Introduzione

Lo scopo principale della comunicazione è lo scambio dell'informazione, essa prende spunto dal verbo latino "communico" che significa mettere in comune, condividere. La necessitá di comunicare è finalizzata alla rimozione di una incertezza che si presenta nell'interlocutore. In particolare, la quantitá di informazoine ceduta durante un processo di comunicazione puó essere misurata proprio dall'ammontare di incertezza che essa ha rimosso nell'interlocutore.

La parola telecomunicazione, invece, antepone il prefisso "tele" (di origine Greco), che significa a distanza, alla parola comunicazione. La telecomunicazione è la scienza che si occupa, quindi, della comunicazione a distanza. Si tratta di una esigenza sentita dall'uomo, cosí come la comunicazione, giá da molto tempo. Tuttavia, la telecomunicazione, a differenza della comunicazione, ha potuto sviluppare i suoi principi solo di recente, quando cioè lo sviluppo tecnologico ne ha permesso il suo avanzamento.

Nella lunga storia dell'uomo la comunicazione è stata, giá a partire dai primi giorni di vita, essenziale per lo sviluppo della societá. Ció infatti ha consentito all'uomo di rimuovere, inizialmente, le incertezze riguardanti il mondo circostante e le conoscenze cosí acquisite gli hanno consentito di generare comportamenti che aumentassero la sua capacitá di sopravvivenza nell'ambiente in cui si è trovato di volta in volta a vivere. E' interessante poi osservare come tali conoscenze siano dipese, ed ancora dipendono, dalle capacitá sensoriali di ciascun individuo.

Il primo problema per l'uomo a riguardo della comunicazione è stata la memorizzazione dell'informazione. L'uomo, dotato di cervello ed organi sensoriali, memorizza dentro di se le conoscenze che man mano acquisiva. In questa fase della storia, ai fini della comunicazione, l'uomo ha reagito sviluppando linguaggi di comunicazione, sicuramente complicati, che gli consentissero lo scambio dell'informazione da individuo ad individuo. Questo periodo storico viene talvolta ricordato come il periodo della cultura orale. Non esistendo altri supporti per raccogliere l'informazione, ogni individuo conserva dentro di se le conoscenze acquisite per poi passarle ai proprio figli.

Il successivo passo ha richiesto diversi millenni e conduce l'uomo alla cosiddetta cultura del manoscritto. Ció presuppone, quindi, la formazione e lo sviluppo di particolari tecniche di scrittura rivolte alla memorizzazione dell'informazione su supporto cartaceo. In questa fase continua ad esistere la cultura orale, essa tuttavia non è piú un esigenza primaria ed andrá scomparendo con il passare del tempo. Si è soliti dire che in questa nuova fase ha inizio la storia. L'uomo ha adesso la possibilitá di raccogliere l'informazione e tramandarla. Ció avviene superando il problema che fin quí gli aveva impedito di memorizzare l'informazione oltre le capacitá massime messe a disposizione del suo cervello. Non a caso ha inizio la storia, le guerre e gli imperi che si susseguono adesso entrano a far parte dei manoscritti dell'uomo e sono cosí conservati fino ai giorni nostri. Tuttavia, l'informazione, anche se adesso viene conservata, é costantemente minacciata.

Tutte le informazioni di una societá sono solitamente raccolte presso le biblioteche e solo poche persone vi accedono per poterne effettuare copie. Gli addetti alle copie dei manoscritti riescono a conservare la storia anche quando a causa di guerre alcune delle copie vengono distrutte. In questa fase lo sviluppo della societá risulta bloccato a causa del continuo ricopiare, ció richiedeva molto tempo.

Quest'ultimo problema è stato poi rimosso dall'importante scoperta della stampa di cui Gutembergh ne fu il predecessore. In questo modo, un ridotto numero di persone puó stampare un elevato numero di copie facilmente ristampabili. Si tratta di unp sviluppo tecnologico che libera notevoli risorse fin qui impegnate nel manoscritto. Puó quindi iniziare una nuova epoca, quella dell'elaborazione.

Tutta l'informazione fin qui conservata inizia ad essere elaborata e ció favorisce un nuovo sviluppo culturale. E' in questo periodo che inizia a nascere l'ingegneria delle telecomunicazioni che inizialmente si occupa del trasporto dell'informazione. Il servizio offerto era orientato al trasferimento dell'informazione su distanza a volte anche elevate ed era basato essenzialmente sul trasferimento dell'informazione mediante lettera.

