



Consiglio Nazionale delle Ricerche

*Concetti telematici generali per il collegamento alla
rete LabNet*

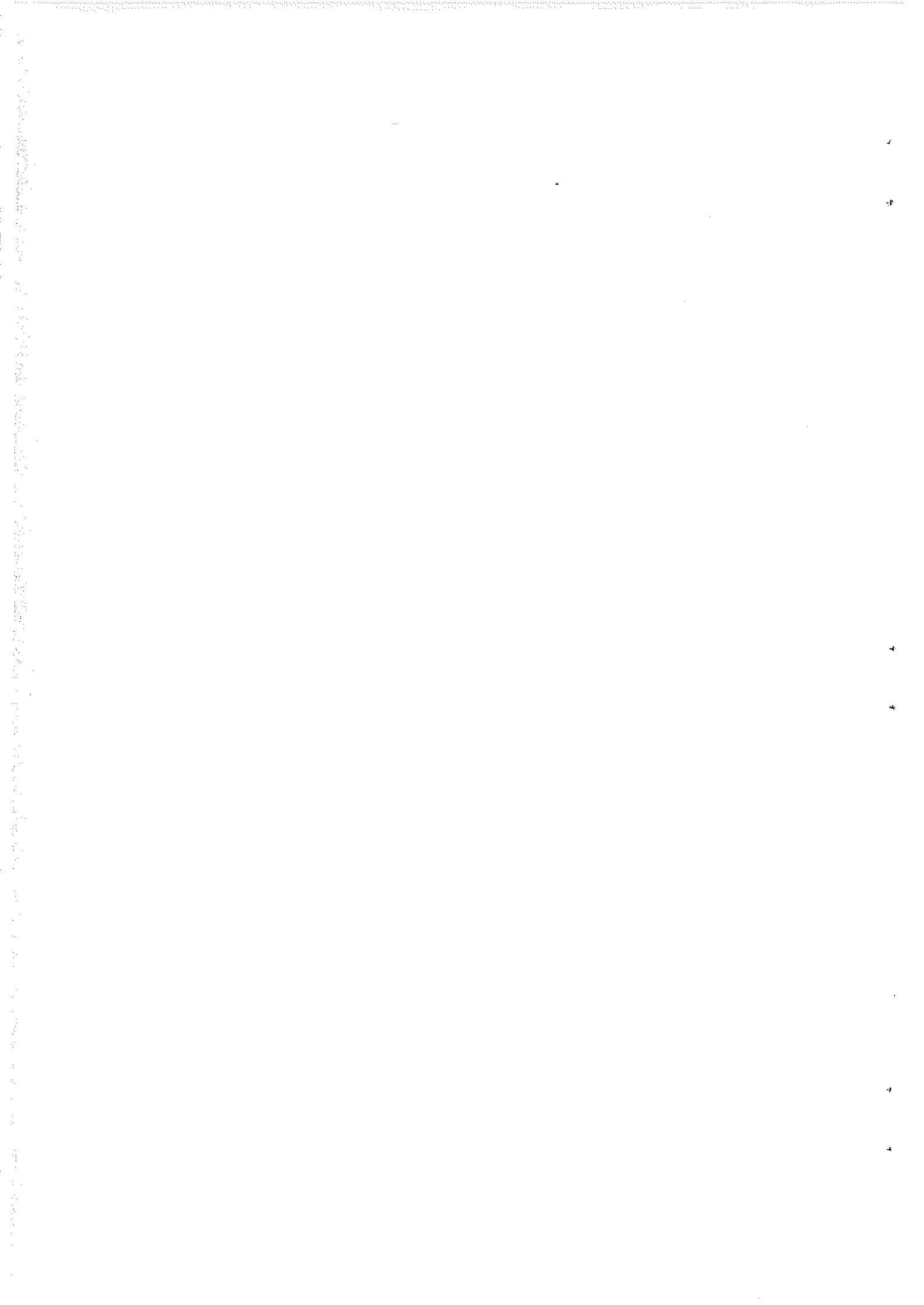
**Rolando Bianchi Bandinelli
Domenico Canino
Cesare Concordia
Vittorio Miori
Luca Trupiano**

Rapporto

CNUCE-B4-1997-027

CNUCE

Pisa



Concetti telematici generali per il collegamento alla rete LabNet

Rolando Bianchi Bandinelli

Domenico Canino

Cesare Concordia

Vittorio Miori

Luca Trupiano

INDICE

INTRODUZIONE	4
Computer e reti	4
LAN, MAN, WAN	4
Local Area Network	5
Metropolitan Area Network	5
Wide Area Network	5
Principi generali di funzionamento	5
La rete Internet	6
La connessione e l'accesso	8
Connessione	8
Connessione tramite linea telefonica	8
Connessione tramite cavi dedicati per trasmissione dati	8
Trasmissione "a vista"	8
Breve analisi dei diversi tipi di connessione	8
L'accesso	9
Accesso tramite gateway	9
Accesso tramite host	9
Accesso diretto	10
Breve analisi dei diversi tipi di accesso	10
GLI INDIRIZZI INTERNET E LA POSTA ELETTRONICA	11
Indirizzi e domini	11
"Interpretiamo" un dominio	12
Modalità client-server	13
La posta elettronica	13
Anatomia di un "Mail Header"	14
Indirizzo errato	15
Mailing Lists	15
Listserver, la mailing list Labnet	16
FTP, ANONYMOUS FTP E FTPMAIL	18
File Transfer Protocol	18
Diversi tipi di FTP Server	20
Binario e ASCII	20
FTPMail	21
La compressione dei dati	22

WORLD-WIDE WEB	24
Cenni storici	24
World Wide Web	24
Il Server	25
Il browser	26
I documenti web e Hypertext Markup Language	26
Cercare Informazioni	27
CGI script	28
Java	28
Le tecnologie <i>Push</i>	29
LABNET	29
Indirizzi utili	31
Glossario	32



Introduzione

Gli argomenti trattati in questa nota, descrivono i concetti principali sulle reti di computer e sul loro utilizzo.

Definiremo le reti in generale e ne illustreremo le caratteristiche, passeremo poi alla rete Internet della quale analizzeremo le modalità di connessione ed accesso e descriveremo i principali applicativi gettando infine uno sguardo alle ultime novità.

Lo scopo è quello di introdurre una persona sprovvista di conoscenze telematiche al mondo virtuale di Internet, mettendolo in grado di iniziare ad utilizzare la rete, questo perché noi siamo convinti che il miglior modo per imparare ad usare le risorse di Internet sia proprio...usare le risorse di Internet.

Computer e reti

Cominciamo col dare la definizione di rete di computer:

Due o più computer autonomi formano una rete se sono fisicamente connessi tra di loro (interconnessi)

Più computer sono detti connessi se ciascuno di essi è in grado di scambiare (inviare e ricevere) informazioni con uno qualsiasi degli altri, in questo caso prendono il nome di *nodi della rete*. Anche le reti possono essere connesse tra di loro formando una rete più grande, in questo caso esse prendono il nome di *sottoreti*. Inoltre affinché una rete possa definirsi tale è necessario che tutti i suoi nodi siano autonomi, ossia ciascun computer che compone la rete deve funzionare perfettamente anche da solo.

LAN, MAN, WAN

Vi sono numerosi criteri di classificazione che si possono adottare per classificare le reti, il più usato (ed anche il più conforme ai nostri scopi) è quello che si basa sulla estensione geografica, cioè sulla distanza massima che c'è tra due nodi della rete. Questo influisce notevolmente sul tipo di tecnologia che si deve adottare per poterle realizzare e di conseguenza sull'utilizzo che della rete si può fare.

Abbiamo allora le *Local Area Network* (reti locali), le *Metropolitan Area Network* (reti metropolitane), *Wide Area Network* (reti geografiche).

Local Area Network

Le Local Area Network (LAN) sono reti i cui nodi tra loro hanno una distanza massima di qualche chilometro; la distanza limitata consente l'impiego di tecnologie che permettono alte velocità di trasmissione dei dati. In genere vengono realizzate connettendo tra di loro i computer presenti in uno stesso edificio, si pensi ad una rete che connette i computer dei vari uffici di una stessa azienda allo scopo di condividere dei programmi software o lo spazio di memoria.

Metropolitan Area Network

Le Metropolitan Area Network (MAN) connettono nodi (o sottoreti) la cui distanza massima sia dell'ordine di poche centinaia di chilometri. Anche in questo caso possono essere impiegate tecnologie che consentono alte velocità di trasferimento dati. Tipico esempio di MAN è una rete regionale che connetta le reti delle vari uffici presenti nei capoluoghi di provincia allo scopo di consentire un rapido scambio di informazioni.

