In particolare, i multiplatori permettono di collegare un certo numero di apparecchiature ad un canale fisico di comunicazione come se ogni apparecchiatura disponesse di un singolo collegamento punto-punto. I multiplatori si usano in coppia, uno a ogni estremità del collegamento. Esistono anche i **multiplatori inversi**, pure utilizzati in coppia, che permettono di vedere un insieme di linee di trasmissione distinte come un unico canale con capacità uguale alla somma di tutte le linee disponibili. I concentratori invece sono dispositivi singoli (non usati in coppia) collegati da un lato ad una linea di comunicazione ad elevata capacità e dall'altro a un certo numero di apparecchi, che possono essere computer, router o altro. Essi sono in grado di gestire in modo indipendente una molteplicità di connessioni provenienti dalla linea ad elevata capacità: per esempio, chi deve assicurare il collegamento a numerosi utenti ognuno dei quali utilizza una linea telefonica o ISDN, invece di installare una grande quantità di modem o di TA ISDN può installare un concentratore, che gestirà le connessioni in arrivo in modo che ogni utente avrà l'impressione di collegarsi ad un modem tutto per lui. Multiplatori e concentratori sono apparecchiature piuttosto costose, e anzi spesso estremamente costose, impiegate soprattutto per collegamenti in rete geografica, anche se esistono pure multiplatori locali.

2.2 **Protocolli di rete**

Un protocollo di rete è un insieme di regole in base alle quali più computer messi in rete possono comunicare tra loro. Il protocollo nasconde le eventuali differenze nel funzionamento interno di ciascun computer, perché fa in modo che tutti comunichino tra loro attraverso il protocollo comune.

Più in particolare, il protocollo è una descrizione formale (cioè indipendente da specifici contenuti) del modo in cui devono essere strutturate e gestite le informazioni che passano da un computer all'altro: per esempio, un protocollo potrebbe stabilire che le informazioni devono viaggiare divise in blocchi di *n* caratteri, e che all'inizio di ogni blocco ci deve essere un identificativo del computer di provenienza, che questo identificativo deve avere quella certa struttura, ecc.

Risulta quindi più chiaro quello che era già stato accennato all'inizio: non si deve credere che per ogni tipo di rete ci sia un unico protocollo che regola tutto il comportamento di quella rete: ci sono invece molti diversi a seconda dello strato interessato (ad esempio: le caratteristiche fisiche del mezzo trasmissivo, l'indirizzamento dei dati al computer giusto, operazioni più vicine all'utente come il trasferimento di file).

La situazione però è complicata dal fatto che ci sono molti protocolli (o più esattamente molti insiemi di protocolli): TCP/IP, OSI, NetBEUI, SNA, XNS, Novell ecc. Due reti o computer con protocolli diversi **relativi allo stesso livello** non possono comunicare direttamente tra loro.

[12] La fonte di tutti questi dati sui vari modem è il catalogo di un produttore, edizione settembre 2000; altri produttori potrebbero avere in catalogo apparecchi con caratteristiche diverse

A proposito dell'OSI, bisogna dire che se come standard e modello delle reti è di grande utilità per comprendere il funzionamento delle stesse, i protocolli basati su OSI hanno avuto scarso successo commerciale, poiché il mondo reale delle reti, e soprattutto di Internet è oggi basato essenzialmente su TCP/IP.

Per quanto riguarda i protocolli, innanzitutto è fondamentale comprendere che ci sono protocolli **senza connessione** e protocolli **con connessione**: nei primi, tra trasmittente e ricevente non viene stabilita una connessione logica con un inizio e una fine ben definiti, ma ogni unità di informazione viene gestita come entità indipendente. Nei secondi, invece, avviene il contrario. IP è un protocollo senza connessione, mentre TCP è orientato alla connessione. Sullo stesso livello del TCP c'è anche il protocollo UDP (User Datagram Protocol), che è senza connessione. Anche alcuni importanti protocolli del livello applicazione, come HTTP (di cui si parlerà dopo), sono senza connessione. Il TCP/IP nel complesso viene in genere considerato come un protocollo senza connessione, anche questo evidentemente è vero solo entro certi limiti. È anche utile notare che ci sono protocolli **affidabili** e protocolli **non affidabili**: questo non significa che i primi funzionino sempre (non c'è niente che funziona sempre) e i secondi non funzionino mai (altrimenti non li userebbe nessuno), ma significa che i primi implementano specifiche tecniche per migliorare l'affidabilità (ad esempio controllo della correttezza dei dati tramite appositi algoritmi), mentre i secondi non lo fanno, e quindi sono più semplici ma anche più esposti alle conseguenze dei vari malfunzionamenti che possono sopraggiungere in una rete. Come vedremo meglio tra poco, una architettura di rete può prevedere una combinazione di protocolli affidabili e non affidabili in modo da raggiungere un risultato complessivo che si può considerare affidabile senza dover rendere affidabile, e quindi più complesso, ogni singolo protocollo.

Qualcuno potrebbe trovare sorprendente che vengano impiegati protocolli non affidabili nel senso descritto sopra, che sembrano un po' troppo rischiosi: questo però è possibile perché l'infrastruttura di comunicazione attualmente usata, soprattutto a livello fisico è estremamente affidabile per cui determina pochissimi errori nei dati, la correzione dei quali viene eventualmente più convenientemente demandata ai protocolli di livello superiore.

Un'altra distinzione importante è quella tra **best effort** e **quality of service** (QoS). I servizi di tipo best effort sono quelli che fanno del loro meglio per riuscire nel loro compito, ma non garantiscono una qualità minima del servizio, ad esempio una velocità minima di trasmissione, mentre quelli di tipo QoS sono appunto quelli che garantiscono comunque all'utente una determinata qualità del servizio, concordata tra gli utenti stessi e il fornitore dei servizi di rete. Gran parte delle le reti attualmente in funzione, e ad anche Internet nel suo complesso, sono di tipo best effort.

La garanzia della qualità del servizio è importante soprattutto nei servizi multimediali: infatti, se bisogna trasferire un file attraverso la rete, quello che importa è che esso arrivi sano e in un tempo complessivo ragionevole, ma ha poca importanza essere sicuri che nel trasferimento non si scenda in nessun momento al di sotto di una certa velocità (ne ha moltissima invece che il file sia trasmesso senza alcun errore). Viceversa se bisogna trasmettere in rete dell'audio o del video, è necessario assicurare una velocità costante di trasmissione, altrimenti i suoni e le immagini saranno riprodotti con fastidiose interruzioni, mentre può essere accettabile qualche limitato errore nella trasmissione che determini solo una piccolissima perdita di qualità.

In particolare, il traffico dati è di tipo impulsivo (*bursty*) caratterizzato da grandissime variazioni con picchi anche elevatissimi accompagnati a momenti in cui esso può essere estremamente basso. Per contro il traffico audio e video[13] comporta un flusso costante di dati, ad una velocità dalla quale dipende direttamente la qualità della riproduzione, nel senso che per ottenere una qualità elevata bisogna trasmettere moltissimi dati per unità di tempo, e quindi si richiede una grande capacità trasmissiva.

Il traffico multimediale si presta particolarmente ad essere realizzato tramite servizi orientati alla connessione, perché questi servizi permettono di riservare in anticipo le risorse necessarie, e quindi di garantire le

[13] Si intende che in questo contesto il traffico multimediale viene contrapposto al traffico dati nel senso di trasferimento di file, interrogazioni su database e simili, ma si tratta sempre di segnali audio e video digitalizzati e inseriti in pacchetti o celle, non di segnali analogici

prestazioni costanti di cui dicevamo. Da notare che, nel caso le risorse richieste non siano disponibili, la rete dovrebbe permettere di rinunciare del tutto al servizio oppure di negoziare una qualità compatibile con le risorse esistenti (per esempio, se non è possibile trasmettere un concerto dal vivo con audio di qualità CD, l'utente potrebbe preferire un ascolto di qualità inferiore piuttosto che nessun ascolto del tutto).

Nessuno pensa che i servizi QoS debbano sostituire quelli best-effort, che per molte applicazioni vanno benissimo (del resto, l'intera Internet si basa oggi su servizi best-effort), ma piuttosto si ritiene che nel futuro le reti dovranno essere in grado di offrire diversi tipi di servizio, a costi diversi, in modo da adattarsi alle diverse esigenze e disponibilità di spesa degli utenti. Tuttavia uno dei principali problemi che devono oggi affrontare i progettisti è come adattare al traffico multimediale Internet, nata per traffico dati puro e quindi completamente basata su servizi senza connessione e di tipo best-effort.

Bisogna qui riprendere un argomento introdotto, senza spiegazioni, nella trattazione delle architetture di rete di livello 1 e 2, quello della gestione dell'accesso al mezzo trasmissivo in una rete broadcast. Vi sono due approcci fondamentali:

- **riconoscimento della collisione** (collision detection), nel quale la stazione che vuole trasmettere trasmette e immediatamente verifica se si è verificata una collisione con un'altra trasmissione in corso, nel qual caso cessa di trasmettere e attende per un certo periodo di tempo prima di ritrasmettere; questa tecnica è tipica della rete Ethernet, è di semplice implementazione anche se può condurre ad un degrado delle prestazioni in condizioni di traffico elevato, ma non consente di assegnare ad un utente della rete una banda garantita; ciò non significa che su una Ethernet, magari a 1 Gb/s, non si possano trasmettere audio e video, ma che non è possibile garantire una qualità costante della riproduzione; ad esempio una videoconferenza trasmessa su una rete del genere potrebbe dare ottimi risultati quando incomincia e soffrire di una qualità scadente più tardi se nel frattempo altri utenti si collegano alla rete per svolgere attività diverse o per partecipare alla stessa videoconferenza
- **accesso arbitrato**, nel quale c'è una supervisione che dà ad ogni stazione il permesso di trasmettere, in assenza del quale la stazione non trasmette; questo concetto viene implementato in almeno due modi diversi:
    - **gettone** (token) come in Token Ring e FDDI, nel quale per la rete circola uno speciale pacchetto detto appunto token, e solo la stazione che possiede quel pacchetto (trattenendolo per un tempo massimo predeterminato) può trasmettere, più o meno come in una tavola rotonda nella quale gli oratori parlano a turno, secondo un ordine e per un tempo massimo predefinito (chi non vuole parlare quando è il suo turno deve aspettare che sia completato il giro per poter poi eventualmente intervenire)
    - **controllo centralizzato** come in 100VG-AnyLAN, nella quale una stazione predeterminata ha il compito di assegnare i permessi di trasmettere a seconda delle richieste che pervengono, come in una tavola rotonda nella quale è il moderatore che dà agli oratori il permesso di parlare in base alle richieste di intervento che questi gli propongono; si tratta di un sistema molto efficace, ma che dipende pesantemente dal buon funzionamento della stazione addetta al controllo

In pratica lo schema della Ethernet è concettualmente inferiore agli altri, ma è semplice ed economico da implementare e di fatto funziona benissimo per il traffico dati, mentre può dare risultati non soddisfacenti (e comunque non può garantire una qualità minima) quando si tratta di traffico multimediale.

Nel campo delle reti geografiche, che sono tutte commutate, la differenza è soprattutto tra servizi senza connessione e servizi orientati alla connessione: questi ultimi sono in generale più adatti al traffico multimediale perché i primi, trattando ogni pacchetto indipendentemente dagli altri, sono più difficilmente in grado di assicurare una qualità minima del servizio. I servizi orientati alla connessione poi tendono ad accoppiarsi alla commutazione di circuito, come avviene in ATM, sotto forma di PVC o SVC. Interessante e significativa è la differenza tra SMDS e ATM: entrambe queste architetture usano lo stesso tipo di cella di 53 byte, ma il primo - senza connessione - è considerato idoneo solo per traffico dati, mentre il secondo - con connessione - è idoneo anche per il traffico multimediale.

Il discorso non sarebbe completo se non parlassimo degli **schemi di indirizzamento**[14]. Qui vogliamo introdurre una nozione del tutto generale e cioè la distinzione di due tipologie fondamentali di schemi di indirizzamento:

- **indirizzamento piatto** (flat): esiste un unico insieme di indirizzi, che sono tutti sullo stesso piano, come un libro in cui i capitoli siano numerati, ma non abbiano suddivisioni interne
- **indirizzamento gerarchico**: gli indirizzi rappresentano una gerarchia, cioè una suddivisione di livello più alto che individua alcuni sottoinsiemi di indirizzi, che poi possono suddividersi ulteriormente, come quando in un libro o in un altro testo (ad esempio questo) i capitoli sono suddivisi in paragrafi, sottoparagrafi ecc.

L'indirizzamento piatto funziona bene quando si devono assegnare pochi indirizzi, ma quando se ne devono usare molti diventa a dir poco scomodo. Basti pensare a quello che succede con la numerazione telefonica, che è a tutti gli effetti uno schema di indirizzamento. La numerazione telefonica segue uno schema gerarchico perché il numero è composto da: indicatore del fornitore di servizi + prefisso del distretto + numero dell'utente. Se la numerazione seguisse uno schema piatto sarebbero necessari (a non considerare le telefonate internazionali) tanti numeri quanto sono le utenze telefoniche in Italia moltiplicati per il numero dei fornitori del servizio, perché ognuno dovrebbe avere la sua numerazione, e quindi avremmo numeri di telefono molto lunghi e difficili da ricordare, perché ogni numero potrebbe essere utilizzato una volta sola. Per di più, se il sistema di comunicazione o di indirizzamento imponesse un limite alla lunghezza dei numeri (e quindi alla quantità di numeri utilizzabili[15]) Per contro, con lo schema gerarchico uno stesso numero, per esempio 9974657, può essere utilizzato più volte associato a diversi prefissi territoriali, ad esempio una volta a 010 e un'altra a 0144[16]. L'indirizzamento gerarchico ha anche il vantaggio di essere "parlante", cioè di poter veicolare informazioni (ad esempio la collocazione geografica di una certa utenza telefonica, deducibile da prefisso) che invece l'indirizzamento piatto non può rappresentare.

2.2.1 **TCP/IP**

Poiché parliamo di Internet, il protocollo che ci interessa in modo particolare è quello su cui Internet si basa, cioè il protocollo TCP/IP, che significa Transmission Control Protocol/Internet Protocol. Si tratta in realtà di un insieme di numerosi protocolli (che viene anche chiamato Internet Protocol Suite = IPS), che prende il nome da due dei principali, cioè appunto IP e TCP. In realtà il nome Internet Protocol Suite sarebbe più appropriato, ma viene usato di raro.

TCP/IP è uno standard di fatto, cioè non è mai stato formalmente avallato dagli organi ufficiali di standardizzazione come l'ISO. Nondimeno, l'importanza commerciale dei protocolli ISO/OSI (es. TP4, X.400 ecc.) è molto piccola in confronto a quella del TCP/IP, la quale, soprattutto con l'avvento di Internet, è divenuta enorme.

Una analisi dettagliata del TCP/IP esula dallo scopo di questo documento; ci limitiamo invece a una breve esposizione delle caratteristiche principali.

Anche il TCP/IP si basa su un modello della rete che la vede composta da diversi strati, e precisamente quattro, che qui di seguito vengono messi a confronto con il modello OSI:

[14] In questo contesto *indirizzamento* significa *assegnazione degli indirizzi* (dall'inglese *addressing*), e non - come spesso nel linguaggio corrente - l'indirizzare qualcuno a una meta, che nel linguaggio delle reti sarebbe il *routing* o *instradamento*, del quale si parlerà dopo

[15] I numeri sono infiniti, ma i numeri di non più di *n* cifre, ad esempio di non più di 6 cifre o 10 cifre **non** sono per nulla infiniti

[16] I prefissi citati esistono davvero, ma il numero è stato scritto assolutamente a caso

TCP/IP

4 Servizi delle applicazioni
3 Trasporto
2 Internetwork
1 Sottoreti

OSI

7 Applicazione
6 Presentazione
5 Sessione
4 Trasporto
3 Rete
2 Data Link LLC MAC
1 Fisico

Come si vede, il TCP/IP può essere considerato una implementazione semplificata del modello OSI. In particolare, il livello 1 TCP/IP corrisponde ai livelli OSI 1 e 2, il 2 corrisponde al 3, il 3 al 4, il 4 ai livelli 5-7. Infatti i compiti dei livelli OSI 5 (sessione, che comprende tra l'altro l'autenticazione degli utenti) e 6 (presentazione, che comprende tra l'altro la gestione dei set di caratteri e la mappatura della tastiera) sono spesso svolti, in TCP/IP, dal livello applicazione, che quindi potrebbe considerarsi corrispondente ai livelli OSI 5-7.

Lo strato delle **sottoreti** è quello in cui non avviene ancora la comunicazione tra una rete e l'altra, e comprende in particolare il livello fisico della trasmissione dei dati (cavi ecc.).
Lo strato dell'**internetwork** provvede alla connessione delle reti e alla consegna dei dati dalla sorgente alla destinazione, individuando, tra l'altro, il percorso appropriato. Contiene, tra l'altro, il protocollo IP.
Lo strato **trasporto** provvede alla comunicazione tra gli utenti della rete, alla sicurezza e al monitoraggio delle transazioni. Contiene, tra l'altro, il protocollo TCP.
Lo strato dei **servizi delle applicazioni** contiene i protocolli orientati alle applicazioni, quelle che l'utente vede effettivamente come servizi resi accessibili tramite la rete: in questa categoria rientrano anche due protocolli di importanza fondamentale come FTP (File Transfer Protocol) e HTTP (Hyper Text Transfer Protocol), sul quale si basa il World Wide Web di cui parleremo oltre. Le applicazioni di cui si parla qui sono applicazioni specifiche per la rete, non qualunque applicazione che gira su un computer collegato in rete e magari usa anche una unità di rete.

Il TCP/IP non prevede nulla per quanto riguarda il livello 1 (corrispondente ai livelli OSI 1 e 2), per cui anche le reti TCP/IP utilizzano protocolli come IEEE 802.3 (Ethernet) e IEEE 802.5 (Token ring), ISDN, ATM ecc. (che sono tutti da classificare ai livelli OSI 1 e 2 - cioè fisico e data link).

I protocolli IP e TCP hanno una importanza particolare: **infatti il primo rende possibile la connessione tra reti, il secondo gestisce il collegamento end-to-end**, cioè **tra un host e l'altro**, indipendentemente da quello che c'è in mezzo (gateway, router, sottoreti ecc.). IP opera utilizzando anche i protocolli di routing che servono ad individuare il percorso necessario per raggiungere un certo host. IP usufruisce dei servizi dei protocolli del livello inferiore, che governano l'accesso al mezzo trasmissivo, e fornisce i suoi servizi a TCP, che si trova al livello superiore. Inoltre allo stesso livello di IP opera il protocollo ICMP (Internet Control Message Protocol), che serve a permettere lo scambio di messaggi di controllo. TCP utilizza i servizi di IP e fornisce i suoi ai protocolli dello strato applicazione, come HTTP ed FTP. È utile notare che alcuni programmi applicativi possono scavalcare il transport layer (e quindi il TCP) e interfacciarsi direttamente con il network layer (IP). IP è un protocollo non affidabile, mentre TCP è affidabile. Al posto di TCP si può anche utilizzare il già citato UDP, che è un protocollo non affidabile e notevolmente più semplice; alcuni protocolli di livello superiore, come FTP, però non funzionano su UDP.

A ogni interfaccia di rete collegata ad una rete TCP/IP (interfaccia che può appartenere ad un computer o ad altri apparecchi, come i router) viene assegnato un indirizzo, detto indirizzo IP, formato da 32 bit; poiché 32

bit corrispondo a quattro byte (cioè a quattro gruppi di otto bit ciascuno), di solito l'indirizzo viene rappresentato non nella poco leggibile forma binaria, ma per mezzo di quattro numeri (corrispondenti al valore numerico dei byte) separati da un punto; ciascuno di questi numeri corrisponde a otto bit, per cui può avere un valore da 0 a 255; gli indirizzi IP possono quindi andare da 0.0.0.0 a 255.255.255.255; un esempio di indirizzo IP può essere 132.132.132.130 (un indirizzo come 132.132.132.**312** è **necessariamente sbagliato**, perché un byte non può avere il valore 312); a un indirizzo IP può anche essere associato un nome, che permette di ricordarlo più facilmente, ad esempio www.regione.liguria.it. Il nome di un host quindi corrisponde sempre ad un indirizzo IP. Di solito il nome di un host è composto da diverse parti, alcune delle quali rappresentano non il singolo host ma il **dominio,** che si può pensare come un insieme di host. Ad esempio nel nome *www.regione.liguria.it* **.it** corrisponde ad un dominio (utilizzato per gli host italiani), mentre **.regione.liguria** corrisponde ad un dominio all'interno del più ampio dominio .it. Così, ad esempio, nel dominio .regione.liguria.it esiste l'host *www.regione.liguria.it* così come l'host *ftp.regione.liguria.it*. Un nome completo di host viene detto FQDN = Full Qualified Domain Name[17]. Si noti che più nomi possono corrispondere ad uno stesso indirizzo IP.

Gli indirizzi IP hanno una loro struttura interna: essi si dividono infatti in due parti, la prima delle quali identifica la rete, mentre l'altra identifica l'host. In base alla suddivisione, si individuano cinque classi di indirizzi. delle quali vengono normalmente utilizzate le prime tre, dette A, B e C. La loro struttura viene esemplificata nella tabella che segue, dove N indica il numero della rete, mentre H indica il numero dell'host.

Classse	**Formato**	**Gamma di indirizzi**	**Numero massimo di host in una rete**
A	N.H.H.H	da 1.0.0.0 a 126.0.0.0	16.777.214
B	N.N.H.H	da 128.1.0.0 a 191.254.0.0	65.543
C	N.N.N.H	da 192.0.1.0 a 223.255.254.0	245

In base a questo schema di indirizzamento risulta subito chiaro, ad esempio, che 193.70.160.233 e 193.70.161.233 sono due host appartenenti a due diverse reti di classe C, mentre 172.34.56.211 e 172.34.57.211 sono due host appartenenti alla stessa rete di classe B.

All'interno di una rete si possono ancora definire più sottoreti, ognuna con un suo schema di indirizzamento. Come già accennato, qui il termine sottorete è utilizzato in un senso particolare che si applica solo nella terminologia IP e non nel senso generale dell'infrastruttura di comunicazione. Su questo argomento, maggiori informazioni si trovano nell'Appendice.

Ma come è possibile mettere in corrispondenza gli indirizzi IP con i nomi di dominio ? A questo scopo, esiste un insieme di tecniche (Domain Name System = Sistema dei nomi di dominio), che permette di mappare gli indirizzi IP su questi nomi. Ci sono infatti dei server appositi, detti **name server** (server dei nomi) che contengono dei database di corrispondenza tra nomi e indirizzi IP: ogni computer collegato ad Internet, o comunque ad una rete con un name server, deve avere indicato nella sua configurazione di rete l'indirizzo IP del name server (evidentemente non può avere indicato il FQDN, perché questo server serve appunto a decodificare il FQDN); ogni volta che dal computer parte una richiesta di collegamento a un qualunque host identificato da un nome, viene in realtà prima interrogato il name server per conoscere l'indirizzo IP di quell'host. Evidentemente un singolo name server non può contenere un database con tutti gli indirizzi IP di Internet: esso però sa a quale name server deve inoltrare le richieste che non può soddisfare direttamente. Se il name server non funziona, è possibile collegarsi agli host della rete solo attraverso l'indirizzo IP. In reti molto piccole, invece del name server, si usa un file chiamato *hosts* (in Windows si trova nella directory di Windows), che contiene una tabella di corrispondenza tra nomi e indirizzi IP, e va aggiornato a mano su ogni computer ogni volta che si aggiunge alla rete un nuovo host: per questo, evidentemente, il file *hosts* viene utilizzato solo in piccole reti.

[17] Questa sigla viene per la verità poco utilizzata

È comunque importante ricordare che sia la rappresentazione decimale degli indirizzi IP sia i nomi degli host sono utili solo per gli esseri umani, mentre i computer utilizzano esclusivamente l'indirizzo in forma binaria (la conversione dalla forma decimale a quella binaria avviene automaticamente, per cui l'utente non deve preoccuparsene).

Si noti che l'indirizzo IP è assegnato **all'interfaccia di rete e non al computer**, per cui un computer può avere più di un indirizzo IP quando ha più di una interfaccia di rete: ciò può avvenire, ad esempio, nei router (che hanno una interfaccia per ciascuna delle reti che collegano), oppure quando un computer ha un indirizzo per la scheda Ethernet su una LAN e un altro assegnato dal provider per il collegamento a Internet. L'indirizzo IP 127.0.0.1 ha un ruolo particolare: corrisponde ad una interfaccia di rete detta *loopback device* che è una interfaccia emulata via software, che serve per poter utilizzare - a scopo di test, sviluppo, esercitazione ecc. - del software di rete anche su un computer che in realtà non è collegato a niente. Inoltre i sistemi operativi di rete più sofisticati permettono di associare più indirizzi IP ad un unica interfaccia fisica (questa tecnica viene detta multihoming).

Altri concetti di grande importanza sono quelli di **porta** e di **socket**, che però verranno esposti nell'appendice (paragrafo 6.2).

2.2.1.1 **IPv6**

Quella che abbiamo descritto finora è la versione 4 del protocollo IP, che quindi è detta anche IPv4. Questa versione, utilizzando indirizzi a 32 bit, permette di assegnare al massimo 2^{32} indirizzi. Questo numero, per quanto enorme, sta diventando insufficiente a causa della grande espansione delle reti, e non solo di Internet come la conosciamo oggi: si prevede infatti vi sarà la possibilità di interagire via rete con un gran numero di apparecchiature di ogni genere, compresi elettrodomestici ed automobili, e quindi anche tutte queste apparecchiature avranno bisogno di indirizzi di rete. Per questo è stata sviluppata la versione 6 di IP, detta appunto IPv6, che utilizza indirizzi a 128 bit, e quindi permette di assegnare addirittura 2^{128} indirizzi. Si prevede quindi che l'attuale IPv4 sarà sostituito dall'IPv6, dopo una fase di transizione in cui apposite tecniche (come il trasporto di IPv4 attraverso reti IPv6 e viceversa) permetteranno una coesistenza del software e hardware IPv6 con quello IPv4.

IPv6 non è più da tempo un protocollo sperimentale, ma un protocollo stabile: tuttavia non si è ancora affermato a causa dell'enorme diffusione e successo di Ipv4, e anche perché il problema dell'esaurimento degli indirizzi non è ancora così immediato. Questo problema però prima o poi andrà risolto, e quindi è inevitabile che IPv6 sostituisca Ipv4, a meno che nel frattempo non venga inventato qualcosa di meglio, il che però non sembra probabile.

Un indirizzo IPv6 è, come abbiamo detto, di 128 bit, che vengono scritti, per nostra fortuna, non in formato binario ma come una serie di 32 interi di 4 bit in formato esadecimale (cioè in base 16). Questi 32 interi poi sono raggruppati a due a due e 16 gruppi così ottenuti sono separati dai due punti (:). Ecco quindi come appare un indirizzo IPv6:

68DA:8909:3A22:FA64:68DA:8909:3A22:FACA

Quando uno di questi gruppo di bit vale 0 (cioè tutti gli 8 bit valgono 0) si scrive uno 0 solo non quattro, ad esempio:

68DA:0:0:FA64:0:0:0:FACA

Uno o più zeri consecutivi possono inoltre essere sostituiti da due caratteri due punti successivi (::), ma solo una volta nell'ambito dell'indirizzo, per cui l'indirizzo dell'esempio precedente può essere scritto

68DA::FA64:0:0:0:FACA

oppure

68DA:0:0:FA64::FACA

ma non così 68DA::FA64::FACA (questa scrittura infatti permetterebbe di sapere quanti zero ci sono in tutto l'indirizzo, ma non quanti corrispondono a ciascuno dei due :: presenti). Naturalmente il :: può anche trovarsi all'inizio o alla fine dell'indirizzo.

Gli indirizzi IPv4 sono rappresentati come indirizzi IPv6 che usano solo gli ultimi 32 bit, per cui i primi 96 sono impostati a 0. Inoltre per gli indirizzi IPv4 è consentito rappresentare i gruppi di 8 bit in notazione decimale puntata, per cui un indirizzo IPv4 può essere rappresentato in questo modo:

::213.75.119.33 (si noti l'uso del punto invece dei due punti).

Come si vede, gli indirizzi IPv6 numerici sono nella maggior parte dei casi ancora più complicati e difficili da ricordare di quelli IPv4, per cui si prevede che gli utenti useranno quasi sempre il nome dell'host e che l'uso degli indirizzi numerici diventerà ancora più ridotto di quanto non sia oggi (l'uso, si intende, da parte degli esseri umani, perché l'hardware e il software continueranno ad usare esclusivamente gli indirizzi numerici).

Le novità di IPv6 non si fermano qui. Tra le cose più significative, c'è il fatto che gli indirizzi hanno intrinsecamente una struttura gerarchica, atta a rappresentare l'autorità di registrazione degli indirizzi, il provider, l'utente, la sottorete dell'utente e l'host all'interno della sottorete.

La struttura dei pacchetti IPv6 inoltre è stata ottimizzata allo scopo di facilitare il lavoro dei router e delle altre apparecchiature, cercando di migliorare così le prestazioni della rete.

2.3 **Interconnessione di reti (internetworking)**

Abbiamo detto che connettendo più computer insieme si forma una rete. Possono così nascere diverse reti indipendenti. A questo punto è facile prenderci gusto e pensare di interconnettere a loro volta queste reti. Questa attività di interconnessione di reti viene detta anche internetworking, ed è oggi un aspetto fondamentale delle attività informatiche, perché, consentendo un collegamento indefinitamente espandibile di sistemi informativi, permette di accedere da ogni punto alla totalità delle informazioni disponibili in rete. Naturalmente Internet è la massima espressione dell'internetworking.

Collegare delle reti non è però una cosa banale, e necessita di apposito hardware e software. Tra i dispositivi necessari ricordiamo in particolare i seguenti, di cui spesso capita di sentire parlare:

- **repeater**, che effettua una connessione al solo livello fisico, ed in pratica si limita a ripetere il segnale per evitare che diventi troppo debole a causa dell'attenuazione prodotta dai mezzi trasmissivi
- **bridge** e **switch** (più sofisticato), che effettuano una connessione al livello data link e **non** hanno un indirizzo di rete
- **router**, che serve a connettere reti che utilizzano lo stesso protocollo a livello 2 del TCP/IP (livello 3 OSI), ad esempio il protocollo IP, e provvede ad instradare i messaggi in modo che vengano inviati alla rete appropriata; i protocolli dei livelli inferiori possono essere diversi, ad esempio Ethernet e X.25; i router hanno un loro indirizzo di rete; essi usano dei particolari protocolli, spesso assai sofisticati, per scambiarsi le informazioni sulla topologia della rete, sulle connessioni utilizzabili, su quelle più vantaggiose, al fine di poter determinare dove instradare i dati per raggiungere una certa destinazione; i router permettono una gestione estremamente sofisticata, ad esempio per quanto riguarda abilitazioni o limitazioni di traffico per particolari utenti

- **switch di livello 3** (layer 3 switch): si tratta di apparecchi di recente introduzione; in pratica questi apparecchi effettuano in hardware (come i tradizionali switch di livello 2) molte operazioni che i router effettuano in software, per cui permettono prestazioni più elevate e costi minori, anche se per ora non consentono una gestione sofisticata come quella dei router; un interessante uso degli switch L3 è quello come **acceleratori di router**: in questo impiego, lo switch non sostituisce il router, ma divide in due segmenti la LAN cui il router è collegato; in un segmento c'è solo il router, nell'altro tutti i computer della LAN; in questo modo al router arriveranno solo i pacchetti a lui destinati, cioè quelli che devono essere inoltrati fuori dalla LAN, sollevandolo dall'incombenza di analizzare tutti i pacchetti che circolano sulla LAN
- **switch di livello 4** (layer 4 switch), recentemente introdotto da alcune case, come Cabletron; si tratta di switch che effettuano in hardware operazioni a livello 4, e quindi - in base alle porte TCP utilizzate - sono in grado di identificare le applicazioni che stanno facendo uso della connessione di rete e quindi di gestire il traffico in base a queste informazioni, ad esempio assegnano ad alcune applicazioni una priorità maggiore rispetto ad altre
- **gateway**, che serve a connettere reti che usano in generale protocolli diversi, ed evidentemente è normalmente più complesso di un router; poiché una rete può usare più protocolli relativi a strati diversi, è possibile che il gateway provveda a interfacciare solo i protocolli relativi a particolari servizi; da notare che spesso si usa il termine gateway anche per indicare sw che effettuano in tempo reale traduzioni di protocollo o formato tra programmi diversi, e non solo tra reti diverse (in seguito saranno dati alcuni esempi); inoltre talvolta, soprattutto in ambito TCP/IP, può essere che il termine venga usato per comprendere anche i router

Come si vede, il routing e lo switching tendono in certo qual modo a convergere: questa convergenza consiste nel fatto che, pur rimanendo le due cose ben distinte (e non si vede come potrebbero non esserlo, almeno finché si rimane in ambito OSI, visto che lo switching avviene a livello 2 e il routing a livello), tendono ad essere implementate con tecnologie analoghe - nel senso che tecnologie switching vengono applicate ai router - sotto la pressione dell'esigenza di aumentare le prestazioni.