Questa soluzione iniziale è peró scadente sotto diversi aspetti. Per la prima volta si inizia a parlare di ritardo di transito, dovuto alla lentezza con cui avviene la consegna dell'informazione. Essa era basata essenzialmente mediante catena di ripetitori umani in cui ogni individuo, cosí come in una staffetta, consegna il proprio testimone (l'informazione) all'addetto che lo segue. I successivi problemi che si ebbero furono invece dovuti alla sicurezza. Molto spesso un addetto rischiava l'intercettazione del nemico oppure veniva attaccato dai briganti. In entrambi i casi l'informazione andava persa. In questa fase di sviluppo culturale, lo studio, frutto dell'elaborazione, consente l'introduzione dei fenomeni elettrici e magnetici che vengono subito impiegati nel settore delle comunicazioni. Una prima applicazione è ad esempio il telegrafo. L'informazione è adesso affidata al campo elettrico, essa è solitamente prima convertita in segnale elettrico e poi trasportata. In tale circostanza sono estremamente importanti ai fini della qualitá gli elementi trasduttori. Il segnale vocale è ad esempio convertito, mediante microfono, in segnale elettrico e quando la tecnologia raggiunge un livello sufficiente ció verrá fatto anche per l'informazione visiva.
Il campo elettromagnetico propagandosi nello spazio trasferisce in maniera quasi istantanea ed in luoghi anche distanti l'informazione che in esso viene confinata. Per fare ció il campo elettromagnetico viene fatto variare nel tempo secondo una specifica legge e poi viene fatto propagare nello spazio fino a destinazione. L'informazione ad esempio contenuta in un segnale vocale risiede nella legge di variazione temporale che descrive alcune grandezze fisiche che interagiscono con l'uomo come ad esempio la pressione acustica che impatta sull'orecchio dell'ascoltatore.
Un ulteriore passo in avanti verso la moderna telecomunicazione è stato suggerito dalla possibilitá di descrivere una legge di variazione temporale con un intervallo limitato di simboli, da ció ne segue che l'informazione puó anche essere vista come una sequenza di numeri. Questo scenario si è presentato con l'avvento dell'introduzione dei calcolatori che meglio interpretano l'informazione sotto forma numerica, tipicamente di natura binaria.
Nel calcolatore confluiscono tutte le esigenze sentite dall'uomo e risolte con il passare del tempo: in esso si presenta infatti il problema della memorizzazione (risolto mediante strutture dati dette file), della elaborazione (il calcolatore ha nell'elaborazione il suo punto di forza) e del trasferimento dell'informazione da calcolatore a calcolatore. La storia sembra ripetersi, all'inizio lo scambio di informazione fra due calcolatori avveniva mediante opportuni supporti magnetici. Successivamente l'informazione è nuovamente affidata al campo elettromagnetico, ora due calcolatori possono scambiare l'informazione tra di loro anche in remoto. Il calcolatore che trasmette l'informazione genera da un segnale analogico un nuovo segnale di tipo numerico ed adatto quindi alla trasmissione dati. Il calcoltore che riceve il segnale numerico ricostruisce, ineve, a partire dal segnale ricevuto, il segnale analogico contenente l'informazione.
Una nuova fase culturale è tutt'ora in corso e nasce dal superamento del supporto cartaceo per la memorizzazione dell'informazione delle informazioni e prevede un uso intenso delle proprietá elettromagnetiche della materia. L'informazione tende man mano ad assumere una forma elettronic, nuovi lingiaggi ne consentono oltre che la memorizzazione anche una rapida ed efficiente consultazione (basti pensare ai linguaggi usati per il web). Quest'ultima fase che oggi viviamo è basata sui solidi pilastri che fin qui l'uomo ha posto nel corso della sua storia:

- capacitá di memorizzazione dell'informazione;
- capacitá di elaborazione dell'informazione;
- capacitá di trasferimento dell'informazione;

Questi tre concetti risultano tra loro legati, infatti: alla capacitá di memorizzazione dell'informazione basata su supporti sempre piú capienti corrisponde una minore importanza del trasferimento in tempo

reale dell'informazione, mentre alla capacitá di trasferire l'informazione in tempo reale corrisponde un minor interesse verso supporti sempre piú per la memorizzazione. L'elaborazione, in base a quanto avviene oggi, sembrerebbe invece confinata ai lati terminali della rete che dunque ha l'importante compito di collegare tra loro i calcolatori elettronici. Altri scenari prevedono, invece, una elaborazione dell'informazione in forma distribuita in cui ogni calcolatore effettua una elaborazione e la condivide con gli altri calcolatori. Appare dunque evidente che anche in questo caso la rete di telecomunicazione svolge un importante compito che è quello dell'interconnessione tra calcolatori. Nel corso di queste note vedremo i principi che sono alla base di una rete di telecomunicazione affrontando anche il problema della costruzione di una rete di telecomunicazione.