Wide Area Network

Le Wide Area Network (WAN), si estendono su grandi distanze (a volte intercontinentali). Sono realizzate connettendo LAN e MAN che si trovano in posizioni geografiche molto distanti tra loro, si pensi alla rete del CNR che collega tra loro le reti degli istituti di ricerca sparsi sul territorio nazionale. La velocità di trasferimento dei dati sulle WAN attualmente è inferiore rispetto a quella che si ha sulle LAN, la tendenza è quella di ricorrere a tecnologie sempre più sofisticate proprio per aumentare la velocità.

Principi generali di funzionamento

In una rete esiste sempre almeno un percorso (cammino) che unisce due nodi, in termini di comunicazione questo vuol dire che da un qualsiasi nodo della rete è possibile comunicare con uno qualsiasi degli altri. Per estendere la rete, aggiungendo un nuovo nodo bisognerà innanzitutto connettere un nuovo computer ad uno dei nodi preesistenti. Chiamiamo questa fase "connessione alla rete".

Una volta fisicamente connesso, il nuovo nodo può comunicare con gli altri, affinché questo sia possibile però è necessario che su esso siano installati dei programmi che implementino:

- 1) *il protocollo di comunicazione*
- 2) *gli applicativi di rete*

Il protocollo di comunicazione definisce l'insieme di regole che il computer deve rispettare per poter interagire con gli altri nodi (come codificare un messaggio, come individuare il destinatario di un messaggio, come ricevere un messaggio ecc.).

Gli applicativi di rete, detti anche programmi di comunicazione, sono programmi software che permettono agli utenti di comunicare. Chiamiamo la fase di utilizzo degli applicativi di rete "accesso alla rete".

La rete Internet

All'inizio degli anni 70 ricercatori statunitensi diedero il via ad un progetto denominato "Internet", lo scopo di questo progetto era di realizzare la connessione di 4 tipi diversi di rete:

- 1) una rete connessa via satellite,
- 2) una rete connessa tramite packet radio,
- 3) ARPANET, la rete dei centri di ricerca e delle università che era stata la prima rete realizzata,
- 4) una LAN di un centro di ricerca della Xerox di Palo Alto.

Si trattava di reti con diversi tipi di connessione; perché la comunicazione fosse possibile era quindi necessario definire, innanzitutto, un protocollo di comunicazione che fosse indipendente dal tipo di connessione usata.

Anche i computer che componevano le quattro reti erano diversi tra di loro e quindi il protocollo avrebbe dovuto essere implementato su diversi tipi di calcolatore. A tale scopo fu perfezionato un insieme di protocolli di comunicazione chiamato :

Transmission Control Protocol/Internet Protocol (TCP/IP)

Fu definita in pratica una sorta di "lingua franca" che permettesse ai calcolatori di comunicare indipendentemente dal modello o dal tipo di connessione.

Una volta definito, TCP/IP fu reso "di pubblico dominio", ovvero chiunque poteva creare dei programmi che lo implementassero senza dover pagare diritti d'uso, o sottostare a particolari vincoli, se non quelli del pieno rispetto delle specifiche, pena la mancata capacità di comunicare.

Artefici dell'immediata popolarità del TCP/IP furono soprattutto i ricercatori del dipartimento di informatica dell'università di Berkeley i quali lo inserirono nella versione 4.2BSD del sistema operativo Unix realizzando anche degli applicativi di rete basati su di esso e poiché questo sistema operativo era stato sviluppato grazie a finanziamenti pubblici anch'esso fu reso di pubblico dominio ed ebbe diffusione immediata. Ben presto la comunità scientifica si accorse delle potenzialità di quel tipo di rete e sempre più nodi si aggiunsero a quelli delle 4 sottoreti originarie.

L'utilizzo prevalente della rete all'inizio era la comunicazione per motivi di lavoro ma ben presto nacquero gruppi di discussione sugli argomenti più vari. Molti centri di ricerca inoltre, realizzarono degli applicativi di rete mettendoli a disposizione di tutta la comunità degli utenti i quali non si limitavano ad usarli ma a loro volta spesso modificavano i programmi apportandovi migliorie e quindi ridistribuivano il nuovo prodotto, nasceva così la "filosofia" di Internet che è quella di contribuire alla evoluzione della rete in maniera collettiva.

Nel 1983 la Defense Information Systems Agency diede vita ad una rete operativa di computer che ne collegava tutte le sedi basata su TCP/IP; contemporaneamente da una costola di ARPANET nasceva MILNET che collegava fra loro i centri di ricerca militari e poco dopo la rete della NASA. Tutte queste reti erano ovviamente interconnesse e la "rete di reti" che ne

nacque prese il nome iniziale di "ARPA internet" e successivamente semplicemente di "Internet".

C'è da aggiungere che sino dagli inizi si associarono al progetto Internet organizzazioni di ricerca inglesi, notevole soprattutto fu l'apporto di questi ultimi alla definizione del TCP/IP, chiaramente fu questo il paese europeo dove per prima si diffuse Internet, seguito via via dagli altri.

La prima istituzione italiana a disporre di un collegamento alla rete è stato l'Istituto CNUCE del Consiglio Nazionale delle Ricerche (CNR) nel 1986. Il collegamento ad ARPANET usufruiva di un canale via satellite che faceva parte della rete SATNET (SATellite NETwork), ed era frutto di un progetto congiunto CNR, TELESPAZIO e ITALCABLE.

L'Italia fu dunque la quarta nazione europea collegata, dopo la Norvegia, l'Inghilterra e la Germania Ovest.

Tutte le specifiche tecniche alla base del funzionamento di Internet, siano esse quelle dei protocolli che degli applicativi di rete sono descritte da documenti chiamati *Request For Comment* (RFC). Questi documenti sono pubblici ed ognuno può consultarli. In base alla filosofia di Internet, cui abbiamo accennato, chiunque può proporre un nuovo RFC; prima che esso sia adottato e diventi operativo però deve essere sottoposto tramite delle particolari procedure ad una discussione (elettronica naturalmente) che ne valuti le caratteristiche e l'opportunità, qualora il parere sia positivo quanto proposto nel documento potrà diventare uno standard di Internet.

La connessione e l'accesso

Ci sono tre fasi che consentono ad un utente di accedere ad una rete, una prima fase nella quale egli dovrà installare sul proprio computer dei programmi che gli consentiranno di connettere il computer alla rete e di poterla utilizzare, utilizzando questi programmi egli dovrà connettere fisicamente il proprio computer ad un nodo della rete, una volta avvenuta la connessione potrà usare la rete.

Esistono vari tipi di connessione ed accesso, di seguito vengono brevemente descritti alcuni tipi di connessione ed accesso alla rete Internet (ma validi in generale).

Connessione

Connessione tramite linea telefonica

Il computer attraverso un modem viene collegato con la linea telefonica, la connessione viene attivata telefonando ad un'altro modem collegato a sua volta ad un computer già connesso ad Internet, si parla in questo caso di *connessione in linea commutata*

Connessione tramite cavi dedicati per trasmissione dati

Tramite particolari interfacce chiamate schede di rete, il computer è collegato ad un cavo che consente il passaggio dei dati, all'altro capo c'è ancora un computer già connesso in rete; i cavi usati possono essere il doppino, il coassiale, le fibre ottiche ecc. .

Trasmissione "a vista"

In questo caso i dati vengono trasmessi "nell'aria" usando particolari interfacce che emanano raggi infrarossi, onde radio, laser oppure microonde.

Breve analisi dei diversi tipi di connessione

Ciascuno di questi mezzi presenta determinate caratteristiche:

la **linea commutata** viene usata generalmente per connessioni di tipo temporaneo, si attiva quando se ne ha bisogno e appena terminato l'accesso si disattiva; attualmente è il mezzo di connessione più economico. I modem in commercio consentono una velocità di trasmissione dati sino a 33.6 Kbps. Esistono tuttavia dei limiti dovuti alle caratteristiche tecniche della rete telefonica pubblica basata sul doppino di rame, che lo rendono inadatto alle moderne esigenze di comunicazione (trasmissione dati, fax, video, voce). È stato perciò

definito un nuovo sistema chiamato *Integrated Services Digital Network* (ISDN) progettato appositamente per integrare i servizi di voce, video e trasmissione dati (digitale). In Italia, attualmente la rete ISDN è in espansione ed il costo di una connessione è in ribasso, ovviamente le prestazioni sono superiori a quelle ottenibili con una connessione alla normale rete telefonica. Nel caso di una connessione ISDN sono necessari particolari apparecchiature (schede di interfacciamento) sul proprio computer e il modem non è più necessario essendo la trasmissione completamente digitale.