2.3.1 **Label based switching (LBS)**

L'idea che sta alla base del LBS è quella di evitare una caratteristica tipica del routing, e cioè la stessa elaborazione, quella necessaria per l'instradamento dei pacchetti, ripetuta per ogni singolo pacchetto. Questo fatto determina un peggioramento delle prestazioni, anche perché tale elaborazione è tutt'altro che semplice e prevede l'esame dell'intero header dei pacchetti. Il LBS invece prevede l'inoltro dei pacchetti lungo circuiti virtuali, inoltro basato sull'elaborazione di semplici etichette a lunghezza fissa (che si possono concepire come header molto semplici): lo switch esamina l'etichetta, la confronta con una tabella di corrispondenze tra etichette e circuiti virtuali, e inoltra il pacchetto lungo il circuito appropriato. Questo permette non solo di migliorare le prestazioni, ma - essendo una tecnologia integrata con quella dei circuiti virtuali - favorisce la differenziazione dei servizi nell'ambito di una stessa rete, e anzi secondo alcuni è questo, più ancora che il miglioramento delle prestazioni, il vero vantaggio di LBS. Non è peraltro inconcepibile l'uso del label based switching in un ambiente connectionless, anche se ben si comprende che questa tecnica non è concepita per prendere decisioni di instradamento basate sul singolo pacchetto, ma su insiemi di pacchetti, cioè appunto quelli identificati da una determinata label (altrimenti finirebbe per somigliare un po' troppo al routing classico).

Le etichette sono gestite con riferimento al singolo link, cioè alla connessione tra uno switch e l'altro (criterio *hop-by-hop*), per cui ogni switch esamina le etichette in entrata e le sostituisce con sue etichette, che saranno esaminate solo dallo switch successivo.

LBS è un concetto molto generale, di cui ci sono numerose implementazioni, nessuna delle quali largamente affermata, per cui possono essere problemi di compatibilità tra apparecchiature. Molti sono gli standard proposti da vari costruttori di apparecchiature di switching e routing, mentre la proposta più indipendente è probabilmente il protocollo MPLS (Multiprotocol Label Switching) elaborato dalla IETF.

2.3.2 **Routing**

Parliamo in questo paragrafo del routing "classico", cioè del routing IP orientato alla commutazione di pacchetto, che è tuttora il cuore di Internet e dell'internetworking in generale. Il routing è quello che ci permette di collegarci a Internet e raggiungere l'host desiderato attraverso le numerosissime reti che compongono Internet.

Per comprendere il funzionamento del routing, si devono fare diverse distinzioni. Bisogna innanzitutto distinguere due diversi tipi di sistemi (nel senso generico di apparecchiature):

- gli **end systems** (sistemi finali) sono quelli che stanno ai due capi della comunicazione, ad esempio il computer dell'utente Internet e quello del sito Web che gli interessa; se gli end systems sono collegati direttamente, e sono quindi sulla stessa rete, non c'è bisogno d'altro e non avviene routing[18]
- gli **intermediate systems** (sistemi intermedi) sono collocati tra gli end systems non direttamente collegati, in modo da permettere appunto questo collegamento per via indiretta; si capisce facilmente che i router sono tutti intermediate systems

Una seconda distinzione è fra tre fondamentali tecniche per guidare l'inoltro:

- **routing by destination address**, in cui la scelta del percorso è determinata dall'indirizzo del sistema di destinazione (secondo regole del tipo: *per raggiungere il tale indirizzo prendi il tale percorso*), ed è quello di cui ci occuperemo nel seguito
- **label based**, che è quello descritto nel paragrafo precedente
- **source routing**, in cui il mittente scrive nell'intestazione del pacchetto l'intero percorso da utilizzare fino alla destinazione; questa tecnica viene usata soprattutto in piccole reti e comunque in reti private, perché altrimenti il percorso potrebbe diventare troppo grande, e non è molto utilizzata su Internet anche perché presenta problemi di sicurezza: infatti chi riuscisse ad intercettare i pacchetti e a modificarli in modo da alterarne il percorso riuscirebbe a dirigerli dove vuole lui senza bisogno di acquisire il controllo sui vari router collocati lungo il percorso.

Una terza distinzione è quella fra il routing **all'interno di un sistema autonomo** (routing interno) e il routing **tra diversi sistemi autonomi** (routing esterno). Questi tipi di routing utilizzano protocolli diversi, che rispondono tra l'altro alla necessità di minimizzare il traffico delle informazioni di routing tra sistemi autonomi, per evitare di occupare molte reti, e magari l'intera Internet con questo traffico a scapito del traffico dati. Non si deve credere che all'interno di un sistema autonomo ogni router possa comunicare direttamente con gli altri sistemi autonomi: questo compito è invece demandato a specifici router, detti **border router**, per cui tutti gli altri router del sistema autonomo hanno bisogno di conoscere solo i percorsi interni al loro sistema autonomo, mentre per quelli esterni devono solo sapere a quale border router inoltrare i pacchetti.

Una quarta distinzione è quella tra **routing statico** e **routing dinamico**. Nel primo i percorsi vengono inseriti a mano in ogni router dagli amministratori del sistema, e non possono essere cambiati se non a mano[19]. Il routing statico ha il vantaggio della sicurezza, poiché le informazioni di routing non possono essere alterate se non dai sistemisti, ed evita la necessità di utilizzare complessi protocolli di routing dinamico. Esso inoltre può

[18] Diciamo in questo modo perché qui parliamo di routing IP, ma naturalmente pur non essendoci routing può esserci switching di livello 2; se gli end systems sono sullo stesso segmento di una LAN commutata non cè neppure switching di livello 2

[19] Si noti che il source routing è un tipo di routing statico, ma il routing statico non si identifica con il source routing: esso infatti può essere realizzato inserendo in ogni router le informazioni di instradamento, e non inserendo tutto il percorso già dall'origine

consentire di ottimizzare al massimo le prestazioni grazie all'abilità dei sistemisti che determinano le regole di instradamento. Per contro, nelle reti grandi è molto complesso da amministrare, ed è molto sensibile ai guasti: se infatti un percorso diventa inaccessibile, ma ne esiste un altro per la stessa destinazione, i router non possono attivarlo automaticamente, ma bisogna attendere l'intervento del sistemista. Nel routing dinamico invece i router si scambiano informazioni attraverso appositi protocolli in modo da aggiornare automaticamente le loro tabelle di instradamento (cioè le tabelle che mettono in corrispondenza le destinazioni con i percorsi). Tra i principali requisiti dei protocolli di routing ci sono: semplicità degli algoritmi utilizzati, e quindi rapidità di elaborazione da parte di router, limitata occupazione di banda per lo scambio delle informazioni, rapidità di convergenza. La **convergenza** è il processo per cui, in seguito ad un mutamento di una topologia della rete (che avviene, ad esempio, quando un collegamento si interrompe o un router si ferma per un guasto) i router arrivano tutti ad avere una rappresentazione corretta della nuova topologia della rete. Prima che sia raggiunta la convergenza, può avvenire che una destinazione appaia irraggiungibile senza in realtà esserlo, con grave disturbo degli utenti della rete. Particolarmente temibili sono i **routing loops** cioè anelli di instradamento, in cui i pacchetti girano in tondo sullo stesso percorso (ad esempio da router A al router B, da B a C e poi da C di nuovo ad A) invece di essere inoltrati a destinazione[20].

Nell'ambito del routing dinamico, ci sono protocolli per il routing interno, i principali dei quali sono RIP e OSPF, e protocolli per il routing esterno, tra cui il principale è BGP. I protocolli, per scegliere i percorsi, utilizzano la nozione di costo di un percorso, costo che può essere valutato in diversi modi, ad esempio come velocità del percorso, o come sua affidabilità, o come numero di passaggi per raggiungere la destinazione, o anche come costo economico. In base ai costi, i protocolli individuano il migliore tra eventuali percorsi alternativi per raggiungere una destinazione.

I vari protocolli di routing si basano su due algoritmi:

- **distance vector** (detto anche algoritmo di Bellman-Ford); prevede che ogni router invii su tutte le sue interfacce un elenco di tutte le destinazioni che può raggiungere e della loro distanza da sé stesso (detto appunto distance vector); il router riceve da tutti gli altri router analoghe comunicazioni, e in base alla fusione di tutti i distance vector costruisce la propria tabella di routing, modificandola poi se nuovi distance vector gli comunicano variazioni nella topologia di rete; questo algoritmo è semplice da implementare, ma genera un notevole traffico di rete, e quando le dimensioni e la complessità della rete crescono, diventa rapidamente molto lento a convergere, per cui è adatto a reti piccole e di semplice topologia; RIP è un protocollo basato sull'algoritmo distance vector
- **link state**; prevede che ogni router invii su tutte le sue interfacce non l'elenco di tutte le destinazioni che può raggiungere, ma solo l'elenco dei suoi collegamenti diretti (appunto link state), attraverso pacchetti detti LSP (Link State Packet); ogni router riceve da tutti gli altri questi LSP, ed in base ad essi ricostruisce la topologia della rete e i percorsi migliori; l'algoritmo link state è semplice de spiegare, ma complesso da realizzare, per cui richiede più potenza di calcolo rispetto al distance vector; tuttavia, sempre rispetto al distance vector, la potenza di calcolo necessaria aumenta molto più lentamente con l'aumentare della complessità della rete, è più veloce a convergere e genera meno traffico; per questo è più indicato per le reti grandi e di topologia complessa; OSPF è un protocollo basato sull'algoritmo link state

2.4 **Sicurezza delle reti, firewalls e proxies**

Un argomento molto importante che entra in gioco soprattutto in seguito all'internetworking è quello della sicurezza. Infatti in un contesto di internetworking, chiunque acceda ad una delle reti interconnesse può in linea di principio accedere a tutte le risorse delle altre reti: spesso però non è opportuno concedere questa libertà di accesso, soprattutto quando alcune parti delle reti sono pubbliche mentre altre sono private, perché vi

[20] I pacchetti IP comunque non girano all'infinito, perchè l'intestazione contiene un campo detto TTL (Time To Live = Tempo di sopravvivenza) che consiste in un numero che viene decrementato di uno da ogni router; quando il valore di TTL diventa 0 il pacchetto viene eliminato; malgrado questo, i routing loop - oltre a rendere impossibile il collegamento ad una certa destinazione - generano comunque una grande quantità di traffico inutile che sottrae banda al traffico rimanente, e quindi sono da evitare in ongi modo

possono essere dati non destinati al pubblico ed inoltre, non essendo possibile identificare preventivamente l'utenza (come invece è più o meno possibile nelle reti private), è maggiore il rischio che qualche sconosciuto si colleghi appositamente per provocare danni (ad esempio cancellazione di files, introduzione di virus ecc.). D'altra parte però è necessario consentire, da ogni rete, quei collegamenti verso l'esterno che sono necessari per l'attività di coloro che utilizzano la rete.

Un tipico esempio di questa situazione è ciò che si verifica nella Regione: una rete privata (la LAN degli uffici regionali e potenzialmente anche la rete geografica) collegata a Internet. Questo comporta la necessità di consentire a tutti coloro che sono sulla rete privata di accedere a Internet, che è una rete pubblica, e nello stesso tempo di impedire a tutti coloro che sono collegati a Internet di accedere alla rete privata.

Per capire quali soluzioni si adottino a questo scopo bisogna innanzitutto introdurre il concetto di proxy, che in sé e per sé non è necessariamente collegato alla sicurezza. Un proxy è un computer dotato di due interfacce di rete e di un apposito software (detto appunto server proxy) che invia su una delle interfacce tutto ciò che riceve sull'altra (ma non il contrario, per cui da questa seconda rete accetta solo ciò che arriva in risposta a una richiesta effettuata dall'interno della prima rete), in modo che da quest'ultima rete non si vedono gli indirizzi IP dell'altra, poiché tutto appare originato dall'IP del proxy. Un dispositivo di questo genere può servire per collegare due reti senza utilizzare un router (superiore in prestazioni ma costoso). Inoltre il proxy può gestire una cache, cioè una zona di memoria (in questo caso su disco) in cui mantiene un insieme di dati più frequentemente utilizzati (nel caso di un proxy per collegamenti Internet si tratterà di solito di pagine WWW), per cui se viene richiesto qualcosa che si trova nella cache il proxy invia questa copia del dato e non quello originale. Infine, se il proxy viene usato per il collegamento a Internet, permette di risparmiare indirizzi IP pubblici: infatti ne basta uno, quello dell'interfaccia Internet del proxy, mentre tutti gli altri computer possono utilizzare indirizzi privati. Il vantaggio deriva dal fatto che gli indirizzi Internet sono limitati, perché ciascuno può essere usato una sola volta su Internet, mentre quelli privati possono essere ripetuti in infinite reti private (purché, naturalmente, lo stesso indirizzo non venga usato più di una volta nella stessa rete).

Un proxy così concepito non migliora di per sé la sicurezza, perché lascia passare qualsiasi connessione tra una rete e l'altra (al massimo può impedire del tutto l'uso di certi protocolli non supportati dal particolare software proxy utilizzato, ma questo spesso non è ciò che si desidera). Se però ad un proxy aggiungiamo la possibilità di controllare i collegamenti, ossia di permettere o impedire in modo selettivo determinati collegamenti, abbiamo un firewall, ossia un computer che serve da protezione per una rete; infatti un firewall, a seconda della configurazione, può ad esempio: consentire a tutti coloro che sono sulla rete privata ogni collegamento verso la rete pubblica, impedire a tutti coloro che sono sulla rete pubblica ogni collegamento verso la rete privata, consentire ogni collegamento esterno a una parte degli utenti della rete privata e solo alcuni tipi di collegamento all'altra parte degli utenti, consentire ad uno o più particolari host della rete pubblica di entrare nella rete privata specificando anche a quali indirizzi IP possono collegarsi, e molto altro ancora. Un firewall quindi consente di regolamentare e tenere sotto controllo ogni aspetto dei collegamenti tra due reti. Di solito come firewall si usa un normale server dotato di due o più interfacce di rete, su cui è installato un sistema operativo di rete e un software specifico per firewall (come sistema operativo, Linux è molto usato per queste applicazioni). Esistono però anche dei firewall hardware dedicati.

Un proxy, faccia o no da firewall, ha anche un interessante effetto collaterale, e cioè quello di migliorare la privacy di chi sta dietro il proxy stesso. Per esempio, se una rete locale privata con cento computer è collegata a Internet attraverso un proxy, su Internet si vedrà solo l'indirizzo Ip del proxy, e quindi non si potrà distinguere quali attività vengono fatte da un certo PC (e da chi lo usa) e quali da un altro. Su Internet vi sono anche molti proxy che vengono messi a disposizione dai provider (a anche qualcuno pubblico), che hanno lo scopo di migliorare le prestazioni, ma dovrebbero anche avere l'effetto di nascondere l'indirizzo di chi si collega (per essere più sicuri è comunque meglio assumere informazioni caso per caso). Anche i servizi di navigazione anonima disponibili su Internet, come Anonymizer (http://www.anonymizer.com) e Spacesurf (http://www.spacesurf.com) sono sostanzialmente dei proxies.

Sui proxies usati come difesa della privacy è bene comunque tenere presenti due cose: chi gestisce il proxy deve essere un soggetto molto affidabile perché il proxy può servire a raccogliere traccia di **tutti** i collegamenti

effettuati dai suoi utenti, per cui se chi gestisce il proxy passa queste informazioni a terzi la privacy è meno tutelata che mai; inoltre, se il proxy è su Internet, la connessione tra la stazione di lavoro dell'utente e il proxy non è in alcun modo protetta né anonima, e quindi qualcuno potrebbe intercettare i dati proprio su tale connessione (questo non vale se il proxy supporta, tra sé e l'utente, connessioni protette tramite sistemi di crittografia, ad esempio Secure-Http, riconoscibile dall'URL https:// che sostituisce il normale http://).

I firewall che abbiamo appena descritto agiscono a livello applicativo (per cui sono a volte classificati nella categoria degli application gateway), perché sono in grado di analizzare e identificare le connessioni a livello applicativo, per cui sanno se è in corso una sessione telnet, o http, o ftp, o la trasmissione di un messaggio di posta elettronica. Ci sono anche dei firewall che agiscono a livello IP, garantendo il *packet filtering*, cioè filtrando i pacchetti secondo certi criteri, e l'*IP masquerading*, cioè il mascheramento degli indirizzi IP di una rete, in modo tale che all'esterno venga visto un solo indirizzo, come del resto avviene anche coi proxy. Questi servizi vengono spesso implementati nei router, ma possono essere realizzati anche da un comune PC dotato del software opportuno (anche qui viene molto usato Linux, che implementa l'IP masquerading direttamente nel kernel). Queste tecniche sono considerate meno sicure rispetto ai proxy/firewall, perché agiscono solo a livello IP e quindi hanno una visione meno completa del traffico di rete.

A un livello di sicurezza ancora minore si colloca la tecnica del *TCP wrapping*, che consiste nel non far arrivare le richieste di connessione TCP direttamente al servizio specifico cui sono destinate (ad esempio il server http), ma a farle passare attraverso un apposito demone, il *TCP wrapper* (di solito chiamato *tcpd*) che consulta dei file di configurazione (chiamati in genere *host_allow* e *host_deny*) che identificano degli indirizzi IP (singoli, o intere reti) a cui quel tipo di connessione deve o non deve essere consentita.

L'installazione di un firewall non risolve infatti tutti i problemi, poiché vi sono i cosiddetti hackers che, utilizzando varie tecniche, tentano - a volte con successo - di collegarsi senza autorizzazione a reti o computer scavalcando le protezioni ivi installate. Gli hackers sono particolarmente abili nello sfruttare errori presenti nei sistemi operativi o negli stessi firewall software, nonché errori commessi dai tecnici nel configurare i sistemi di protezione. Bisogna anche dire che gli hacker a volte sono teppisti o criminali che cercano di impossessarsi in modo illecito di informazioni che non avrebbero diritto di conoscere, magari allo scopo di utilizzarle per ulteriori attività illecite (si pensi a un mafioso che cerca di accedere ai dati della polizia), ma spesso si limitano a dimostrare la loro abilità tecnica nel campo delle reti senza provocare alcun danno reale. A volte poi le azioni degli hacker hanno un significato politico di difesa del diritto alla libertà di parola e alla riservatezza contro i tentativi di controllo dell'informazione in rete e delle comunicazioni personali.

Sempre in tema di sicurezza, bisogna ricordare la tecnologia delle **reti private virtuali** (VPN = Virtual Private Networks), per cui un certo numero di utenti comunicano in modo riservato su di una rete pubblica, che quindi fisicamente viene nello stesso momento utilizzata anche da altri utenti. In linea generale, questa tecnologia prevede che i pacchetti della VPN vengano crittografati e imbustati, completi di header e dati, in altri pacchetti, rispetto ai quali questi pacchetti appaiono come semplici dati. Questa tecnologia è detta **tunnelling**, perché i dati della VPN viaggiano attraverso la rete pubblica come nascosti in un tunnel.

Un pericolo per la privacy recentemente venuto all'attenzione degli utenti, ma non ancora noto come meriterebbe, è quello dello **spyware**. I programmi spyware sono software che svolgono attività del tutto "oneste" e magari funzionano anche in modo ottimo e sono di grande utilità, ma contemporaneamente inviano sulla rete al produttore o ad un altro soggetto informazioni sull'attività dell'utente che li sta usando e, quello che è peggio, lo fanno all'insaputa di quest'ultimo[21]. Ad esempio, tra i programmi indicati come spyware vi sia il noto Go!zilla, un eccellente software per la gestione del download di file: secondo alcune fonti, Go!zilla oltre a scaricare i file richiesti dall'utente, invia su Internet al sito del produttore l'indicazione di quali file vengono scaricati, con evidente violazione del privacy dell'utilizzatore. Abbiamo scritto "pare" perché naturalmente è difficile sapere con certezza quali software sono davvero spyware e quali no, e neppure si può escludere che

[21] Alcuni programmi dichiarano esplicitamente di inviare a terzi determinate informazioni, per cui non possono considerarsi spyware, dal momento che si suppone che chi li usa accetti questa pratica. Taluni software però sono stati criticati per il fatto che danno tali informazioni in modo incompleto o ambiguo.

qualche software venga falsamente accusato di essere spyware, con grave danno di immagine. Naturalmente lo spyware non può fare alcun danno a chi non è collegato in rete e usa il programma su un computer stand alone. Se l'utente invece si trova dietro un proxy/firewall, quello che avviene dipende da come opera lo spyware: se trasmette le informazioni utilizzando una normale connessione http dovrebbe riuscirci (a meno che l'operazione non comporti qualcosa di strano nell'url utilizzata o nei dati trasmessi tale da venire bloccato dal firewall), se invece usa protocolli proprietari e/o delle porte TCP particolari un firewall ben configurato dovrebbe bloccare l'operazione.

Molti software spyware rientrano nella categoria dell'**adware**, cioè del software di uso gratuito che visualizza, in una parte dello schermo, messaggi pubblicitari. In linea di principio, un software può benissimo essere adware senza essere spyware e viceversa, tuttavia risulta che di fatto numerosi adware sono stati accusati di essere spyware[22]: il collegamento tra le due categorie sta nel fatto che chi gestisce gli annunci pubblicitari ha interesse a conoscere quali sono le abitudini degli utenti nel loro uso di Internet in modo da poter fare apparire messaggi adatti ai presumibili interessi di questi ultimi. Degno di nota è il fatto che è stato diffusamente accusato di essere spyware un software su cui si basano molti programmi adware, quello della Aureate, che poi ha cambiato nome in Radiate. Per conoscere quale sia attualmente la politica della Radiate riguardo alla privacy si può consultare il sito ufficiale http://www.radiate.com. Un sito che sembra molto informato sullo spyware è http://www.grc.com.

Altre informazioni attinenti la privacy su Internet si trovano nell'ambito della trattazione dei cookies.

3. INTERNET

Descritta in generale, Internet è una cosa così semplice da essere quasi banale: si tratta semplicemente di una interconnessione di reti estesa a tutto il mondo, e basata sul protocollo TCP/IP. In questo modo, ogni computer collegato alla rete può accedere, in linea di principio, a quanto si trova su qualunque altro computer collegato[23].

Internet quindi è il risultato della connessione di molte reti, ognuna delle quali sarebbe in grado di funzionare anche come rete indipendente. Queste reti sono della più varia tipologia: si va da piccole LAN fino alle più grandi reti geografiche estese a diversi continenti (ad esempio la rete Sprint). L'importante è che ogni rete possa collegarsi con le altre attraverso il TCP/IP, anche se internamente utilizza un protocollo diverso.

Ogni rete che fa parte di Internet è quindi collegata a tutte le altre: questo non vuol dire che sia collegata **direttamente** a tutte le altre; per esempio, è possibile che la rete A sia collegata direttamente alla rete B, mentre sia collegata alla rete C solo passando attraverso le reti Y e Z. Questa disposizione dei collegamenti è indipendente dalla distanza dei computer che si devono collegare: per esempio, trovandosi a Genova e utilizzando un computer collegato a Video On Line, che utilizzava la rete Interbusiness, per consultare il catalogo delle biblioteche dell'Università di Genova, che si trova sempre a Genova ma su un server collegato alla rete GARR, i messaggi effettuavano il percorso Genova-Parigi-Ginevra-Segrate-Genova. Quindi su Internet la velocità e l'affidabilità del collegamento è in generale indipendente dalla distanza dei computer che si collegano.

Importantissimo è il fatto che Internet **non ha una struttura gerarchica**. Non c'è il computer centrale di Internet: i computer e le reti collegate si differenziano per prestazioni e volume di traffico gestito, ma sono tutti parti della rete allo stesso livello.

[22] Per quanto risulta all'autore del testo tuttavia non è finora incorso in accusa un adware molto importante e diffuso, e cioè Eudora in modalità sponsored.

[23] Qui non trattiamo delle origini e della storia di Internet. L'argomento è trattato in modo particolarmente chiaro ed esaustivo in *[Advanced 1999]*

Spesso si sentono cose catastrofiche sulle prestazioni e sull'affidabilità di Internet: si dice che i collegamenti sono troppo lenti, che non si riesce a lavorare e così via. La cose non stanno però proprio così: molte volte infatti le prestazioni sono più che accettabili. Il vero problema è che, proprio per la natura decentrata di Internet, **non è possibile garantire un livello minimo di prestazioni**. Ciascuno infatti può al massimo garantire per ciò che dipende da lui, ma le prestazioni di cui gode l'utente finale dipendono anche da tutte le reti utilizzate per connettersi ad un determinato host. Per questo esse sono molto variabili a seconda dell'host al quale ci si collega e anche da un'ora all'altra della giornata. Diverso è il caso dell'utente che si è rivolto ad un provider che fornisce un servizio mediocre o cattivo: in questo caso le prestazioni saranno sempre scarse, perché limitate proprio dal punto di accesso alla rete.

3.1 **Collegarsi a Internet**

Essere collegati a Internet significa quindi utilizzare un computer collegato ad una rete che fa parte di Internet.

Ci sono però diversi tipi di collegamento, che possono essere classificati in base al tipo di connessione oppure in base all'architettura del collegamento.

In base al tipo di connessione abbiamo:

- il collegamento in linea commutata (cioè con la normale linea telefonica), che è quello utilizzato soprattutto dall'utenza privata, oppure con linea ISDN (linea telefonica digitale), o tramite servizi pubblici a commutazione di pacchetto (X25 o frame relay) o ADSL che sono collegamenti **non permanenti**, che sussistono solo finché è in corso una chiamata o che comunque utilizzano un particolare circuito di una rete pubblica
- il collegamento **permanente**, con linee affittate di trasmissione dati o comunque tramite collegamenti punto-punto (cioè riservati a collegare due punti specifici e fissi), che è utilizzato soprattutto dalla grande utenza, e comunque **da chi ha una intera rete da collegare** (anche se è possibile collegare reti ad Internet tramite ISDN o linea commutata)

In base al tipo di architettura abbiamo

- il collegamento **in emulazione di terminale**, in cui il computer utilizzato non è uno dei computer della rete (e non ha indirizzo IP), ma solo il terminale di un altro computer, che è quello effettivamente collegato alla rete e su cui si trovano i programmi utilizzati; questo collegamento era utilizzato, fino a poco tempo fa, soprattutto in linea commutata, ma può trovarsi anche su una rete, quando un computer della rete si collega a un altro in emulazione di terminale; spesso tramite questo collegamento ci si collegava a una BBS che a sua volta era collocata su una rete collegata a Internet, per cui talvolta si aveva a disposizione solo un sottoinsieme dei servizi di Internet; si può dire che questo tipo di collegamento sia pressoché sparito, tuttavia potrebbe avere ancora un ruolo (marginale) per l'accesso remoto a reti private a loro volta collegate a Internet
- il collegamento **TCP/IP**, in cui il computer utilizzato è a tutti gli effetti uno dei computer collegati in rete (con tanto di indirizzo IP), e su cui girano realmente i programmi utilizzati; questo tipo di collegamento viene ora normalmente usato anche su linea commutata, attraverso i protocolli SLIP (ora in disuso) e PPP, che sono protocolli di livello Data Link che permettono di supportare IP e TCP su linea seriale, che può essere una linea telefonica o anche una linea dedicata punto-punto o ancora una rete ottica SONET ad alta velocità (PPP over SONET); di solito il collegamento TCP/IP viene usato per supportare un'architettura detta **client-server**, che per la sua importanza verrà illustrata in un paragrafo a parte; **attenzione alle confusioni**: soprattutto in ambiente Unix è anche possibile, anche non frequente che il computer, con il suo bell'indirizzo IP, si colleghi ad un altro computer come terminale (attraverso il protocollo Telnet o un altro simile), e poi utilizzi, per la navigazione su Internet, dei programmi client che si trovano su questo secondo computer; questa è evidentemente una cosa **completamente diversa** dall'emulazione di terminale di cui si parlava nel punto precedente

- il **mail gateway**, sistema raramente usato (almeno come unico tipo di accesso a Internet), con il quale si ha accesso direttamente al solo servizio di posta in rete, e ad altri servizi, come il trasferimento di file, solo tramite la posta, ossia mandando messaggi con una specifica sintassi a server di posta che fanno da gateway, cioè da intermediario, verso altri servizi, come trasferimento di file o interrogazione di database; se si accede alla posta tramite una connessione TCP/IP, in linea di principio si potrebbe avere accesso a tutti i servizi di Internet, e quindi le limitazioni sono dovute al provider, mentre se si accede alla posta tramite altri sistemi, allora non si ha in realtà alcun tipo di collegamento a Internet se non indiretto tramite il mail gateway[24]

A proposito del collegamento TCP/IP ci sono in giro degli equivoci: in particolare, esso viene identificato con l'uso di sistemi operativi e software grafici. Questa però è una illusione derivante dal fatto che in ambiente DOS/Windows quasi sempre il supporto TCP/IP e i relativi software applicativi sono destinati all'ambiente grafico, appunto Windows, mentre più raramente viene utilizzato il TCP/IP sotto DOS, che è l'ambiente a carattere (non grafico) di gran lunga più diffuso. In realtà è del tutto possibile l'uso di TCP/IP - e di tutti i protocolli del livello applicazione che vedremo in seguito - anche in un ambiente a carattere, come ad esempio Unix.

Notiamo che quanto vale per l'accesso a Internet realizzato con sistemi diversi da un collegamento permanente punto-punto vale più in generale per l'accesso remoto ad una rete: può trattarsi ad esempio dell'accesso ad una rete aziendale da parte di un ufficio periferico o di un dipendente in trasferta. In tutti questi casi, la rete a cui ci si collega (sia essa una rete privata o Internet) deve mettere a disposizione un apposito nodo, detto *access server* o *terminal server*, per autenticare chi tenta di collegarsi e stabilire poi il collegamento. L'access server può essere un normale server dotato di apposito software o anche di una macchina specializzata per questo uso. In particolare, vi sono access server che integrano concentratore, modem e router.

Il collegamento utilizzato dagli uffici regionali è un collegamento permanente TCP/IP.

3.1.1 **Architettura client-server**

L'architettura client-server è un tipo di interazione tra computer in cui un programma, detto client, che si trova su un computer (detto anch'esso client), invia richieste ad un altro programma, detto server, che si trova su un altro computer (detto anch'esso server), e ne riceve le risposte; in questa architettura non c'è colloquio permanente tra client e server: tra la ricezione di una risposta e una eventuale successiva richiesta non succede nulla, e sulla rete non c'è alcun traffico (anche se nel collegamento in commutata la linea telefonica è occupata e gli scatti corrono); in linea di principio qualsiasi computer, purché dotato dei software adeguati, può fare indifferentemente da client e da server (anzi, un client sw può collegarsi ad un sw software che sta girando sullo stesso computer), ma in pratica i server devono essere computer molto più potenti ed affidabili dei client, perché devono fornire servizi a molti utenti; parlando di client e server software, si può ben immaginare che un client non può formulare qualsiasi tipo di richiesta, e che non può inviare quelle che è in grado di formulare a qualsiasi server: invece ogni client e ogni server sono progettati per colloquiare attraverso uno o più ben determinati protocolli, ognuno dei quali serve per uno scopo specifico; si tratta di protocolli come i già citati FTP, HTTP e altri (es. POP3 e SMTP per la posta elettronica, NNTP per le news); un client può colloquiare solo con i server che utilizzano lo stesso protocollo, oppure può interagire con altri programmi tramite dei gateway; ovviamente nulla vieta che un particolare sw sia stato scritto in modo da fare da client per diversi servizi: in questo caso equivale a un insieme di diversi client distinti

L'architettura client-server si contrappone a quella basata sui terminali collegati con un host centrale che gestisce tutta l'elaborazione, mentre il terminale si occupa di accettare gli input degli utenti e di visualizzare

[24] Alcuni servizi di mail gateway, utilizzati appunto per il trasferimento di file o l'interrogazione di basi dati, sono comunque disponibili su Internet per gli utenti dotati del normale collegamento TCP/IP. Quelli di trasferimento file possono essere molto comodi, perché il file desiderato viene inviato direttamente in allegato ad un messaggio di posta, senza che l'utente debba collegarsi ad un server ftp o http e seguire lo scaricamento.

l'output (che può essere grafico o a carattere) prodotto dal sistema centrale. Attualmente si usano di solito non dei veri terminali, ma dei normali computer dotati di software in grado di emulare vari tipi di terminali e di collegarsi all'host via rete (i veri terminali non hanno un indirizzo di rete). Questa tecnica è comunemente usata per connettersi soprattutto a sistemi Unix o a mainframe, ed eseguirvi programmi o comandi del sistema operativo. Rispetto al client-server, questa tecnica pone un maggior carico elaborativo sul sistema centrale, che quindi deve essere ben dimensionato; per contro, come posti di lavoro sono spesso sufficienti anche vecchi computer, purché dotati del software di rete e di un emulatore di terminale, che è un programma molto leggero. Tutti i programmi e i dati risiedono sul server, per cui quando questo o la rete non funzionano è completamente impossibile lavorare; ciò comporta però il vantaggio che sul client non si devono fare installazioni e configurazioni di programmi complicati, e non c'è rischio di perdere dati.