I flussi informativi nella rete

Nel seguente paragrafo verrá utilizzato un livello di astrazione molto generale, ció è dovuto alle tante reti di telecomunicazioni che oggi esistono ed alle esigenze, sempre piú diverse, che essa soddisfa. Una rete di telecomunicazione puó essere vista come un insieme di nodi interconnessi tra loro, alcuni di questi nodi sono detti nodi di frontiera o nodi terminali. Essi si trovano, infatti, ai lati della rete e ne costituiscono appunto la terminazione. Altri nodi, invece, partecipano in maniera piú attiva alla rete poiché sono interessati da flussi in ingresso prodotti da altri nodi terminali oppure da nodi anch'essi interni alla rete, essi si dicono nodi intermedi.
La rete di telecomunicazione interconnette, quindi, i nodi terminali tra di loro mediante i nodi interni. Per i nodi terminali i flussi di informazione immessi nella rete sono generati da opportuni trasduttori che trasformano in segnale elettrico le grandezze fisiche dell'ambiente circostante; oppure, i flussi informativi si presentano sotto forma di grandezze fisiche giá acquisite e per questo memorizzate su appositi supporti di memorizzazione. Qualora la rete è interconnessa ad un'altra rete, mediante un nodo terminale, il flusso informativo in ingresso puó essere costituito da flussi informativi che in quel momento stanno circolando sull'altra rete.
Per i nodi interni i flussi informativi in ingresso, invece, sono costituiti dai flussi informativi immessi dai nodi terminali della rete oppure dai flussi informativi di altri nodi interni. Talvolta presso i nodi terminali della rete sono raccolti dispositivi di trasduzione che riportano il segnale di partenza alla sua forma originaria e dispositivi per la memorizzazione. I nodi terminali della rete hanno l'importante compito di interfacciarsi con l'utente, l'operatore umano costituisce la vera sorgente di informazione. Esso, infatti, guida il funzionamento dei trasduttori che cosí acquisiscono l'informazione; crea l'informazione traducendola in formato elettrico; elabora oppura apporta modifiche all'informazione; seleziona l'informazione per il trasferimento da un nodo ad un altro.
Non tutte le reti di telecomunicazione offrono tutte le funzioni appena citate, la tendenza a cui oggi giorno si assiste è quella che spinge tutte le funzioni appena viste in un unico contenitore, il web browsing. Vista la varietá dei dispositivi che puó trovarsi nei pressi di un nodo terminale, è piú corretto dire che un applicazione utilizza la rete per scambiare flussi informativi con i dispositivi presenti presso il nodo di rete. In questo scenario appare quindi evidente l'importanza di offrire un servizio essenziale che è appunto la telecomunicazione. La qualitá del servizio fornita da una rete di telecomunicazione è caratterizzata da alcuni parametri. Per i flussi informativi di natura binaria si considerano:

- la probabilitá di errore *P(e)*, misura la frazione di bit del flusso informativo che giungono con valore errato;
- la frequenza di cifra o ritmo binario f_c, misura il numero di bit al secondo che vengono ricevuti da una entitá in fase di trasmissione;
- il ritardo di transito Δ, rappresenta il tempo che intercorre dal momento in cui il bit lascia l'applicazione all'istante di tempo in cui esso raggiunge l'entitá in ricezione;

Per i flussi informativi di natura analogica si considerano invece:

- il rapporto segnale rumore *SNR* tra la potenza del segnale e quella del rumore di fondo;
- la banda massima dei segnali che la rete di telecomunicazione consente di trasferire da un applicazione ad un'altra;
- il ritardo del segnale, inteso come il tempo necessario affinché il segnale di origine arrivi a destinazione;