Il **cavo dedicato** è utilizzato generalmente per connessioni di tipo permanente, le velocità di trasmissione dati sono molto elevate e con l'impiego delle tecnologie più avanzate arrivano all'ordine delle centinaia di Mbps.

La connessione tramite **trasmissione "a vista"** non è oggi molto usata, essa tuttavia presenta delle caratteristiche di elevata velocità di trasmissione dati (paragonabile a quelle ottenibili con i cavi) anche se può risentire di fenomeni atmosferici. Numerosi passi avanti si stanno facendo per quanto riguarda l'uso del satellite.

L'accesso

Una volta effettuata la connessione si possono attivare le procedure per accedere ad Internet. Possiamo individuare tre tipi principali di accesso.

Accesso tramite gateway

E quello che avviene quando ci si connette ad una rete che non usa TCP/IP.

Nell'insieme delle reti che costituiscono Internet, ve ne sono alcune che non usano il TCP/IP ma altri protocolli ed altri applicativi di rete. Queste sottoreti sono collegate ad Internet tramite particolari computer chiamati *gateway* il cui compito è quello di ricevere i dati inviati da un nodo della sottorete in base al particolare protocollo di comunicazione, trasformarli in dati TCP/IP e inviarli al destinatario; naturalmente farà anche il lavoro contrario provvedendo a trasformare e instradare i dati provenienti da Internet. Di fatto i gateway compiono operazioni di "traduzione" da un protocollo all'altro. Questa operazione di traduzione avviene "dietro le quinte"; l'utente, una volta connesso, userà i normali applicativi di rete senza preoccuparsi se, quando trasferisce i dati, lo fa all'interno della stessa rete oppure in un'altra via Gateway. La rete forse più grande collegata ad Internet tramite gateway è quella di America On Line (AOL) che conta milioni di utenti; altri esempi sono le reti UUCP e le reti di BBS quali FidoNet ecc..

Accesso tramite host

Questo tipo di accesso si ha quando la connessione permette di accedere ad un computer (host) in modalità terminale e l'accesso ad Internet avviene tramite gli applicativi di rete presenti sull'host. Per questo tipo di accesso può essere sufficiente un normale programma di emulazione terminale

Accesso diretto

Con questo tipo di accesso il computer diventa un vero e proprio nodo della rete Internet. Sul computer una volta stabilita la connessione è possibile usare tutti gli applicativi di rete installati.

Breve analisi dei diversi tipi di accesso

È importante notare come l'accesso sia indipendente dal tipo di connessione usata.

L'accesso tramite gateway è in genere il più limitato, ossia quello che consente di utilizzare meno le risorse di Internet; non sempre infatti è possibile creare dei gateway che riescano ad integrare completamente tutti i servizi "interni" alla rete originaria.

L'accesso tramite host consente in teoria il pieno utilizzo dei servizi della rete Internet, ma comporta una serie di inconvenienti. Ad esempio qualora si volesse copiare un file da un computer remoto lo si dovrebbe fare in due fasi: in un primo momento sull'host e successivamente, usando un altro programma di comunicazione, lo si può trasferire sul proprio computer. Ancora: qualora non si abbia la possibilità di installare programmi sull'host, ogni volta che si vuole utilizzare un nuovo applicativo oppure aggiornarne uno vecchio si deve fare richiesta all'amministratore dell'host ecc. .

L'accesso diretto consente il pieno utilizzo dei servizi di Internet, gli unici limiti sono dati dalla potenza del computer (intesa sia come capacità di calcolo che come disponibilità di memoria). Grazie ai protocolli *SLIP* e *PPP* è possibile effettuare accesso diretto utilizzando anche connessioni in linea commutata.

Gli indirizzi Internet e la Posta Elettronica

Indirizzi e domini

Quando qualcuno inizia a usare degli applicativi di rete per accedere ad Internet si trova disorientato nel leggere gli indirizzi, ovviamente è una questione di abitudine; è utile però cercare di capire come sono fatti e quali informazioni possono fornirci. Partiamo subito con un esempio di indirizzo di posta elettronica:

mario@labnet.cnuce.cnr.it

Possiamo dividere la stringa in tre parti:

- _ il nome dell'utente: mario
- _ il simbolo "@"
- _ il vero e proprio indirizzo della macchina su cui Mario riceve la posta: "labnet.cnuce.cnr.it", questa parte prende il nome di *Fully Qualified Domain Name* (FQDN) o più semplicemente domain name (noi vi faremo riferimento spesso come "dominio")

Il dominio non è il vero identificativo di un nodo nella rete Internet. Ogni computer collegato ad Internet viene identificato univocamente dal suo:

indirizzo IP (IP sta per Internet Protocol)

Ogni nodo Internet ha almeno un indirizzo IP (in alcuni casi può averne più d'uno) che è costituito da 4 numeri compresi tra 0 e 255 separati dal simbolo "." ad esempio quello della macchina su cui Mario riceve la posta è:

131.114.1.171

ciascuna parte di un indirizzo IP (ad es "131") prende il nome di *octet*.

Gli indirizzi IP e i domini non sono assegnati arbitrariamente, a farlo è un'organizzazione chiamata Network Information Center (NIC) la quale si comporta seguendo determinati criteri. Quando una azienda o una organizzazione vuole connettere la propria rete ad Internet, richiede al NIC un insieme di indirizzi IP; gli indirizzi ottenuti saranno assegnati ai nodi della rete. Inoltre propone un nome per il proprio dominio e se accettato quello diventerà il nome con il quale verrà conosciuta la sua sottorete.

Diverso è il caso del singolo utente che si rivolge ad un Internet Service Provider (ISP) per accedere ad Internet; in questo caso sarà il provider a fornire l'indirizzo IP all'utente prendendone uno tra quelli ad esso concessi dal NIC.

È assolutamente indifferente riferirsi ad un nodo della rete Internet con l'indirizzo IP o con il FQDN. Poiché per una persona è molto più semplice ricordare un nome che un numero, esiste un "meccanismo" chiamato Domain Name System (DNS) che associa automaticamente al FQDN l'indirizzo IP corrispondente. Il DNS è realizzato tramite dei

particolari nodi della rete (Name Server) che hanno il compito di tradurre i domini in indirizzi IP. Se qualcuno invia un messaggio a Mario via e-mail indirizzandolo con la stringa del nostro esempio, avviene “dietro le quinte” un colloquio preliminare tra il computer del mittente ed il Name Server per tradurre il FQDN in indirizzo IP; una volta che la traduzione è avvenuta, il messaggio viene effettivamente spedito.

Il meccanismo di assegnazione degli indirizzi e domini sta per essere modificato anche se attualmente non si è giunti ad un accordo su come dovrà essere né su chi dovrà occuparsene.

“Interpretiamo” un dominio

Tornando all'esempio precedente cerchiamo di “interpretare” il dominio: labnet.cnuce.cnr.it per cercare di capire come sono fatti i domini. A tale scopo prendiamo in considerazione le singole stringhe separate dal “.” che compongono il FQDN a partire dall'ultima fino ad arrivare alla prima. L'ultima stringa del nostro dominio è:

it

che sta per: Italia. Questo serve a far capire che il computer che stiamo indirizzando è connesso ad una delle reti italiane che fanno parte di Internet, in sostanza sta (quasi certamente) in Italia. L'ultima stringa di un dominio può quindi essere la sigla internazionale che individua una nazione es:

fr (Francia),
jp (Giappone),
uk (Regno Unito),
ca (Canada),

ma ci sono anche dei casi particolari come:

“edu” che individua i computer delle università o degli istituti di ricerca degli Stati Uniti,
“com” per le compagnie commerciali,
“gov” individua i siti appartenenti al governo statunitense,
“net” individua le organizzazioni amministrative della rete,
“mil” siti militari dell'esercito statunitense,
“org” altre organizzazioni.

Queste stringhe prendono il nome di *Top Level Domain* (TLD). Per effetto dell'esplosione commerciale di Internet da più parti si propone di aumentare i TLD introducendone altri quali “firm”, “home” etc., entro breve tempo si potrà sapere quali nuovi TLD saranno introdotti e soprattutto chi li assegnerà e in base a quali criteri.

Abbiamo visto che poiché il nostro indirizzo termina con “it” il computer fa parte di una delle reti collegate ad Internet che si trovano in Italia, ma di quale rete si tratta?