Come si vede, client-server e terminali hanno ognuno vantaggi e svantaggi. A complicare il quadro, c'è poi il fatto che ci possono essere diversi tipi di client: ci sono client pesanti, cioè programmi che compiono localmente molte elaborazioni e sono grossi e complessi, oppure leggeri, cioè programmi che compiono localmente pochissime elaborazioni, demandandole quasi tutte al server. Quali sono le elaborazioni che compie localmente un client leggero ? Devono essere almeno l'accettazione degli input dell'utente e la visualizzazione delle risposte del server. Questi però sono anche precisamente i compiti dei terminali ! Come si vede, più un client è leggero e più tende ad assomigliare a un terminale, o ad un emulatore di terminali.

Attualmente si assiste ad una ripresa di interesse per i terminali opportunamente rivisitati ed aggiornati. Ciò è dovuto soprattutto al fatto che gestire una rete con molti computer ognuno dotato di elevata capacità di elaborazione, di programmi complicati e di dati importanti rischia di essere molto costoso, mentre contemporaneamente è cresciuta la capacità elaborativa dei server e l'affidabilità delle reti. Si sente quindi parlare continuamente del **network computer**, che è un computer (non un puro terminale) che però se disconnesso dalla rete non può fare quasi nulla, perché è progettato per scaricare dalla rete quasi tutto ciò che esegue: si tratta per lo più di programmi client, che quindi operano in collegamento con server remoti. Il network computer puro non ha memorie di massa, per cui accede ai dati solo attraverso la rete. Non mancano inoltre le tecnologie, come Hydra e WinFrame, per utilizzare in emulazione di terminale i server Windows NT, finora refrattari a questo tipo di collegamento.

Per ora è difficile dire quanto successo commerciale avranno queste soluzioni (è indubbio che il network computer non ha ancora avuto il successo che i suoi propugnatori da anni si attendono), ma sembra certo che la tendenza attuale sia in linea generale rivolta ai client leggeri e alla semplificazione della gestione e manutenzione delle stazioni di lavoro.

3.2 **Servizi di Internet e delle reti**

Dopo aver visto la struttura e il funzionamento di Internet, vediamo a cosa serve e come si usa.

Internet offre diversi servizi, per ognuno dei quali si deve utilizzare un apposito software (di fatto, come si diceva, ci sono software che sono stati progettati per servire a più di una funzione: questo non toglie che si tratti di funzioni ben distinte).

In realtà, le cose che descriveremo nel seguito si possono fare, e spesso si fanno, anche sulle reti locali, ma nella spiegazione verranno evidenziati soprattutto gli aspetti più legati a Internet. Parleremo però anche della condivisione di unità per la sua notevole importanza pratica, anche se si tratta di un servizio utilizzato quasi esclusivamente sulle reti locali.

Nelle spiegazioni seguenti, se non diversamente specificato, si farà riferimento a un collegamento TCP/IP. **I protocolli corrispondenti ai vari servizi appartengono allo strato applicazione.**

3.2.1 **Emulazione di terminale**

Come abbiamo detto, anche quando il collegamento a Internet avviene tramite un computer dotato del suo indirizzo IP, è sempre possibile attivare un software di emulazione per collegarsi, tramite il protocollo **telnet** (o più raramente, soprattutto per collegarsi ai mainframe IBM o agli AS/400 sempre IBM, il protocollo **tn3270**) a server che supportano questo tipo di collegamento, che è considerato piuttosto superato, soprattutto perché meno amichevole per l'utente non esperto. Si tratta di solito di servizi già funzionanti da molto tempo e collegati direttamente a Internet senza modifiche, a volte residenti su mainframe. Comunque il fatto che esteticamente siano meno piacevoli e meno facili da usare rispetto al Web non vuole certamente dire che anche questi collegamenti non possano rivelarsi anche molto utili. Anzi, diversi software con interfaccia telnet tuttora utilizzati sono oltremodo agili ed efficienti, una volta che si sia acquisito un minimo di addestramento all'uso. Interessante è anche il fatto che i client telnet richiedono pochissime risorse al computer sul quale girano, per cui possono essere utilizzati anche su computer non solo superati, ma addirittura obsoleti e inutilizzabili per quasi ogni altro scopo.

3.2.2 **Posta elettronica**

La posta elettronica permette di spedire e ricevere messaggi a e da qualsiasi altro utente che abbia accesso al servizio. Di solito l'utente, anche quando dispone di un collegamento permanente, utilizza un client di posta elettronica che si collega un server specializzato sia per ricevere che per inviare i messaggi, utilizzando quasi sempre i protocolli SMTP per il trasporto e POP3 per ottenere dal server la propria posta. Più esattamente, i messaggi arrivano al server, e poi vengono prelevati dal client, che li trasferisce sul computer dell'utente; i messaggi scritti da quest'ultimo vengono trasferiti dal client al server che poi provvede all'inoltro. In genere l'account di posta del mittente e del destinatario non si trovano sullo stesso server, per cui il server SMTP deve inoltrare il messaggio utilizzando dei server intermedi.

Ogni utente è identificato da un indirizzo di posta elettronica, che è protetto da password in modo che nessuno possa leggere la posta degli altri.

Gli indirizzi di posta elettronica hanno la struttura: *nomeutente@nomehost*, ad esempio *beni.culturali@regione.liguria.it*.

Le funzioni di base dei programmi di posta elettronica sono, ovviamente, quelle di inviare e ricevere posta. Comunque i programmi attuali permettono, tra l'altro, di rispondere ad un messaggio senza dover digitare l'indirizzo del mittente, e citando automaticamente il testo del messaggio cui si risponde, inviare copia di un messaggio a più utenti (*carbon copy*), inoltrare ad altri un messaggio ricevuto (*forward*), salvare i messaggi su disco. I client più moderni consentono inoltre di suddividere i messaggi in diverse cartelle (dette di solito *folders*), e anche di filtrare automaticamente i messaggi in arrivo in base a determinati parametri (ad esempio il soggetto o il mittente), inserendoli automaticamente nella cartella desiderata.

I client email inoltre permettono di cancellare i messaggi non più utili, cosa indispensabile se non si vuole riempire il disco di vecchi messaggi; bisogna però ricordare che in quasi tutti i client i messaggi cancellati non vengono subito eliminati fisicamente ma, per consentire di rimediare a cancellazioni involontarie, vengono spostati in una speciale cartella detta di solito *Trash*, per cui quando si vuole eliminare fisicamente un messaggio bisogna poi rimuoverlo anche da questa cartella. Per recuperare effettivamente lo spazio disco, poi, è necessario effettuare il compattamento delle cartelle, cosa che si fa utilizzando un semplice comando presente nei menu di tutti i client.

Tra i più noti programmi di posta elettronica citiamo Eudora, Pegasus, Pine.

Attraverso la posta elettronica è anche possibile inviare files di qualunque genere (ad esempio un documento di Word) non come parte del messaggio, ma come allegati (attach) allo stesso. Il client del destinatario, alla ricezione, provvede a salvare su disco il file allegato. Si tratta, come è evidente, di un servizio estremamente utile: tuttavia si deve cercare di evitare di inviare files molto grandi (diciamo superiori a 1 Mb), o almeno inviarli solo se è davvero necessario, perché questi file provocano molto traffico sulla rete, possono riempire oltre misura il server, e inoltre possono anche venire rifiutati dal server stesso. Può anche accedere che alcuni utenti non ricevano correttamente gli allegati, o per limiti del loro client, o perché non avviene in modo appropriato la codifica dei dati nel passaggio tra i diversi sistemi, ma questo ormai è diventato piuttosto raro.

Negli ultimi anni è emerso un fenomeno molto fastidioso legato alla posta elettronica, cioè quello dello **spamming**, che consiste nell'invio non sollecitato di messaggi di posta elettronica di carattere commerciale, che per lo più pubblicizzano siti porno oppure offerte alquanto sospette (prestiti a tassi eccezionali, lavori che fanno arricchire in un attimo ...). Alcuni indirizzi sono presi di mira dallo spamming in misura maggiore di altri, al punto che può accadere che uno riceva molti più messaggi dagli **spammers** (cioè quelli che praticano lo spamming) che non dai suoi reali interlocutori, con grande perdita di tempo sua e spreco di banda sulla rete. Gli spammer raccolgono indirizzi da ogni fonte possibile, ad esempio siti web, liste di discussione ecc. Ci si può chiedere che vantaggio traggano le ditte da questo genere di pubblicità che quasi sempre genera soprattutto fastidio nei destinatari, ma bisogna tener conto del fatto che il costo unitario di un messaggio di posta è molto basso, quindi evidentemente agli spammers basta una piccola percentuale di esiti positivi perché l'attività sia redditizia, Come ci si può difendere dagli spammers? In vari modi, nessuno dei quali però è infallibile: ad esempio una cosa utile evitare di rendere pubblico il proprio indirizzo di posta elettronica, il che però spesso non è possibile, o almeno non completamente (ora che è facile procurarsi indirizzi email gratis, conviene sempre averne almeno due, ed evitare accuratamente di renderne pubblico almeno uno, in modo da essere abbastanza sicuri che rimanga libero dallo spamming). In molti sistemi di posta è possibile bloccare gli indirizzi degli spammers, in modo che messaggi provenienti da loro vengano automaticamente cancellati (attenzione a non bloccare per sbaglio anche qualche indirizzo "onesto"!), ma questo è meno efficace di quanto si possa pensare perché gli spammers cambiano continuamente i loro indirizzi, per cui succede che lo stesso messaggio arrivi ogni giorno da un indirizzo diverso. Non contesti di questo, talvolta gli spammers utilizzano abusivamente l'indirizzo di qualche persona ignara, che così rischia di venire bloccato dai destinatari dei messaggi. Infine è possibile bloccare interi server di posta specializzati nello spamming, ma neanche questo ha sempre successo perché gli spammers non utilizzano solo questi server specializzati, ma anche dei server di uso generale (a volte succede anche qualche server innocuo finisca per sbaglio nella lista nera di quelli da bloccare, determinando l'impossibilità di ricevere messaggi utili finché l'inconveniente non venga scoperto e risolto). Molto importante è anche ricordare che non si deve mai rispondere ai messaggi degli spammers, anche quando contengono l'indicazione di come scrivere in modo da non ricevere ulteriori messaggi: queste indicazioni servono in realtà agli spammers per individuare meglio gli indirizzi effettivamente attivi. Se si risponde, si dimostra con ciò che l'indirizzo viene utilizzato da qualcuno, e quindi esso diventa più prezioso agli occhi degli spammers, per cui la conseguenza più probabile di queste risposte è quella di vedersi arrivare ancora più messaggi di prima.

3.2.3 **News**

Il sistema delle news è simile alla posta elettronica, perché permette di ricevere e spedire messaggi. Qui però è necessario collegarsi ad appositi server, che ospitano i newsgroup (gruppi di discussione), che raccolgono messaggi di argomento omogeneo: chi si collega può vedere tutti i messaggi degli iscritti al newsgroup, e rispondere al gruppo o - tramite la posta elettronica a un iscritto in particolare.

Ciascun newsgroup è duplicato su moltissimi server nel mondo: di solito i server delle news (newsserver) non permettono l'accesso a chiunque, ma solo agli utenti dell'organizzazione che li gestisce (ad esempio un provider per i suoi abbonati). Per questo è possibile che alcuni newsgroup si trovino su certi server e non su altri. I messaggi più vecchi vengono eliminati dal server, con scadenze diverse per ciascun server: l'utente può comunque salvare qualunque messaggio sul proprio computer, mentre, se non li ha salvati in tempo, non può più recuperare i messaggi eliminati dal server.

I client di posta elettronica permettono di compiere con facilità tutte le operazioni necessarie per usare i newsgroup: visualizzare la lista dei newsgroup disponibili, iscriversi a uno o più di essi, visualizzare i messaggi, salvarli su disco, rispondere al newsgroup o all'indirizzo di posta elettronica dell'autore ecc. Tra i client più noti citiamo Agent e Microsoft News.

3.2.4 **FTP**

Il protocollo FTP permette di trasferire file, e quindi è di fondamentale importanza per trasferire sul proprio computer programmi e dati disponibili su Internet.

Per utilizzare questo protocollo è necessario un client FTP che si colleghi a un server FTP: il protocollo gestisce tutti i comandi necessari per individuare, scaricare, copiare, cancellare ecc. i file del server. Per collegarsi a un server FTP bisogna avere un identificativo utente: di solito i server FTP pubblici consentono l'accesso a tutti quelli che si identificano come *anonymous* ma solo per scaricare files, e qualche volta anche per inviarli al server, non per modificare, cancellare o eseguire i files del server. Queste operazioni invece sono possibili a specifici utenti abilitati a fare ciò da parte dell'amministratore del server.

Una volta che si è collegati al server, si vede un panorama simile a quello che si vede sul proprio computer, cioè elenchi di files e di directory. I migliori client FTP grafici (ad esempio WS FTP e CuteFTP) consentono di effettuare tutte le operazioni di individuazione e copia dei file con estrema facilità, utilizzando una interfaccia utente simile a quella del File Manager di Windows. Ci sono anche dei client FTP a carattere che mettono a disposizione una interfaccia simile a quella del prompt del DOS.

3.2.5 **Gopher**

Su Internet ci sono moltissime cose, per cui c'è bisogno di qualche mezzo per organizzare e reperire le informazioni.

Il Gopher è un sistema di menu e sottomenu, che permette di sfogliare facilmente pagine di informazioni, e anche di saltare ad una pagina prestabilita, così come avviene nei menu dei vari programmi, che elencano tutti i comandi che si possono dare al programma.

A differenza del WWW (di cui parleremo dopo), il Gopher ha una struttura gerarchica: ci sono menu di livello superiore, poi sottomenu, poi altri sottomenu: per questo è meno flessibile del WWW ed è oramai pressoché scomparso, anche se può darsi che su Internet ci sia ancora qualche server gopher.

I client gopher permettono di scorrere i menu, saltare ad altri menu su altri server, visualizzare testi e scaricare file che compaiano come voci di menu. Tra i client gopher citiamo Wsgopher e Hgopher. Questi client già alcuni anni fa erano stati quasi del tutto soppiantati dai browser WWW (v. seguito) che possono essere usati anche come client Gopher, e a maggior ragione oggi hanno solo più un interesse storico[25].

3.2.6 **WWW**

Passiamo finalmente a trattare il WWW, di cui oggi tanto si parla. Il WWW, che significa World Wide Web, cioè *ragnatela mondiale* è il più potente sistema di reperimento di informazioni oggi disponibile.

[25] Anche se forsi pochi lo sanno, ci sono anche i collezionisti di vecchi programmi che fanno un'opera molto utile perché conservano memoria di quello che c'è stato nel mondo dell'informatica in cui le novità si susseguono a ritmo vorticoso. I client gopher dovrebbero essere piuttosto appetibili per tali collezionisti.

Il WWW fu ideato verso il 1990 al CERN di Ginevra da Tim Berners-Lee, allo scopo di fornire un mezzo adeguato per gestire l'immensa quantità di informazioni presenti su una rete non gerarchica come Internet. In precedenza, il modello dei sistemi informativi era stato centralizzato (un computer centrale con molti terminali collegati), e quindi i mezzi adatti a questa architettura, soprattutto l'emulazione di terminale, mal si prestavano ad una architettura distribuita, nella quale per di più convivono i prodotti hardware e software più diversi.

Pertanto il WWW fu concepito per essere un sistema

- non gerarchico, cioè che non facesse riferimento a un centro, a una fonte privilegiata, ma consentisse di accedere a tutte quelle disponibili secondo criteri determinati dalle esigenze dell'utente e non dalla struttura dei sistemi informativi (di qui il riferimento alla ragnatela)
- non proprietario, cioè che non dipendesse da strumenti hardware e software, o da protocolli e linguaggi esclusivi di uno specifico fabbricante
- multimediale, cioè in grado di gestire documenti di diversa natura (testo, immagini, suono)

Si può ben dire che questi obiettivi siano stati raggiunti al punto che oggi il WWW è un sistema informativo grandioso, che permette di accedere a informazioni in ogni campo del sapere in modo uguale per tutti coloro che vi hanno accesso, indipendentemente dalla localizzazione o dall'uso di prodotti di una certa marca piuttosto che di un'altra. Una caratteristica del WWW, infatti, è quella di nascondere completamente le particolarità informatiche dei computer utilizzati: quando ci colleghiamo a pagine WWW, vediamo le informazioni ivi contenute e i collegamenti tra esse in modo indipendente dal fatto che il computer sia Compaq o Sun, o che il sistema operativo sia Unix o NT. Quello che si richiede è solo che vengano utilizzati i protocolli e gli standard propri del WWW.

Allo scopo di evitare confusioni è bene fare un precisazione sulla multimedialità: con l'orientamento sempre più commerciale e di massa che ha acquistato Internet, vi sono alcuni che tendono a vedere il WWW sul modello della televisione e a identificarlo con i documenti non testuali, come immagini, filmati, suoni ecc. In realtà il WWW è una struttura di organizzazione e reperimento dei documenti che di per sé non determina con quali documenti si riempie la struttura, né tantomeno con quali scopi. Quindi nella ragnatela del WWW possono benissimo esserci (e ci sono, di fatto) documenti esclusivamente testuali (e quindi fruibili anche con computer privi di qualsiasi supporto multimediale), e sarebbe un errore gravissimo pensare che essa sia destinata sono a finalità commerciali o di intrattenimento gestite da poche organizzazioni (come accade con la TV): in realtà il WWW offre immense possibilità di fruizioni a fini di studio, ricerca e documentazione (come presumibilmente era nelle intenzioni dei suoi ideatori).

Il WWW si compone di diversi elementi, che ora - per la loro importanza - descriveremo singolarmente.

3.2.6.1 **Il protocollo HTTP**

Il WWW ha bisogno innanzitutto di un protocollo comune che permetta alle macchine collegate di interagire in modo appropriato per la realizzazione degli scopi che il WWW stesso si propone.

Questo protocollo è denominato HTTP, cioè Hyper Text Transfer Protocol (Protocollo di trasferimento di ipertesto), e si colloca nello strato applicazione del TCP/IP, cioè al livello più alto, più vicino all'utente finale, e quindi non si occupa delle caratteristiche degli strati più bassi, per esempio della connettività fisica, del routing ecc. L'instaurazione di una sessione HTTP presuppone che tutti questi elementi funzionino correttamente.

Tutte le operazioni che verranno descritte di seguito, effettuate nell'ambito del WWW, presuppongono che il colloquio tra i computer interessati avvenga secondo quanto previsto dal protocollo HTTP. Il contenuto di questo colloquio è una serie di richieste e risposte che hanno come scopo il trasferimento di determinati oggetti dal server (o da un altro host cui il server ha accesso) al client: questi oggetti sono poi in ultima analisi dei

files. Si noti che se il file, invece di essere un documento HTML (v. seguito) è, per esempio, un programma eseguibile, esso in genere non viene visualizzato ma solo trasferito sul computer dell'utente: il risultato finale è lo stesso che se si fosse utilizzato il protocollo FTP, ma il modo per arrivarci è diverso, poiché non viene utilizzato FTP, ma http, e quindi sull'host deve essere presente un server WWW e non un server FTP. Inoltre il fatto che il file venga semplicemente trasferito e non visualizzato o trattato in altro modo dipende dal client: è concepibile un client HTTP che, ad esempio, rifiuta di trasferire programmi eseguibile (anche se i client reali non lo fanno), mentre un client FTP ha come unica attività quella di trasferire files (si può dire che FTP è orientato al file system, mentre HTTP è orientato al documento).

La versione di HTTP attualmente utilizzata è la 1.0, ma ultimamente è stata rilasciata la 1.1, che presenta notevoli miglioramenti (dovrebbe essere più efficiente e garantire prestazioni migliori).

3.2.6.2 **Il linguaggio HTML**

Adesso che abbiamo il protocollo per far comunicare il client e il server, abbiamo bisogno anche di un insieme di regole per creare i documenti che vengono scambiati in questa comunicazione.

Questo insieme di regole è il linguaggio HTML (Hyper Text Markup Language), che non è un vero linguaggio di programmazione come il C e il Pascal, ma un linguaggio, di facile apprendimento, che permette di definire pagine associando al testo dei *tag*, cioè dei comandi che determinano come deve apparire un certo testo e che operazioni consente di fare.

La cosa fondamentale dei documenti HTML non è però il poter determinare l'aspetto del testo, ma è la loro struttura ipertestuale: essi possono contenere dei **link**, cioè delle parti di testo che, con un comando apposito (di solito cliccando su di esse con il mouse) permettono di accedere ad altre entità. Ad esempio, in un documento HTML possiamo inserire un elenco di nomi, e fare in modo che ciascun nome sia un link alla fotografia della persona corrispondente, per cui l'utente cliccando sul nome *Claudia Schiffer* vedrà apparire la fotografia appropriata.

La potenza di questa tecnica deriva anche dal fatto che le "entità" cui si può riferire un link sono molteplici:

- altri documenti HTML
- files di testo
- files di qualunque genere (che, a seconda del tipo di file e della configurazione del client, vengono riprodotti, eseguiti o salvati sul computer dell'utente), che possono anche essere immagini, suoni, filmati ecc.
- menu Gopher
- newsgroup
- indirizzi di posta elettronica
- sessioni in emulazione di terminale (telnet e tn3270)

Il fatto che un link possa riferirsi ad altri documenti HTML, che a loro volta possono contenere altri link ad altri documenti HTML è proprio ciò che consente di effettuare una consultazione non gerarchica, ma a ragnatela.

Il documento HTML, per suo conto, oltre ai link può contenere immagini o altri elementi multimediali - che vengono visualizzati o riprodotti insieme al documento, e - ovviamente - testi di qualsiasi tipo e lunghezza. Il contenuto di una particolare pagina dipende dallo scopo per cui è stata realizzata: in particolare, vi sono pagine che sono interessanti soprattutto per i link (ad esempio indici, elenchi), altre soprattutto per le immagini, altre soprattutto per i testi.

Un documento HTML poi può contenere dei moduli (forms) in cui l'utente può inserire dati per dare il via a determinate operazioni, per esempio una ricerca in una banca dati o una transazione commerciale. Il risultato di queste operazioni viene presentato come documento HTML, e può contenere altri forms, link ecc. Infine un documento HTML può determinare l'esecuzione di programmi in linguaggio Java. Sui forms e sul Java sarà detto di più successivamente.

A questo punto dovrebbe essere ancora più chiaro quello che avevamo osservato all'inizio della trattazione del WWW: anche se esso nell'opinione corrente è strettamente associato alla grafica e al multimediale, immagini, video, suoni, sono solo alcuni dei documenti che si possono gestire attraverso il WWW, ossia sono solo nodi della ragnatela, ma non la ragnatela stessa. Il fatto, poi, che i documenti HTML vengano visualizzati in formato grafico dipende dai client, e di per sé non determina il trasferimento attraverso la rete di files grafici, né alcuna elaborazione grafica sul server.

Il linguaggio HTML è giunto alla versione 4.0. Sono poi in progetto numerose estensione del HTML come lo conosciamo ora, ed in particolare XML (Extended Markup Language). Scopo di XML è quella di consentire una elaborazione dei documenti molto più sofisticata di quella possibile finora. Altre soluzioni allo studio sono il DHTML (Dynamic HTML) per la creazione di documenti dinamici, che cioè si modificano a seconda delle azioni dell'utente e il HEITML, dotato di speciali istruzioni per consentire con facilità l'accesso a basi dati.

3.2.6.3 **Gli URL**

Abbiamo visto che i documenti HTML possono contenere link ad altre entità: si capisce facilmente che queste entità devono essere accessibili sulla rete (o, come caso limite, sul computer su cui gira il client WWW). Ma per poter essere raggiunte devono essere identificabili in modo univoco; inoltre l'identificativo dovrebbe essere facilmente gestibile dall'utente, e non solo comprensibile dalla macchina.

Gli identificativi utilizzati nell'ambito del WWW sono detti URL, cioè Uniform Resource Locator.

Un URL è diviso in due parti:

- **prefisso**, che identifica il tipo di collegamento; la tabella che segue mostra i prefissi più importanti e il loro significato

http://	sessione http (documenti HTML)
ftp://	sessione ftp (trasferimento file)
wais://	sessione WAIS (ricerca informazioni)
gopher://	sessione Gopher
mailto:	invio posta elettronica
news:	collegamento a un newsgroup
telnet://	emulazione terminale (VT 100)
tn3270://	emulazione terminale (TN 3270, utilizzata con i mainframe IBM)
nfs://	sessione che utilizza il protocollo WebNFS (per nulla diffuso)

- **identificativo dell'oggetto**, che consiste nel nome di un host (nome o indirizzo IP), eventualmente seguito dalla specificazione di una particolare entità (ad esempio www.regione.liguria.it, www.geocities.com/CapeCanaveral/3616); bisogna porre attenzione al fatto che **le maiuscole e le minuscole sono significative, cioè sono considerati caratteri diversi** (tranne che negli indirizzi di posta elettronica), per cui www.Regione.Liguria.it è diverso da www.regione.liguria.it (ed è sbagliato).

Ecco alcuni esempi di URL (tutti reali) in forma completa:

http://www.netcom.com

http://lcweb.loc.gov
http://www.nectec.or.th
http://www.geocities.com/CapeCanaveral/3616
ftp://ftp.winsite.com/pub/win/w95/
ftp://ftp.cnr.it
ftp://ftp.cdrom.com
gopher://val-dor.cc.buffalo.edu
mailto:informatica@regione.liguria.it
news:comp.internet.library
telnet://access.usask.ca
tn3270://vm.cineca.it

Notiamo che spesso il nome degli host fa riferimento al tipo di server, per cui spesso i nomi dei server WWW cominciano con www, quelli dei server ftp con ftp ecc., ma questo è solo un aiuto alla memoria degli utenti, mentre è assolutamente irrilevante dal punto di vista della rete.

Notiamo inoltre che è errato citare una risorsa Internet con il solo nome dell'host (per esempio ftp.cnr.it), perché questo non permette di sapere con certezza (ma, al più, di supporre) quale protocollo viene supportato da quell'host.

Infine un accenno al sistema dei nomi di dominio. Possiamo dire che un **dominio** è un insieme di indirizzi e che ci possono essere sottodomini all'interno di un dominio. La parte più a **destra** di un nome di host identifica il dominio di primo livello: il dominio di primo livello corrisponde al paese in cui si trova l'host, tranne che per gli USA, in cui identifica il tipo di organizzazione. Ecco alcuni domini di primo livello:

gov	governo USA
mil	organizzazioni militari USA
com	organizzazioni commerciali
net	reti
org	altre organizzazioni (in particolare quelle non commerciali)
edu	organizzazioni educative
int	organizzazioni internazionali

ar	Argentina
at	Austria
au	Australia
be	Belgio
br	Brasile
ca	Canada
ch	Svizzera
cl	Cile
cn	Cina
co	Colombia
cu	Cuba
de	Germania
dk	Danimarca
ee	Estonia
es	Spagna
fi	Finlandia
fr	Francia
hr	Croazia
hu	Ungheria
il	Israele
in	India

ir	Iran
is	Islanda
it	Italia
jp	Giappone
kr	Corea
mx	Messico
my	Malaysia
mz	Mozambico
nl	Olanda
nz	Nuova Zelanda
pe	Peru
pl	Polonia
pk	Pakistan
ro	Romania
ru	Russia
se	Svezia
sg	Singapore
sk	Slovacchia
ua	Ucraina
uk	Gran Bretagna
us	USA
uy	Uruguay
ve	Venezuela
za	Sudafrica

Spesso si crede che i primi sei domini elencati siano riservati agli host USA, mentre ciò è vero solo per gov e mil (mentre per contro esiste il dominio us, anche se poco utilizzato).

Ci sono alcuni criteri che **di fatto** vengono **di solito** (quindi non sempre) utilizzati per dare il nome agli host delle organizzazioni, ed in particolare degli enti commerciali, che tendono ad avere la struttura seguente (ci riferiamo qui ai server WWW):

www.[nome dell'organizzazione].[com oppure org negli USA o nome del paese altrove]

Com si riferisce ad enti commerciali, *org* ad altri enti, in particolare quelli senza scopo di lucro.

Ad esempio:

http://www.ferrari.it
http://www.ford.com
http://www.yamaha.com
http://www.pentax.com
http://www.oclc.org

Ci sono comunque parecchi nomi che non corrispondono a questo schema. Attenzione inoltre alle omonimie, che possono condurre a gravi equivoci: se due enti si chiamano entrambi *xyz*, non potranno avere entrambi un host denominato *www.xyz.com*, perché non possono esserci due host con lo stesso nome, per cui l'host *www.xyz.com* corrisponderà ad uno solo di essi. È quindi necessario controllare che si tratti proprio dell'*xyz* che interessa a noi.

3.2.6.4 **Il client WWW**

Il client WWW è un programma importantissimo, perché è quello che gira sul computer dell'utente, ed ha il compito di gestire il colloquio con il server WWW, ed in particolare di scaricare i documenti HTML richiesti dall'utente, visualizzarli interpretando correttamente i tag HTML, e inviare nuovamente al server le ulteriori richieste eventualmente formulate dall'utente. Esso viene di solito denominato **browser.**

È molto importante comprendere che il colloquio tra il client e il server si riduce a inviare richieste e a ricevere risposte: ogni coppia richiesta-risposta sta a sé, anche se l'utente sta effettuando una unica attività che richiede molteplici richieste. Ad esempio, l'utente potrebbe fare ricerche in una banca dati, e inviare una ricerca, e poi la stessa con un piccola variazione: per il server si tratta di due ricerche indipendenti.

Spesso l'utente, ottenuto il documento desiderato, si sofferma poi ad esaminarlo: non si deve credere che durante tutto questo tempo ci sia un collegamento con il server WWW; l'utente sta esaminando la copia del documento trasferita sul suo computer locale e visualizzata dal client. Il compito del server si è esaurito con l'invio del documento appropriato. Anche la grafica viene gestita completamente dal client.

Spesso avviene inoltre che il documento sia una lista di link, per cui esso viene scaricato, e a partire di lì si attivano connessioni con altri server: in tutte queste connessioni il server da cui era stata originariamente prelevata la lista non ha alcun ruolo, poiché si tratta di richieste effettuate **direttamente** dal client ai server compresi nella lista che si trova nel file trasferito sul sistema locale.

Il client WWW in senso stretto riconosce solo il protocollo HTTP: di fatto i client WWW comunemente utilizzati sono programmi particolarmente sofisticati e ricchi di funzionalità, per cui ne supportano anche altri: tutti supportano le sessioni FTP, quasi tutti anche le news, il Gopher e l'invio di posta elettronica; qualcuno (in particolare Netscape Navigator 2.0 e versioni successive) la gestione completa della posta elettronica. Molti ora supportano i linguaggi Java e Javascript, di cui parleremo oltre. Alcuni poi forniscono servizi aggiuntivi, ad esempio la creazione di pagine WWW (Navigator e Oracle Power Browser) o il funzionamento come server Web (Power Browser). Inoltre i browser possono visualizzare direttamente non solo file HTML, ma anche files di testo e immagini in formato GIF e JPEG (gli ultimi browser supportano inoltre il nuovo formato grafico PNG); inoltre possono utilizzare programmi esterni per visualizzare o riprodurre numerosi altri tipi di file, in particolare quelli multimediali.

I client WWW più diffusi attualmente sono: Netscape Navigator, Microsoft Internet Explorer e Opera, cui si aggiungono altri prodotti meno diffusi, come Turbo Browser, Neoplanet ed Enigma Browser. Vanno citati anche alcuni browser ancora disponibili ma non più sviluppati da anni, in particolare NCSA Mosaic (una tempo il più importante), e altri che sono scomparsi, come Oracle Power Browser e Cello (uno dei primi browser grafici, diffuso nei tempi remotissimi prima del 1995; aveva una grafica molto graziosa ma, a differenza degli altri, oggi è pressoché inutilizzabile perché non supporta i form). Tutti questi programmi funzionano in ambiente grafico (Windows, Macintosh o Unix/X-Window); ci sono però anche client a carattere, tuttora utilizzati, per lo più sotto Unix, come Lynx, ben noto a coloro che utilizzavano già il Web prima dell'avvento generalizzato dei client grafici (detto così sembra una cosa remota, ma questo avveniva quattro o cinque anni fa).

3.2.6.5 **Il server WWW**

Da un punto di vista strettamente tecnico, parlare del server WWW (o server Web) è più interessante che parlare del client, perché il server svolge funzioni più complesse e raffinate. In realtà, la grandissima maggioranza degli utenti di Internet mette le mani solo sui client, e non sui server: tuttavia sapere come funzionano questi ultimi serve almeno a capire meglio anche quello che si fa coi client, e poi - a proposito di mettere le mani sui server Web - diremo alla fine qualche altra cosa interessante ...