In una nota applicazione come la telefonia, basata su flussi informativi di natura analogica, la rete di telecomunicazione puó trasportare sia flussi informativi di natura numerica che flussi informativi di natura analogica. Nel primo caso (rete per flussi numerici) il flusso informativo subisce una conversione dal formato analogico a quello numerico, quindi viene immesso nella rete. Lo stesso segnale, giunto a destinazione, subisce una nuova conversione, questa volta dal formato numerico a quello analogico.
Nel secondo caso (rete per flussi analogici), invece, si è soliti dire che la rete è essa stessa di tipo analogico in quanto tratta i soli segnali analogici. In questo caso il ritardo di transito del segnale è trascurabile poiché i vari nodi interni della rete non hanno la possibilitá di memorizzare il flusso informativo che viaggia. Essi si limitano soltanto a passare il flusso informativo al successivo nodo della rete.
Quando due entitá di applicazione hanno una stringa A di bit da scambiarsi puó essere opportuno modificare la stringa informativa senza variarne il contenuto. In altre parole, se T è un nodo trasmittente e se A è la stringa da trasferire, si puó operare su A mediante funzione $B=f(A)$ affinché la nuova stringa abbia un numero di bit inferiori rispetto ad A. Il nodo ricevente R applicherá quindi sulla stringa B la funzione $A=f^{-1}(B)$ in maniera tale da ricostruire la stringa A di partenza. Quando ció avviene si dice di effettuare una compattazione del flusso informativo, un requisito importante della compattazione è l'esistenza della funzione $f^{-1}(.)$, l'inversa di $f(.)$.
In altri scenari, invece, si decide di rinunciare alla invertibilitá della funzione $f(.)$. In tal caso, in fase di ricezione, l'entitá R si troverá davanti al problema della scelta per la funzione di inversione. In molti casi si sceglie una funzione inversa che realizza un segnale A' diverso da A ma che differisce di poco da quest'ultimo. Tale scostamento non è in alcuni casi recepito dall'interlocutore finale e permette di abbassare ulteriormente la frequenza di cifra richiesta al trasferimento. In questo caso si dice di operare una funzione di compressione del flusso informativo. La compressione, a differenza della compattazione, prevede un deterioramento del flusso informativo abbassandone ulteriormente i requisiti di qualitá.
Ad esempio, un segnale vocale analogico, percepito dall'uomo in una banda di frequenza che va dai 20Hz ai 20KHz, è comprimibile in una banda lorda di 4KHz. Ció avviene nelle applicazioni telefoniche senza tra l'altro compromettere il contenuto informativo e l'intelleggibilitá del suo contenuto. Altri esempi di compressione sono invece lo standard MPEG per i flussi video ed il formato MP3 per i flussi audio. Mentre, esempi di compattazione possono invece essere riconosciuti in applicazioni note come WINZIP e WINRAR, qui l'informazione e quindi il segnale di partenza viene ricostruito in maniera del tutto uguale e precisa a quello effettivamente compattato.
La variabilitá del flusso informativo si misura dal grado di intermittenza GI, definito come il rapporto tra il valore di picco ed il valore medio del ritmo binario, $GI=f_c max/f_c medio$. Quando GI=1 la sorgente si dice non intermittente. Alcune entitá di applicazione richiedono un GI~10 come ad esempio i servizi di consultazione telematica, altre applicazioni richiedono invece un GI~2 come ad esempio i servizi di compressione vocale. E' chiaro quindi che la diversitá delle entitá di applicazione richiede al

progettista della rete di telecomunicazione di soddisfare i requisiti di qualitá che vengono chiesti dalle entitá.
In passato era molto forte il concetto che collocava presso un terminale di rete un'unica applicazione, quest'ultima regolamentata da un opportuno protocollo. Non a caso iniziavano a nascere diverse reti di telecomunicazioni: la rete telefonica, la rete televisiva, la rete di calcolatori, etc… Ognune di queste reti era caratterizzata da particolari esigenze, dettate quindi dai diversi requisiti delle applicazioni. Lo sviluppo di cosí tante reti, tra loro separate, è stato senz'altro imposto dalla distanza temporale che intercorre dall'introduzione di una rete ed un'altra. L'assenza di protocolli in grado di gestire la comunicazione per piú entitá tra loro diverse favorí quindi lo sviluppo delle cosiddette reti dedicate ad un servizio.
Proprio per le diverse esigenze che le entitá di applicazione richiedevano alla propria rete non fu possibile aggregare in una rete piú servizi. Le reti dedicate risultavano incapaci di soddisfare le esigenze di qualitá del servizio di altri protocolli di applicazione. Per questo motivo, le reti per un certo tempo sono state tenute divise.
Solo quando la creazione di nuove reti dedicate a nuove applicazioni inizio a costare di piú del riutilizzo delle attuali reti in circolazione si inizió a riciclare le reti esistenti affinché esse consentissero il funzionamento di nuovi protocolli (ci fú in questo periodo l'introduzione dell'uso della rete telefonica per l'applicazione fax e l'uso della rete televisiva per l'applicazione teletext). Solo di recente si è andara man mano consolidando l'affermazione di nuove tecnologie che rendono possibile la creazione di un'unica rete in grado di fornire una qualitá adeguata a tutti i protocolli di applicazione al momento esistenti.