A questa domanda risponde la seconda stringa che incontriamo (sempre procedendo da dx a sx) e cioè:

cnr

Da questa sigla possiamo capire che si tratta della rete del Consiglio Nazionale delle Ricerche. Come dice il nome il CNR ha carattere nazionale, la sua rete unirà quindi le reti degli istituti di ricerca sparsi sul territorio. La stringa

cnuce

dice in quale delle reti "interne" al cnr si trova il computer, precisamente in quella dell'istituto CNUCE. Finalmente incontro l'ultima stringa prima del simbolo "@" che mi dice il nome vero è proprio della macchina su cui Mario riceve la posta e cioè labnet.

Se esaminiamo quindi l'indirizzo:

armstrong@osiride.hq.nasa.gov

possiamo capire che il computer indirizzato si trova all'interno di una delle reti degli organi governativi statunitensi e precisamente in quella che collega gli istituti della NASA; a questo punto hq potrebbe significare headquarter (quartier generale) e il computer dove Mr. Armstrong riceve la posta si chiama osiride.

Non sempre tuttavia le cose sono così chiare, in quanto spesso gli acronimi usati possono essere fraintesi o oscuri; è comunque buona norma cercare di scegliere nomi che identifichino il più possibile il dominio con l'entità a cui appartiene.

Modalità client-server

Attualmente quasi tutti i servizi di Internet sono progettati ed implementati secondo il modello "*Client/Server*". In base a questo modello ciascun servizio è diviso in tre parti:

il programma *client*

il programma *server*

il protocollo di comunicazione fra client e server.

Chi intende usare un servizio usa un programma client, chi fornisce un servizio usa un programma server; il protocollo consente ai due tipi di programmi di comunicare.

Ad esempio nel caso del trasferimento di un file, il client viene usato per collegarsi al server che gestisce l'archivio dove è memorizzato il particolare file. Terminata l'operazione il client può essere disattivato, il server ovviamente continuerà a restare attivo in attesa di una nuova richiesta di accesso. Riferendoci a quanto detto a proposito delle connessioni, il server generalmente avrà una connessione permanente, mentre il client potrà essere implementato su computer con una connessione temporanea. Client e server usualmente stanno su computer diversi, ma su uno stesso computer possono essere attivi allo stesso tempo programmi client e programmi server.

La posta elettronica

Lo scambio di messaggi di posta elettronica (e-mail) è oggi uno dei principali modi di utilizzo della rete Internet. La semplicità di utilizzo dei moderni programmi di posta elettronica, ne ha diffuso enormemente l'uso al di fuori dei confini nei quali veniva adoperata sino a qualche anno fa, cioè la comunità scientifica. Un indirizzo Internet di posta elettronica (Internet e-mail address) ha una forma di questo tipo:

- _ nome dell'utente
- _ simbolo "@"
- _ dominio

Ci rendiamo subito conto di come una tale stringa contenga praticamente tutte le informazioni necessarie per individuare univocamente il destinatario. È importante dire che il destinatario di un messaggio non è necessariamente una persona; esistono infatti dei programmi (ne vedremo in seguito esempi) che accettano i comandi via e-mail.

Anatomia di un "Mail Header"

Una e-mail ha una struttura specifica, definita dal RFC-822. Il testo vero è proprio è preceduto da un certo numero di righe che costituiscono il cosiddetto "mail header". Queste righe a prima vista possono sembrare oscure, tuttavia poche di esse hanno un effettivo interesse per l'utente.

```
From romano@governo.it Sat May 25 10:06:01 1996
Received: from governo.it by house.gov with SMTP id AA21901
(4.1/SMI for bill.clinton@whitehouse.gov); Sat, 25 May 96 17:05:56 -0200
Date: Sat, 25 May 96 17:05:56 -0100
From: The President
Message-Id: <9105252105.AA06631@hq.it>
To: bill.clinton@whitehouse.gov
Subject: Meeting
```

Hi Bill .. we have a meeting next week in Lione. Please don't forget this time.

Analizziamo l'esempio:

- _ Le prime tre linee, non sono molto interessanti, ci dicono il vero indirizzo da cui la e-mail proviene, infatti questo può non coincidere con quello a cui si dovrebbe rispondere,
- _ La linea che inizia con "Date:", contiene il giorno e l'ora in cui il messaggio è stato inviato.
- _ Nella linea che inizia con "From:", troviamo in genere il nome di chi spedisce la e-mail, ciascuno può però decidere se inserire altri dati invece del nome.
- _ La linea che inizia con "Message-ID:", contiene informazioni di scarso interesse per l'utente;

- _ La linea che inizia con "To:" contiene il destinatario (o la lista dei destinatari) del messaggio; può essere presente anche una linea che inizia con "Cc:" con uno o più indirizzi aggiuntivi.
- _ Nella linea che inizia con "Subject:" troviamo una breve frase che sintetizza il contenuto del messaggio.

L'ordine esatto con il quale le righe di un mail header si possono presentare e anche il numero delle stesse varia in relazione ai programmi o ai computer usati.

Indirizzo errato

Quando un indirizzo di posta elettronica non è corretto, perchè è sbagliato il dominio oppure l'utente destinatario non esiste, il mittente riceverà una e-mail contenente un messaggio di errore, che spiegherà anche i motivi del "mancato inoltro".

```
From: Mail Delivery Subsystem
Date: Sat, 25 May 91 16:45:14 -0400
To: mario@labnet.cnuce.cnr.it
Cc: Postmaster@bande.desinnee.fr
Subject: Returned mail: User unknown
```

```
----- Transcript of session follows -----
While talking to bande.desinnee.fr:
>>> RCPT To:
<<< 550 ... User unknown
550 lsimpson... User unknown
```

Questo messaggio spiega a Mario che l'utente "lsimpson" non esiste e quindi ovviamente la e-mail non ha raggiunto la destinazione. Come si può leggere una copia del messaggio (Cc:) è stata inviata al gestore della posta relativa al dominio del destinatario. Nel corpo della e-mail sarà presente anche il messaggio che Mario ha inviato.

Mailing Lists

Una *mailing list* è uno strumento che consente ad un gruppo di persone di discutere (spesso le mailing list vengono indicate anche come "gruppi di discussione") o scambiarsi informazioni via e-mail. Per poter partecipare ad una mailing list bisogna essere degli "iscritti", l'iscrizione avviene inserendo il proprio indirizzo e-mail in una lista che contiene gli indirizzi delle persone che ne fanno parte (membri).

La lista degli iscritti è memorizzata su un calcolatore ed è associata ad un particolare indirizzo e-mail. Ogni volta che uno dei membri invia un messaggio di posta elettronica a quel particolare indirizzo, un programma provvede a distribuire il messaggio a tutti gli iscritti.

All'interno di una azienda i cui uffici sono in rete, in genere ci sono mailing list dove vengono inviate le circolari amministrative, oppure mailing list fra i partner di uno stesso gruppo di lavoro dislocati in uffici diversi. Allo stesso modo sulla rete Internet esistono migliaia di gruppi di discussione che trattano argomenti che vanno dall'educazione dei minori, alla programmazione in assembler, al calcio.

Le mailing list sono gestite da un amministratore di lista (*owner*) il quale principalmente deve:

redigere un documento che indichi chiaramente quale argomento viene trattato nella lista e in quali modalità, questo documento viene indicato con il nome di *charter*
assicurarsi che il tutto funzioni correttamente dal punto di vista tecnico
garantire che la discussione rimanga *on topic*, ossia che gli interventi siano attinenti l'argomento per cui la lista è stata realizzata

Le mailing list possono essere:

aperte: l'iscrizione è libera, ciascuno scrive direttamente a tutti gli altri;
chiuse: è possibile iscriversi solo in base a determinate condizioni;
moderate: l'iscrizione può essere libera o no, esiste un "moderatore" che riceve tutte le e-mail e decide quali sono pertinenti e quali scartare inviando a tutti gli altri solo le prime, in queste liste spesso owner e moderatore sono la stessa persona.

Listserver, la mailing list Labnet

Tutte le liste di una certa dimensione sono gestite tramite dei particolari programmi chiamati, listserver che automaticamente amministrano la lista. Un qualsiasi listserver svolge almeno i seguenti compiti:

- _ accetta nuove iscrizioni/cancella gli indirizzi delle persone che non vogliono più far parte della lista,
- _ invia dei file a chi chiede informazioni sulla lista,
- _ archivia le e-mail e periodicamente prepara dei file con digest dei messaggi scambiati

Questi compiti vengono eseguiti in risposta a determinati comandi inviati tramite e-mail.