Per impiantare un server WWW ci vuole innanzitutto un computer (che ovviamente deve avere i collegamenti di rete appropriati, se si vuole che qualcuno vi si colleghi), che deve essere una macchina piuttosto potente, perché deve fornire servizi a molti utenti: per questo scopo vengono usati dei server compatibili PC, cioè con

CPU Intel o compatibile, come il Pentium o il Pentium Pro (a volte con più CPU, e sempre con abbondante RAM), oppure macchine con altre architetture più potenti e costose, come i Digital con CPU Alpha (e altre macchine con CPU Alpha), gli IBM RISC, i SUN, gli HP, i Silicon Graphics e altri. L'uso dei mainframe come server Web è raro, comunque esiste anche del software apposito per questo tipo di macchine. In particolare, le ultime versioni dei mainframe IBM sono alquanto orientate al WWW. Da tenere presente, per contro, che si può realizzare un server WWW già con una macchina da meno di dieci milioni.

Il computer deve poi essere equipaggiato con un sistema operativo: può andare bene qualunque sistema operativo predisposto per la rete, anche Windows95, ma di solito, per avere più affidabilità e prestazioni migliori, si usano sistemi concepiti espressamente per i server di rete; i due attualmente più utilizzati sono Microsoft Windows NT, che si è diffuso soprattutto negli ultimi tempi, e Unix, che è il sistema operativo di Internet per eccellenza (di Unix vengono usate le varie implementazioni: AIX della IBM, Digital Unix della Digital, SunOS e Solaris della Sun, HP-UX della HP, e Linux, che è addirittura gratuito e sta ottenendo sempre maggior successo per le sue qualità); si possono trovare anche MacOS (sui server Apple Macintosh) e Novell Netware.

Infine ci vuole un programma specifico per supportare il protocollo HTTP: questo programma è detto anch'esso - come il computer - server Web. I più diffusi server Web sono probabilmente quelli della NCSA, della Netscape e della Microsoft. Esiste un notevole numero di server Web disponibili gratuitamente, come ad esempio NCSA e Apache. Il server Web ha, come ovvio, il compito di rispondere alle richieste dei client Web, secondo le modalità previste dal protocollo HTTP.

3.2.6.5.1 **Applicazioni interattive e interrogazione di basi dati: CGI, ASP, PHP, JDBC, Servlet, cookies**

A questo punto, la lista sembrerebbe finita, ma non è così: il client e il server Web possono fare molte cose, ma non tutte quelle che a volte sarebbero desiderabili. In particolare, a volte sarebbe utile poter interagire, sempre restando in ambito WWW, con programmi diversi dal solito server Web: ad esempio, abbiamo visto che i documenti HTML possono contenere i forms, che sono moduli in cui l'utente può inserire dei dati. È chiaro che questi dati possono essere utilizzati per gli scopi più diversi, ad esempio iscriversi a qualcosa, comprare qualcosa, compilare un questionario, ricercare in una banca dati, ma d'altra parte il client non può farne alcun uso se non passarli al server Web, che a sua volta non è certamente in grado di svolgere, con questi dati, qualunque compito che possa essere desiderato dall'utente (ricordiamo che un server Web è in ultima analisi un file server, cioè ha la funzione di trasferire file da un computer all'altro). Il problema è quindi quello di realizzare, in ambiente WWW, applicazioni interattive tra le quali hanno particolare importanza quelle che prevedono l'accesso a database, e che possono essere usate nei contesti più diversi, dagli acquisti online alla consultazione dei cataloghi di biblioteche.

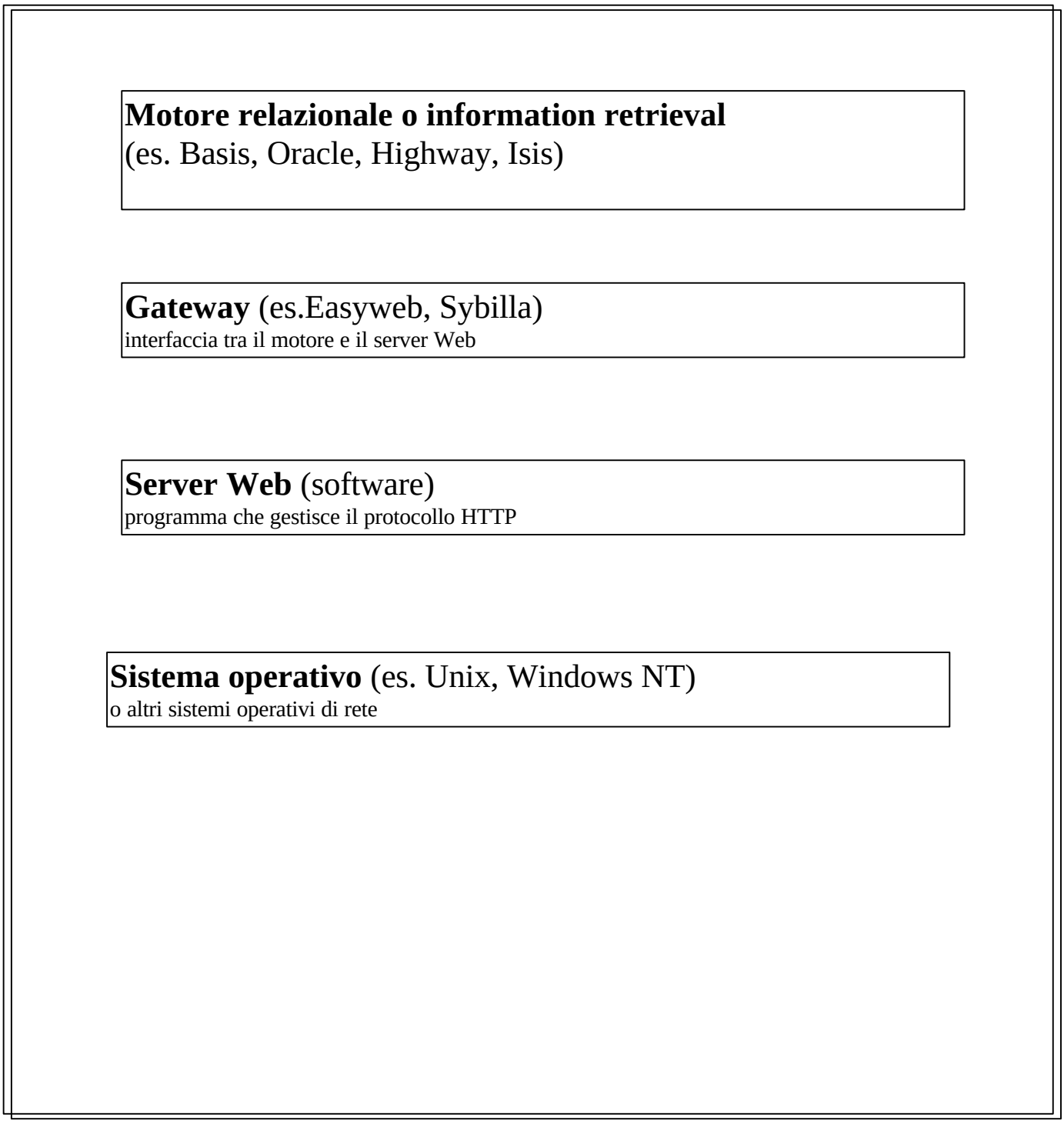
È dunque necessario un programma che da una parte possa interagire con il server Web, e dall'altra possa fare operazioni diverse: per consentire la realizzazione di tali programmi è stata creata una specifica detta Common Gateway Interface (CGI), che determina le modalità di interazione tra il server Web e l'altro programma, che viene detto gateway perché permette la connessione tra l'ambiente HTTP e l'esterno (naturalmente non bisogna confondere questo gateway con il gateway tra due reti).

I programmi CGI si trovano normalmente sulla stessa macchina del server Web, e di solito sono scritti in linguaggio Perl, Tcl, C o Python, ma di per sé possono essere scritti quasi con qualunque linguaggio di programmazione. Si tratta di programmi a tutti gli effetti, che quindi possono effettuare direttamente qualsiasi operazione, dalla più semplice alla più complessa (a seconda della bravura di chi li ha scritti). Tuttavia, quando le operazioni da effettuare sono particolarmente impegnative, soprattutto ricerche complesse su banche dati, si preferisce scrivere un programma che non effettui direttamente tali operazioni, ma che comunichi da una parte con il server Web, e dall'altra con programma specializzato (ad esempio un programma di gestione

di database relazionali) già reperibile in commercio. Non è peraltro impossibile che un programma sia stato scritto in modo da potersi interfacciare direttamente con il server Web senza bisogno del gateway.

Quanto sopra descritto estende immensamente le potenzialità del WWW perché consente di effettuare, sempre in ambiente WWW, qualunque operazione resa possibile dai programmi con cui si può interfacciare il server Web.

Ecco uno schema degli strati di software presenti su una macchina server Web in cui sia stato realizzato quello che abbiamo descritto **(ATTENZIONE: chi legge questo documento con Winword deve passare in visualizzazione layout di pagina per vedere lo schema, altrimenti vedrà solo una pagina bianca)**:



A volte la struttura può essere anche più complessa. È particolarmente interessante il caso in cui viene usato un server WAIS (Wide Area Information Server), che è un software che consente di interrogare in modo trasparente (cioè con una interfaccia unica), banche dati diverse anche su host diversi in rete geografica tramite il protocollo di search and retrieval ANSI Z39.50: in questo caso si può utilizzare un gateway WWW-Wais e un altro gateway Wais-motore di gestione del database (con una architettura a sei strati invece che a quattro).

Come accennato, i programmi CGI non servono solo a interrogare banche dati, ma anche per molti altri scopi: un impiego molto diffuso è quello per visualizzare in un documento HTML il contatore degli accessi al documento stesso.

La tecnica CGI ha alcuni limiti, ad esempio il fatto che il programma CGI deve essere rieseguito ad ogni nuova richiesta per la quale esso è necessario, con evidente discapito delle prestazioni. Ci sono in realtà altre tecniche per ottenere gli stessi risultati, ma per ora non sempre hanno la generalità di applicazione delle specifiche CGI, per cui - ad esempio - funzionano solo con certi server WWW e non con altri. Tra queste tecniche, negli ultimi anni ha avuto notevole successo quella delle **Active Server Pages** della Microsoft. Si tratta innanzitutto di una tecnica proprietaria, ossia esclusiva della Microsoft, che funziona solo sui server dotati del sistema operativo Windows NT (ora anche Windows 9x) e dei server WWW della Microsoft (non c'è invece alcun particolare vincolo per quanto riguarda il client), per cui è inutilizzabile in tutto il vasto ambiente dei sistemi Unix e comunque con tutti i prodotti non Microsoft. Questa tecnologia è orientata essenzialmente all'accesso ai database, e prevede l'uso di particolari pagine HTML che devono avere l'estensione .asp invece delle solite .htm o .html. Queste pagine, oltre a normali tag HTML contengono anche degli script, cioè dei programmi scritti normalmente in VBScript o Jscript (si tratta di linguaggi di programmazione concepiti originariamente per l'uso sui client, e che sono descritti più in dettaglio nel paragrafo successivo). Il server, ogni volta che viene richiesta da un client una pagina con l'estensione .asp non la invia immediatamente, ma esegue prima gli script contenuti nella pagina e poi invia al client l'output risultante dall'esecuzione. La tecnologia ASP, poi, mette a disposizione i mezzi per accedere facilmente a database (ad esempio tramite query SQL) tramite i linguaggi sopra citati. Più in dettaglio, l'architettura ASP prevede che il server comunichi, tramite il protocollo ISAPI, con il modulo ASP vero e proprio, il quale - a differenza di quanto avviene con i programmi CGI - non viene attivato come un processo separato, per cui dovrebbe conseguirne un incremento delle prestazioni. All'interno del modulo ASP poi si distingue uno *scripting host* che ha il compito di sovrintendere all'esecuzione degli script e comunica con uno o più *scripting engines*, che sono i veri e propri interpreti degli script. Gli *scripting engines* possono essere di terze parti, e quindi consentire l'uso di qualsiasi linguaggio (ad esempio ne esiste già uno per il Perl), e anche di più linguaggi nell'ambito di uno stesso documento ASP. Notiamo infine che l'ASP è una implementazione dello standard COM (proposto, guarda caso, dalla Microsoft) appunto per architetture ad oggetti - ossia che utilizzano molteplici componenti anche distribuiti sulla rete e interfacciati secondo modalità predefinite - da utilizzare in ambito WWW. Per lo stesso scopo vengono proposti anche altri standard, tra cui il principale è attualmente il CORBA, che non è proprietario ed è sostenuto da altri produttori tra cui Netscape.

ASP è quindi una tecnologia assai interessante e ingegnosa, che ha però il notevole limite di essere legata a prodotti di uno specifico produttore. Può darsi che in futuro, forse anche grazie alle estensioni del HTML di cui si diceva in precedenza, si arrivi ad uno standard più universale per la creazione di applicazioni complesse ed in particolare per l'accesso ai database.

Esiste una tecnologia concettualmente simile ad ASP, che funziona sotto Unix ma potrebbe essere implementata in qualunque ambiente, non è proprietaria e il cui software è distribuito gratuitamente: si tratta del linguaggio PHP, che può funzionare come un normale CGI oppure come componente del diffusissimo Web server Apache, che è distribuito gratuitamente (http://www.apache.org). Col PHP si possono scrivere dei programmi che vengono inclusi nella pagina HTML, per cui, come nell'ASP, l'utente vede il risultato dell'esecuzione del programma. Normalmente il PHP viene utilizzato per l'accesso ai database, che viene effettuato utilizzando le API dei database server supportati, che sono già molto numerosi, e poiché vengono interrogati via rete non devono necessariamente girare sotto Unix (tra questi è infatti incluso anche Microsoft SQL Server, che gira sotto NT).

Altre soluzioni prevedono la creazione di programmi Java (v. seguito per maggiori informazioni sul Java) che accedono ai database tramite le API JDBC (Java Data Base Connectivity) e sono divisi in diverse componenti, alcune delle quali vengono scaricate dal browser WWW ed eseguite sulla macchina client, mentre altre vengono eseguite lato server.

Tra le tecnologie basate su Java recentemente è emersa quella dei **servlet**, che sono programmi Java creati utilizzando delle particolari librerie in modo da facilitare la scrittura di procedure destinate all'esecuzione in ambiente WWW lato server e possono interagire col Web server in due modi: o utilizzando un web server che supporta in modo nativo i servlet, come il Java Web Server della Sun (a sua volta scritto in Java) oppure attraverso un componente software detto Java Servlet Engine che funge da interfaccia tra un web server e i servlet. Attualmente esistono numerosi Servlet Engine per i principali web server su diverse piattaforme.

È importante notare che tutte le soluzioni basate su Java sono intrinsecamente portabile su diverse piattaforme perché lo è appunto il Java in se stesso.

Si deve comunque osservare che le differenti ed incompatibili tecnologie di cui si parlava si collocano solo sul lato server, perché in ogni caso quello che viene inviato al client è normale HTML, per cui l'utente Internet non deve avere per questo aspetto alcuna preoccupazione e può accedere a pagine ASP così come attivare programmi CGI o Servlet con qualunque browser.

Abbiamo detto che di solito gli utenti non hanno l'occasione di mettere le mani su un server WWW; tuttavia anche su una qualsiasi macchina con Windows 9x si possono facilmente installare server Web, FTP, di posta elettronica, proxies e altro che possono poi colloquiare con i rispettivi client anche se questi ultimi girano sulla stessa macchina (in cui deve essere installato e configurato il TCP/IP). Se si vuole, si possono anche installare dei CGI. Lo stesso, ovviamente può essere fatto con Linux. Tutto ciò è molto utile per chi vuole studiare meglio il funzionamento dei server.

Una tecnica legata ai siti Web interattivi è quella dei cookies (biscottini), che sono piccoli file di testo inviati dal server e contenenti varie informazioni di interesse del server stesso, che poi possono a loro volta essere di nuovo inviati dal client al server. Dal punto di vista tecnico i cookies rispondono ad una esigenza molto importante, e cioè quella di tenere traccia della sessione di lavoro dell'utente (cosa impossibile da realizzare direttamente attraverso un protocollo senza connessione come http), evitandogli ad esempio di immettere nuovamente la sua password ad ogni pagina prelevata da un sito. Nell'ambito dei servizi commerciali però i cookies permettono anche di tenere traccia della navigazione dell'utente per cercare di ricostruire le sue preferenze, propensioni al consumo ecc., cosa che a molti, non a torto, riesce sgradita, anche perché i cookies contengono stringhe di testo leggibili, ma di solito (non sempre) assolutamente incomprensibili se non a chi abbia conoscenza dettagliata del funzionamento dell'applicazione che ne fa uso (il che naturalmente non è quasi mai il caso), per cui non si sa quasi mai bene quali informazioni vengano effettivamente trasmesse.

I cookies non possono essere letti da tutti i server ma solo da quello che lo ha inviato o al massimo da altri server dello stesso dominio. Questo però non risolve il problema della privacy perché (a parte la possibilità che le informazioni raccolte da un server vengano passate ad altri) ci sono server che fanno servizio per molti soggetti, che quindi condividono tutte le informazioni raccolte: è il caso soprattutto dei server che visualizzano banner pubblicitari, che appaiono in un gran numero di pagine, ma non originano dallo stesso server che ospita la pagina, bensì da server specializzati, che fanno servizio per ditte diverse acquisendo dati utili per il marketing.

I browser permettono di bloccare i cookies, oppure possono richiedere all'utente conferma se accettare o no ogni cookie (cosa, questa, che può diventare noiosa perché alcuni siti ne inviano a decine). Il problema è però che se si rifiutano i cookies alcuni siti possono diventare solo parzialmente utilizzabili o anche inutilizzabili del tutto (ad esempio non permettendo di andare oltre la pagina iniziale). Ci sono diversi software che permettono quanto meno di monitorare o gestire i cookies (ad esempio impostando delle regole su quali accettare e quali rifiutare), in modo da cercare di tenerli sotto controllo. Alcuni di questi software sono in effetti dei proxies che accettano i cookies, evitando quindi gli inconvenienti che derivano da rifiutarli, ma poi li eliminano evitando che possano essere letti dopo la fine della sessione di lavoro. Lo stesso effetto si ottiene cancellando a mano tutti i cookies dal proprio hard disk: dopo questa operazione il computer apparirà ai servi come se si collegasse per la prima volta.

Ricordiamo per finire che non tutti i cookies hanno questi scopi un alquanto sgradevoli: molti rispondono effettivamente all'esigenza tecnica per la quale sono nati, e cioè quella di mantenere lo stato della sessione, per cui si possono trovare anche siti privi di qualsiasi carattere commerciale.

3.2.6.6 **I linguaggi Java e Javascript**

I programmi CGI possono fare molte cose ma non sono la soluzione ideale per ogni uso. In particolare, hanno l'inconveniente di dover girare sul server, per cui non si possono installare troppi programmi CGI per moltissimi usi diversi, in quanto ciò aggraverebbe il carico di lavoro del server, con conseguente degrado delle prestazioni.

Si pensi inoltre all'uso di un programma per validare l'input dell'utente: usando un programma CGI i dati dell'utente dovrebbero comunque andare al server, che avrebbe sempre il carico dell'elaborazione necessaria, anche quando i dati si rivelassero da scartare.

In altri casi non è possibile eseguire il programma sul server, perché si dovrebbe intervenire su operazioni gestite dal client, ad esempio la grafica.

D'altra parte, far eseguire i programmi sul client comporta dei problemi: innanzitutto l'utente dovrebbe disporre di un esercito di programmi, molti dei quali servirebbero solo occasionalmente. In secondo luogo, ci vorrebbe una versione di ogni programma per ogni diverso tipo di computer e sistema operativo, proprio all'opposto degli obiettivi del WWW, uno dei quali è la trasparenza del sistema informativo rispetto alle particolarità dei prodotti hardware e software utilizzati. Inoltre i responsabili dei siti WWW non potrebbero mai essere sicuri del fatto che chi consulta il loto sito disponga del software necessario.

Per risolvere questi problemi è stato ideato il linguaggio Java, che sta trovando sempre maggiori applicazioni.

Il Java è un linguaggio di programmazione completo (molto più complesso del HTML), derivato dal C++, e progettato per essere indipendente dalla macchina (cioè nessuna istruzione del Java fa riferimento alla particolari caratteristiche di un certo computer o di un certo sistema operativo). Un programma in Java, detto applet, una volta scritto, viene compilato per generare un file in uno speciale formato detto bytecode, che non è direttamente eseguibile e che viene tenuto sul server. In un documento HTML può poi essere inserito un riferimento ad un applet Java. Quando un client richiede il documento, riceve insieme ad esso anche il bytecode dell'applet: a questo punto, se il client supporta il Java, entra in funzione un interprete che esegue il bytecode. L'interprete viene detto JVM = Java Virtual Machine (macchina virtuale Java), ed ha lo scopo di isolare completamente l'esecuzione dalle specificità dell'ambiente operativo (è ovvio che invece la JVM è specifica dell'ambiente, per cui una JVM per Unix non funziona sotto NT, e una per la CPU Alpha non funziona con la CPU PowerPC). In questo modo un programma Java può essere eseguito su qualunque client dotato della JVM, senza aggravare in alcun modo il carico del server.

Gli applet Java possono servire a molti scopi, ad esempio a validare dati e ad effettuare qualunque elaborazione che non debba necessariamente essere eseguita sul server. Spesso vengono utilizzati per creare effetti grafici che non sarebbero possibili con il solo HTML e a rendere i documenti HTML più interattivi.

L'utente non deve preoccuparsi dell'esecuzione degli applet Java: se il suo client WWW supporta il Java, essi vengono automaticamente eseguiti quando il documento HTML li richiede.

Si può dire che quasi tutti gli attuali client WWW supportino il Java, e comunque questo vale per i due più diffusi, cioè Explorer e Netscape.

Java, comunque, è un linguaggio di programmazione completo, che può servire anche per creare applicativi a sé, non destinati all'esecuzione in ambito WWW. Il vantaggio di un programma scritto in Java è che funzionerà su qualsiasi computer e sistema operativo per i quali sia disponibile una JVM.

La home page ufficiale del Java, attraverso la quale si possono trovare tutte le informazioni sull'argomento è: http://www.javasoft.com.

Malgrado la somiglianza del nome, non bisogna confondere Java e Javascript. Quest'ultimo è un linguaggio simile al Java ma più semplice, utilizzato per scrivere piccoli programmi che si trovano direttamente nel documento HTML, e che servono agli stessi scopi degli applet Java. Questi programmi, detti *script*, vengono eseguiti dal browser senza l'uso della JVM. Attualmente le specifiche di Javascript, promosso soprattutto dalla Netscape, sono ancora in corso di definizione, e il linguaggio è supportato solo da Netscape Navigator e da Microsoft Internet Explorer. La Microsoft, per gli stessi scopi di Javascript, propone VB Script, derivato dal Visual Basic, e supportato per ora solo da Internet Explorer, cui ha poi aggiunto Jscript, una versione di Javascript quasi del tutto compatibile con quella originale. Non bisogna dimenticare che Javascript e VB Script sono prodotti proprietari, molto meno supportati del Java, e che quindi possono rendere certe funzionalità del documento HTML non disponibili a chi non dispone di certi browser.

3.2.7 **WAP, UMTS e l'Internet portatile**

Negli corso del 1999 è emerso prepotentemente un tema del quale peraltro si parlava già da tempo, e cioè l'uso di Internet non attraverso normali PC, desktop o portatili, ma attraverso dispositivi sempre portatili ma più compatti e leggeri, tra i quali in primo luogo i telefoni cellulari. Questi ultimi vengono da molto tempo utilizzati per accedere a Internet tramite un PC portatile utilizzando la rete cellulare invece della rete telefonica fissa per collegarsi al numero telefonico del provider (così come del resto ci sono reti wireless, cioè senza cavi, tra normali computer). Ora invece si vogliono impiegare i telefonini GSM, opportunamente perfezionati, e altri dispositivi di analogo ingombro, detti perciò **handheld** (cioè *tenuti in mano*) per accedere direttamente a servizi Internet, quali la posta e i vari servizi informativi presenti sul Web.

A questo scopo un consorzio di produttori formato da Ericsson, Nokia, Motorola e Unwired Planet ha definito il protocollo WAP (Wireless Application Protocol), che non ha lo scopo di creare una specie di Internet parallela basata sul WAP invece che sugli altri protocolli, ma di interfacciare con Internet i dispositivi handheld. Questo interfacciamento è necessario perché tali dispositivi non hanno le risorse hardware e software per utilizzare direttamente nella loro forma originale le risorse Internet. Ci possono essere peraltro servizi WAP nativi destinati espressamente a questo genere di dispositivi.

Il collegamento tra l'ambiente WAP e Internet avviene attraverso dei gateway che traducono in HTTP le richiede inviate dai dispositivi WAP e le inviano ai web server e poi traducono le risposte dei web server in termini comprensibili ai dispositivi handheld. In particolare, i documenti HTML vengono codificati in due linguaggi appositamente creati per questi dispositivi, e cioè HDML (Handeld Device Markup Language) e WML (Wireless Markup Language). La semplicità di questi linguaggi ha consentito di creare browser molto piccoli (anche 20 kb) in grado di essere facilmente incorporati anche in dispositivi con limitate risorse. Esiste anche un linguaggio di scripting client side, funzionalmente analogo al Javascript, detto Wml script.

L'accesso ad Internet tramite i dispositivi handheld, e soprattutto via WAP è attualmente un tema di gran moda, anche sull'onda del successo dei telefoni cellulari, e nel 2000 i produttori vi si sono buttati con grande impegno, ma sull'effettiva utilità di questi servizi è forse lecito nutrire ancora qualche dubbio, soprattutto perché il WAP è una specie di versione ridotta e semplificata di Internet (anche se vi sono dei siti così pesanti e pletorici che dalla semplificazione hanno più da guadagnare che da perdere). Più promettente sembra la nuova tecnologia UMTS che garantisce una larghezza di banda molto maggiore e quindi ben altra flessibilità e disponibilità di servizi, anche multimediali. Sarà comunque interessante vedere quanti utenti si rivolgeranno a questi dispositivi per effettiva necessità professionale o personale (che in molti casi vi può certamente essere) e

quanti solo sull'onda della moda, con dubbi risultati sotto il profilo economico, della tranquillità personale e anche della privacy.

Informazioni sul WAP si possono trovare nel sito http://www.wapforum.com.

3.2.8 **Servizi e protocolli multimediali**

Come abbiamo già accennato, il problema di gestire il traffico multimediale su Internet sta destando grande attenzione. Si tratta di un problema non banale, perché Internet non è stata affatto progettata per questo tipo di traffico, eppure è necessario utilizzarla, trattandosi dell'unica infrastruttura universale di comunicazione (è ovvio che supportare servizi multimediali in una rete privata progettata ex novo per questo solo scopo sarebbe molto più facile).

Un servizio molto appetibile per gli utenti è il trasporto della voce su reti IP (VoIP = Voice over IP), che permette di fare telefonate utilizzando una rete IP invece di una rete telefonica, e quindi di usare una LAN invece di una rete telefonica interna e - cosa ancora più interessante - di telefonare tramite Internet invece di fare telefonate interurbane o internazionali, con grandissimo risparmio economico. La realizzazione di servizi VoIP può essere realizzata in diversi modi, con o senza l'uso di appositi gateways tra la rete IP e la parte di infrastruttura di comunicazione deputata la solo trattamento della voce digitalizzata. Nella maggior parte dei casi, attualmente VoIP su Internet fornisce risultati piuttosto modesti per l'incapacità della rete di garantire il flusso di dati a velocità costante che sarebbe richiesto per una corretta riproduzione dell'audio. La maggior causa di problemi è la parte di rete che va dalla sede dell'utente alla centrale telefonica (*local loop*): naturalmente questo non riguarda quei fortunati che dispongono di connessioni dirette ad alta velocità, i quali possono comunque incontrare inconvenienti dovuti a rallentamenti in una qualsiasi altra parte di Internet.

Sono stati elaborato numerosi protocolli per la gestione del traffico multimediale, che però non sono ancora molto diffusi. A titolo di esempio, ne citiamo qui alcuni:

- **H.323** è una architettura d'insieme per servizi multimediali che utilizza nel suo ambito diversi protocolli, tra cui i soliti IP e TCP e alcuni degli altri elencati di seguito, e può tra l'altro essere usata anche per VoIP[26]
- **RCP** (Real Time Control Protocol) è un protocollo di livello 4 che ha lo scopo di integrare le funzionalità di TCP e UDP per le specifiche esigenze del traffico multimediale, soprattutto la sincronizzazione tra diverse sorgenti e la costanza nel flusso dei dati.
- **RTP** (Real Time Transport Protocol) ha funzionalità simili a quelle di TCP ma fornisce servizi adatti alla gestione del traffico multimediale, in particolare per monitorare e controllare il ritardo dei pacchetti; è considerato particolarmente adatto all'uso con RSVP (v. seguito)
- **RTSP** (Real Time Streaming Protocol) è un protocollo di livello applicazione (livello 7) che presenta analogie con HTTP; come HTTP serve per trasferire file, ed in particolare documenti composti da diversi elementi, RTSP server per attivare il trasferimento sincronizzato di dati da diverse sorgenti (per esempio l'audio e il video di un concerto). RTSP ha anche il suo prefisso per gli URL, peraltro finora poco o nulla usato, che è rtsp:// quando si appoggia su un protocollo affidabile di livello 4 (tipicamente TCP) e rtspu:// quando si appoggia, sempre a livello trasporto, su un protocollo non affidabile (tipicamente UDP).
- **RSVP** è stato progettato per garantire la prenotazione di risorse e la QoS in reti TCP/IP native; deve essere usato con applicazioni che supportato RTP, e deve inoltre essere supportato sia dalle stazioni al punto finale della connessione che dai router; si tratta di un protocollo che potrà assumere grande importanza in corrispondenza con l'emergere dell'esigenza della QoS

[26] Per una descrizione più dettagliata si veda *[Advanced 1999]*

3.2.9 **Altri servizi**

Oltre a quelli citati, ci sono diversi altri servizi (corrispondenti naturalmente a specifici protocolli) che vengono utilizzati su Internet e sulle reti, e che possono assumere grande importanza:

- **Ping** è un protocollo che riporta informazioni sulla rete: la sua funzione fondamentale è di verificare se un certo host è raggiungibile, restituendo un eco dei pacchetti inviati a quell'host, ma molti programmi Ping permettono di acquisire altre informazioni, ad esempio di conoscere il percorso da fare per raggiungere un certo host (cosa utile per capire meglio l'architettura della rete); ping utilizza il già citato protocollo di messaggistica ICMP
- **condivisione di unità**: viene utilizzata quasi esclusivamente sulle reti locali, per problemi di prestazioni e di sicurezza, anche se non è escluso che possa essere impiegata anche su Internet; appositi protocolli consentono ad un computer di vedere una unità remota (cioè un disco di un computer collegato in rete o una stampante) come una unità locale, secondo le convenzioni del suo sistema operativo; ad esempio, un computer DOS/Windows potrebbe vederla come k: oppure z:, una macchina Unix come /mnt/netdrives/discok ecc.; se le prestazioni della rete sono buone, l'utente non avverte quasi nessuna differenza rispetto all'uso di unità locali[27]; spesso i protocolli di condivisione di unità permettono a ogni macchina di fare sia da client che da server, per cui un utente può vedere come unità le macchine altrui, mentre gli altri possono fare lo stesso sulla sua; tuttavia nelle reti non molto piccole le unità condivise sono essenzialmente quelle di appositi server; la condivisione di unità è molto utile perché permette facilmente a più utenti di utilizzare gli stessi dati (ad esempio lavorare sullo stesso documento o sullo stesso database), e anche di gestire facilmente i salvataggi dei dati, che vengono effettuati dai sistemisti direttamente dai server di rete con dispositivi efficienti ma costosi (come i drive per nastri DAT) che non si possono certamente installare su tutte le postazioni degli utenti (i quali per di più hanno la tendenza a trascurare del tutto i salvataggi); recentemente hanno avuto successo alcune configurazioni specializzate per aumentare facilmente la capacità di archiviazione della rete locale senza dover sempre acquistare dei nuovi server o intervenire su quelli esistenti per aggiungere nuove unità: la fascia alta di queste soluzioni è rappresentata dalle **San** (*Storage Area Network*), che è un insieme di server e dispositivi di immagazzinamento (storage) connessi tramite una rete Fibre Channel in modo da ottimizzare le prestazioni nel trasferimento di grandi quantità di dati, mentre la fascia economica (economica, si intende, nell'ambito dei server di rete) è costituita dai **Nas** (*Network Attached Storage*), che sono computer specializzati per l'uso come file server dotati di grande quantità di spazio disco (centinaia di Gb) e di un sistema operativo semplificato, dovendo servire a un solo scopo, che vengono di solito gestiti via rete (non hanno tastiera e monitor) tramite un browser web[28], per cui, oltre che relativamente economici, sono anche semplici da gestire e affidabili; indichiamo alcuni protocolli per condivisione di unità:
    - **NFS**, usato tradizionalmente in ambiente Unix; in Windows vengono utilizzati a volte dei client NFS per collegarsi a unità condivise su server Unix, mentre l'uso di server NFS in Windows è quanto meno raro; NFS è un protocollo proprietario della Sun, per cui i software NFS sono di solito alquanto costosi
    - **SMB** è un protocollo che più o meno svolge gli stessi compiti di NFS, ed era utilizzato nell'ambiente LAN Manager di Windows; ora però viene utilizzato moltissimo anche con TCP/IP, in particolare nell'applicativo Samba (ambiente Unix), che è gratuito e sta avendo grande successo, soprattutto con Linux, dove è ormai diventato il prodotto più diffuso per la condivisione di unità
- **DHCP** permette di distribuire informazioni di configurazione tramite un server centralizzato, che viene interrogato da un apposito client; è molto utilizzato nelle reti locali per assegnare tramite questo server gli indirizzi IP e la configurazione di DNS e gateway, tuttavia il suo uso potrebbe estendersi anche alla distribuzione di altre informazioni, anche in rete geografica[29]
- **SNMP** è un protocollo che permette la gestione remota dei dispositivi di rete