Il principio di raggruppamento e stratificazione

Una rete di telecomunicazioni deve poter offrire il trasferimento di un flusso informativo da un punto della rete ad un altro. La complessitá di tale problema è stata allora ripartita in piú sottofunzioni affinché il problema del trasferimento di informazione risultasse piú semplice da attuare e da gestire.
Quando un sistema ha una struttura a strati è anche piú semplice variare l'implementazione dei servizi forniti da uno strato. La stratificazione riduce quindi la complessitá progettuale, in questo modo ogni entitá di rete viene modellata mediante una sequenza ordinata di sottosistemi detti strati. A ciascuno strato competono uno o piú protocolli, la comunicazione si svolge quindi attraverso gli strati adiacenti della pila protocollare, fino a giungere l'entitá di destinazione. Ogni strato comunica con quello adiacente mediante messaggi detti unitá dati o PDU (*protocol data units*).
Inoltre, per garantire un certo livello di indipendenza funzionale, ciascuna funzione di strato deve ignorare le modalitá con cui funziona un protocollo di strato. In questo modo è possibile cambiare un protocollo di strato con uno di pari livello senza inficiare una disfunzione dello strato sottostante. Funzioni tra loro simili ma svolte da differenti protocolli vengono raggruppate in insiemi funzionali omogenei. Infine, le funzioni che appartengono ad un certo insieme costituiscono, per gli insiemi gerarchicamente inferiori, un arricchimento rispetto alle funzioni di strato che li sono collocate. Tali funzioni, infatti, realizzano mediante funzioni di strato superiore altre funzioni piú complicate.
Ciascuna entitá di livello è collegata virtualmente all'entitá di pari livello di un sistema omologo. Le entitá alla pari appartengono allo stesso strato di rete. Allo strato funzionale piú alto compete l'interazione conl'uomo, esso sfrutta tutti i servizi offerti dagli strati a lui sottostante. La comunicazione si svolge mediante protocollo di comunicazione, esso stabilisce e formalizza le regole per il corretto svolgimento della conversazione. Ad ogni strato puó essere associato un *service access sap* (SAP), si tratta di punti di accesso attraverso la quale passano le unitá dati PDU. Nel corso di queste note definiremo le funzioni piú importantu che possono essere eseguite da uno strato di rete,

parleremo quindi della funzione di multiplazione o multiplexing, della funzione di controllo di errore, della funzione di controllo del flusso e della funzione di commutazione.

Le risorse di rete

Affinché le entitá della rete funzionino a dovere è opportuno che i nodi della rete utilizzino le risorse appartenenti alla rete. Quando affronteremo le funzioni di rete saranno definite con maggiore chiarezza le risorse che caratterizzano la rete. In una rete ci sono grandezze dedicate ad un uso esclusivo, a carico cioè di un singolo terminale, e risorse condivise, dedicate cioè all'uso di piú entitá di rete. Ad una risorsa puó essere assegnato un carico di lavoro, la portata media della risorsa stabilisce il numero medio di unitá di lavoro che essa riesce a svolgere nell'unitá di tempo considerato. Si definisce, poi, capacitá di una risorsa il numero massimo di unitá di lavoro che la risorsa riesce a svolgere nell'unitá di tempo. Per misurare l'efficienza di utilizzo della risorsa si introduce il concetto direndimento R definito come il rapporto fra la portata media e la capacitá massima: *R=portata media/capacitá*.
Data l'esistenza di risorse condivise bisogna allora prevedere degli opportuni meccanismi per la gestione di tali risorse. Questi meccanismi dovranno poi risolvere i problemi dovuti alle chiamate parallele verso un'unica risorsa, le cosiddette contese. Esistono strategie di assegnazione a domanda e strategie di assegnazione con pre-assegnazione. Nella richiesta di assegnazione a domanda la risorsa è data ad una sola entitá dopo che quest'ultima ne abbia fatto domanda. Solitamente la risorsa è assegnata all'entitá richiedente per un breve istante di tempo, successivamente essa è assegnata all'entitá che segue nella coda di attesa. La coda di attesa, come ben si intuisce, introduce inevitavilmente un ritardo dovuto all'attraversamento della coda.
Una diversa soluzione della contesa di una risorsa condivisa è l'assegnazione con pre-assegnazione individuale. La risorsa è assegnata ad una sola entitá che quindi la impegna affinché essa possa svolgere l'intero lavoro che gli è stato assegnato. Una differente strategia di assegnazione della risorsa è la pre-assegnazione collettiva, in questo caso la risorsa viene assegnata ad un insieme di entitá accreditate al suo utilizzo. Ciascuna entitá accede alla risorsa presentando una domanda di richiesta ed impegnando quindi la risorsa per un fissato intervallo di tempo.
Le contese, invece, possono essere risolte secondo la modalitá orientata alla perdita oppure secondo una modalitá orientata al ritardo. La modalitá orientata al ritardo colloca le domande di richiesta in una coda di attesa, quella orientata alla perdita rifiuta ogni domanda di assegnazione della risorsa quando questa è giá impegnata. Per meglio apprezzare le qualitá di una rete basata sulla gestione delle contese mediante modalitá orientata al ritardo si considera il tempo di permanenza della domanda all'interno della coda di attesa. Mentre, per valutare le qualitá di una rete basata sulla gestione delle contese orientata alla perdita si misura la probabilitá che la risorsa risulti impegnata quando giá assegnata.