Nel caso ad esempio della lista labnet il cui indirizzo è

labnet@cnuce.cnr.it

la richiesta di iscrizione va effettuata inviando una e-mail a:

listserv@listserv.cnuce.cnr.it

con il subject vuoto e come unica riga di testo:

SUBSCRIBE labnet

cioè la parola chiave "SUBSCRIBE" che indica quello che si vuole fare ed il nome della lista a cui ci si vuole iscrivere. Una volta ricevuto il comando il listserver aggiunge l'indirizzo

Internet (prendendolo dall'header della mail) del mittente alla lista dei membri. Se da questa lista si vuole essere esclusi bisognerà inviare allo stesso indirizzo il comando:

SIGNOFF labnet

in questo caso il server provvederà a togliere l'indirizzo del mittente dalla lista. Ci sono numerosi altri comandi che consentono di fare richieste particolari, di sapere l'elenco degli iscritti quando la lista è aperta etc. Per conoscerli tutti basta inviare un comando al server al solito con subject vuoto e con nel corpo della lettera il comando

help

il file che vi verrà inviato in risposta conterrà tutte le informazioni. È sempre consigliabile prima di iscriversi ad una lista inviare un comando di help, la risposta potrebbe contenere informazioni preziose.

Attenzione quindi, le e-mail vanno inviate all'indirizzo della lista e i comandi a quello del listserver.

FTP, Anonymous FTP e FTPMail

File Transfer Protocol

File Transfer Protocol (FTP definito da RFC 959) è il metodo principale per trasferire i file sulla rete Internet. L'applicativo FTP è basato sul modello client/server, il server è chiamato "FTP server" ed il computer sul quale esso è installato prende il nome di sito FTP. Compito del FTP server è quello di gestire un archivio di file inviandoli e/o ricevendoli ai/dai client che ne fanno richiesta collegandosi.

Vediamo quali passi bisogna effettuare per copiare un file da un FTP server al proprio computer (o viceversa):

1. Attivare il proprio FTP client,
2. Fornire al client l'indirizzo del server FTP,
3. Fornire al client la propria login di accesso al server FTP,
4. Fornire al client la password di accesso al server FTP,
5. Fornire la directory a cui si vuole accedere (opzionale),
6. Collegarsi,
7. Copiare il (o i) file.

L'indirizzo di un server FTP è dato dal FQDN oppure dall'indirizzo IP del computer sul quale è installato. Per poter accedere all'archivio di file, è necessario essere riconosciuti dal server ed il meccanismo di identificazione si basa sulla user ID e sulla password. Se si conosce, si può inserire nel client anche la localizzazione esatta del file all'interno dell'archivio da cui si vuole copiare.

Se si conosce l'archivio a cui ci si collega, la ricerca di un file è semplice; ma se ci si collega per la prima volta ad un archivio magari molto grande la cosa può risultare problematica.

Gli archivi gestiti da FTP server sono strutturati ad "albero" (più propriamente a grafo). Collegandosi ad un server senza fornire la directory si accede alla radice di questo albero, da essa si dipartono i rami all'interno dei quali sono dislocati i file. Non esiste un modo standard di disporre le informazioni all'interno di un server FTP, esistono però degli accorgimenti generalmente adottati dagli Information Providers che possono facilitare le cose a chi accede.

Per prima cosa quasi tutti i server forniscono una schermata all'inizio del collegamento che contiene delle informazioni utili. Come esempio riportiamo un estratto della schermata iniziale del FTP server del Dartmouth College:

```
220-  
220-      Welcome to the Dartmouth College FTP Server.  
220-  
220-      Unauthorized or illegal use of this archive is prohibited.  
220-      Send comments or complaints to ftp-admin@ftp.Dartmouth.EDU.  
220-
```

220- *Your hostname and userid are logged for all transfers.*
220- *If you object to that, please disconnect now.*
220-
220 *webster.dartmouth.edu FTP server (Version wu-2.4.2 Wed Feb*
220 *21 12:40:40 EST 1996) ready.*
230-
230- *This archive includes contributions from Dartmouth College students,*
230- *faculty, and staff. For details, please refer to the README files*
230- *located in most directories.*
230-
230- *Dartmouth Macintosh software, including Blitzmail and Fetch, are in*
230-
230- */pub/software/mac*
230-
230- *You can use the directories /incoming and /outgoing to exchange files*
230- *with Dartmouth College users. Please do not abuse this service.*

Dal testo si hanno informazioni su come raggiungere via e-mail il gestore del server, quale è la politica di controllo degli accessi adottata, chi fornisce il software presente sul server e come localizzarlo.

Per quanto riguarda la localizzazione del software sul server, viene suggerito di leggere i file "README" presenti nelle varie directory. Questi file contengono informazioni sul contenuto della directory stessa ed è questo il secondo accorgimento generalmente adottato nei server FTP per aiutare chi accede ad orientarsi. Continuando con lo stesso sito FTP vediamo un file README:

This is the Dartmouth College anonymous FTP directory. It contains the following items. If a person's name is listed after an item, you may contact that person at Firstname.Lastname@Dartmouth.EDU. Otherwise, send comments or complaints to ftp@ftp.Dartmouth.EDU.

6300+ This archive of programs for the AT&T 6300 computers has
 moved to summit.whoj.edu in /pub/6300+.

Dante The Dartmouth Dante Project manuals. (Dante)

Exceptions Computer Files for "Exceptions are the Rule." (Joel Levine)

Hyperbooks Hypercard versions of literary classics. (Espen Koht)

prologue-users Information for users of Prologue Systems, Inc., ticketing
 systems. (Alan Fahrner)

protein The Secretary Pathways Protein Database. (Jon Rothblatt)

security Security and password work. (Matt Bishop)

software Various Dartmouth-written software. (Rich E. Brown)

Ci sono dunque i nomi dei file presenti nella directory seguiti da una breve descrizione, ovviamente nel caso delle directory la descrizione è relativa al contenuto della stessa.

È bene ribadire che la maggior parte dei server adottano tali accorgimenti, ma può capitare di collegarsi a server dove l'unico aiuto è quello dato dai nomi dei file.

Diversi tipi di FTP Server

I server FTP possono essere privati oppure pubblici. In entrambi i casi come detto, per accedere è necessario identificarsi tramite una User ID ed una password; ma mentre nel caso dei server privati le due parole chiave sono conosciute solo dagli autorizzati, per tutti i server pubblici esiste una parola standard da fornire come user ID ed è:

anonymous

mentre come password nei server pubblici è buona norma inserire l'indirizzo di posta elettronica di chi accede; questo per consentire al gestore di tenere traccia di chi utilizza il server. I server FTP pubblici sono detti Anonymous FTP server.

Ci sono tantissimi Anonymous FTP server e in alcuni di essi il numero degli accessi arriva ad essere talmente alto da rendere lentissima ogni operazione di trasferimento file. Quando questo avviene sistematicamente, si ricorre alla replica delle informazioni contenute in questi server, su altri siti FTP; questi prendono allora il nome di "mirrors". Programmi particolari provvedono ad effettuare aggiornamenti automatici dei mirrors, ogni volta che c'è una variazione sul server originario.

Binario e ASCII

Chi trasferisce un file da un computer ad un altro copiandolo si aspetta che le due copie siano identiche; affinché questo sia possibile nel trasferimento di file via FTP, è necessario adottare un piccolo accorgimento.

I file che possono essere memorizzati su un computer (e quindi trasferiti via rete) si possono raggruppare in due categorie:

- file di testo: sono quei file che possono essere letti correttamente con qualsiasi text editor;
- file binari: appartengono a questa categoria i file eseguibili, i file che codificano immagini o suoni, i file che codificano testi strutturati in modo particolare e che possono essere letti solo con particolari tipi di text editor, ecc.;

Vediamo che differenze ci sono nel trasferimento dei due tipi di file.

1. Supponiamo di voler trasferire un file di testo via FTP. È noto che ogni tipo di computer ha un suo "set di caratteri" cioè un suo modo di rappresentare i caratteri con dei codici alfanumerici; per intenderci la lettera "a" che compare sul monitor viene registrata nella memoria di un calcolatore con una sequenza di cifre e lettere e questa sequenza può essere diversa da calcolatore a calcolatore. Si può verificare allora la situazione in cui alcuni codici

che il server interpreta in un modo, possono essere interpretati in maniera diversa dal client dell'utente. L'utente quindi trasferendo un file di testo via FTP rischia di ritrovarsi con un documento non perfettamente leggibile e comunque diverso da quello originario. In questi casi può essere allora necessario modificare la rappresentazione del codice perché i dati restino gli stessi.

2. Esaminiamo ora il caso di un file binario, per esempio un eseguibile. Tale file, una volta copiato, deve essere "eseguito" correttamente,; non si deve quindi correre il rischio di modificare qualcosa rendendolo inservibile. In questo caso è necessario che il file venga memorizzato sul client esattamente come sul server.