[27] Si noti che non è escluso che un server cui si collega via Internet usi dei dati o dei programmi collocati su unità di rete visbili però a lui solo, e non all'utente Internet

[28] Dovrebbe essere ovvio, a questo punto, che il software incorporato nei Nas include un server HTTP e dei programmi CGI o di altro genere ma comunque aventi la stessa funzione

[29] Il client DHCP di Win 9x si chiama winipcfg, quello normalmente usato in Linux è dhclient

- **LDAP** permette una gestione unitaria di tutte le informazioni riguardanti una rete
- **X-Window** viene comunemente considerato una interfaccia grafica per Unix (un po' come Windows per il DOS), ma in realtà, anche se di fatto viene utilizzato essenzialmente in ambiente Unix, è un protocollo client-server che regolamenta il colloquio tra un server e un client grafico, finalizzato appunto alla gestione della grafica; il client e il server possono anche trovarsi su macchine diverse, collegate in rete; in realtà, comunque, utilizzare un client X-Window per colloquiare con un server via Internet è quanto meno inconsueto
- **Finger** è un protocollo che permette di sapere chi è collegato ad un certo host in un dato momento
- **NTP** permette di ricevere l'ora esatta da un apposito server (per lo stesso scopo ci sono anche altri protocolli meno precisi e sofisticati)
- **PICS** serve per la selezione e la classificazione delle risorse Internet
- **WebNFS** è una estensione di NFS progettata per consentire ad un client WWW di accedere ad unità remote in modo più veloce e flessibile di quanto non consenta il protocollo HTTP; esso comporta l'introduzione di un nuovo prefisso per gli URL, cioè nfs://; questo protocollo è stato presentato già dal 1996 ma non ha ancora avuto in pratica alcuna diffusione
- **Archie** è un protocollo che, interrogando appositi server con un apposito client, permette di effettuare ricerche sui file presenti nei server FTP (ora però tali ricerche sono possibili anche con interfaccia Web, in taluni casi tramite un gateway WWW-Archie, ma spesso con altri sistemi più perfezionati, per cui Archie ha attualmente una importanza piuttosto scarsa)
- **Z39.50** è un protocollo client-server orientato alla connessione, che viene utilizzato per interrogare basi dati diverse in modo standardizzato e con possibilità di ricerca molto sofisticate; attualmente è molto importante nel campo bibliografico, anche se può essere utilizzato in diversi contesti (altre informazioni sull'argomento si trovano nella dispensa sulle basi dati)
- **controllo remoto**; vi sono applicazioni client-server che permettono, attraverso il client, di prendere il controllo di un computer remoto (che può essere accessibile attraverso una rete TCP/IP, una rete di altro genere o anche un semplice cavo - soluzione, quest'ultima ovviamente non praticabile su Internet) su cui sia stato installato un apposito server; tramite il client si può vedere il desktop del computer remoto, eseguire applicazioni, accedere alle risorse di rete cui quest'ultimo ha accesso e cosi via; l'esecuzioni di applicazioni generiche via rete geografica può porre dei problemi di prestazioni perché tali applicazioni non sono state progettate per questo ambiente, ma il controllo remoto può essere molto utile per la gestione della macchina o per la copia o il prelevamento di file; su Internet è diventato celebre il programma Back Orifice: questo programma (e altri dello stesso genere) è costituito da un server che viene installato di nascosto (ad esempio mascherato all'interno di un altro programma apparentemente innocuo) sulla macchina da controllare, per cui poi attraverso un client si può esaminare tutto il contenuto di tale macchina, modificarlo o distruggerlo all'insaputa dell'utente; tuttavia Back Orifice, al di là dell'uso scorretto con cui ha esordito, è semplicemente un valido programma di controllo remoto, che potrebbe benissimo essere usato in modo lecito (ossia da chi ha il diritto di controllare un certo computer), tanto che recentemente è stato ufficialmente rilasciato proprio come programma di questo tipo, che può essere impiegato sia in rete locale che in rete geografica
- **condivisione file**; vi sono programmi come il celebre Napster e altri, tra cui GNUtella, che consentono la condivisione di file tra numerosi utenti in rete; il principio su cui si basano è che ogni utente rende visibile su Internet una parte delle proprie unità di rete con alcuni file che vuole rendere pubblici; il software di gestione provvede ad indicizzare le informazioni su questi file in modo da renderli accessibili e ricercabili da tutti (questa è una descrizione generale, i dettagli dell'implementazione possono variare notevolmente); come noto, Napster è stato messo sotto accusa perché favorirebbe la violazione del copyright, poiché tra i dati resi pubblici in questo modo vi sarebbero soprattutto brani musicali copiati illecitamente da CD audio o altre fonti; l'accusa, come ben si comprende, non riguarda la tecnologia in sé, ma il suo uso, per cui indipendentemente da come si concluderà la vicenda di Napster, è da prevedere che questo genere di servizi continueranno a svilupparsi

3.3 **Cose utili da fare su Internet**

Abbiamo visto il funzionamento e i mezzi tecnici per utilizzare Internet, mezzi che del resto corrispondono a vari servizi disponibili.

Ora vediamo la cosa più direttamente dal punto di vista dell'utente finale, per descrivere cosa si può fare con Internet e che vantaggi se ne possono ricavare.

Parlando in generale, su Internet si possono produrre ed acquisire tutti i tipi di informazioni che possono essere trattati tramite computer. Ecco qualche esempio concreto di cose che si possono fare:

- ricevere e dare informazioni tramite messaggi di posta elettronica o nei newsgroup
- partecipare a discussioni e dibattiti, sempre con la posta e i newsgroup
- procurarsi testi e documenti di argomento tecnico, filosofico, letterario, giuridico ecc.
- procurarsi informazioni sui prodotti di qualsiasi ditta
- procurarsi informazioni sulle attività di qualsiasi organizzazione
- procurarsi fotografie, disegni, filmati ecc.
- consultare i cataloghi delle biblioteche più importanti del mondo
- mettere online documenti HTML creati da noi
- procurarsi (lecitamente) programmi, spesso gratuiti
- fare nuove conoscenze
- cercarsi la fidanzata o il fidanzato
- mettere a disposizione di altri i programmi scritti da noi
- fare acquisti
- giocare
- .....

Naturalmente chi usa Internet a casa propria può fare quello che vuole, ma chi usa Internet in ufficio deve farlo solo per attività utili al lavoro d'ufficio. Pertanto alcune delle attività elencate sopra, come giocare e cercarsi la fidanzata, difficilmente potranno rientrare nei doveri d'ufficio di qualcuno, e quindi difficilmente potranno essere lecitamente effettuate utilizzando il collegamento Internet della Regione. È comunque utile notare che, per convenzione ormai universalmente accettata, tutti i messaggi che una persona invia a una lista di discussione o a un newsgroup (v. seguito) sono considerati inviati a titolo personale, anche se riguardano il lavoro, salvo che non sia detto espressamente che non lo sono.

Si deve anche ricordare che l'Amministrazione, nei limiti di quanto previsto dalle leggi vigenti, può monitorare l'uso che viene fatto di Internet per poter comprendere se ne viene fatto un uso improprio.

Si deve dire che per poter effettivamente utilizzare molte delle risorse di Internet bisogna sapere l'inglese: per chi non lo sa, potrebbe essere l'occasione per impararlo; del resto, nella maggior parte dei casi, si tratta di un inglese molto semplice e non letterario. Sarebbe invece un errore gravissimo limitarsi alle risorse in italiano, che sono una minima parte di quelle disponibili.

3.3.1 **Come orientarsi su Internet**

Un problema di Internet è che ci sono troppe cose, e può essere difficile individuare proprio quella che fa per noi.

Per individuare la cosa giusta, la prima condizione è avere una qualche idea di quale sia, la seconda è utilizzare i mezzi che effettivamente Internet mette a disposizione, adottando le tecniche appropriate.

Si può anche *navigare* (cioè passare da una risorsa all'altra) più o meno senza scopo, ma questo può essere utile le prime volte, tanto per farsi una idea di quello che c'è, e per fare pratica nell'uso dei browser, ma presto questo diventa noioso e inconcludente. Si deve quindi innanzitutto individuare qualche cosa di utile da fare,

qualche informazione interessante da reperire, e quindi navigare con un obiettivo. Naturalmente non è necessario avere un obiettivo estremamente specifico: ad esempio, potremmo desiderare di curiosare nel catalogo di qualche biblioteca lontana, oppure vedere il catalogo dei mainframe IBM, o fare una discussione di filosofia, o trovare il testo del protocollo HTTP, o cercare testi letterari in formato elettronico, o vedere cosa c'è sulla giurisprudenza in materia sanitaria, o cercare un programma che sostituisca il file manager di Windows che non ci piace, oppure un compilatore gratuito per il linguaggio C e anche un manuale per lo stesso linguaggio.

Se non si sa la URL esatta che si vuole (e di solito anche se la si sa, perché spesso ce ne sono anche molte altre interessanti), è bene individuare qualche sito di riferimento per una certa materia, che serva di base per ulteriori ricerche: per quasi ogni cosa si trova qualche sito del genere.

È pressoché indispensabile prendere nota delle cose interessanti che si trovano, invece di affidarsi alla propria memoria. I browser WWW consentono di creare degli elenchi di URL scelti a piacere (detti bookmark), elenchi che si possono poi usare per collegarsi agli URL che contengono. Si può anche creare una pagina HTML che contenga gli URL, pagina che poi si può aprire con il proprio browser proprio come se fosse su un server remoto. È pero consigliabile gestire questi dati con un database, che è molto più potente e flessibile di una semplice lista di URL: ad esempio si può usare una semplice tabella di Access, o almeno un foglio di Excel. Ricordarsi di associare ad ogni URL una o più parole chiave, per individuare l'argomento, e una breve descrizione.

Se però si parte proprio da zero, ci vuole qualche punto di partenza ancora più generale, come spieghiamo meglio nel seguito.

3.3.1.1 **Portali e liste**

Esistono liste generali di URL, di solito organizzate (più o meno bene) per materia o soggetto, che si possono scorrere per individuare quello che interessa.

Eccone alcuni esempi:

http://ivory.lm.com/~mundie/CyberDewey/CyberDewey.html
http://www.geocities.com/CapeCanaveral/3616
http://www.100hot.com

Il primo URL elencato, CyberDewey, è organizzato secondo la Classificazione Decimale Dewey, utilizzata di solito per il materiale bibliografico, e questo rende molto più agevole l'orientarsi tra le liste. Il secondo, The Internet Index, è invece organizzato per soggetto.

Il difetto delle liste è che per cercare qualche cosa bisogna scorrerle, il che può diventare pesante e complicato se le liste sono molto grandi o se non sono molto bene organizzate.

Quelle che abbiamo citato sono liste **generali,** che abbracciano qualsiasi argomento. Ci sono poi anche liste **specializzate**, che comprendono risorse relative ad un argomento specifico. Gli esempi sono innumerevoli, ed evidentemente non si possono citare qui. Chi ne ha bisogno le può trovare partendo dalle liste generali, oppure dai motori di ricerca, di cui si parla nel prossimo paragrafo.

Più recentemente le varie liste e guide si sono evolute dando origine al concetto di **portale** (dall'inglese *portal*). I portali, che hanno avuto un grandissimo sviluppo, sono siti che si propongono - come dice il nome stesso - come punti di accesso di facile uso all'intera Internet, e tipicamente propongono: indici sistematici di risorse, motore di ricerca, link diretti a servizi informativi di interesse generale, come notiziari, borsa, previsioni del tempo, servizi aggiuntivi spesso gratuiti, tra i quali il più diffuso è l'email. Molti motori di ricerca (v. seguito) e anche provider propongono dei portali, che se ben realizzati possono essere molto utili, purché si tenga conto

che le possibilità offerte dai portali non possono mai esaurire tutte le risorse disponibili su Internet. Tra i portali inoltre ci sono differenze qualitative che spesso si trovano nel tipo di classificazione adottato per le risorse Internet, classificazione che in alcuni casi è veramente dettagliata e rigorosa, mentre in altri è più approssimativa.

Alcuni esempi di portali, oltre a quasi tutti (per non dire tutti) i motori di ricerca che verranno citati in seguito e quindi non ripetiamo qui, sono:

http://home.netscape.com
http://www.dada.it
http://www.dmoz.com
http://digiland.iol.it
http://www.kataweb.it
http://www.infinito.it
http://www.jumpy.it
http://www.terra.com.br

Si tenga presente che si tratta di un settore in grande espansione, per cui appaiono continuamente nuovi siti di questo genere.

3.3.1.2 **Motori di ricerca e subject gateways**

Invece di scorrere le liste, si possono anche fare ricerche su banche dati che contengono informazioni sugli URL disponibili: si possono quindi cercare, ad esempio, tutte le pagine WWW che contengono certe parole. Il risultato della ricerca viene presentato sotto forma di pagina HTML, per cui ci si può collegare immediatamente agli URL trovati.

Ecco i principali motori di ricerca, che contengono riferimenti a molti milioni di URL (il numero è notevolmente variabile da un motore all'altro):

http://www.altavista.com (Altavista)
http://www.yahoo.com (Yahoo)
http://www.excite.com (Excite)
http://www.lycos.com (Lycos)
http://www.hotbot.com (HotBot)
http://www.google.com (Google)
http://www.looksmart.com (Looksmart)
http://cuiwww.unige.ch/w3catalog (CUI W3 Catalog)
http://gnn.com/gnn/wic/wics/index.html (Wic)
http://metasearch.com (Metasearch)
http://pubweb.nexor.co.uk/public/cusi/doc/list.html (CUSI)
http://webcrawler.com/ (Webcrawler)
http://www.eureka.it/~cesare/italynet.html (Italy on the Net)
http://www.lib.umich.edu/chhome.html (Argus ClearinHouse)
http://www.linkstar.com/linkstar/ (Link Stars)
http://www.nerdworld.com/ (Nerd World Media)
http://www.nova.edu/Inter-Links/search/search.html (Inter-Links)
http://www.opentext.com (Open Text)
http://www.shiny.it/sseek/index.html (ShinySeek)
http://www.stpt.com (Starting Point)

Oltre ai veri e propri motori di ricerca, vi sono i **metamotori,** cioè interfacce che permettono di effettuare una unica ricerca su più motori, ad esempio:

http://www.thebighub.com (The Big Hub[30])
http://www.infozoid.com (Infozoid)
http://www.mamma.com (Mamma)

Una citazione merita anche Nlightn (http://www.nlightn.com), anche se non particolarmente noto, perché indicizza non solo URL, ma anche cataloghi di biblioteche e pressoché ogni altra fonte informativa. Alcuni dei suoi servizi però sono a pagamento.

Da ricordare anche i motori/portali specializzati nei siti di un determinato paese, ad esempio:

http://www.arianna.it (Arianna - Italia)
http://www.virgilio.it (Virgilio - Italia)
http://web.de (Web.de - Germania)
http://www.ole.es (Olé - Spagna)

Spesso i siti dei motori di ricerca mettono a disposizione anche liste piuttosto ricche e bene organizzate; in questo si distingue particolarmente Yahoo.

I motori di ricerca in sé e per sé funzionano benissimo, ed hanno dei tempi di risposta molto veloci, anche se devono gestire milioni di accessi al giorno, tuttavia non sono la soluzione di tutti i problemi. Il loro inconveniente è che di solito trovano troppe cose, che è poi impossibile esaminare tutte, e che a volte non solo realmente rilevanti per la ricerca: ad esempio, può accadere che una ricerca abbia in risposta 50.000 documenti, in molti dei quali le parole ricercate ricorrono in contesti diversi da quelli che interessavano. La vera ragione di questo problema non è il gran numero di documenti HTML esistenti, ma il fatto che essi, a differenza di quanto accade, ad esempio, ai libri delle biblioteche, non vengono indicizzati con parole chiave standardizzate, ma vengono ricercati con il testo completo o con il titolo, che usano un linguaggio libero, per cui, una stessa parola può ricorrere con molti significati diversi.

Ciò non vuol dire che i motori di ricerca siano inutili, basta usarli nel modo appropriato, e specialmente seguire questi criteri:

- restringere la ricerca il più possibile, quindi cercare più parole e non una sola (a meno che sia già una parola molto specifica), e provare con diverse combinazioni di parole
- quando vengono restituiti moltissimi documenti, non cercare di scorrere tutta la lista, ma provare subito a connettersi agli URL che sembrano particolarmente interessanti, in modo da raggiungere presto qualche fonte di informazione più specifica (in genere i motori presentano i risultati della ricerca in ordine di rilevanza per l'obiettivo della ricerca stessa, il che è molto utile, anche se non è detto che l'ordinamento corrisponda davvero a quello che interessava all'utente)

Per rimediare a questi inconvenienti vi sono archivi, detti in genere **subject gateways**, che indicizzano a mano e non automaticamente una selezione di siti, normalmente di argomento omogeneo, utilizzando non i termini stessi contenuti nei documenti indicizzati, ma un apposito linguaggio di indicizzazione, ad esempio un tesauro o un sistema di classificazione. Un esempio importante è il SOSIG, specializzato nelle scienze sociali, che corrisponde all'URL http://www.sosig.ac.uk e che utilizza un sofisticato tesauro specializzato. Il prezzo che si paga per i vantaggi dei subject gateways è una copertura minore, dovuta all'impiego dell'intervento umano per la scelta e indicizzazione delle risorse, intervento che può anche determinare inclusioni o esclusioni opinabili o comunque effettuate in base a criteri e interessi che non sono gli stessi dell'utente.

Un punto molto importante da tener presente è che la maggior parte dei motori di ricerca e portali, anche se gratuiti, sono servizi commerciali (questo non vale per i subject gateways, che hanno origine piuttosto nell'ambito accademico e della ricerca) che si mantengono attraverso la pubblicità che visualizzano, attraverso

[30] In precedenza denominato The Internet Sleuth (http://www.isleuth.com)

l'offerta di servizi aggiuntivi a pagamento oltre a quelli gratuiti, e attraverso la raccolta di informazioni utili per il marketing. Questa raccolta avviene attraverso i cookies, che permettono di mettere in correlazione un utente con sue scelte significative dal punto di vista della propensione al consumo (ad esempio i link pubblicitari seguiti) e attraverso la raccolta diretta di dati dagli utenti, fatta attraverso la compilazione online di moduli per la registrazione necessaria, ancorché gratuita, per l'accesso ad alcuni servizi. In tutto questo di per sé non c'è nulla di illegale, e in un certo senso si può considerare come un prezzo da pagare per avere a disposizione gratis numerosi servizi di indubbia utilità. Si tratta però anche di un aspetto che può risultare alquanto sgradevole e pericoloso per la privacy, soprattutto se si pensa alla possibilità che informazioni diverse e ciascuna insignificante se presa a sé vengano incrociate e collegate. La soluzione più radicale per evitare questi inconvenienti è ovviamente quella di non utilizzare i servizi che danno loro origine, ma poiché questo comporta anche la rinuncia a molte cose comode e utili, per non dire quasi indispensabili (come i motori di ricerca), si possono adottare alcuni accorgimenti per ridurre i rischi di schedature più o meno sinistre:

- non utilizzare sempre lo stesso sito, ma molti (ad esempio, una volta Altavista, un'altra Lycos, una terza Kataweb ecc.), in modo da disperdere le informazioni che si lasciano
- rifiutare i cookies (questo però, come abbiamo già visto, spesso compromette la possibilità di usare il sito)
- utilizzare software di controllo e monitoraggio dei cookies
- cancellare i cookies dal proprio hard disk tra una sessione di lavoro e l'altra
- utilizzare dei proxy o dei servizi di navigazione anonima (v. paragrafo sulla sicurezza)
- non cliccare sui link pubblicitari, o farlo solo quando si è realmente interessati
- utilizzare i servizi che richiedono una registrazione solo quando sono realmente utili
- quando ci si registra, inserire solo le informazioni obbligatorie; spesso infatti è richiesto di indicare solo il proprio nome, cognome, recapito, età o poco più (a volte anche meno), mentre altre informazioni su lavoro, hobbies, interessi sono facoltative
- se possibile, utilizzare alternativamente più servizi equivalenti per una stessa attività (ad esempio più indirizzi emai, o più servizi di spedizione fax via Internet ecc.)
- evitare di registrarsi quando vengono richieste troppe informazioni
- leggere il documento, che in genere viene messo a disposizione sul sito, che illustra la politica seguita nel trattamento dei dati; se si tratta di siti italiani, dovrà esserci un riferimento alla legge 675/1996

Interessante è notare che coloro che si collegano a Internet tramite un firewall sono in certa misura più tutelati rispetto a coloro che si collegano direttamente, perché su Internet viene visto solo l'indirizzo del firewall senza possibilità di distinguere quelli dei singoli posti di lavoro. Il firewall peraltro, a meno che non sia impostato secondo criteri di sicurezza particolarmente severi, lascia passare i cookies, e non impedisce di fornire propri dati in occasione della registrazione a qualche servizio o in altri casi. È poi evidente che chi gestisce il firewall può raccogliere in quella sede dati su tutte le attività di coloro che si collegano a Internet tramite il firewall stesso.

3.3.1.3 **Altri servizi**

Un altro tipo di servizio di orientamento che ha avuto una certa diffusione è costituito da quei siti in cui si possono porre direttamente domande ad esperti umani, invece che a software. Può trattarsi di una risorsa da tenere in considerazione, soprattutto per ricerche che non sono facili da condurre con altri mezzi. Tra i numerosi siti di questo genere citiamo:

http://www.help.com
http://www.askme.com

Un altro servizio che citiamo qui perché tutto sommato rientra nella categoria dei servizi di *reference*, anche se non ha direttamente Internet come oggetto, e che è emerso recentemente è quello dei siti che ospitano recensioni di prodotti di vario genere (elettronica, auto, moto ecc.), immesse direttamente dagli utenti. Sono risorse informative interessanti, anche se il livello qualitativo delle recensioni è molto variabile. Alcuni siti di questo genere trattano prodotti di ogni genere, ad esempio http://www.deja.com (che è diventato un sito di

recensioni ma per fortuna ha mantenuto il suo bellissimo e utilissimo motore di ricerca sui newsgroup), altri sono specializzati, come http://www.photographyreview. com, dedicato alle attrezzature fotografiche.

Una certa importanza hanno acquisito i cosiddetti *drive virtuali* (ad esempio http://www.xdrive.com/). Si tratta di siti WWW che mettono a disposizione dell'utente una certa quantità di spazio disco (di solito non superiore a 100 Mb) non per creare un sito Web, ma per immagazzinarvi propri file come si farebbe su una unità disco locale. A questi dati su può accedere solo tramite password, per cui non sono pubblici, ma accessibili solo alle persone che conoscono la password. I dati vengono gestiti con una interfaccia Web che consente di inviarli, prelevarli, cancellarsi, rinominarli ecc. I fornitori di questi servizi, molti dei quali sono gratuiti, assicurano la massima tutela della privacy. I drive virtuali possono essere utili per chi deve accedere ai dati da computer diversi, magari in viaggio, oppure come metodo di salvataggio più sicuro dei tradizionali floppy, e che può comunque essere affiancato ad altri come i nastri ecc. per un supplemento di tranquillità. Si deve tenere presente, però, che - indipendentemente dal livello di riservatezza garantito dal gestore del servizio - se i dati viaggiano in chiaro tra il computer dell'utente e il sito del gestore possono essere intercettati appunto in quella fase.

3.3.2 **Come saperne di più su Internet**

Un vantaggio di Internet è che contiene anche tutte le informazioni su se stessa, dalle più elementari fino al testo di tutte le specifiche tecniche. Di seguito sono elencati alcuni siti che contengono informazioni **tecniche** su Internet (quelli che contengono informazioni utili per la navigazione sono indicati nella sezione 3.3.1):

http://www.internic.net (Internic)
http://www.w3.org (Home page del WWW)
http://www.rfc-editor.org (tutte le RFC)
http://www.internet.com (informazioni varie, link)
http://internettrafficreport.com (traffico su Internet)
http://www.ietf.org (Internet Engineering Task Force)
http://www.amdahl.com/internet/events/inet25.html (Storia e informazioni sulla rete)
http://www.iol.it/internet/index.html (Informazioni tecniche)
http://www.jce.it (Software, provider, documentazione)
http://www.nis.garr.it (GARR-NIS)
http://java.sun.com (Tutto sul linguaggio Java)
http://www.sci.kun.nl/thalia/guide/index.html#page-stats (Gestione di siti Web)
http://WWW.Stars.com (The Web Developer's virtual library)
http://www.browserwatch.com/ (browsers)
http://www.serverwatch.com/ (servers)
http://www.traceroute.org (traceroute)

I primi tre siti elencati sono fondamentali. In particolare, www.rfc-editor.org contiene il testo di tutte le specifiche tecniche relative al TCP/IP e a Internet (questi documenti sono denominati Request for Comments = RFC).

Osserviamo di passaggio che Internet è anche una immensa riserva di informazioni su tutto ciò che riguarda l'informatica. Oltre ai siti delle ditte di informatica piccole e grandi, con tutte le informazioni tecniche sui prodotti, possono essere interessanti i siti dei rivenditori di hardware e software, perché permettono di farsi facilmente una idea complessiva di ciò che offre il mercato. Tra i tanti siti di questo genere ne citiamo due a titolo di esempio:

http://www.warehouse.com (USA)
http://www.chl.it (Italia)

3.3.3 **Liste di discussione**

Passando ad argomenti di interesse più generale, trattiamo di uno dei principali servizi di Internet, cioè le liste di discussione.

Le liste di discussione raccolgono gruppi di persone che si scambiano messaggi di posta elettronica. Alla lista corrisponde un indirizzo di posta elettronica: ogni messaggio inviato a questo indirizzo viene automaticamente inviato a tutti gli iscritti alla lista. Ogni iscritto riceve tutti i messaggi inviati dagli altri iscritti. In questo modo ognuno può comunicare, con un minimo sforzo, con numerose altre persone.

Le liste di discussione sono meno spettacolari del WWW, ma quasi altrettanto utili: spesso permettono di scambiare opinioni con persone di ogni parte del mondo, a volte con veri esperti di una determinata materia, e di ottenere nel giro di poche ore informazioni che altrimenti richiederebbero innumerevoli lettere o telefonate.

Le liste possono essere moderate o non moderate: nelle liste non moderate ogni messaggio inviato alla lista viene senz'altro distribuito ai membri; in quelle moderate viene invece esaminato prima dal moderatore, che in genere si limita bloccare i messaggi fuori argomento o quelli offensivi e simili.

La maggior parte delle liste sono gestite con appositi software (come Listserver, Listprocessor, Majordomo), che permettono anche di ottenere automaticamente informazioni come l'elenco degli iscritti, altre invece sono gestite a mano dal moderatore. Dal punto di vista dell'utente questo fa poca differenza, poiché egli usa comunque il suo client di posta elettronica, esattamente come se comunicasse con una singola persona.

Non citiamo qui alcuna lista in particolare, perché ce ne sono su quasi ogni argomento concepibile. Citiamo invece alcune URL da cui si possono ricavare informazioni sulle liste esistenti:

http://tile.net/tile/listserv/viewlist.html
http://www.neosoft.com/internet/paml/
http://www.liszt.com

Tra una lista e l'altra c'è molta differenza nel numero dei messaggi, che può anche essere molto alto. Per questo non è bene iscriversi a molte liste contemporaneamente: è più consigliabile iscriversi a una per volta, e osservare qual è il traffico di ciascuna, altrimenti si rischia di dover passare la vita a scaricare messaggi di posta elettronica.

È bene fare qualche osservazione sul comportamento che bisogna tenere nell'uso della posta elettronica. Bisogna ricordare, in particolare, che nessuno è obbligato a rispondere ai nostri messaggi, né tantomeno a rispondere immediatamente, per cui se qualcuno non ci risponde non bisogna continuare a tormentarlo con solleciti. Se però l'argomento ci sta a cuore, non c'è nulla di male a sollecitare la risposta una volta o due, anche perché il nostro interlocutore potrebbe semplicemente avere dimenticato di risponderci, o anche aver cancellato per sbaglio il nostro messaggio.

3.3.4 **Newsgroup**

Come abbiamo già accennato, i Newsgroup sono per certi aspetti simili alle liste di discussione, nel senso che sono raccolte di messaggi su un certo argomento, accessibili a tutti i partecipanti al gruppo. Per partecipare ai newsgroup si usa un client apposito, che permette di collegarsi al server delle news, esaminare l'elenco dei newsgroup, selezionarne uno, vedere la lista dei messaggi, leggerli, rispondervi e inviarne di nuovi.

I newsgroup che esistono (decine di migliaia) sono replicati sui vari newsservers sparsi per il mondo. In genere questi server permettono l'accesso solo ai membri o ai clienti dell'organizzazione che li gestisce (tipicamente un provider per i suoi clienti).

I newsgroup sono organizzati in gerarchie, e sottogerarchie, identificate dalla prima parte del nome: ad esempio, la gerarchia *comp* raccoglie i newsgroup che trattano di informatica; all'interno di *comp* si trovano la gerarchia *comp.os* che raccoglie i newsgroup che trattano di sistemi operativi, al gerarchia *comp.ai* sull'intelligenza artificiale, e molte altre. Ci sono molte gerarchie, ma le principali sono probabilmente *comp*, *sci* (scienza), *misc*, e *alt*. Le ultime due comprendono gruppi su qualunque argomento: vediamo se indovinate di che cosa tratta, ad esempio, la gerarchia *alt.sex*. Se non lo indovinate, ecco la risposta: i newsgroup di *alt.sex* sono normalmente di argomento pornografico, e comprendono anche la pedofilia e altre perversioni sessuali: essi, insieme a un insieme di siti Web non piccolissimo in assoluto, ma che rappresenta una percentuale minima di tutti quelli presenti sulla rete, sono i principali responsabili della cattiva immagine di Internet che viene a volte diffusa dai mezzi di informazione. In realtà, chiunque abbia una qualche conoscenza di Internet comprende facilmente che la rete è ben altro che un club di pedofili, poiché il materiale destinato a questo genere di utenti sarà pure abbondante ma incide per una parte piccolissima su tutto quello che è accessibile[31]. L'argomento merita però qualche ulteriore approfondimento soprattutto per le campagne di isteria giornalistica che di tanto in tanto si scatenano sull'argomento, come è avvenuto, ad esempio nell'ottobre 2000 in seguito alla scoperta di un traffico di materiale per pedofili proveniente dalla Russia e che si appoggiava appunto a Internet. Si tratta, per l'appunto, di campagne di isteria giornalistica che non derivano tanto dalla necessità di denunciare e colpire questi crimini contro i bambini, ma piuttosto dall'intenzione di dare di Internet una immagine a forti tinte e del tutto superficiale, e anzi distorta, appunto come amano fare i giornalisti[32], cioè l'immagine di una specie di territorio senza regole in cui avvengono ogni sorta di cose terribili. In realtà Internet non è un territorio senza regole, perché esistono tutti gli strumenti giuridici necessari per colpire i reati commessi tramite l'uso della rete e perché in molti casi i colpevoli lasciano probabilmente più tracce agendo in rete che non passandosi di mano per strada foto di bambini o bustine di droga, oltre al fatto non ci sono prove che su Internet avvengono più cose terribili di quante ne avvengono nel resto del mondo. Ci sarebbero, anzi, motivi per dire che di cose terribili ne avvengono meno: su Internet trovano voce infinite persone e associazioni impegnate a fare del bene nel campo della cultura, della solidarietà, della pace, dello sviluppo e che altrimenti non avrebbero potuto farsi sentire allo stesso modo, mentre quelli che operano il male preferiscono certamente agire nell'ombra piuttosto che in un luogo pubblico come Internet[33]. Ma allora può darsi che le ipocrite campagne dei giornalisti e di quelli che stanno loro dietro abbiano il vero scopo di fare tacere tutte queste voci alternative e non conformiste, prendendo a pretesto i crimini di pochi per imporre la censura a tutti, e non già quello di difendere i bambini, ai quali per lo più nessuno pensa: come mai non si fa altrettanto chiasso per i bambini sfruttati come lavoratori nel Sud del mondo?