Le funzioni dello strato fisico

Lo strato fisico occupa la posizione piú bassa all'interno dello schema a strati che modella una enititá di rete. Il suo compito è quello di spostare l'informazione da un punto della rete ad un altro e la sua efficienza costituisce pertanto un buon presupposto per la realizzazione di una rete. Non vi è un scambio di informazione se non vi è un adeguato strato fisico. Per questo motivo lo strato fisico realizza un collegamento punto-punto con l'entitá coinvolta nel processo di scambio dell'informazione e pone le basi necessarie a tutte le altre funzioni di strato superiore.
Per spostare l'informazione da un punto della rete ad un altro occorre dunque un canale di comunicazione in cui convogliare l'informazione. Le modalitá con cui tale canale puó essere realizzato sono diverse e comprendono, in maniera del tutto generale, canali con propagazione libera (antenne) e canali con propagazione guidata (cavi coassiali e fibra ottica).

Quando due punti della rete sono tra loro collegati si dice che tra essi esiste un canale logico o collegamento logico. Tale collegamento ammette come ingresso un flusso informativo e ne effettua la trasmissione avvalendosi delle proprietá di propagazione del canale e di opportuni blocchi funzionali.
Il modem effettua operazioni di modulazione (in fase di trasmissione) e operazioni di demodulazione (in fase di ricezione), tale operazione consente di portare il segnale informativo, tipicamente detto inbanda base, alla banda del canale di comunicazione. L'operazione inversa di demodulazione riporta, quindi, il segnale ricevuto alla sua banda originaria. Per tale motivo il segnale informativo da trasposrtare non puó essere a banda illimitata ma deve risultare confinato all'interno di un certo intervallo di banda.
La comunicazione non è comunque ideale, al segnale ricevuto *y(t)'* si aggiunge inevitabilmente un rumore di fondo che distrurba il flusso informativo ricevuto. Tale disturbo viene quantificato mediante il rapporto segnale-rumore SNR. Altro aspetto da considerare è il ritardo dovuto alla propagazione del segnale, tipicamente indicato con il simbolo Δ. La qualitá del canale di comunicazione limita poi il ritmo binario con cui i bit del flusso informativo sono scambiati Infine, la probabilitá di errore *P(e)* tiene traccia degli errori nei bit trasferiti.
Le prime reti di telecomunicazioni realizzate sono state quelle di tipo analogico, solo dopo l'introduzione del calcolatore i collegamenti logici sono stati poi adattati alla trasmissione numerica. Infine, con l'affermarsi delle nuove tecnologie ed in particolar modo delle fibre ottiche, è stato possibile realizzare collegamenti fisici dalle elevate prestazioni in cui il segnale informativo si propaga a frequenze ottiche (con le fibre ottiche è possibile creare collegamenti numerici il cuiritmo binario è dell'ordine dei Terabit al secondo Tb/s).
Le moderne tecnologie, inoltre, sfruttando il doppino telefonico, ideato inizialmente per segnali analogici nella banda da 300Hz a 4KHz, riescono a realizzare collegamenti fisici il cui ritmo binario puó raggiungere alcune decine di Megabit al secondo Mb/s. Tale valore viene raggiunto sfruttando, mediante opportuni modulatori, la banda che eccede i primi 4KHz dedicati alla telefonia. Se invece vengono utilizzati i primi 4KHz della banda analogica del doppino telefonico si realizzano collegamenti fisici con ritmo un binario di circa 60 Kbit/s. E' interessante notare come le caratteristiche del collegamento fisico finiscono poi per determinare le limitazioni alla qualitá del servizio offerto da una rete ai suoi utenti.