A ciascuno di questi due tipi di file è stato associata una "modalità" di trasferimento. Per il trasferimento di un file di testo si adotterà la modalità:

ASCII

se invece si deve trasferire un file binario la modalità da adottare sarà:

Binary

Prima di eseguire il trasferimento del file bisogna dunque stabilire quale modalità usare in base al tipo di file, e predisporre il client ad adottarla. Se viene adottata la modalità ASCII, client e server FTP adottano un meccanismo che "traduce" il file che viene inviato, dal codice del computer sul quale è installato il server a quello del computer destinatario. Il sistema si basa su un codice intermedio chiamato "standard 8 bit NVT-ASCII" ed ha il seguente funzionamento: ogni qualvolta il server invia un documento, esso traduce i dati dal proprio codice al codice intermedio, il client poi traducendo dal codice intermedio a quello della macchina destinataria, ricostruisce il documento originale. Se viene adottata la modalità Binary, invece, il trasferimento avviene senza che nulla venga modificato. I client più avanzati possono a volte riuscire a scegliere automaticamente la modalità di trasferimento.

FTPMail

Esiste la possibilità di accedere via e-mail ad un FTP server tramite dei particolari programmi chiamati "FTPMail". È possibile inviare a questi programmi dei comandi per chiedere loro di accedere ad un sito FTP ed eseguire delle operazioni. Il risultato di queste operazioni sarà spedito al richiedente ancora tramite e-mail. Se ad esempio la richiesta è quella di trasferire un file dal server sul proprio computer, il programma di FTPMail provvederà ad accedere il sever e spedire all'indirizzo del richiedente il file inserendolo in una e-mail.

Uno dei programmi di FTPMail più usato è installato sul computer il cui dominio è doc.ic.ac.uk; l'indirizzo da usare per attivare il programma di FTPMail è:

ftpmail@doc.ic.ac.uk

Per avere informazioni su come usarlo bisogna inviare una e-mail all'indirizzo specificato con subject vuoto e nel corpo del messaggio la sola parola:

help



World-Wide Web

World-Wide Web , è lo strumento più nuovo per la diffusione delle informazioni sulla rete, l'applicativo che ha innescato la vera e propria esplosione di Internet avvenuta negli ultimi anni.

Cenni storici

World-Wide Web , chiamato anche semplicemente *web*, o *WWW*, o *W³*, nasce da un progetto iniziato nel marzo del 1989, al CERN (Conseil Européen pour la Recherche Nucléaire) , ed il suo ideatore si chiama Tim Berners-Lee. Il progetto iniziale era quello di realizzare un semplice sistema per trasmettere documenti e per permettere la comunicazione tra i fisici dislocati nei diversi paesi, utilizzando la rete Internet.

Nell'estate del 1991 i risultati (le specifiche ed i programmi) del "WWW program" vengono diffusi su Internet attraverso le newsgroup. Ben presto molti si accorgono dell'importanza di questi risultati e mentre il CERN lentamente si "defila", dapprima NCSA (National Center for Supercomputing Applications) e successivamente diverse compagnie commerciali partendo dall'idea iniziale, realizzano strumenti sempre più sofisticati per utilizzare il web cominciando ad introdurre l'idea di sfruttamento commerciale di Internet.

World Wide Web

Gli applicativi che consentono di accedere al web sono organizzati secondo il modello client/server. I client sono chiamati *browser*, i server sono chiamati "WWW server" oppure "HTTP server" e i computer su cui sono installati sono chiamati siti web oppure siti www (il particolare protocollo di comunicazione HTTP è definito da RFC 1945).

Le innovazioni fondamentali introdotte dal web sono principalmente tre:

1. possibilità per l'Information Provider, di organizzare le informazioni in maniera "ipermediale" e distribuita;
2. utilizzo di un'unica interfaccia grafica per accedere a diversi tipi di dati: (documenti testuali, immagini, animazioni, video e suoni);
3. il fatto che in un unico client (il browser) siano "inglobati" i client dei principali servizi di Internet.

Il concetto di ipermediale si basa su quello di ipertesto, un ipertesto è un normale testo che contiene al suo interno dei collegamenti ("link") con altri testi; leggendo al computer un ipertesto con un semplice "clic" del mouse si può passare da un testo all'altro. Ad un documento ipermediale, oltre ai testi, possono essere collegati immagini, animazioni e suoni. I link possono riferirsi a file (documenti) che si trovano sullo stesso sito, oppure a file memorizzati su un qualsiasi altro nodo della rete; questo consente di realizzare dei documenti

“distribuiti”, ossia documenti in cui le varie parti possono fisicamente trovarsi in luoghi diversi; le varie parti di cui è composto un documento ipermediale, sono chiamate anche “pagine web”.

Chi accede ad un HTTP server tramite un browser, può passare da una pagina web all'altra, semplicemente seguendo i link con il mouse, ossia con un "clic" sulle parole (o immagini) inserite nel testo e appositamente evidenziate (questo è ciò che viene indicato col termine di “navigazione” o *browsing*). Per l'utente è completamente trasparente il fatto che nel seguire un link, la pagina visualizzata si trovi fisicamente su un computer diverso da quello acceduto, inoltre i documenti ipermediali possono essere letti senza la necessità, come avveniva su Internet prima dell'avvento del WWW, di ricorrere ad un programma diverso per ogni tipo di dato (uno per le immagini, un altro per i suoni etc) e quindi senza necessità per l'utente di possedere particolari conoscenze tecniche.

Elemento di innovazione non meno importante degli altri due, è la possibilità di poter accedere con il browser anche a server che non sono HTTP server; più precisamente tutti i browser consentono l'accesso a :

FTP server,

NNTP server (i server attraverso i quali vengono distribuite le news di Internet),

GOPHER server (un sistema per l'accesso ai documenti basato su menù gerarchici).

Inoltre, i browser più sofisticati, hanno una propria interfaccia per i programmi di e-mail e per gli altri principali applicativi di Internet. Quindi, se prima erano necessari più programmi per usare i servizi di Internet, ora ne esiste uno solo che li raggruppa tutti e questo può dare una idea del perché molto spesso oggi, si confonde il WWW, con Internet stessa.

Il Server

L'indirizzo di un HTTP server ha la seguente forma:

`http://nome.e_dominio_del.server:porta/percorso/da/seguire/`

dove per nome.e_dominio_del.server si intende il FQDN oppure l'indirizzo IP del sito web. Da notare che la stringa inizia con l'indicazione del particolare protocollo di comunicazione usato; gli indirizzi aventi questa forma prendono il nome di Uniform Resource Locator (URL).

Ad esempio l'URL del web server del CNR di PISA è il seguente:

`http://www.pi.cnr.it`

I server possono essere pubblici o privati; quelli pubblici sono liberamente accessibili, quelli privati generalmente sono costituiti da una parte pubblica e da una parte alla quale si accede con una procedura di riconoscimento, che prevede l'introduzione di una user ID e di una password.

Accedendo ad un server, senza specificare il percorso, si accede ad una pagina di default chiamata “home page” del server, ma occorre precisare che questo termine è spesso utilizzato anche per indicare la prima pagina web di un documento ipermediale. In un server tuttavia, non troviamo solo dei documenti ipermediali in lettura, ma è possibile accedere a pagine costruite in

modo da permettere all'utente di "interagire" con il server stesso. Ad esempio, vi possono essere delle bacheche in cui ognuno può lasciare delle informazioni, oppure, nel caso di siti commerciali, vi può essere la possibilità di ordinare un articolo; queste pagine web prendono il nome di FORM. Una FORM è una sorta di "scheda" elettronica e una volta terminata la sua compilazione basterà "cliccare" con il mouse l'apposito "bottone" ed i dati inseriti saranno inviati ad una applicazione integrata con il server (CGI) che provvederà ad elaborarli .

Il browser

Il primo browser è stato il Mosaic del CERN, più tardi divenuto "NCSA Mosaic". Attualmente esistono molti tipi di browser, il più diffuso è "Netscape Navigator" della Netscape Inc., ma si sta prepotentemente diffondendo anche "Internet Explorer" della Microsoft; anche i browser più sofisticati comunque sono disponibili gratuitamente sulla rete. Tuttavia possono esistere sostanziali differenze tra i browser, per la loro diversa implementazione delle più recenti versioni del linguaggio HTML. Utilizzare un browser è molto semplice ed intuitivo e, oltre che per "navigare" sulla rete, il browser può essere usato anche "in locale", come un normale visualizzatore di documenti ipemediali.