A parte questo, possiamo dire che le potenzialità dei newsgroup siano analoghe a quelle della posta elettronica, e a volte sorprendenti: ad esempio, tra i partecipanti al newsgroup *sci.psichology.consciousness* si trovano Marvin Minsky e David Chalmers che sono tra i più celebri studiosi dell'argomento.

A seconda delle preferenze personali si può avere predilezione per le liste di discussione o per i newsgroup. Un vantaggio dei newsgroup è che non intasano la casella di posta elettronica, per cui un newsgroup è probabilmente preferibile ad una mailing list con moltissimi messaggi. Uno svantaggio è che sul server a cui si ha accesso potrebbe non esserci il gruppo che interessa (anche se questo caso è ora piuttosto raro, perché i server raccolgono decine di migliaia di gruppi), e che i messaggi più vecchi vengono eliminati dal server, per cui l'utente non ha modo di recuperarli se non li ha salvati per tempo.

[31] Qui parliamo di pedofilia e comunque di crimini sessuali: come tutti sanno, esistono poi anche numerosi siti a luci rosse che, benché non siano particolarmente degni di elogio, non sembra che rientrino in questi casi, almeno se si giovano dell'opera di persone adulte e consenzienti. Anche questi siti comunque sono solo una modesta parte di Internet.

[32] In uno dei servizi sull'argomento, mentre il giornalista parlava dei siti per pedofili, le immagini mostravano schermi di computer che visualizzavano siti del tutto innocui, come Yahoo e altri: in questo modo il pubblico televisivo che non usa Internet direttamente (che in Italia è la maggioranza) finirà per associare questi siti frequentatissimi - e per estensione l'intera Internet - con l'idea della pedofilia e delle attività illecite; difficile dire se questo fosse proprio l'effetto che si voleva ottenere, o se derivi solo dalla superficialità degli autori del servizio

[33] Tutti i problemi di privacy che vengono segnalati a proposito di Internet fanno capire che la rete, se da un lato può essere sfruttata per molti scopi criminosi, dall'altro non è l'ambiente ideale per chi fa cose che è meglio tenere nascoste

In genere però i motori di ricerca indicizzano anche il contenuto dei newsgroup. C'è anche un motore specializzato in ricerche sui newsgroup, all'url

http://www.deja.com

Si noti che con questi mezzi è possibile ricercare non solo che cosa è stato detto su un certo argomento, ma anche che cosa ha detto una determinata persona in tutti i suoi interventi. Bisogna quindi sempre ricordare che tutto ciò che mettiamo in un newsgroup, o anche in una mailing list, **è pubblico**, e quindi può essere poi conosciuto e raccolto da chiunque: attenzione quindi a non inserire cose che si vorrebbe mantenere private, oppure messaggi di contenuto offensivo o calunnioso.

3.3.5 **Scaricare programmi e dati**

Il trasferimento di file da un host di Internet sul proprio computer è una attività che può essere di grande interesse: infatti in questo modo ci si può procurare programmi (anche distribuiti gratuitamente) e informazioni utili, contenute ad esempio in file di testo o di immagini.

Come abbiamo accennato, è possibile salvare sul proprio computer, sotto forma di file, messaggi di posta elettronica, news e pagine HTML. Qui però ci occupiamo dell'attività espressamente rivolta al trasferimento di file attraverso il protocollo FTP.

Il client FTP permette di accedere facilmente ai server FTP, scorrendone le directory in modo simile all'uso del File Manager di Windows per esaminare il contenuto del proprio disco.

Ma che cosa si trova sui server FTP ? Non bisogna credere che si possano scaricare liberamente programmi commerciali, anche se ci sono server illegali che lo permettono: si possono invece scaricare programmi gratuiti e programmi shareware. I programmi gratuiti sono quelli che, per vari motivi, non vengono distribuiti commercialmente, e vanno dalle piccole utility fino ai sistemi di sviluppo (come il compilatore GNU C) e ai sistemi operativi completi, come il già citato Linux e FreeBSD, che forse ne sono l'esempio più clamoroso. I programmi shareware possono essere liberamente utilizzati a scopo di valutazione, ma poi si deve pagare la licenza all'autore, altrimenti l'uso diventa illegale. Bisogna sempre prestare attenzioni alle condizioni d'uso dei singoli programmi: **si tenga presente che alcuni sono gratuiti per uso personale, ma soggetti al pagamento se utilizzati in ambito aziendale o della pubblica amministrazione**.

Ecco un elenco dei maggiori siti FTP; a quelli con il prefisso ftp:// si accede direttamente con una sessione ftp, cioè viene subito presentata la directory come se si trattasse di una unità locale, mentre quelli con il prefisso http:// hanno una interfaccia Web che rende più comoda la consultazione; tutti i server elencati contengono materiale dei più diversi generi

ftp://cnuce_arch.cnr.it (CNUCE - contiene Linux)
ftp://ftp.cc.ukans.edu (University of Kansans)
ftp://ftp2.cc.ukans.edu (University of Kansans)
ftp://ftp.cdrom.com (Walnut Creek - contiene Linux e FreeBSD)
ftp://ftp.cnr.it (CNR)
ftp://ftp.funet.fi (Funet - contiene Linux)
ftp://ftp.ncsa.uiuc.edu (NCSA)
ftp://ftp.pht.com (Pacific HigTech - contiene Linux)
ftp://ftp.sunet.se (Sunet)
ftp://sunsite.unc.edu (University of North Carolina - contiene Linux)
ftp://tsx-11.mit.edu (MIT - contiene Linux)
ftp://ftp.winsite.com (Winsite - software per Windows)
ftp://ftp.sunsite.com (una delle raccolte più vaste)

Poiché la quantità di file disponibile sui server FTP è immensa, ci sono dei motori di ricerca specializzati per la ricerca di programmi; eccone tre che funzionano molto bene

http://www.davecentral.com
http://www.freshmeat.net
http://www.winfiles.com
http://www.download.com
http://www.hotfiles.com
http://www.filez.com (ricerca sui nomi di file)
http://ftpsearch.lycos.com
http://www.volftp.mondadori.co
http://www.jumbo.com
http://www.shareware.com

Da ricordare anche che è possibile scaricare file anche dai siti delle compagnie informatiche, che però in genere non hanno raccolte generali, ma software di loro produzione, utilities, aggiornamenti, drivers, documentazione e anche prodotti completi. Anche in questi casi, bisogna sempre leggere attentamente le condizioni e limitazioni per l'uso dei singoli prodotti.

Non basta però fare attenzione a non incorrere in violazioni di copyright: si deve infatti tener presente che alcuni programmi, se vengono installati in modo improprio o se presentano dei malfunzionamenti, possono compromettere gravemente il funzionamento del computer e per di più - se questo avviene in ufficio - intralciare gravemente il lavoro anche di diverse persone. I programmi che andrebbero installati con estrema cautela, perché pericolosi se difettosi, male installati o male utilizzati sono soprattutto i seguenti:

- i programmi che manipolano le partizioni e il boot sector: sono potenzialmente i più pericolosi perché possono anche portare alla perdita di tutto il contenuto dell'hard disk
- gli editor di disco a basso livello: permettono di modificare a mano aree non accessibili come file, cioè la tabella delle partizioni e il boot sector; un minimo errore può rendere inutilizzabile l'intera unità
- i programmi che permettono di controllare l'accesso al PC o particolari risorse tramite password o altro: possono rendere l'accesso impossibile a tutti, anche a chi ha stabilito le limitazioni
- i programmi di diagnostica, riparazione o deframmentazione del filesystem: possono provocare danni al filesystem invece di risolverli, e quindi determinare la perdita totale o parziale del contenuto dei dischi; **mai installare uno di questi programmi in un sistema operativo diverso da quello per il quale è stato progettato** (quindi mai usare un programma di diagnostica per Windows 3.x sotto Windows 95)
- i programmi che sostituiscono l'interprete di comandi originale del sistema operativo, che è command.com sotto DOS, progman.exe (Program Manager) sotto Windows 3.x e explorer.exe (Explorer, da non confondere con Internet Explorer); possono rendere il sistema instabile o difficile da utilizzare: di solito si può rimediare, ma si provoca comunque una perdita di tempo
- i sistemi operativi: anch'essi possono, oltre a non funzionare, arrecare danni al preesistente contenuto del disco; su Internet sono disponibili molti sistemi operativi**:** Linux e FreeBSD sono di qualità eccellente, ma non sono semplici da installare, per cui gli errori dell'installatore possono determinare innumerevoli guai

Altri programmi sono meno rischiosi: tuttavia prima di utilizzarli su dati importanti è bene fare sempre una copia di riserva di questi dati, per il caso che un malfunzionamento del programma li danneggiasse.

Con tutto ciò, non si vuole dissuadere dal provare programmi diversi dai soliti: anzi, in questo modo si acquisiscono conoscenze che non di rado si rivelano anche utili per il lavoro. Bisogna solo adottare le opportune cautele, ad esempio:

- evitare di installare programmi potenzialmente rischiosi per puro passatempo, ma farlo solo se si prevede di trarne una effettiva utilità

- leggere accuratamente la documentazione, ed accertarsi di aver capito bene gli scopi e il funzionamento del programma; accertarsi anche di avere compreso bene le eventuali incompatibilità hw e sw, la procedura di installazione, i rischi eventualmente già indicati nella documentazione
- se possibile, assumere informazioni da altri utenti, magari utilizzando le mailing list e i newsgroup di Internet; cercare di capire se si tratta di un programma sconosciuto o di un programma noto da anni in tutto il mondo e collaudatissimo: è ovvio che non comporta particolari rischi l'uso di certi programmi scaricabili da Internet utilizzati da lungo tempo senza nessun problema, come i compattatori Pkzip, ARJ, LHArc, Winzip, oppure gli antivirus McAfee; in genere non sono particolarmente rischiosi i software gratuiti diffusi dalle grandi compagnie di software, come la Microsoft, le quali ovviamente non hanno molto interesse a provocare guai a milioni di utenti
- un programma dal funzionamento impeccabile può diventare una bomba in mano a un pasticcione: **non sopravvalutate le vostre capacità di utilizzare programmi complicati o pericolosi !** (ma neanche lasciatevi paralizzare davanti al computer: bisogna prima imparare e poi fare le cose, invece di farle senza avere imparato oppure rifiutarsi sia di imparare che di fare)
- è anche utile tener presente che non tutti i sistemi operativi sono ugualmente vulnerabili dai programmi malfunzionanti o male utilizzati: tra quelli della Microsoft, il DOS è il più vulnerabile, seguito da Windows 3.x, Windows 95 e Windows NT; gli Unix normalmente sono molto sicuri nei riguardi dei programmi malfunzionanti, ma non nei riguardi dei comandi errati, soprattutto se si fa il login come root (amministratore del sistema); ad esempio in Linux e in altri Unix il comando *rm * -r* dato da root nella directory principale elimina tutto il contenuto di tutte le unità !

Un problema a parte è quello dei **virus**, che non sono poi così frequenti (altrimenti sarebbero già saltati per aria quasi tutti i computer del mondo), ma che è pur sempre possibile incontrare, anche con gravi conseguenze. Contro i virus si devono usare due tipi di precauzioni: innanzitutto cercare di scaricare il software preferibilmente da siti noti e ufficiali, che sono certamente più controllati rispetto a questo pericolo; in secondo luogo installando un antivirus, che si può benissimo scaricare da Internet, scegliendolo ovviamente tra quelli lecitamente utilizzabili; ricordarsi di installare anche gli aggiornamenti periodici, altrimenti l'antivirus presto diventa inutile.

Si tenga presente che **l'effetto dei virus si esplica solo quando viene eseguito un programma che li contiene** (può essere un eseguibile binario o anche una macro di Winword o eventualmente un altro script), per cui tutte le notizie di messaggi di posta elettronica che contengono dei virus che si attivano alla semplice apertura del messaggio sono solo invenzioni ! Naturalmente se il messaggio ha un programma eseguibile come allegato, questo può contenere un virus, ma per attivarlo bisogna eseguire il programma, non basta aprire il messaggio.

3.3.6 **Usare il WWW**

Molto su come si usa il WWW è già stato detto, e in questa sezione non si potranno aggiungere cose straordinarie, perché il WWW, a differenza di altri servizi orientati a scopi ben determinati (come FTP, che è orientato solo al trasferimento di file), è una sorta di infrastruttura che può essere utilizzata per le esigenze più diverse.

Cercheremo comunque di illustrare qualche concetto di applicazione generale.

3.3.6.1 **Consultare pagine**

Naturalmente lo scopo immediato dei collegamenti WWW è permettere di visualizzare documenti HTML. Questi documenti vengono spesso anche chiamati pagine WWW. Ciò non deve far pensare a qualche connessione con le pagine stampate in formato A3 o A4: pagina è solo un sinonimo di documento, per cui una pagina può avere qualsiasi lunghezza.

Le pagine possono avere contenuti molto diversi: alcune possono essere importanti soprattutto od esclusivamente per i link, altre per le immagini che contengono, altre per il testo, altre per più di un aspetto.

Quando si trova una pagina interessante, e che magari contiene molto testo che richiede tempo per la lettura, è bene salvarla in un file sul proprio computer, in modo da averla sempre a disposizione, perché nulla garantisce che sarà nuovamente accessibile in futuro: il responsabile del sito Web potrebbe sempre decidere di eliminarla.

I browser salvano la pagina in formato HTML, quindi con tutti i tag che potrebbero rendere difficile la lettura, a meno di non utilizzare poi nuovamente un browser per visualizzare il file salvato (cosa possibile senza alcuna difficoltà). Alcuni browser permettono di salvare solo il testo senza i tag. È anche possibile che il documento ricevuto dal browser via Internet sia già un semplice file di testo, che quindi viene in ogni caso salvato come tale.

Un'altra soluzione è quella di effettuare, con i normali comandi di Windows, il taglia e incolla tra il browser e un'altra applicazione, per esempio Winword.

Per quanto riguarda le immagini, esse possono essere salvate, separatamente dal resto del documento, in un formato grafico, per esempio GIF, per poter poi essere utilizzate con altre applicazioni (**attenzione che non si tratti di immagini coperte da copyright !** in caso di dubbio, consultare il responsabile del sito).

Spesso le pagine contengono un link all'indirizzo di posta elettronica del Webmaster, cioè dell'amministratore del sito: ci si deve rivolgere a lui per problemi o commenti sul sito Web in senso stretto, per esempio link errati, organizzazione confusa, pagine poco leggibili, troppe immagini o troppo poche, ma di solito non per commenti sui contenuti. Per commenti di questo genere spesso vengono presentati link ad altri indirizzi.

Ovviamente non si possono dare indicazioni generali su che uso fare di quello che si trova nei documenti HTML, perché tutto dipende dall'attività per la quale questi documenti sono stati ricercati.

3.3.6.2 **Interrogare banche dati**

Tra le attività che si possono fare attraverso il WWW, c'è anche l'interrogazione di banche dati. Evidentemente le banche dati disponibili possono riguardare qualunque argomento. Tra quelle di interesse più generale (oltre ai motori di ricerca di URL e di software di cui abbiamo già parlato) ci sono i cataloghi delle biblioteche: infatti su Internet si può accedere facilmente ai cataloghi di tutte le più importanti biblioteche del mondo. Ecco qualche URL di riferimento:

http://sunsite.berkeley.edu (Biblioteche, testi elettronici, documentazione, Internet)
http://www.cilea.it/Virtual_Library/ (Biblioteca Virtuale Lombarda, molti link)
http://www.aib.it/ (Associazione Italiana Biblioteche, ricchissimo elenco di link)
http://lcweb.loc.gov (Library of Congress e molti link)
http://library.usask.ca/hywebcat/ (WebCAT, elenco di cataloghi WWW)
http://www.lib.uiuc.edu (Biblioteca dell'University of Illinois, link a cataloghi, testi ecc.)
http://www.nerdworld.com/nw40.html (raccolta di link a biblioteche)
http://www.regione.liguria.it/conosc/biblio/intro.htm (Pagine della Regione Liguria sulle biblioteche)
http://opac.regione.liguria.it/cgi-win/hiweb.exe/a3 (Catalogo delle Biblioteche Liguri)
http://opac.sbn.it (OPAC del Servizio Bibliotecario Nazionale)
http://www.iccu.sbn.it/ (ICCU, con molte informazioni su SBN e altro)
http://www.sba.unige.it (Università di Genova)
http://csisbn1.csi.it (OPAC SBN, molto evoluto tecnicamente)

I primi cinque siti elencati sono particolarmente ricchi e bene organizzati.

Per effettuare le ricerche con profitto è necessario avere qualche nozione generale sulle ricerche online, anche perché ora molte banche dati consentono ricerche estremamente raffinate. Innanzitutto è bene aver presente che più condizioni di ricerca (ad esempio parole) possono essere combinate con l'uso degli **operatori booleani,** come spiegato di seguito:

immaginiamo che A e B siano due condizioni di ricerca, ad esempio due parole che si vogliono ricercare; possiamo associare A e B con i seguenti operatori:

- **and** = A and B significa che devono essere presenti sia A che B
- **or** = A or B significa che deve essere presente almeno uno tra A e B, o anche entrambi
- **not** = not A significa che A non deve essere presente, mentre A not B equivale ad A and (not B) e significa che deve essere presente A ma non B

Fondamentale è capire che l'operatore and restringe la ricerca (la ricerca A and B non può mai restituire **più** dati della ricerca del solo A, e quasi sempre ne restituisce meno), mentre l'operatore or la allarga (la ricerca A or B non può mai restituire **meno** dati della ricerca del solo A, e quasi sempre ne restituisce di più). Per questo la ricerca in or su grandi banche dati deve essere effettuata con molta cautela.

È anche importante la distinzione tra ricerca per stringa intera (stringa significa semplicemente successione di caratteri) e per parole: ad esempio bisogna accertarsi se il programma interpreta la condizione di ricerca "Come sapere tutto su Internet e vivere felici" (che potrebbe essere il titolo di un libro) come un tutto unico, oppure come "come and sapere and tutto ..." o addirittura come "come or sapere or tutto ...".

Per molti può essere istintivo cercare stringhe intere (ad esempio tutto il titolo citato sopra), ma spesso è più produttivo cercare per singole parole combinate opportunamente: questo infatti rende la ricerca indipendente dall'ordine con cui compaiono le parole nel titolo o documento da cercare, e permette di provare più combinazioni di termini, in particolare quando non si è sicuri della forma esatta con cui la stringa compare nel documenti. La ricerca per stringhe intere è utile quando si vuole individuare quella determinata stringa della cui forma si è ben certi.

Bisogna anche abituarsi a provare ricerche alternative quando la prima non ha l'esito soddisfacente. Per esempio, se la ricerca del titolo citato sopra non ha esito, potrebbe essere dovuto al fatto che esso è in realtà "Come sapere tutto su Internet e vivere contenti", oppure che una parola è stata inserita in modo errato nel database. Provare ricerche differenti consente invece di neutralizzare le possibili varianti di una stringa e gli errori di inserimento dei dati.

3.3.6.3 **Posta, agenda e calendario**

Grande sviluppo hanno avuto recentemente altri tipi di servizi interattivi, spessissimo offerti gratuitamente (questo è possibile perché il finanziamento avviene tramite la pubblicità mostrata - peraltro di solito in modo abbastanza discreto - agli utenti). Si tratta sempre di servizi che si basano su dei gateway tra il server web e altri software ad esempio di database, di email o altro.

Citiamo innanzitutto quei servizi che offrono caselle gratuite di posta elettronica consultabili tramite un browser web, ad esempio:

http://www.hotmail.com
http://www.telemail.it

e moltissimi altri associati a portali e/o motori di ricerca (Yahoo, Lycos, Deja ecc.). Qualcuno potrebbe stupirsi di non trovare citato il classico Rocketmail, ma questa ditta è stata assorbita da Yahoo, per cui ora Rocketmail serve solo a coloro che avevano in precedenza un account, ma non è più possibile crearne di nuovi.

Questi servizi hanno il vantaggio di permettere l'accesso alla posta con qualunque browser, senza bisogno di installare e configurare un client di posta, e da qualunque computer collegato a Internet, compresi quelli che si possono trovare in biblioteche o Internet Cafe. Nello stesso tempo hanno lo svantaggio di richiedere il collegamento online per tutto il tempo in cui si lavora con la posta: non si possono, ad esempio, scaricare i messaggi, leggerli offline e preparare le risposte sempre offline, perché non si sta usando un client di posta (naturalmente si possono salvare i messaggi sul proprio computer, ma come semplici file di testo o HTML). I vari servizi si differenziano per velocità, spazio disponibile per conservare i messaggi (che vengono tutti immagazzinati sul server del provider), funzionalità aggiuntive (ad esempio dimensioni degli allegati, possibilità di scaricare posta da altri account, beninteso conoscendone la password, e altro).

Sempre in tema di posta, vi sono molte possibilità di creare e gestire, sempre gratis, le proprie mailing list su qualunque argomento desiderato. In genere il gestore del servizio aggiunge messaggi pubblicitari di due o tre righe in calce a ogni messaggio che circola nella lista. Tutte le attività di gestione e configurazione di una lista avvengono di solito via Web. Alcuni esempi sono:

http://www.egroups.com
http://www.listbot.com
http://www.coollist.com

Altri invece offrono la possibilità di creare dei forum, che sono sostanzialmente analoghi ai newsgroup ma sono distinti dai newsgroup della usenet,e possono essere consultati solo via web collegandosi al server di chi fornisce il servizio. Tra i principali siti di questo genere citiamo

htp://www.insidetheweb.com

Più recentemente sono sorti siti che svolgono funzioni di agenda degli appuntamenti, calendario e simili, ad esempio

http://www.when.com
http://www.anyday.com
http://www.magicaldesk.com

Questo genere di servizi è di diffusione più recente, per cui è verosimile che nel prossimo futuro si aggiungano molti nomi alla lista. Da notare che anche i soliti portali hanno cominciato a offrire calendari e agente. Vantaggi e svantaggi sono analoghi a quelli dei servizi di email.

4. INTRANET

In questo periodo, accanto a Internet, si parla sempre di più delle Intranet. Si tratta di un concetto molto semplice, e cioè dell'applicazione delle tecnologie di Internet su una rete privata, cioè chiusa, non accessibile al pubblico.

Tipico esempio di una rete privata è una rete aziendale (come quella della Regione), che viene utilizzata dal personale dell'azienda o dell'organizzazione che la detiene e non è accessibile al pubblico. Evidentemente nulla vieta di installare su una rete di questo genere un server Web, o un server di posta elettronica SMTP/POP3, affinché le persone che hanno accesso alla rete, dotate degli opportuni cliente, vi si possano collegare per visualizzare i documenti disponibili sul server.

Dal punto di vista strettamente tecnico una intranet non si differenzia da Internet se non per la maggiore semplicità dei collegamenti, mentre i vari server ovviamente funzionano sempre allo stesso modo ed effettuano le stesse operazioni. Dal punto di vista organizzativo, le intranet sono interessanti soprattutto perché permettono l'uso dell'interfaccia WWW, che è particolarmente comoda ed intuitiva. Attraverso questa interfaccia possono essere messi a disposizione, in una struttura chiara e di semplice utilizzo, documenti di

ogni genere e, attraverso l'uso della tecnologia CGI, banche dati anche diverse (e prodotte con programmi eterogenei) con una interfaccia unica. Inoltre una Intranet viene gestita da un unico soggetto, che quindi può garantire un livello determinato di prestazioni, e con ciò rendere possibili - in ambiente WWW - anche attività molto interattive, come l'inserimento di dati in un database, che su Internet potrebbero essere troppo negativamente condizionate dall'incertezza delle prestazioni.

5. CREARE PAGINE HTML

Questa sezione ha lo scopo di fornire le principali indicazioni a chi deve creare pagine HTML. In particolare, verranno illustrati gli elementi principali del linguaggio HTML, e forniti suggerimenti sul modo migliore di organizzare le pagine affinché siano utili ed efficaci per chi le consulta.

In questa sede si fa riferimento alla versione 3.2 di HTML, anche se la versione corrente è la 4. La versione 3.2 è molto adatta per imparare il linguaggio, mentre il passaggio alla 4 non è particolarmente difficile. Comunque per il testo ufficiale della versione corrente delle specifiche HTML bisogna fare riferimento al sito del W3 Consortium, http://www.w3.org.

5.1 **Creare pagine HTML**

Chi conosce, anche per sentito dire, linguaggi di programmazione come il Pascal e il C, si consoli: HTML è infinitamente più semplice di tali linguaggi. Chi non li conosce, si consoli ugualmente: HTML è molto semplice, e alla portata di tutti, sia pure con un poco di studio e di esercizio.

Innanzitutto diciamo che un documento HTML è un file di testo, senza caratteri di controllo e formattazione (a differenza, ad esempio, di un documento Winword). Un documento HTML è composto da due categorie di elementi: il **testo** e gli **elementi** (elements o tags). Il testo richiede poche spiegazioni: è ciò che deve essere visualizzato letteralmente sulla pagina, per esempio "Benvenuti nel sito WWW della Regione Liguria", "Asino chi legge" o qualunque altra cosa. **Bisogna ricordare che gli a capo del testo vengono ignorati (tranne che nei casi che vedremo in seguito), e devono invece essere indicati al browser tramite appositi comandi**. Gli elementi invece sono i comandi previsti dal linguaggio HTML, che devono essere interpretati dal browser affinché si produca l'effetto desiderato dal creatore della pagina.

Gli elementi hanno due funzioni:

- determinare il formato con cui appare il testo (es. corsivo, neretto, a lista ecc.)
- rendere possibili i link, cioè collegamenti ad altri oggetti, ad esempio altri documenti HTML; un link appare all'utente come un testo o una immagine cliccando sui quali con il mouse il browser richiede il documento all'host appropriato

Gli elementi vengono posti insieme al testo della posizione appropriata (un po' come i comandi punto del vecchio Wordstar). Essi si distinguono dal testo perché sono posti tra parentesi uncinate: <>. La struttura completa di un elemento è: <ELEMENTO ATTRIBUTO=valore [ATTRIBUTO=valore] ...> ... </ELEMENTO>. Al nome dell'elemento possono seguire uno o più attributi che, a seconda del valore ad essi assegnato, modificano - per l'uno o l'altro aspetto - l'effetto del tag. Il valore deve essere posto tra virgolette, che possono essere semplici (') o doppie ("). Se il valore contiene a sua volta virgolette, esse dovranno essere del tipo non utilizzato per racchiudere il valore. L'elemento ha efficacia finché non si incontra la chiusura (appunto </ELEMENTO>). Alcuni elementi però non prevedono la chiusura.

Evidentemente, a differenza del testo, che è libero, gli elementi devono essere scritti seguendo quanto indicato dallo standard HTML, altrimenti, anche se i browser sono tolleranti e riescono ad ignorare moltissimi errori, ed in particolare ignorano l'assenza di molti elementi indicati come obbligatori dallo standard HTML,

verranno prodotti prima o poi effetti imprevedibili, diversi a seconda del browser, ma certamente non quelli desiderati dal creatore della pagina. **Per questo è assolutamente necessario che le pagine HTML non contengano errori**, mentre non è necessario che siano particolarmente complesse: chi non si sente sicuro può limitarsi a scrivere il minimo indispensabile per dare luogo a una pagina corretta.

Vediamo quindi la struttura generale di un documento HTML:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4 Final//EN">
<HTML>

<HEAD>
<META NAME="qualunque" CONTENT="vario">
<TITLE>Titolo della pagina</TITLE>
</HEAD>

<BODY>
Testo testo
testo
     testo
</BODY>

</HTML>
```

Le righe vuote sono libere: qui sono state inserite nel modo che vedete solo per leggibilità.

L'elemento <!DOCTYPE ...>, che non viene mai visualizzato, è obbligatorio (anche se in realtà i browser ignorano la sua assenza), e può essere utile per i programmi che eseguono operazioni di validazione del documento; esso può assumere forme diverse: nella forma adottata qui, indica un documento conforme alla specifica HTML versione 4 (la più recente).

L'elemento <HTML> ... </HTML> indica l'inizio e la fine del codice HTML. L'elemento <HTML> è **obbligatorio**, perché in sua assenza i browser potrebbero non riconoscere che si tratta di un documento HTML e prenderlo per un semplice file di testo, visualizzando così elementi e testo assieme.

Segue l'**intestazione**, identificata dall'elemento <HEAD> ... </HEAD>. Per la verità, l'omissione di questo elemento viene ignorata dai browser. Il contenuto dell'intestazione è comunque molto importante. Tra i numerosi elementi utilizzabili per l'intestazione (che si possono inserire anche in assenza dell'elemento <HEAD>), i principali sono:

<META NAME="*valore*" CONTENT="*valore*">; questo elemento è facoltativo, e non viene mai visualizzato. Esso serve a contenere informazioni **sul** documento di cui fa parte, e viene utilizzato, tra l'altro, da alcuni programmi che indicizzano i documenti per renderli ricercabili dai motori di ricerca di cui abbiamo parlato. Per la precisione, l'elemento <META> contiene una serie di coppie name-content, i cui valori sono liberi. In certi casi è consigliabile utilizzare l'attributo HTTP-EQUIV a cui assegnare il valore *keyword* (parola chiave), e assegnare all'attributo CONTENT, come valore, dei termini (possibilmente in inglese) che identifichino il contenuto del documento. Il risultato potrebbe essere, ad esempio: <META HTTP-EQUIV="keyword" CONTENT="Liguria,health service,hospitals">. L'esempio, ovviamente, si riferisce ad una ipotetica pagina che descriva i servizi sanitari in Liguria.

<TITLE> ... </TITLE>; questo elemento è obbligatorio e anche molto utile. Il testo inserito in questo elemento viene visualizzato come titolo della finestra del browser, ed inoltre è molto utile per identificare rapidamente il contenuto del documento. Come titolo andrebbero scelte delle espressioni brevi e chiare, non generiche. Ad esempio: <TITLE>Le biblioteche in Liguria</TITLE>

All'intestazione segue il **corpo** del documento, delimitato dall'elemento <BODY> ... </BODY>, che è indispensabile. Il corpo del documento può contenere solo testo, oppure testo ed elementi, o anche solo elementi.

L'elemento <BODY> ha, tra gli altri due attributi importanti:

- BACKGROUND (uso: <BODY BACKGROUND=[nome file immagine]>) serve per identificare una immagine, normalmente in formato GIF, che il browser visualizzerà come sfondo del documento; questo può permettere di ottenere degli effetti molto piacevoli, ma sono necessarie alcune cautele: utilizzare immagini piccole (il browser riproduce automaticamente l'immagine in modo da coprire tutto lo sfondo) per non rallentare il caricamento, non utilizzare immagini che rendano poco visibile il testo o altri elementi del documento; ricordare che questo attributo ha solo una funzione estetica, e quindi non deve essere utilizzato a discapito della funzionalità
- BGCOLOR (uso: <BODY BGCOLOR=#numero esadecimale>) determina il colore dello sfondo, senza che debba essere scaricato alcun file di immagine (per cui, dove possibile, andrebbe preferito a BACKGROUND); il valore dell'attributo è un numero esadecimale, cioè in base 16, (**sempre** preceduto da #) di sei cifre (ricordiamo che le cifre possibili nei numeri esadecimali, oltre a 0-9, sono A, che corrisponde al decimale 10, B=11, C=12, D=13, E=14, F=15); le prime due cifre corrispondono al colore rosso, le seconde due al verde e le ultime due al blu. Il valore 000000, cioè con tutti i colori al minimo, produce il nero, quello FFFFFF, cioè con tutti i colori al massimo, produce il bianco; FF0000 è un rosso puro, 00FF00 un verde puro, 0000FF un blu puro. Ovviamente le combinazioni possibili sono innumerevoli, ed è bene che chi è interessato all'uso di questo attributo faccia direttamente un po' di prove. Si tenga presente che è consigliabile evitare i colori violenti, e che non tutti i browser producono lo stesso effetto: ad esempio Netscape ed Explorer sono impeccabili, ma NCSA Mosaic, anche nella recente versione 3.0, produce effetti sgradevoli con alcuni colori. A proposito poi dei numeri esadecimali, ricordiamo che ogni cifra rappresenta un coefficiente per cui moltiplicare potenze della base (**16, non 10 !**) crescenti da destra verso sinistra. Per esempio, 7B5 corriponde, in notazione decimale, a $(7x16^2)+(11x16^1)+(5x16^0)$ = (7x256)+(11x16)+(5x1) = 1792 + 176 + 5 = 1973.

Aver compreso la struttura generale del documento HTML è la cosa fondamentale: tutti i particolari degli elementi e degli attributi non sono difficili da imparare, e si assimilano con la pratica.

Vediamo ora i principali elementi, **che possono essere utilizzati sia nell'intestazione che nel corpo del documento.**

Ci sono innanzitutto gli elementi che determinano il formato:

- <PRE> ... </PRE> non applica alcun formato al testo, ma mantiene la formattazione (con gli a capo ecc.) che ha il testo di per sé; utile quando si vuole trasferire un file di testo preesistente in un documento HTML senza fare altre modifiche: in questo modo si possono preparare pagine HTML in brevissimo tempo, anche se il risultato di solito non è particolarmente attraente dal punto di vista della grafica
- <P ALIGN=right | center | left> ... </P> indica l'inizio e la fine di un paragrafo; all fine di ogni paragrafo, il browser manda a capo il testo, e alcuni browser inseriscono una riga vuota; l'attributo align determina l'allineamento del paragrafo (a destra, centrato o a sinistra)
- <H1> ... </H1> gli elementi H1 - H6, che hanno la stessa sintassi (ad esempio <H4> ... </H4>) servono per creare intestazioni di varia misura: H1 è molto grande, le altre via via più piccole
- <CENTER> ... </CENTER> è previsto dal HTML 3.2, ma non in versioni più vecchie, tuttavia già da prima era riconosciuto da tutti i principali browser, per cui si può usare tranquillamente; esso centra tutto quanto contenuto tra l'apertura e la chiusura
- <B> ... </B> inizio e fine del grassetto
-
 manda a capo il testo

- <HR SIZE=*n*> disegna una riga attraverso lo schermo; l'attributo SIZE (facoltativo) determina lo spessore della riga; l'aspetto esatto della riga dipende dal browser; l'elemento <HR> permette di ottenere effetti molto eleganti senza dover ricorrere a immagini grafiche
- <IMG SRC="*nome del file*"> elemento di grande importanza, perché permette di inserire nella pagina l'immagine contenuta nel file grafico di cui si indica il nome: è questo elemento che permette di visualizzare, ad esempio, fotografie e disegni di qualsiasi genere; si deve tenere presente che
    - il file dell'immagine dovrebbe essere in formato GIF, a meno che non ci siano ragioni speciali per usarne un altro, perché questo formato viene riconosciuto dai browser, che possono visualizzare direttamente l'immagine; altri formati potrebbero richiedere programmi esterni per la visualizzazione, che quindi sarebbe impossibile per chi non dispone di programmi adatti
    - la presenza di file grafici aumenta di molto la quantità di dati che viene trasferita dal server al client, e quindi il tempo per visualizzare la pagina: per questo bisogna evitare l'**abuso** (non l'uso) di elementi grafici (l'argomento sarà ripreso in seguito)
    - molto importante è l'attributo ALT, che permette di specificare un testo che viene visualizzato se l'immagine non è disponibile o mentre non è ancora stata scaricata (ad esempio: <IMG SRC="genova.gif" ALT="Panorama di Genova">; l'elemento ALT andrebbe **sempre** utilizzato perché consente di utilizzare il documento indipendentemente dalle immagini, persino con un browser a carattere
- <UL> ... </UL>
- <OL> ... </OL>
- <MENU> ... </MENU> gli elementi <UL>, <OL> e <MENU> servono a creare liste puntate o numerate (simili a quelle create da Winword), e quindi sono particolarmente utili ogni volta che il documento deve contenere un elenco; la sintassi di questi tre elementi è uguale: ogni voce dell'elenco è identificata dall'elemento <LI>, e può poi contenere qualunque altro elemento; <UL> sta per *unordered list* (lista non ordinata), in cui ogni voce è preceduta da un punto, <OL> sta per *ordered list* (lista ordinata), in cui ogni voce è preceduta da un numero progressivo o da una lettera), <MENU> è simile a <UL> ma viene reso da qualche browser in modo più compatto: molti però lo rendono esattamente come l'elemento <UL>, per cui non è molto utile utilizzarlo; le liste possono essere nidificate, cioè una lista può contenerne un altra; ecco un esempio dell'uso delle liste:

```
<UL>
<LI>Primo elemento della lista non ordinata
<LI>Secondo elemento della lista non ordinata

<OL>
<LI>Unico elemento della lista ordinata
</OL>

<LI>Terzo elemento della lista non ordinata
</UL>
```

- <TABLE> ... </TABLE> questo elemento serve per creare tabelle, incorniciate o no, che spesso rendono la pagina molto più ordinata ed elegante; si noti che la grafica delle tabelle viene prodotta localmente dal browser, per cui non richiede il trasferimento di grossi file grafici; se la tabella è grossa c'è però bisogno di un certo tempo di elaborazione, soprattutto sui computer più lenti; l'elemento <TABLE> ha l'attributo BORDER, che se presente determina la visualizzazione delle cornici della tabella; all'interno dell'elemento <TABLE> bisogna poi specificare altri elementi che determinano l'impostazione ed il contenuto della tabella:
    - <TH> ... </TH> intestazione della tabella, che appare in testa alla stessa; l'attributo COLSPAN permette di specificare su quante colonne della tabella l'intestazione si deve estendere; per esempio, se abbiamo una tabella con quattro colonne, e vogliamo che l'intestazione si estenda in larghezza su tutte e quattro useremo la forma: <TH COLSPAN=4>Tabella molto interessante</TH>

- <TR> ... </TR> indica l'inizio e la fine di una riga della tabella; all'interno di <TR> viene inserito l'elemento
    - <TD> ... </TD> che delimita i dati di ciascuna cella della tabella; <TD> può contenere semplice testo, immagini, link ecc.

Ecco un esempio di tabella

```
<TABLE>
<TH COLSPAN=2>Tabella di esempio</TH>
<TR>
<TD>Prima cella della prima riga</TD>
<TD>Seconda cella della prima riga</TD>
</TR>
<TD>Prima cella della seconda riga</TD>
<TD>Seconda cella della seconda riga</TD>
</TR>
</TABLE>
```

- e sono due nuovi elementi introdotti nella versione 3.2 di HTML; i nomi stanno rispettivamente per *superscript* e *subscript*, e il loro effetto è intuitivo: il testo compreso in questi elementi appare rispettivamente un poco più in alto o un poco più in basso del testo normale. Interessante è il fatto che questi elementi possono essere annidati per disporre il testo per così dire "a scaletta"; ad esempio:

```
Testo normale
<SUP>pi&ugrave; in alto
   <SUP>ancora pi&ugrave; in alto</SUP>
un po' pi&ugrave; in basso </SUP>
di nuovo testo normale
```

- <A></A> è un tag veramente fondamentale, perché è quello che permette di gestire i link ad altri documenti, o a specifici punti all'interno dello stesso o di altri documenti. Il principale attributo di <A> è HREF, il cui valore è appunto il documento a cui il link rimanda. Il valore può essere qualunque URL, e quindi il nome di un file HTML, di un file di testo, di una immagine, o di qualunque altro file. Il testo tra l'apertura e la chiusura del taf, cioè tra <A> e </A> diventa quello visualizzato dal browser come link cliccabile. Se il valore è preceduto dal carattere #, allora si riferisce ad uno specifico punto all'interno di un documento (se non è specificato il nome del documento, allora è sottinteso il riferimento a quello corrente). Questi punti a cui può rimandare un link devono essere stati predisposti tramite l'uso di <A> con l'attributo NAME, il cui valore è un nome qualsiasi a cui far riferimento nell'ambito di un attibuto HREF. Un altro attributo molto importante è TARGET, che serve a specificare dove può essere aperto il documento puntato dal link. In particolare, il valore "_blank" determina l'apertura del documento in una nuova finestra del browser, senza chiudere il documento di partenza. Questo è utile soprattutto quando il documento di partenza contiene molti link, per cui l'utente può trovare comodo tenerlo sempre aperto per scegliere via via i vari link. Questa tecnica dovrebbe comunque essere usata solo quano è effettivamente utile, perché può confondere gli utenti meno esperti ed inoltre l'apertura di molte finestre del browser può creare problemi ai computer più lenti. Qualche esempio chiarirà meglio l'uso dell'elemento <A>:

```
<A HREF="http://www.regione.liguria.it">Clicca qui per la home page
della Regione Liguria</A>

<A HREF="http://www.myhost.org/data.htm">Pagina dei dati</A>

<A NAME="Capitolo 1"></A>
....
<A HREF="#Capitolo 1">Vai al capitolo 1</A>
```

```
<A HREF="ftp://ftp.regione.liguria.it" TARGET="_blank">Sito FTP della
Regione Liguria</A>
```

Molto importante è il trattamento dei caratteri speciali, cioè di quelli con il codice ASCII superiore a 127. Se si introduce uno di questi caratteri in un documento HTML il browser visualizzerà invece un puntino o comunque un altro carattere. Bisogna invece utilizzare una codifica particolare, secondo la tabella riportata qui di seguito (è in inglese perché è stata trasferita da un testo in inglese in modo da non doverla digitare tutta a mano). **ATTENZIONE: al codice indicato nella tabella deve SEMPRE seguire il punto e virgola** (ad esempio: Á)

Name	Syntax	Description
Aacute	&Aacute	Capital A, acute accent
Agrave	&Agrave	Capital A, grave accent
Acirc	&Acirc	Capital A, circumflex accent
Atilde	&Atilde	Capital A, tilde
Aring	&Aring	Capital A, ring
Auml	&Auml	Capital A, dieresis or umlaut mark
AElig	&AElig	Capital AE dipthong (ligature)
Ccedil	&Ccedil	Capital C, cedilla
Eacute	&Eacute	Capital E, acute accent
Egrave	&Egrave	Capital E, grave accent
Ecirc	&Ecirc	Capital E, circumflex accent
Euml	&Euml	Capital E, dieresis or umlaut mark
Iacute	&Iacute	Capital I, acute accent
Igrave	&Igrave	Capital I, grave accent
Icirc	&Icirc	Capital I, circumflex accent
Iuml	&Iuml	Capital I, dieresis or umlaut mark
ETH	&ETH	Capital Eth, Icelandic
Ntilde	&Ntilde	Capital N, tilde
Oacute	&Oacute	Capital O, acute accent
Ograve	&Ograve	Capital O, grave accent
Ocirc	&Ocirc	Capital O, circumflex accent
Otilde	&Otilde	Capital O, tilde
Ouml	&Ouml	Capital O, dieresis or umlaut mark
Oslash	&Oslash	Capital O, slash
Uacute	&Uacute	Capital U, acute accent
Ugrave	&Ugrave	Capital U, grave accent
Ucirc	&Ucirc	Capital U, circumflex accent
Uuml	&Uuml	Capital U, dieresis or umlaut mark;
Yacute	&Yacute	Capital Y, acute accent
THORN	&THORN	Capital THORN, Icelandic
Szlig	&szlig	Small sharp s, German (sz ligature)
aacute	&aacute	Small a, acute accent
agrave	&agrave	Small a, grave accent
acirc	&acirc	Small a, circumflex accent
atilde	&atilde	Small a, tilde
aring	&aring	Small a, ring
auml	&auml	Small a, dieresis or umlaut mark
aelig	&aelig	Small ae dipthong (ligature
ccedil	&ccedil	Small c, cedilla
eacute	&eacute	Small e, acute accent
egrave	&egrave	Small e, grave accent
ecirc	&ecirc	Small e, circumflex accent

euml	&euml	Small e, dieresis or umlaut mark
iacute	&iacute	Small i, acute accent
igrave	&igrave	Small i, grave accent
icirc	&icirc	Small i, circumflex accent
iuml	&iuml	Small i, dieresis or umlaut mark
eth	&eth	Small eth, Icelandic
ntilde	&ntilde	Small n, tilde
oacute	&oacute	Small o, acute accent
ograve	&ograve	Small o, grave accent
ocirc	&ocirc	Small o, circumflex accent
otilde	&otilde	Small o, tilde
ouml	&ouml	Small o, dieresis or umlaut mark
oslash	&oslash	Small o, slash
uacute	&uacute	Small u, acute accent
ugrave	&ugrave	Small u, grave accent
ucirc	&ucirc	Small u, circumflex accent
uuml	&uuml	Small u, dieresis or umlaut mark
yacute	&yacute	Small y, acute accent
thorn	&thorn	Small thorn, Icelandic
yuml	&yuml	Small y, dieresis or umlaut mark
reg	&reg	Registered TradeMark
copy	&copy	Copyright
trade	&trade	TradeMark
nbsp		Non breaking space

Si può anche usare la codifica numerica (i codici numerici **non** sono quelli ASCII ma quelli previsti dallo standard ISO 8859-1). **Anche qui ricordarsi del punto e virgola.**

Reference	Description
&#00- &#08	Unused
&#09	Horizontal tab
&#10	Line feed
&#11 - &#31	Unused
&#32	Space
&#33	Exclamation mark
&#34	Quotation mark
&#35	Number sign
&#36	Dollar sign
&#37	Percent sign
&#38	Ampersand
&#39	Apostrophe
&#40	Left parenthesis
&#41	Right parenthesis
&#42	Asterisk
&#43	Plus sign
&#44	Comma
&#45	Hyphen
&#46	Period (fullstop)
&#47	Solidus (slash)
&#48- &#57	Digits 0-9
&#58	Colon
&#59	Semi-colon
&#60	Less than

&#61 Equals sign
&#62 Greater than
&#63 Question mark
&#64 Commercial at
&#91 Left square bracket
&#92 Reverse solidus (backslash)
&#93 Right square bracket
&#94 Caret
&#95 Horizontal bar
&#96 Acute accent
&#97- &#122 Letters a-z
&#123 Left curly brace
&#124 Vertical bar
&#125 Right curly brace
&#126 Tilde
&#127 -   Unused
&#161 Inverted exclamation
&#162 Cent sign
&#163 Pound sterling
&#164 General currency sign
&#165 Yen sign
&#166 Broken vertical bar
&#167 Section sign
&#168 Umlaut (dieresis)
&#169 Copyright
&#170 Feminine ordinal
&#171 Left angle quote, guillemot left
&#172 Not sign
&#173 Soft hyphen
&#174 Registered trademark
&#175 Macron accent
&#176 Degree sign
&#177 Plus or minus
&#178 Superscript two
&#179 Superscript three
&#180 Acute accent
&#181 Micro sign
&#182 Paragraph sign
&#183 Middle dot
&#184 Cedilla
&#185 Superscript one
&#186 Masculine ordinal
&#187 Right angle quote, guillemot right
&#188 Fraction one-fourth
&#189 Fraction one-half
&#190 Fraction three-fourths
&#191 Inverted question mark
&#192 Capital A, acute accent
&#193 Capital A, grave accent
&#194 Capital A, circumflex accent
&#195 Capital A, tilde
&#196 Capital A, dieresis or umlaut mark
&#197 Capital A, ring
&#198 Capital AE dipthong (ligature)

&#199 Capital C, cedilla
&#200 Capital E, acute accent
&#201 Capital E, grave accent
&#202 Capital E, circumflex accent
&#203 Capital E, dieresis or umlaut mark
&#204 Capital I, acute accent
&#205 Capital I, grave accent
&#206 Capital I, circumflex accent
&#207 Capital I, dieresis or umlaut mark
&#208 Capital Eth, Icelandic
&#209 Capital N, tilde
&#210 Capital O, acute accent
&#211 Capital O, grave accent
&#212 Capital O, circumflex accent
&#213 Capital O, tilde
&#214 Capital O, dieresis or umlaut mark
&#215 Multiply sign
&#216 Capital O, slash
&#217 Capital U, acute accent
&#218 Capital U, grave accent
&#219 Capital U, circumflex accent
&#220 Capital U, dieresis or umlaut mark
&#221 Capital Y, acute accent
&#222 Capital THORN, Icelandic
&#223 Small sharp s, German (sz ligature)
&#224 Small a, acute accent
&#225 Small a, grave accent
&#226 Small a, circumflex accent
&#227 Small a, tilde
&#228 Small a, dieresis or umlaut mark
&#229 Small a, ring
&#230 Small ae dipthong (ligature)
&#231 Small c, cedilla
&#232 Small e, acute accent
&#233 Small e, grave accent
&#234 Small e, circumflex accent
&#235 Small e, dieresis or umlaut mark
&#236 Small i, acute accent
&#237 Small i, grave accent
&#238 Small i, circumflex accent
&#239 Small i, dieresis or umlaut mark
&#240 Small eth, Icelandic
&#241 Small n, tilde
&#242 Small o, acute accent
&#243 Small o, grave accent
&#244 Small o, circumflex accent
&#245 Small o, tilde
&#246 Small o, dieresis or umlaut mark
&#247 Division sign
&#248 Small o, slash
&#249 Small u, acute accent
&#250 Small u, grave accent
&#251 Small u, circumflex accent
&#252 Small u, dieresis or umlaut mark

&#253 Small y, acute accent
&#254 Small thorn, Icelandic
&#255 Small y, dieresis or umlaut mark

Risulta quindi evidente che se in un documento HTML si vuol fare apparire del codice HTML non si può introdurlo direttamente, perché in questo caso il browser lo interpreterebbe invece di visualizzarlo, ma bisogna utilizzare la codifica vista sopra. Ad esempio se si vuole che il browser **visualizzi** il testo

In HTML è previsto l'elemento <B></B>

si dovrà scrivere in HTML in questo modo:

```
In HTML &egrave; previsto l'elemento &lt;B&gt&lt;/B&gt;
```

Bisogna anche dedicare qualche accenno alle cosiddette **estensioni del HTML**, cioè elementi non previsti dallo standard. Il problema principale delle estensioni è che spesso sono molto utili, ma nulla assicura che vengano riconosciute da tutti i browser: in particolare, molte di esse sono riconosciute solo o da Netscape Navigator o da Microsoft Internet Explorer.

Numerose estensioni permettono di gestire in modo più sofisticato alcuni aspetti della visualizzazione: per esempio, l'elemento <BLINK> ... </BLINK> riconosciuto da Netscape Navigator rende lampeggiante il testo che contiene. I browser che non supportano queste estensioni le ignorano semplicemente, senza dare luogo ad altri inconvenienti, per cui in linea generale si possono anche utilizzare. Non bisogna però che il documento sia progettato in modo da diventare inutilizzabile se tali elementi non vengono riconosciuti; ad esempio, nel caso del <BLINK> non si dovrà mettere una indicazione come "Fai clic sul testo che lampeggia", perché chi non ha Netscape non saprebbe dove fare clic.

Fino a poco tempo fa la principale delle estensioni allo standard HTML erano i **frames,** supportati in origine solo da Netscape Navigator, poi anche da Microsoft Internet Explorer, che ora fanno parte dello standard HTML ufficiale. La tecnica dei frame permette di dividere il video in più parti, ciascuna delle quali può visualizzare un documento HTML indipendentemente dalle altre: questo permette, ad esempio, di mantenere fissa in un riquadro del video una lista di link a modo di menù, mentre la pagina cui il link fa riferimento viene di volta in volta visualizzata in un altro riquadro. Quella dei frames è una tecnica molto potente, anche se complica la realizzazione della pagina. Tuttavia un documento basato sui frames non sarà utilizzabile con le più vecchie versioni dei browser; questo fino a poco tempo fa rendeva quasi indispensabile realizzarne una versione alternativa per coloro che invece utilizzano questi vecchi programmi (magari perché dotati di computer obsoleti), per cui le versioni precedenti di questo documento consigliavano molta cautela nell'utilizzare i frames per le pagine WWW della Regione. Ora però la situazione è cambiata, perché il supporto dei frames è ormai presente da molto tempo (molto, si intende, in relazione alla velocità di evoluzione delle tecnologie Internet) per cui non c'è più così tanto rischio di tagliare fuori una grande quantità di potenziali utenti. Resta comunque la raccomandazione di utilizzare i frames non per essere alla moda, ma solo quando ciò è veramente utile per la consultazione del nostro sito. **Deve comunque essere chiaro che i frames non sono più una estensione proprietaria di HTML, ma fanno parte dello standard ufficiale.**

Vediamo dunque come si creano i frames in HTML. Notiamo innanzitutto che è necessario creare non più uno solo, ma diversi documenti HTML: uno principale, che determina le dimensioni, la posizione e i nomi dei frames che si intende visualizzare, e uno per ciascuno di questi frames. Sono quindi sempre necessari tanti documenti quanti il numero di frames più uno. Il nome che dovrà comparire nell'URL è quello del documento principale (un URL ovviamente può puntare anche a un frame, che è pur sempre un documento HTML, il quale in questo caso verrà visualizzato a piena finestra come se fosse un documento indipendente; questo talvolta può essere utile, mentre altre volte quello che appare risulta difficilmente comprensibile).

Il documento principale, dopo l'intestazione, conterrà, invece del tag <BODY>...</BODY>, il tag <FRAMESET>...</FRAMESET>. Questo tag serve per stabilire come i frame devono essere distribuiti sullo schermo. A questo scopo si usano gli attributi ROWS e COLS: come facilmente si comprende, il primo determina l'altezza che deve occupare ciascun frame, il secondo la larghezza. Almeno uno di questi attributi **deve** sempre essere utilizzato, ma si possono usare anche entrambi. I possibili valori di questi attributi sono una lista di numeri o percentuali separati da virgole. La lista può contenere anche il simbolo *, che significa: *tutto il resto dello schermo*. I numeri rappresentano dimensioni in punti, mentre le percentuali rappresentano percentuali di larghezza (per COLS) o altezza (per ROWS) dello schermo. Qualche esempio chiarirà meglio la cosa:

```
<FRAMESET ROWS="40%,30%,*">
```

significa che lo schermo dovrà contenere tre frames: il primo dovrà estendersi per il 40% dell'altezza, il secondo per 30% e il terzo per il resto;

```
<FRAMESET COLS="25%,100,*">
```

significa che lo schermo dovrà contenere anche qui tre frames: il primo occuperà il 25 per cento della larghezza, il secondo 100 punti e il terzo tutto il resto.

A questo punto bisogna determinare che cosa deve contenere ciascun frame. A ciò si provvedere con il tag <FRAME>, che non prevede l'elemento di chiusura, ed è sempre contenuto all'interno di <FRAMESET>. Devono esserci tanti tag <FRAME> quante erano le parti dello schermo definite nell'ambito di <FRAMESET> (tre in entrambi gli esempi riportati sopra). <FRAME> supporta diversi attributi, di cui i principali sono: SRC (obbligatorio), che determina quale documento deve essere visualizzato nel frame, e NAME, che è un nome attribuito al frame che serve per fare riferimento ad esso (di seguito sarà spiegato in che circostanze questo avviene). Ecco un esempio

```
<FRAMESET COLS="25%,100,*">
     <FRAME SRC="esempio1.htm" NAME="Esempio 1">
     <FRAME SRC="esempio2.htm" NAME="Esempio 2">
     <FRAME SRC="esempio3.htm" NAME="Esempio 3">
</FRAMESET>
```

In questo esempio lo schermo viene diviso in tre parti: nella prima verrà visualizzato il documento esempio1.htm, nella seconda il documento esempio2.htm, nella terza il documento esempio3.htm. Si possono anche indicare meno tag frame delle parti in cui è stato suddiviso lo schermo: in questo caso alcune parti rimarranno vuote. Ovviamente i documenti da visualizzare nelle varie parti dello schermo non dovranno essere scelti a casaccio, ma con riguardo allo spazio disponibile: non si dovrà quindi visualizzare in una striscia ai bordi dello schermo un documento progettato per svilupparsi in larghezza.

Ovviamente un documento HTML visualizzato in un frame potrà contenere dei link che puntano ad altri documenti qualsiasi disponibili su Internet: dove verranno visualizzati questi documenti ? Possiamo determinarlo attraverso l'attributo TARGET. Se questo attributo non è presente, il nuovo documento verrà visualizzato nel frame di partenza. Se l'attributo TARGET ha il valore "_blank", il nuovo documento verrà visualizzato in una nuova finestra del browser. Ma se TARGET ha come valore uno dei nomi prima specificati nell'attributo NAME dell'elemento <FRAME>, ecco che il nuovo documento verrà visualizzato nel frame identificato da quel nome.

Tornando all'esempio di prima, possiamo supporre che il documento esempio1. htm contenga un elenco di link, e che esempio2.htm contenga una schermata di spiegazioni. In esempio1.htm potrebbe esserci un link della forma:

```
<A HREF="http://www.anyhost.com/something.html"  TARGET="Esempio 2">Qui c'&egrave; da vedere qualcosa</A>
```

Cliccando su questo link si otterrà l'effetto che nel frame Esempio 2 il documento originario (esempio2.htm) verrà sostituito dal documento something.html che si trova sull'host www.anyhost.com, mentre conitnuerà a riamanere visualizzato sia quanto si trova nel frame Esempio 1 (nel nostro caso la lista dei link), sia quanto si trova anche nel frame Esempio3.

Le informazioni date sopra sul linguaggio HTML non sono certo complete, ma dovrebbero essere sufficienti a dare un quadro degli elementi essenziali del linguaggio: chi padroneggia bene questi non dovrebbe avere difficoltà ad estendere la sua conoscenza agli ulteriori dettagli.

5.2 **Manuale di stile**

Quello che vogliamo fornire qui è un manuale di stile, cioè un insieme di indicazioni sul modo di realizzare documenti HTML belli ed utili, redatto tenendo in particolare considerazione le esigenze del sito WWW della Regione, ma utile anche al di fuori del contesto regionale.

Si deve tener presente che la Regione sostiene delle spese per mantenere il suo sito WWW: queste spese si giustificano se i documenti in esso inseriti sono di interesse pubblico, cioè utili per chi li consulta, e non vuoti, oziosi, insulsi, oppure disordinati o inutilmente complicati.

In questo campo, spesso non si possono dare indicazioni così esatte e determinate come quando si tratta dello standard HTML in senso stretto: sono, piuttosto, un insieme di consigli e di criteri.

I principali consigli potrebbero essere riassunti come segue:

- **SCRIVETE CHIARAMENTE IL CODICE HTML.** Voi stessi o qualcun altro potreste trovarvi a correggere e modificare il codice HTML: anzi questo succede nella maggior parte dei casi, perché di solito non si crea un documento per lasciarlo immobile per l'eternità. Ora, lo standard HTML non specifica come vanno inseriti gli elementi, ad esempio se uno per riga, o molti su di una riga, poiché il risultato finale è lo stesso. Tuttavia è bene inserirli in modo ordinato e uno per riga, in modo da rendere il codice più leggibile e quindi più facile da esaminare e da modificare. Ad esempio, invece di scrivere:

  ```
  <P><B>Pagina della Regione Liguria</B></P><BR><HR><CENTER>Testo centrato</CENTER><A HREF="http://lcweb.loc.gov">Library of Congress</A>
  ```

  è meglio scrivere