Funzione di controllo della trama, il framing

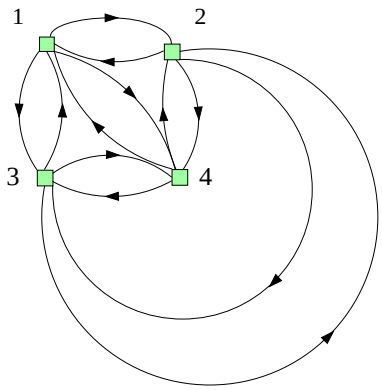
In base a quanto finora detto, lo strato fisico dovrebbe immettere, all'interno di un collegamento logico, il flusso informativo. Tale flusso non viene tuttavia trasmesso con continuitá ma viene spezzato in piú trame. Questa suddivisione in trame del flusso informativo, come vedremo piú avanti, è necessaria alla funzione di controllo dell'errore. Lo strato fisico effettua, dunque, sul flusso informativo una funzione di controllo della trama, talvolta detta framing. Un primo metodo per il controllo della trama consiste nell'aggiunte di pause all'interno del flusso informativo che deve essere trasmesso. Tuttavia, a cuasa dell'aleatorietá dei tempi di transito nella rete, tale approccio viene per questo motivo scartato. Ció infatti richiederebbe al trasmettitore di inviare al ricevitore un certo numero di bit e di attendere quindi la pausa imposta. Dall'altro lato, invece, quello del ricevitore, si riceveranno i bit inviati dal trasmettitore e si attenderá la pausa imposta. Se dunque uno o piú bit stanno ancora attraversando la rete poiché ritardati dai tempi di attraversamento aleatori della stessa, questi potrebbero giungere presso il ricevitore quando quest'ultimo è impegnato a rispettare una pausa imposta. In tal caso i bit giunti con ritardo vengono scartati e l'informazione subisce un degrado (ad esempio, trasmettitore e ricevitore potrebbero accordarsi con questi tempi: 5 secondi di trasmissione/ricezione e 2 secondi di pausa).
Una alternativa piú valida è il metodo del character count o conteggio del carattere. Si tratta di inserire nel flusso informativo un certo numero di bit aggiuntivi, tali bit aggiuntivi indicano al ricevitore quanti

bit compongono la trama o messaggio. In questo modo il ricevitore puó distinguere l'inizio e la fine del blocco inviato dal trasmettitore. I bit aggiuntivi sono ovviamente rimossi dal ricevitore in fase di ricostruzione del messaggio. Anche questa tecnica non è peró affidabile, infatti: l'errore che il canale puó introdurre sul flusso informativo potrebbe interessare proprio il character count ed in tal caso il ricevitore non saprebbe piú riconoscere con esatteza l'inizio e la fine delle successive trame.
Un'altra tecnica di framing prevede l'utilizzo di opportune stringhe di bandiera che delimitano l'inizio e la fine della trama. Tali stringhe di bandiera possono essere diverse oppure tra loro uguali. Questo metodo garntiche, in caso di errore in ricezione, la possibilitá di ritrovare l'allineamento nei confronti delle trame successive. In altre parole, una particolare stringa indica al ricevitore l'inizione della trama. Una nuova stringa oppure sempre la stessa stringa citata prima potrebbe poi indicare la fine della trama precendete e quindi l'inizio di una nuova trama. Ad esempio, il messaggio contenuto nella TRAMA 1 viene preceduto dalla stringa AAA. Tale stringa ha l'effetto di allertare il ricevitore affinché quest'ultimo si predisponga alla ricezione della TRAMA 1. La successiva stringa AAA decreta quindi la fine del messaggio di TRAMA 1 e stabilisce l'inizio per il messeggio di TRAMA 2.
La stringa di bandiera non indica la lunghezza della trama ma ne delimita quindi l'inizio e la fine. Lo svantaggio di tale procedura risiede nel fatto che occasionalmente nella trama si possono verificare stringhe di bit uguali a quelle di bandiera. Ció comporterebbe una errata interpretazione della trama. Si potrebbe allora vietare al livello superiore di generare tali caratteri ma ció violerebbe il proncipio di indipendenza funzionale ed allora viene usata una ulteriore tecnica detta tecnica del riempimento di carattere. Stabilita la stringa di bandiera si individua una sua possibile sottostringa (tipicamente la metá dei caratteri dell'intera stringa di bandiera ma ció non costituisce una regola fissa). Quando il flusso informativo viene analizzato per la trasmissione si inserisce la restante parte della sottostringa di bandiera laddove essa è effettivamente necessaria in fase di framing. Se invece la sottostringa di bandiera individuata rappresenta effettivamente una stringa del messaggio di trama si fa seguire quest'ultima da una ulteriore stringa il cui compito è quello di indicare al ricevitore la reale esigenza di tale stringa.
Un approccio simile è poi applicato anche al caso in cui il flusso informativo è costituito da una stringa di bit invece che da caratteri. Anche in questo caso si stabilisce una particolare stringa di bit di bandiera e se ne individua una sua sottostringa. Quando il trasmettitore scorre la stringa di bit da inviare nel canale logico quest'ultimo intercetta tutte le sottostringhe di bandiera e se esse appartengono effettivamente alla trama fa seguire tali sottostringhe da un bit di ignora. Supponendo ad esempio di adottare come stringa di bandiera la sequenza di bit 1101 scegliamo come sottostringa la sequenza di bit 110 e come bit di ignora il bit 0. Il messaggio di trama da inviare nel canale logico è costituito dalla stringa di bit 1010110. Il trasmettitore aggiunge quindi i bit alla trama modificandola in questo modo: la sequenza 1101 (stringa di bandiera) stabilisce l'inizio della trama, quindi seguono i bit 10101100 e per finire di nuovo la stringa di bandiera 1101.
Un altro metodo è quello in cui lo strato fisico delega al canale logico la risoluzione del problema. Il canale logico potrebbe allora prevedere tre simboli per la trasmissione dati di cui uno solo di questi è usato come elemento di separazione per le trame. Un metodo siffato costituisce peró una soluzione poco efficiente poiché sacrifica un simbolo alla informazione piuttosto che arricchire l'alfabeto usato per codificare il flusso informativo.