I documenti web e Hypertext Markup Language

La progettazione di un documento ipermediale per il web si articola in due livelli:

progettazione della "struttura" dell'intero documento, cioè quante e quali parti comporranno il documento e quali saranno i link , tenendo presente che un documento ipermediale per sua natura può, in qualsiasi momento, essere arricchito con ulteriori informazioni;

progettazione della struttura delle informazioni all'interno delle pagine web; cioè stabilire quale sarà il formato del testo quali immagini inserire nella pagina, quali frasi "linkare" e come evidenziarle, se inserire animazioni etc.

Per realizzare le pagine web viene utilizzato il formalismo HTML. Una pagina web visualizzata con un normale editore di testi risulta essere una pagina di testo con degli "elementi HTML" inseriti. Quello che segue è l'esempio di un banale documento HTML:

```
<HTML>
<HEAD>
<TITLE>Un Semplice Documento</TITLE>
</HEAD>
<BODY>
<H1 align="center"> Esempio di Pagina Web</H1>
```

Hello world!

<HR>

</BODY>

</HTML>

Si provi per esercizio a scrivere in un file ASCII il testo dell'esempio e visualizzarlo con il proprio browser...

Il formalismo HTML si evolve parallelamente all'evolversi del web, attualmente "HTML Level 3.2" è la versione ufficiale. La diversità delle possibili applicazioni e l'enorme diffusione dello strumento Worl-Wide Web, ha incoraggiato però singole persone e grandi gruppi commerciali (Netscape e Microsoft in testa), a fornire alla comunità di Internet sempre nuovi strumenti in alcuni casi prescindendo dalla procedura di standardizzazione. Il pericolo quindi è che il tentativo di far avanzare gli standard in maniera coordinata ed uniforme possa essere destinato a fallire, superato dalla continua rincorsa dei vari produttori di software per il Web verso caratteristiche in grado di attirare i clienti verso i propri programmi, a scapito dell'universalità del linguaggio.

Cercare Informazioni

Naturalmente la disponibilità gratuita del software, la relativa semplicità della sua utilizzazione e la facilità di realizzazione dei documenti ipermediali, ha fatto sì che in poco tempo siano sorti tantissimi siti web, con la conseguente enorme crescita delle informazioni offerte sulla rete. Oggi sotto web si può trovare di tutto, dalla divina commedia in inglese con testo originale a fronte, sino ai testi dei sermoni dei pastori della chiesa Anabattista. Non esiste nessun "indice generale del web", orientarsi in questo mare di informazioni è un'impresa a dir poco ardua. Esistono tuttavia degli strumenti che possono aiutare:

i "motori di ricerca" (*search engines*)
gli "indici di rete"

I search engines, sono dei particolari software che periodicamente inviano "agenti software" in giro per il web. Gli agenti sono programmi che visitano le pagine che incontrano e ne memorizzano il contenuto registrando in un archivio gli indirizzi delle pagine visitate ed una parte o tutto il testo contenuto in esse. Questi archivi sono associati a particolari server HTTP collegandosi ai quali un utente può interrogare l'archivio inserendo delle parole chiave che riguardano l'argomento che lo interessa in una *form*. A questo punto viene attivato un programma che cerca nell'archivio le parole chiave inserite, è visualizza, per le informazioni trovate, in un documento ipermediale le seguenti informazioni:

il titolo
una parte del contenuto
il link al sito originario del documento

All'utente basterà usare il mouse per accedere ai documenti originali. E' possibile per un information provider, chiedere di essere inserito nel percorso degli agenti di un motore, rendendo i documenti da lui forniti soggetti a possibile ricerca. Il grosso difetto iniziale dei motori di ricerca era l'enorme "rumore" prodotto dalle ricerche, cercare una parola come "school" poteva avere come risultato alcune centinaia di migliaia di link a documenti, oggi vi sono dei motori di ricerca molto sofisticati (vedi paragrafo sugli indirizzi utili) che consentono di "raffinare" la ricerca inserendo delle condizioni che delimitino il più possibile l'ambito nella quale le parole chiave dovranno essere cercate.

Una precisazione: spesso collegandosi ad un sito si possono trovare dei motori di ricerca locali, questi effettuano ricerche solo sui documenti del server in cui sono installati, non vanno quindi confusi con quelli illustrati sopra.

Gli indici di rete sono dei veri e propri indici divisi per argomenti.

Collegandosi ad un HTTP server che gestisce un indice di rete, viene visualizzata una home page su cui sono indicati tutti gli argomenti trattati dai documenti archiviati. Scegliendo un argomento, si accede ad una pagina con dei sottoargomenti e così via sino a quando non si trova il link a ciò che interessa; a questo punto si passa al sito contenente il documento. In genere questi indici vengono aggiornati su segnalazione degli utenti o degli information provider.

Nessuno dei due strumenti è da considerarsi completo, d'altra parte il continuo crescere del numero dei siti web ed l'incremento dei documenti memorizzati nei server, renderebbe obsoleto un eventuale indice "generale" solo dopo poche ore.

CGI script

Le pagine web oltre che dei puri contenitori di informazione possono essere delle "interfacce" tramite le quali si ottengono informazioni. Un esempio può essere quello dei motori di ricerca che abbiamo descritto nel paragrafo precedente. Una volta compilata la form per una ricerca l'utente attiva un programma che cerca le parole e trasforma i risultati in una nuova pagina HTML che viene visualizzata sul browser. Un programma di questo tipo prende il nome di *CGI script*. Il CGI viene eseguito sul server, solo il risultato dell'esecuzione viene inviato al client. Per realizzare un CGI si può usare in teoria un qualsiasi linguaggio di programmazione, in genere la scelta dipende dal sistema operativo del computer su cui è installato il server.

Java

Negli ultimi due anni le pagine web si sono notevolmente raffinate sia dal punto di vista grafico con animazioni o altri effetti, sia dal punto di vista dell'interazione con l'utente. Questo principalmente per la nascita e/o lo sviluppo di particolari linguaggi di programmazione con i quali è possibile creare programmi che vengono eseguiti quando con un browser si accede ad una pagina web. Il linguaggio che ha avuto maggiore diffusione è Java (altri linguaggi sono Javascript, ActiveX). Con Java si possono creare dei particolari programmi chiamati *applet*, quando una pagina HTML che contiene un riferimento ad un applet viene acceduta con un browser, l'applet viene scaricato sul computer dell'utente che visita la pagina e qui viene eseguito in contemporanea con la visione della pagina.

Attenzione: il programma viene eseguito qualunque sia il sistema operativo del computer con il quale l'utente legge la pagina web, una delle principali caratteristiche di Java è infatti il suo essere *multiplatforma*. Per motivi vari (non ultima la sicurezza: scaricare sul proprio computer un programma ed eseguirlo può essere pericoloso, potrebbe ad esempio generare virus) le cose che si possono realizzare tramite applet sono (ad oggi) limitate, ma Java è un linguaggio di programmazione a tutti gli effetti e con esso è quindi possibile realizzare delle vere proprie applicazioni. In particolare è possibile realizzare delle applicazioni "distribuite" dove uno o più

applet Java eseguiti sul computer di un utente comunicano con applicazioni Java eseguite contemporaneamente sul server. Notevoli sono dunque le potenzialità di Java, anche se è ancora presto per dire se riuscirà ad esprimerle.

Le tecnologie *Push*

La "tradizionale" modalità di distribuzione delle informazioni su Internet basata, come detto, sul modello client/server prende anche il nome di tecnologia *pull*. Gli esempi di applicativi che abbiamo descritto sino ad ora (FTP, web browser) sono basati sul pull; sono evidenti alcuni limiti di questa tecnologia: è necessario ad esempio conoscere l'indirizzo del server su cui ci sono le informazioni che cerchiamo per poterle leggere, inoltre se le informazioni variano spesso si è obbligati a controllare periodicamente il server per avere le novità.