  ```
  <P>
  <B>Pagina della Regione Liguria</B>
  </P>
  <BR>
  <HR>
  <CENTER>Testo centrato</CENTER>
  <A HREF="http://lcweb.loc.gov">Library of Congress</A>
  ```

  Il risultato è identico, ma la seconda versione è molto più leggibile

- **CREATE PAGINE HTML SOLO SE POI VALE LA PENA CONSULTARLE**. Detto così, sembra il colmo della banalità: invece in Internet si trovano moltissime pagine di contenuto informativo trascurabile, che più o meno equivalgono alla dichiarazione "Ci sono anch'io !", e che certamente nessuno andrebbe a consultare una seconda volta (tranne forse la mamma e la fidanzata dell'autore). Sembra quasi che qualcuno creda che qualsiasi sciocchezza acquisti interesse solo per il fatto di essere trasferita in

HTML e messa su Internet, ma è improbabile che coloro che usano Internet per le proprie attività siano dello stesso parere. Bisogna quindi aver cura affinché **ogni singola pagina** HTML sia interessante ed utile per le informazioni che reca: In particolare, non bisogna suddividere su più pagine cose che potrebbero stare senza problemi su una sola, perché in tal caso almeno alcune delle pagine saranno di contenuto troppo scarso.

- **ORGANIZZATE BENE I LINK TRA LE VOSTRE PAGINE.** Quando si mettono su Internet molte pagine, bisogna poi organizzarle nel modo più opportuno, ossia più facilmente utilizzabile da parte degli utenti. L'organizzazione avviene inserendo dei link che permettano di passare da una pagina all'altra, definendo così uno o più percorsi tra cui l'utente si può muovere. L'insieme dovrebbe essere strutturato in modo che ogni pagina costituisca una unità informativa dotata di significato autonomo: invece spesso si vedono gerarchie di pagine in cui le informazioni significative si trovano solo al penultimo o all'ultimo livello, per cui la prima pagina fa poco più che rimandare alla seconda, la seconda si limita a rimandare alla terza, la terza dice due o tre cose di poco conto e rimanda alla quarta, la quarta infine rivela il vero e proprio contenuto del sito. Una organizzazione di questo genere è assolutamente da evitare ! Bisogna invece creare una *home page*, che sia la prima cui si accede e che dia una idea precisa del contenuto del sito (qui per *sito* non intendiamo quello della Regione, ma piuttosto l'insieme di documenti messi a disposizione da una singola Struttura o Dipartimento). Nulla vieta poi di inserire diversi livelli di link, ma è necessario che ad ogni livello si trovi qualche cosa di utile. È anche possibile mettere sul server pagine cui non rimanda alcun link: questo andrebbe fatto solo se è reso opportuno da ragioni molto particolari, perché tali pagine non sono accessibili se non indicando direttamente il loro URL oppure attraverso link posti fuori dal sito, e quindi costituiscono a tutti gli effetti altrettanti siti indipendenti, anche se fisicamente sono poste nella stessa directory dello stesso computer. Quando poi si crea una pagina molto lunga che tratta di argomenti diversi, è molto utile inserire delle ancore (<A NAME=*nome*>) e un indice all'inizio della pagina che contenga i link alle diverse ancore

- **ANDATE PIANO CON LA GRAFICA E IL MULTIMEDIALE !** Abbiamo visto che un documento HTML può contenere non solo testo, ma anche immagini, musica, video ecc.: di queste sorgenti di informazione, a parte il testo, le immagini sono di gran lunga le più usate. In alcuni casi lo scopo principale del documento HTML è quello di visualizzare una immagine: si pensi, ad esempio, ad un documento destinato a mostrare paesaggi o riproduzioni di opere d'arte. Su questo non ci sono particolari considerazioni da fare. Altre volte le immagini vengono inserite soprattutto a scopo decorativo, o comunque non sono il principale contenuto informativo della pagina: in questi casi ci vuole molta cautela, perché le immagini corrispondono a file in formato grafico che possono anche essere piuttosto grossi, e che devono essere trasferiti dal server al computer dell'utente. Questo aumenta notevolmente il tempo di scaricamento della pagina, e può essere molto fastidioso, soprattutto quando il collegamento è lento, e quando alla fine si scopre che non valeva la pena aspettare tanto, perché tutte quelle immagini (e qualche volta anche tutto il resto della pagina ...) sono del tutto inutili. In effetti, si vedono pagine piene di icone, bandierine, disegnini, figurine e molto altro, che poi non contengono alcunché di interessante. Inoltre le pagine con molta grafica ed elementi multimediali, indipendemente dai loro eventuali meriti, sono spesso di difficile consultazione per i ciechi che utilizzano sintetizzatori vocali i quali sono in grado di interpretare il testo ma non certo le immagini. Ci vuole quindi grande sobrietà nell'uso di immagini al solo scopo decorativo. Se comunque si creano pagine ad alto contenuto grafico, è bene, se possibile, preparare di ciascuna pagina una versione senza grafica o con pochissima grafica, appunto per coloro che dispongono di un collegamento lento.

- **SIATE UN PO' MEGALOMANI**. Le indicazioni date finora sono soprattutto cautele da seguire e cose da evitare. Questo non dovrebbe impaurire: anzi, nel realizzare pagine HTML bisogna essere capaci a pensare in grande, e non avere paura ad aggiungere dati, testi, files, link, e qualunque altra cosa, **purché di reale contenuto informativo**. Bisogna ricordare che, una volta che abbiamo la possibilità di mettere online documenti HTML, stiamo usando più o meno le stesse tecnologie di coloro che creano le pagine dei siti della Microsoft o della Netscape, per cui non c'è motivo perché non dobbiamo essere bravi come loro (e anche di più), e perché le nostre pagine non siano altrettanto consultate delle loro (questo però è un po' più difficile). Della megalomania (giusta) fa parte anche il far conoscere le proprie pagine: Internet è grande,

per cui se si vuole che qualcuno venga a consultare le nostre pagine bisogna rendere note il più possibile; tra i mezzi per fare questo, ricordiamo: annunciarle in liste di discussione e newsgroup di argomento attinente a quello delle pagine; comunicarlo a chi gestisce pagine di argomento affine, chiedendo di mettere un link alle nostre pagine; inserire l'URL nei motori di ricerca (Yahoo, Lycos ecc.), cosa che si può fare facilmente e gratuitamente.

6. APPENDICE

In questa appendice sono raccolte alcune informazioni di carattere più strettamente tecnico, che comunque si raccomanda di leggere ugualmente

6.1 **Informazione e computabilità**

Dei due aspetti teorici fondamentali che stanno alla base della scienza degli elaboratori, cioè la **teoria dell'informazione** e la **teoria della computabilità**, qui accenniamo principalmente al primo. Nell'ambito della **teoria matematica dell'informazione** viene tra l'altro definito il concetto di **bit**, che ricorre continuamente in ogni discorso di argomento informatico. La teoria dell'informazione (proposta originariamente nel 1946 da Claude E. Shannon) considera **l'informazione come scelta tra alternative**: quando il verificarsi di un evento consente di determinare quale tra diverse alternative possibili si è verificata, e quindi diminuisce la quantità di incertezza, allora si può dire che l'evento codifica informazione. Più specificamente, consideriamo un insieme di simboli (che potrebbe essere qualsiasi evento che si conviene di considerare un simbolo), ognuno dei quali ha la stessa probabilità di ricorrere: allora diciamo che la quantità di informazione di ogni simbolo è

$$I = \log n$$

dove n è il numero dei simboli impiegati e log è il logaritmo in base 2. Il bit è la quantità di informazione codificata dalla scelta di uno tra due simboli (log2 = 1, poiché in qualunque base il logaritmo della base è uno). L'informazione totale trasmessa da un messaggio con N simboli sarà

$$I = N \log n$$

Se i simboli non sono equiprobabili, l'informazione di un messaggio, essendo p_n la probabilità di ricorrenza del simbolo n-esimo, sarà

$$H = - (p_1 \log p_1 + ... + p_{n-1} \log p_{n-1} + p_n \log p_n)$$

Si può dimostrare che la quantità di informazione trasmessa da un messaggio è massima quando i simboli sono equiprobabili. Infatti, in quel caso è massima l'incertezza: se le probabilità di ricorrenza sono diverse, possiamo sfruttare questo fatto per prevedere i simboli successivi, e quindi diminuire l'incertezza. Per esempio, nella lingua italiana le lettere e combinazioni di lettere non hanno tutte la stessa probabilità, per cui sappiamo che dopo una *l* è più probabile che appaia una *a* piuttosto che non una *z*.

Non bisogna confondere il **bit** con il **byte**: il byte è un raggruppamento di 8 bit, trattati come un tutto unico, ed utili, ad esempio, per la codifica dei caratteri. Un byte può assumere 256 valori diversi.

Altri importantissimi risultati della teoria dell'informazione riguardano la capacità dei canali di trasmissione e la codifica. In particolare, Shannon ha dimostrato che, in presenza di disturbi, è sempre possibile ridurre quanto si vuole il tasso di errore nella trasmissione, aumentando però la capacità del canale. Inoltre è sempre possibile trovare una codifica ottimale, cioè che renda minima la quantità di dati da trasmettere (questi peraltro sono teoremi non costruttivi, cioè dicono che esiste una codifica con certe caratteristiche, ma dicono quale sia

esattamente). Per quanto riguarda, poi, la capacità del canale in presenza di rumore (condizione che vale per qualunque canale reale), essa è determinata dalla **formula di Shannon-Hartley**:

$$C = W(1 + \frac{S}{N})$$

dove C è la capacità del canale, W l'ampiezza di banda e la frazione S/N il rapporto segnale/rumore.

Questi semplici concetti matematici sono alla base di una teoria di enorme potenza, sulla quale si basa la scienza dei calcolatori, ed in particolare tutto quello che riguarda la codifica e la trasmissione dei dati, e quindi le reti di computer.

Per quanto riguarda la computabilità, possiamo dire che in termini generali un calcolo è un insieme di operazioni **determinate in modo univoco** da regole tali che da un certo stato di cose si passi ad un altro (ad esempio, da un problema al risultato, dagli ingredienti alla torta ecc.). La teoria della computabilità indaga la natura e i limiti di ciò che è computabile. Il matematico inglese Alan Turing ha definito le cosiddette *macchine di Turing*, che non sono apparati fisici, materiali, ma sono modelli astratti di dispositivi computazionali. Una macchina di Turing può avere un certo numero di stati interni, e leggere e scrivere informazioni da un nastro (anch'esso da concepire in senso astratto), suddiviso in caselle ognuna delle quali può contenere un simbolo. La macchina inoltre è dotata di regole di transizione degli stati. Essa parte da un certo stato e, in base ai simboli letti e alle regole di transizione degli stati, arriva ad un altro, che può comportare - ad esempio - la lettura o scrittura di simboli da nastro (ed equivale alla fine della computazione). La *tesi di Church-Turing* (Alonzo Church era un logico e matematico inglese) afferma che ogni problema computabile in senso intuitivo è computabile da una macchina di Turing: questa però è una tesi e non un teorema, perché non è stato possibile dimostrarne la verità (e neppure la falsità). Infatti un limite della teoria della computabilità è che finora non è stato possibile definire un concetto rigoroso di computabilità, e questo evidentemente non permette di accertare i limiti della stessa. Ci sono comunque problemi sensati e non computabili: questo è evidentemente analogo all'esistenza di proposizioni indecidibili nell'ambito dei linguaggi formali (teorema di Gödel). Si comprende facilmente, quindi, che la teoria della computabilità è di notevole interesse dal punto di vista filosofico, proprio perché indaga sulla struttura e la portata di alcune forme della conoscenza.

6.2 **Porte e socket**

Probabilmente molti conoscono le **porte** fisiche del computer, che sono connettori cui si attacca qualche apparecchio che comunica con il computer, ad esempio una stampante o un modem. Le potre vengono identificate dal sistema operativo con dei particolari nomi, ad esempio, in DOS/Windows, com1, com2, lpt1 ecc. Ebbene, **le porte di cui parliamo qui non sono queste** (quindi attenti alla confusione), ma sono, in un certo senso, una estensione dello stesso concetto.
In ambiente TCP/IP, si dice che ogni server software del livello applicazione ascolta (cioè si aspetta richieste) su una porta, che non è una porta fisica del computer, ma una entità astratta (che si potrebbe definire una porta logica) che rappresenta un possibile canale di comunicazione, un punto di accesso ai servizi di una applicazione. Le porte sono definite al livello trasporto, e quindi sono di competenza del protocollo TCP (per questi si parla anche di porte TCP). Ai vari servizi (la cui natura è stata illustrata nelle sezioni precedenti) sono attribuite delle porte di default; di seguito si indicano le principali porte di default:

FTP	>	21 per il controllo e 20 per i dati
TELNET	>	23
GOPHER	>	70
HTTP	>	80
POP3	>	110
WAIS	>	210

È possibile configurare i server in modo che ascoltino su porte diverse da quelle di default, ma in questo caso la porta dovrà essere espressamente specificata nella richiesta. Per esempio, sull'host www.qualcosa.it potrebbero essere attivi due server WWW (protocollo HTTP), uno sulla porta di default, l'altro sulla porta 47. Per collegarsi al primo si userà l'indirizzo nella forma: *http://www.qualcosa.it*, per il secondo bisognerà usare la forma *http://www.qualcosa.it:47*. Si può anche dare indicazioni ad un client di interrogare la porta su cui c'è un server diverso da quello normale: nella maggior parte dei casi questo viene fatto per errore, ma talvolta si fanno queste operazioni con i client telnet a scopo di test. Ad esempio, per verificare se sull'host di cui sopra i vari server rispondono correttamente, si può usare il comando *telnet://www.qualcosa.it:80* oppure *telnet://www.qualcosa.it:47* ed esaminare i messaggi che vengono visualizzati sullo schermo. Si noti che in questo caso è necessario specificare anche la porta 80 perché non è quella di default per le sessioni telnet.

Il **socket** è una entità astratta che identifica una specifica connessione attiva, e deriva dalla combinazione di porta e indirizzo IP. Per esempio, se dal mio computer sto scaricando la posta elettronica (utilizzando la porta 110 e l'indirizzo IP del server da cui la ricevo) e contemporaneamente sto scaricando un file (utilizzando la porta 21 e l'indirizzo IP del server), ho attivi due socket. Si può concepire la lettura e la scrittura da e in un socket come analoga all'I/O da file.

Si noti che il socket così definito costituisce una violazione del modello OSI, perché è una entità definita in base ad elementi di livello 3 (l'indirizzo IP) e di livello 4 (la porta), andando quindi contro al principio della separazione dei livelli.

6.3 **Analogico e digitale**

Si sente continuamente parlare della differenza, e anzi della contrapposizione, tra *analogico* e *digitale*: ad esempio le registrazioni musicali possono essere analogiche o digitali, un apparecchio può essere analogico o digitale, persino un calcolatore può essere analogico o digitale (un tempo calcolatori analogici erano impiegati in ambito tecnico e scientifico). Inoltre di solito il digitale viene presentato come superiore all'analogico, e non a torto: ad esempio una registrazione musicale digitale in genere ha più dinamica e minore distorsione di una analogica.

Vediamo però più da vicino che cosa significano questi termini. Essi si applicano ad un segnale (ad esempio un segnale elettrico), oppure ad un apparecchio che elabora segnali (ad esempio un calcolatore o un riproduttore musicale). Un segnale **analogico** è quello che può variare in modo continuo, ossia può assumere un qualunque valore in un certo intervallo: per esempio, un segnale elettrico analogico che può variare in tensione da 0 a 10 V potrebbe assumere il valore di 1 V, 2V, 2,073 V, 2,0735 V, 2,07358 V e così via. Se il segnale analogico deve rappresentare qualcosa, per esempio se si tratta di segnale musicale, esso varia seguendo in modo esatto (in linea di principio) le variazioni dell'informazione che viene rappresentata. Un segnale **digitale** invece è un segnale discreto, ossia non può assumere tutti i valori compresi in un intervallo, ma solo alcuni valori ben distinti. Per arrivare a rappresentare una certa informazione, per esempio un segnale musicale, tramite un segnale digitale, sono necessari due procedimenti: il **campionamento** e la **quantizzazione**. Il campionamento consiste nel rilevare, ad intervalli regolari, ossia con una certa determinata frequenza, il valore del segnale originale. Il risultato del campionamento è una lista di valori che rappresentano lo stato del segnale originale in momenti distinti. A questo punto si effettua la quantizzazione: ciascun valore ottenuto con il campionamento viene trasformato nel più vicino di una lista predeterminata di valori numerici ammissibili. Il risultato di tutta l'operazione è appunto un insieme di numeri che vengono elaborati in luogo del segnale originale.

I parametri che determinano il risultato del campionamento e della quantizzazione sono la **frequenza di campionamento** e **il numero di bit di quantizzazione**.

La frequenza di campionamento è il numero di volte al secondo in cui viene registrato il valore del segnale originale. La massima frequenza fedelmente rappresentabile è la metà della frequenza di campionamento; quest'ultima viene detta **frequenza di Nyqvist**. Aumentare la frequenza di campionamento oltre la frequenza

di Nyqvist (il doppio della frequenza del segnale che si vuole rappresentare) è in generale inutile perché non si ricava nessuna informazione in più. Invece a volte è utile, per motivi pratici, rinunciare a riprodurre una parte del segnale originale, riducendo la frequenza di campionamento. Ad esempio, la frequenza di campionamento per i CD audio è di 44,1 KHz (cioè il campionamento avviene 44.100 volte al secondo), il che consente di riprodurre frequenze di 22,05 KHz, già oltre la soglia dell'udibile che è di 20.000 Hz. La DCC (cassetta audio digitale) permette di utilizzare diverse frequenze di campionamento, tra cui quella di 32 KHz, che consente di registrare una maggior durata di musica, ma riproducendo segali audio fino a 16.000 Hz, quindi già con una qualche perdita di qualità. Gli apparati telefonici digitali sono progettati per riprodurre in modo accettabile (e non certo a livello Hi-Fi) solo il parlato, che arriva fino a circa 3 Khz; essi dovrebbero quindi utilizzare una frequenza di campionamento di soli 6 Khz, ma in realtà utilizzano la frequenza di 8 KHz perché, allo scopo di far viaggiare senza interferenze su di una stessa linea più comunicazioni riservano una larghezza di banda di 4 KHz per ciascuna.

Quando si diminuisce la frequenza di campionamento al di sotto della frequenza di Nyqvist e non si prendono altre precauzioni, non si ha solo un degrado più o meno rilevante della qualità del segnale digitalizzato, ma anche la comparsa di artefatti, ossia di segnali del tutto estranei a quelli originali e che si mescolano a questi ultimi. Questo fenomeno viene detto **aliasing** [34]. Per evitarlo bisogna, prima del campionamento, eliminare dal segnale tutte le componenti di frequenza superiore alla metà della frequenza di Nyqvist, cosa che si ottiene attraverso un filtro passa-basso, che per l'occasione viene anche detto filtro anti aliasing.

Il numero di bit di quantizzazione determina il numero di valori (numerici) distinti che possono essere assegnati ad ogni campione. Ad esempio una quantizzazione a 2 bit mette a disposizione due bit, ciascuno dei quali può - per definizione - assumere due valori: avremo in tutto quindi 4 possibili valori. È quindi evidente che una quantizzazione a 2 bit permette di rappresentare solo un segnale molto semplice, oppure di rappresentare un segnale più complesso in modo molto approssimativo. In generale, il numero di valori che si ottengono da n bit è pari 2^n . Ad esempio con una quantizzazione a 8 bit si possono avere 256 valori distinti (2^8), con una a 16 bit 65536 (2^{16}) e così via. Si comprende facilmente che mentre c'è un limite oltre quale è inutile aumentare la frequenza di campionamento (cioè il doppio della massima frequenza che vuole riprodurre), non c'è un limite oltre al quale è **in linea di principio** inutile aumentare i bit di quantizzazione, perché in questo modo si ottiene una quantizzazione via via sempre più fedele.

A questo punto ci accorgiamo di una cosa sorprendente: **digitalizzare, cioè campionare e quantizzare, equivale a discostarsi dal dato originale, ossia ad introdurre un errore** ! In particolare, l'**errore di quantizzazione** è la differenza tra il valore originale di un campione e il valore ad esso attribuito dopo la quantizzazione: è chiaro questo valore non potrà, in generale, essere diverso da zero[35]. Ma allora la superiorità del digitale sull'analogico è forse solo un mito creato per scopi commerciali ? Non è così (anche se nel campo musicale vi sono alcuni, pochi in verità, che sostengono la superiorità delle registrazioni analogiche). Infatti il digitale ha alcuni vantaggi che nella maggior parte dei casi rendono praticamente irrilevanti gli errori che esso si porta dietro:

- poiché il segnale viene tradotto in un insieme di numeri è più facile effettuare su di esso delle opportune elaborazioni atte a correggere certi difetti, ad esempio ad eliminare il rumore (cioè il segnale indesiderato che si mescola a quello che si intende invece trattare), oppure a trasformare il segnale per scopi tecnici o estetici (si pensi al fotoritocco digitale o alla musica elettronica)
- il segnale digitale è intrinsecamente più robusto, perché per trasmetterlo e riprodurlo correttamente è sufficiente poter riconoscere quale dei due valori possibili assume un certo bit, e questo si può fare anche se fisicamente (ad esempio dal punto di vista elettrico) il segnale è degradato; questo non è possibile evidentemente con il segnale analogico; si tratta di un punto di particolare importanza, perché nelle situazioni reali il segnale subisce sempre un certo degrado

[34] Un tempo, soprattutto quando si cominciava a parlare di audio digitale, questo termine veniva storpiato in italiano con il termine **alea**, ma ora sembra che fortunaamente questo uso sia stato abbandonato

[35] L'errore di quantizzazione può anche essere descritto come **rumore di quantizzazione**; infatti l'effetto è lo stesso che si avrebbe aggiungendo un segnale originale un certo rumore

- è possibile determinare i parametri di digitalizzazione in modo che gli errori siano tali da non avere più rilievo pratico: ad esempio, per restare negli esempi musicali, si può stabilire una frequenza di campionamento che permetta di riprodurre tutte le frequenze udibili dall'orecchio umano; si può aumentare in modo opportuno il numero di bit di quantizzazione ecc.; vi sono in effetti ancora alcuni campi in cui l'analogico è superiore al digitale: in particolare, la fotografia digitale non riesce ancora a pareggiare, per qualità dell'immagine, la fotografia analogica, cioè quella tradizionale su pellicola; si tratta però non di limitazioni di principio, ma di limitazioni dovute alle attuali tecnologie di ripresa e visualizzazione, e anche all'esigenza di contenere i costi in limiti commercialmente accettabili, per cui è pensabile che in futuro la foto digitale riuscirà ad eguagliare e superare la qualità di quella su pellicola

6.5 **Analisi del segnale. Traformata di Fourier. Trasformata di Laplace.**

L'analisi del segnale effettuata per mezzo di strumenti matematici, e in particolare di quelli dell'analisi matematica, è uno degli aspetti fondamentali della teoria delle comunicazioni, benché possa sembrare un po' astrusa. Infatti solo in questo modo si possono individuare chiaramente le proprietà del segnale e quelle dei canali trasmissivi.

Il segnale può essere considerato sotto diversi aspetti. In particolare può essere considerato come **funzione del tempo**, perché ad ogni istante del tempo corrisponde un certo valore del segnale. Questo è studio del segnale nel **dominio del tempo** può comportare operazioni piuttosto complesse, per cui spesso si preferisce invece studiare il segnale nel **dominio della frequenza**, ossia come funzione della frequenza.

Qui diamo qualche accenno a due degli strumenti di analisi del segnale nel dominio della frequenza, e cioè la **trasformata di Fourier** [36] e la **trasformata di Laplace**.

Spesso è interessante sapere qual è il contenuto energetico delle componenti del segnale. Una funzione variabile nel tempo può infatti venire scomposta in componenti di diversa frequenza (i musicisti sanno bene che un suono è composto di una fondamentale e delle sue armoniche, che sono suoni di frequenza superiore). Questa scomposizione si ottiene tramite un'altra tecnica matematica ideata da Fourier, la serie di Fourier. Ci si può chiedere: quanta energia è contenuta in ogni componente in frequenza ? In termini più informali: quale componente è più debole e quale è più forte ? Questa analisi viene detta analisi spettrale.

A questa domanda si risponde appunto tramite la trasformata di Fourier, che è una funzione che ha come dominio la frequenza e come codominio l'energia. Attrvaerso la trasformata di Fourier una funzione nel dominio del tempo viene trasfornata in una funzione nel dominio della frequenza. Essa quindi fornisce la **risposta in frequenza** di un sistema di elaborazione del segnale (chi abbia letto qualche rivista di hi-fi avrà notato che nelle prove degli apparecchi audio viene appunto riportata la risposta in frequenza).

La trasformata di Fourier della funzione nel dominio del tempo f(t) è definita da:

$$F\{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{jwt} dt$$

Il simbolo ω che sta in esponente ad e[37] è appunto la frequenza di cui si vuole conoscere il contenuto energetico. Mentre comprendere la trasformata di Fourier nelle sue linee generali non è molto difficile, il suo

[36] Sarebbe un errore credere che questa trasformata sia dovuta a qualche esperto di elettronica o telecomunicazioni: il barone Jean Baptiste Jospeh Fourier visse dal 1768 al 1830, non sapeva nulla di computer e si occupava invece di ricerche sulla trasmissione del calore, nell'ambito delle quali ottenne i suoi risultati matematici

[37] Ricordiamo che *e* è un numero reale trascendente che è definito nel modo seguente:

uso pratico nell'analisi del segnale richiede padronanza dell'analisi matematica. Esistono appositi strumenti che effettuano automaticamente l'analisi di Fourier di un qualsiasi segnale.

Citiamo infine un caso particolare: lo spettro di una costante, ossia di un segnale che non varia nel tempo, è la funzione delta, che è un impulso (ideale) di durata nulla concentrato alla frequenza zero, che è appunto la frequenza di una costante.

La trasformata di Laplace invece è definita nel modo seguente:

$$F(s) = \int_0^{\infty} f(t)e^{-st}dt$$

dove *s* è definito come: $\sigma + j\omega$. Questa grandezza è anche chiamata **frequenza complessa**, per cui può dire che la trasformata di Laplace è una rappresentazione del segnale nel dominio della frequenza complessa-

Se X(s) è la trasformata dell'ingresso di un sistema e Y(s) la trasformata dell'uscita, il rapporto tra le due, cioè H(s) = Y(s) / X(s) è detto **funzione di trasferimento**, ed è la rappresentazione del comportamento del sistema (chi si occupa di fotografia sa che il principale parametro per la valutazione delle prestazioni degli obiettivi è proprio una funzione di trasferimento, e precisamente la funzione di trasferimento della modulazione).

Se l'ingresso è un impulso unitario, si ha che X(s) = 1 e quindi H(s) = Y(s), perciò si può affermare che **la funzione di trasferimento è la trasformata della risposta all'impulso**.

6.5 **Subnet e calcolo della netmask**

Si diceva che una rete IP può a sua volta essere divisa in insiemi di indirizzi IP detti sottoreti. Vediamo qualche cosa di più su questo argomento. A questo scopo, dovremo trattare gli indirizzi IP nella loro forma binaria, ossia rappresentandoli come insiemi di 32 bit.

Sia dato innanzitutto un indirizzo IP qualunque. Sia data poi una **netmask**, che è un indirizzo IP con una struttura particolare, ossia tutti gli 1 a sinistra e tutti gli 0 a destra. Il numero di zeri viene scelto liberamente ed è quello che determina le dimensioni della sottorete (ovviamente però bisogna dimensionare la sottorete in modo che non sia più ampia di tutta la rete; bisogna anche evitare di definire sottoreti che si sovrappongono tra loro). Ecco alcuni esempi:

indirizzo di rete (si tratta di una rete di classe B):

10001001	00011111	01010101	10111001
137	31	85	185

netmask:

11111111	111111111	11111111	11110000
255	255	255	240

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

A questo punto vogliamo sapere quali sono il **primo indirizzo** (detto **network address**) e l'**ultimo indirizzo** (detto **broadcast address**) della sottorete che ha il netmask specificato e a cui appartiene l'indirizzo dato.

Per sapere il network address si effettua un **bitwise and** tra l'indirizzo e la subnet mask. Il bitwise and è una operazione per la quale i due indirizzi vengono confrontati bit per bit: il risultato è 1 se entrambi i bit sono a 1, e 0 negli altri casi (se si interpreta 0 come falso e 1 come vero è evidente che questo corrisponde ad applicare la funzione di verità della congiunzione). Con i valori dell'esempio, il risultato è il seguente:

10001001	00011111	01010101	10110000
137	31	85	176

Per ottenere invece il broadcast address si confrontano sempre l'indirizzo e la subnet mask, ma come risultato si scrive 1 quando si incontra un bit a 0 nella subnet mask, mentre per il resto si procede come nel caso precedente (in termini di funzioni di verità, si tratta di una implicazione materiale che ha come antecedente i valori della subnet mask: infatti l'implicazione materiale è sempre vera quando l'antecedente è falso). Nel nostro caso il risultato sarà il seguente:

10001001	00011111	01010101	10111111	
137	31	85	191	

La sottorete individuata comprende quindi 16 indirizzi IP, da 137.31.85.176 a 137.31.85.191.

6.6 **Interrupt, microkernel e macchine virtuali**

In questo paragrafo ci proponiamo di illustrare qualche nozione in più sui sistemi operativi. Osserviamo innanzitutto che tutti i sistemi operativi evoluti si basano sulla possibilità delle moderne CPU di funzionare in due distinte modalità, dette **modo monitor** e **modo utente**. Il modo monitor è una modalità di funzionamento privilegiata, nella quale vengono eseguite in particolare le istruzioni di accesso diretto all'hardware e alle risorse di sistema: queste istruzioni vengono eseguite dal solo sistema operativo, mentre i programmi utente girano appunto in modo utente[38]. In questo modo il sistema operativo ha il pieno controllo del sistema, e può quindi impedire ad un programma mal scritto (o scritto in modo malizioso) di interferire con il funzionamento degli altri programmi o del sistema operativo stesso.

Più in dettaglio, i sistemi operativi evoluti sono **interrupt driven**: quando un programma ha bisogno dei servizi del sistema operativo ad esempio per scrivere su disco genera un **interrupt**, cioè interrompe il suo funzionamento per richiedere i servizi del sistema operativo, con una operazione detta **chiamata di sistema** (system call). Ogni interrupt è identificato da un numero, per cui quando una system call genera un interrupt l'hardware viene commutato in modo monitor e il sistema operativo, in base al numero dell'interrupt individua la procedura da eseguire per venire incontro alla richiesta del programma utente. Terminata l'operazione, la CPU viene di nuovo commutata in modo utente e il programma utente prosegue la sua esecuzione.

Come ben si comprende, le chiamate di sistema avvengono continuamente, e il gran numero di passaggi di modo di funzionamento della CPU determina già di per se un certo carico computazionale, che si traduce in perdita di tempo. Per cercare di ovviare a questo inconveniente si sono studiati dei sistemi operativi, per lo più in ambito di ricerca, basati sull'idea del microkernel. Come abbiano visto, il kernel è il nucleo del sistema operativo, quella parte che esegue le vere e proprie operazioni tipiche del sistema operativo, ed in particolare la risposta alle chiamate di sistema. Un microkernel è un kernel di dimensioni ridotte, che gestisce il minimo

[38] Uno dei principali limiti del DOS è che non prevede la distizione tra modo utente e modo monitor perché progettato per le CPU Intel 8086/88 che non avevano queste due modalità di funzionamento. La conseguenza è che in DOS tutti i programmi hanno pieno accesso all'hardware, per cui possono fare ogni genere di operazioni pericolose e non di rado bloccare il sistema, anche se questa caratteristica viene a volte sfruttata, specialmente nei giochi, per ottenere migliori prestazioni

indispensabile per assicurare il buon funzionamento del sistema, in modo da ridurre la necessità di ricorrere alle chiamate di sistema. Evidentemente il rischio è di ridurre troppo il kernel e di dare troppo spazio ai programmi utente, con conseguenti rischi per l'affidabilità del sistema.

Un'altra interessate tecnologia, che ha il suo principale rappresentate nel sistema VM, destinato ai mainframe IBM, è quella dei sistemi operativi a macchine virtuali. In questa architettura il s.o. non interagisce direttamente con i programmi utente, ma crea appunto una o più *macchine virtuali* (virtual machines), che sono dei computer non realizzati materialmente in hardware, ma emulati in software. I programmi che girano in una macchina virtuale la vedono come il loro hardware, e non hanno notizia del vero hardware e del vero sistema operativo. In ogni macchina virtuale creata dal VM viene fatto girare un sistema operativo utente, che in genere è il CMS, nell'ambito del quale a loro volta vengono eseguiti i programmi utente. Diverse macchine virtuali possono funzionare contemporaneamente e far girare diversi sistemi operativi. Per l'ambiente PC si può citare il software Virtual Platform della VMware, che gira sotto Linux, permettendo a quest'ultimo di agire da sistema operativo host per macchine virtuali, nelle quali possono girare altri sistemi operativi come Windows 98 ed NT.

La tecnica delle macchine virtuali è difficile da implementare, ma è interessante perché realizza il massimo della separazione tra il sistema operativo e i programmi utente.

7. **BIBLIOGRAFIA**

Le edizioni sono elencate in ordine alfabetico di titolo. La chiave riportata tra parentesi quadra prima del titolo viene utlizzata all'interno del testo per i riferimenti alle pubblicazioni.

[ADSL 1999] ADSL/VDSL principles : a practical and precise study of asymmetric digital subscriber lines and very high speed digital subscriber lines / Dennis Rauschmayer. - Indianapolis : Macmillan Technical Publishing, c1999. - ISBN 1-57870-015-9.

[Advanced 1999] Advanced Internet technologies / Uyless Black. - Upper Saddle River : Prentice Hall, c1999. - ISBN 0-13-759515-8.

[Architetture 1998] Architetture di instradamento per Internet / Bassam Halabi. - Milano [etc.] : McGraw-Hill, 1998. - ISBN 88-386-0477-0.

[Automaten 1995] Automaten Sprachen Berechenbarkeit / Peter Sander, Wolffried Stucky, Rudolf Herschel. - 2. Auflage, durchgesehene Auflage. - Stuttgart : Teubner, 1995. - ISBN 3-519-12937-X.

[Broadband 1997] Broadband networking : ATM, SDH and SONET / Mike Sexton, Andy Reid. - Boston ; London : Artech House, c1997. - ISBN 0-89006-578-0.

[Building 1999] Building high-speed network / Tere' Parnell. - Berkeley [etc.] : Osborne/McGraw-Hill, c1999. - ISBN 0-07-211858-X.

[Buildings 1999] Building switched networks : multilayer switching, QoS, IP multicast, network policy, and service level agreements / Darryl P. Black. - Reading (Massachussets) : Addison-Wesley, c1999. - ISBN 0-201-37953-8.

[Computer 1996] Computer networks / Andrew S. Tanenbaum. - 3. ed. - Upper Saddle River : Prentice Hall, c1996. - ISBN 0- 13-349945-6.

[Comunicazioni 1992] Comunicazioni elettriche : corso di telecomunicazioni / Paul H. Young. - [Milano] : Jackson, 1992. - ISBN 88-256-0358-4.

[Creare 1996] Creare un server Internet con Unix / George Eckel. - Bresso : Jackson, 1996. - ISBN 88-256-0940-X.

[Elaborazione 1996] Elaborazione numerica dei segnali / Alan V. Oppenheim, Roland W. Schafer. - 11. ed. - Milano : Franco Angeli, 1996. - ISBN 88-204-3006-1.

[Elementi 1983] Elementi di comunicazione digitale / Garry J. Marshall. - Milano : Hoepli, c1983. - ISBN 88-203-1360-X.

[Internetworking 1999] Internetworking / Mario Baldi, Pietro Nicoletti. - Milano [etc.] : McGraw-Hill, 1999. - ISBN 88-386-0780-X.

[Internetworking 1997] Internetworking technologies handbook / Merilee Ford, H. Kim Lew, Steve Spanier, Tim Stevenson. - [San Jose (California)?] : Cisco Press ; Indianapolis : New Riders, c1997. - ISBN 1-56205-603-4.

[Ipv6 1997] IPv6 / Silvano Gay. - Milano : McGraw-Hill, 1997. - ISBN 88-386-3209-X. - Titolo della copertina: Guida a IPv6.

[Manuale 1995] Il manuale TCP/IP / Uyless Black. - Milano : McGraw-Hill, 1995. - ISBN 88-386-0346-4.

[Principles 1994] Principles of signals and systems / Fred J. Taylor. - New York [etc] : McGraw-Hill, 1994. - ISBN 0-07-113706-8.

[Programmare 1996] Programmare Internet / Kris Jamsa, Ken Cope. - [Milano] : Mondadori Informatica, 1996. - ISBN 88-7131-741-6.

[Remote 1998] Remote access networks : PSTN, ISDN, ADSL, Internet and Wireless / Chander Dhawan. - New York [etc.] : McGraw-Hill, c1998. - ISBN 0-07-016774-5.

[Reti 1995] Reti locali : dal cablaggio all'internetworking / Silvano Gay, Pietro Nicoletti. - L'Aquila : Scuola Superiore Reiss Romoli, 1995.

[Segnali 1998] Segnali e sistemi / M. L. Meade, C. R. Dillon. - Bresso : Jackson Libri, 1998. - ISBN 88-256-1327-X.

[Sistemi 1995] Sistemi operativi / A. Silberschatz, P. Galvin. - 4. ed. - Milano [etc.] : Addison-Wesley, 1995. - ISBN 88-7192-059-7.

[Teoria 1971] La teoria matematica della comunicazione / Claude E. Shannon, W. W. Weaver. - Milano : ISEDI, 1971.

Per mantenersi aggiornati su quanto riguarda Internet e le reti è indispensabile anche consultare le riviste, che riportano le notizie più recenti. In pratica, ormai qualsiasi rivista di informatica contiene in ogni numero qualche articolo su Internet. Qui riportiamo però i dati di alcuni periodici più specializzati: i primi tre hanno un taglio più divulgativo, anche se non mancano articoli di approfondimento tecnico, mentre il quarto, diffuso solo in abbonamento, è decisamente più tecnico e contiene servizi che non è facile trovare altrove. I primi tre periodici contengono un CD-ROM in ogni numero, mentre *Login* fornisce in genere alcuni CD in omaggio con l'abbonamento, e altri allegati ad alcuni numeri.

Inter.net. - Vigano di Gaggiano : Systems Comunicazioni - Mensile
http://www.systems.it/

Internet news. - Milano : Tecniche Nuove. - Mensile.
http://inews.tecnet.it/

Login : building the information highway. - Ponsacco : Infomedia. - Bimestrale.
http://www.infomedia.it/Login/

Alcuni periodici dedicati alla programmazione che riservano ampio spazio alla programmazione in ambiente di rete (e sono quindi per loro natura di livello alquanto tecnico, anche se accessibili anche non professionisti):

Dev : developing software solutions. - Ponsacco : Infomedia. - Mensile
http://www.infomedia.it/DEV

Io programmo. - Rende : Edizioni Master. - Mensile.
http://www.gol.it/ioprogrammo

Un periodico di informatica generale che dedica molta attenzione non solo agli aspetti tecnici, ma anche a quelli giuridici e sociali di Internet è:

MC Microcomputer. - Roma : Technimedia. - Mensile.
http://www.mclink.it/

Approfondimenti tecnici di notevole livello, anche per quanto riguarda l'hardware si trovano anche su

PC professionale : guida indipendente al personal computing. - [Milano] : Mondadori Informatica. - Mensile. - ISSN 1122-1984.
http://pcpro.mondadori.com/

Un'altra utilissima fonte di informazione, alla quale a volte non si pensa, sono i cataloghi a stampa e i siti Internet delle ditte produttrici o distributrici di apparecchiature di rete. I cataloghi a stampa sono particolarmente comodi da consultare a scopo didattico (indipendtemente dal fatto che poi si voglia acquistare qualcosa o no), anche perché alcuni di essi, oltre alla presentazione dei prodotti, contengono interessanti esposizioni e chiarimenti sui concetti tecnici relativi ai vari prodotti.