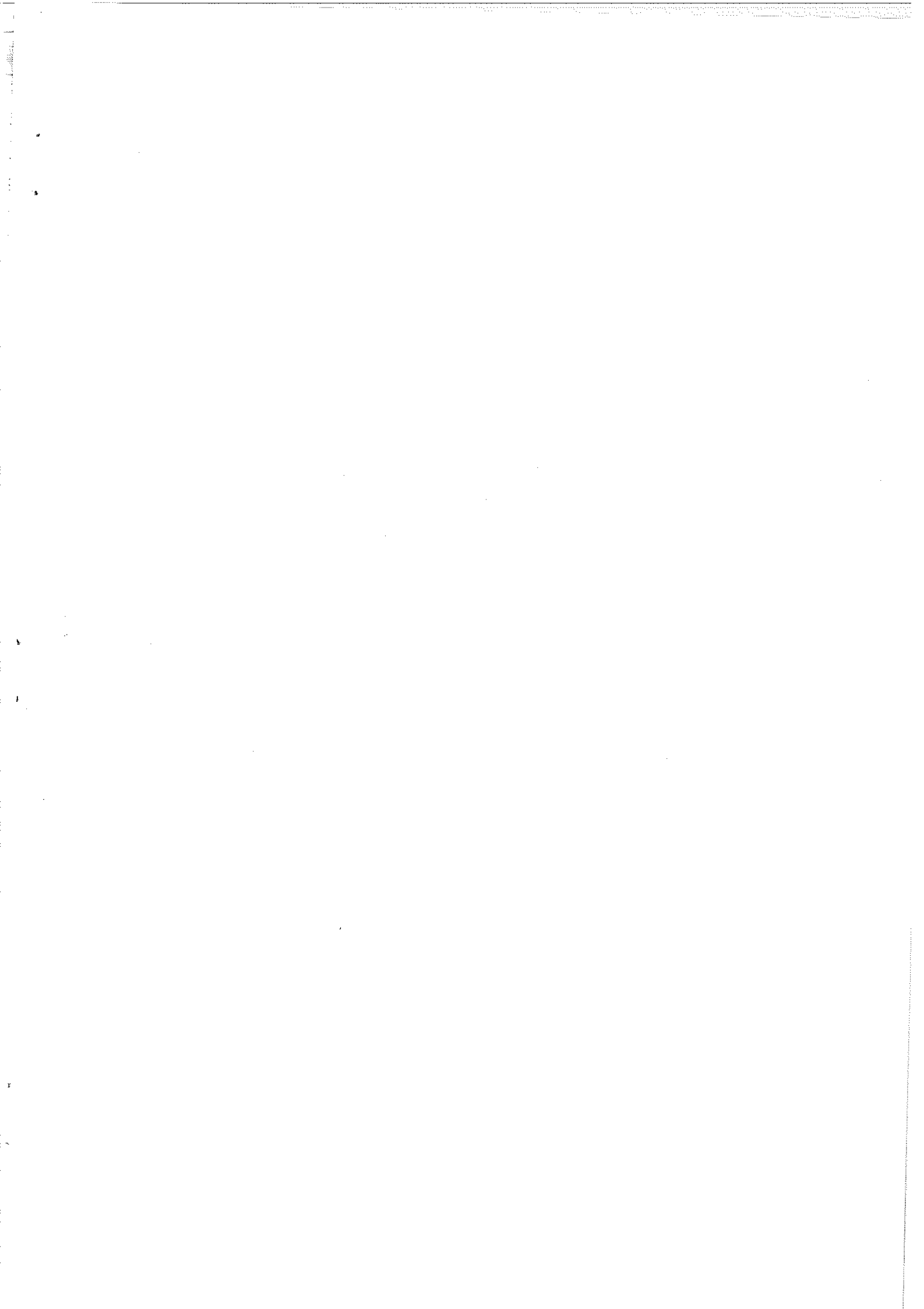
Da qualche tempo si stanno diffondendo applicativi di rete basati su una modalità praticamente opposta alla pull. L'idea è che siano le informazioni a raggiungere l'utente e non viceversa. Questa modalità è chiamata *push*. In pratica gli applicativi push mettono a disposizione dell'utente una serie di "canali" attraverso i quali può ricevere informazioni, l'utente sceglie quelli che lo interessano e inizia a ricevere informazioni direttamente sul proprio computer. Dall'altra parte del canale c'è ovviamente un server che "spinge" le informazioni verso tutti gli utenti "abbonati". Nonostante applicativi di questo tipo siano comparsi su larga scala da meno di un anno molti fornitori di informazioni anche prestigiosi (CNN, Wall Street Journal, Wired) hanno attivato server push, sono inoltre comparsi numerosi canali tematici.

Anche se concettualmente diversi push e pull possono comunque convivere nello stesso applicativo di rete e anzi la tendenza è proprio quella di integrare le due tecnologie.

LabNet

Infine due parole sulla rete LabNet.

I laboratori si collegheranno ad Internet utilizzando la connessione in linea commutata ed accesso diretto, sul computer dovrà essere installato il software di connessione fornito è ciascun computer dovrà quindi essere corredato di modem per la chiamata. Per l'accesso saranno forniti un programma per la posta elettronica, un browser web ed un FTP client. Per i dettagli sull'installazione e l'utilizzo di questi programmi vi rimandiamo ai manuali ed alle esercitazioni.



Indirizzi utili

Di seguito sono riportati alcuni indirizzi di rete utili per saperne di più sugli argomenti accennati.

Per maggiori informazioni sugli RFC:

<http://www.cis.ohio-state.edu/hypertext/information/rfc.html>

Un ricco glossario dei termini di rete si trova al seguente URL:

<http://www.rirr.cnuce.cnr.it/Glossario/ghpage.html>

Una pagina web contenente i principali motori di ricerca ed indici di rete:

<http://www.rirr.cnuce.cnr.it/>

Un motore di ricerca abbastanza sofisticato:

<http://www.hotbot.com>

Per informazioni su HTML, CGI e sul web:

<http://www.w3.org/>

Informazioni su Java

<http://java.sun.com>

Informazioni su Javascript

<http://www.netscape.com>

Informazioni su ActiveX

<http://www.microsoft.com>

Alcuni applicativi (player) push:

Pointcast (<http://www.pointcast.com>)

Marimba (<http://www.marimba.com>)

Netcaster (<http://www.netscape.com>)

Due indirizzi di server FTPMail:

ftpmail@grasp.insa-lyon.fr

ftpmail@sunsite.unc.edu

Glossario

bps	(bit per second). Unità di misura per la velocità di trasmissione seriale (es. per i modem e le porte seriali di un PC/Macintosh).
Browser	Programmi per la navigazione e la visualizzazione di pagine multimediali. I browser WWW (pagine HTML) sono Mosaic, Cello, Netscape, ecc. .
Client	Computer che richiede una operazione ad un altro computer (detto server), attende la risposta, la interpreta, eventualmente la elabora e infine la presenta all'utente. Anche una applicazione (un programma) può essere definita client.
Compressione	Trattamento cui viene sottoposto un file per ridurne le dimensioni di memorizzazione e/o di trasmissione. La compressione deve essere eseguita secondo un protocollo, per essere ripetibile.
Dial-up	Collegamento via modem e linea telefonica.
Download	Operazione di trasferimento dati da un computer remoto a uno locale.
FTP	(File Transfer Protocol). E' un servizio Internet che serve per trasferire file tra computer. Viene usato per recuperare un file reso disponibile su un sistema remoto.
Gateway	Apparecchiatura elettronica utilizzata per connettere 2 reti molto diverse: eseguono la conversione dei protocolli.
GOPHER	E' un protocollo che permette di consultare un database remoto.
HOST	Questo e' il nome che viene dato ad una macchina che supporta un sito Internet.
HTML	(Hyper Text Mark-up Language). Linguaggio per la definizione di pagine ipertestuali e multimediali. E' il linguaggio con cui sono scritte le pagine WWW.
HTTP	E' il protocollo che rende disponibili le pagine ipertestuali che costituiscono il World Wide Web.
ISDN	(Integrated Services Digital Network). La tecnologia emergente che viene offerta già da un po' di tempo dalle compagnie telefoniche di tutto il mondo. ISDN trasporta sia la voce che i servizi di rete digitale utilizzando un solo mezzo di trasmissione.
Modem	(Contrazione di MODulatore/DEMudolatore). Apparecchiatura che serve per convertire un segnale digitale proveniente dal computer con un segnale analogico in grado di transitare sulla

rete telefonica tradizionale (modulazione). Naturalmente effettua anche l'operazione inversa da linea telefonica a computer (demodulazione). comunicazione. Lo standard è definito dal CCITT (ora ITU).

- URL (Uniform Resource Locator). Con questo acronimo s'identifica un "indirizzo Internet" in modo simbolico, prescindendo dal tipo di servizio. E', praticamente, quanto dovete avere per contattare un sito WWW o prelevare un file via FTP, GOPHER.
- SLIP (Serial Line IP) Un protocollo usato per permettere connessioni TCP/IP tramite linee seriali (linee telefoniche o cavi RS-232). SLIP è definito in RFC 1055.
- PPP (Point-to-Point Protocol). Definito in RFC 1661, è un metodo standard che consente di implementare più protocolli su connessioni punto-punto.
- userID Il nome con il quale un utente è conosciuto da un computer. Tale nome deve essere unico ed è assegnato dall'amministratore quando l'utente viene registrato nel computer.
- Newsgroup Gruppi di discussione della rete Usenet