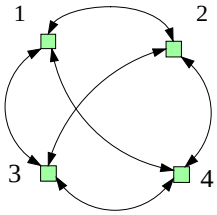
Fondamenti di topologia

La topologia di una rete è la configurazione logica dei collegamenti esistenti tra i vari terminali della rete. Essi interagiscono tra di loro scambiandosi l'informazione necessaria e qualora i terminali coinvolti nello scambio dell'informazione non sono reciprocamente collegati essi riescono ugualmente a comunicare sfruttando le varie funzioni di strato e avvalendosi, quindi, della topologia della rete.

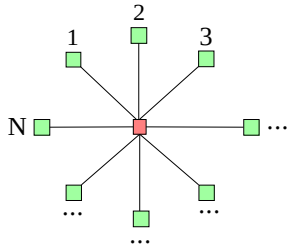
L’insieme dei terminali in grado di scambiarsi direttamente l’informazione costituisce la topologia della rete. Dalla topologia della rete discendono importanti proprietà come i costi, l’affidabilità, l’espandibilità e la complesità della rete. La topologia più naturale a cui pensare è detta a maglia e costituisce una metodologia di costruzione troppo onerosa. In tale topologia ciascuna utenza (nodo terminale) è collegata in maniera diretta alle altre utenze. Se ad ogni nodo terminale associamo un cavo per la ricezione ed uno per la trasmissione, ciascuno di questi diretto verso altri nodi della rete, saranno allora necessari *N(N-1)* collegamenti o canali fisici. Ad esempio:



Se invece, si considera un unico canale logico che collega due nodi terminali, dedicato quindi sia alla trasmissione dati che alla ricezione, il numero di collegamenti da effettuare scende ad *N(N-1)/2*.

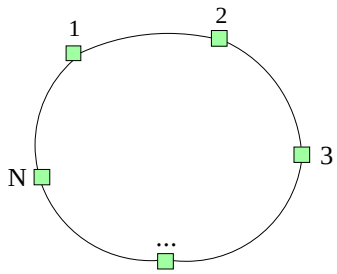


Dunque il numero dei collegamenti richiesti da una topologia a maglia (completa) ha una crescita quadratica. Una soluzione decisamente più conveniente è costituita dalla topologia a stella. In tale configurazione tutti i nodi terminali della rete sono collegati ad un nodo centrale detto centro-stella oppure hub di rete. Adesso il numero di collegamenti fisici ha una crescita lineare ogni qualvolta si decide di aggiungere un nuovo nodo terminale.

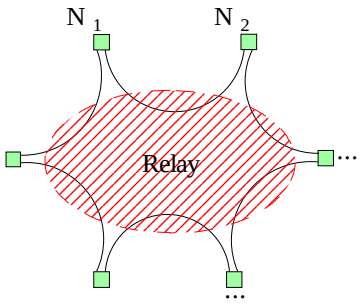


Una tale configurazione fornisce tuttavia alcuni vantaggi e svantaggi da considerare in fase di progetto. Se un nodo terminale della rete finisce fuori uso, in una topologia a stella, l’intera rete non risente di

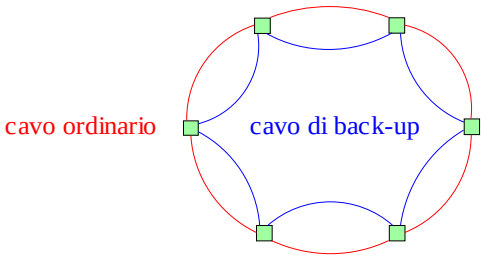
tale guasto cosicchè ai nodi terminali rimasti in funzione sarà ancora possibile scambiare reciprocamente, attraverso il centro-stella, flussi informativi. Tuttavia, a cuasa di un eccessivo traffico nella rete, il nodo centrale può risultare spesse volte in sovraccarico e ciò potrebbe portare al blocco delle richieste di connessione. Altra possibile topologia è quella ad anello in cui tutti i nodi terminali della rete sono connessi tra loro a formare un ciclo chiuso. La trasmissione avviene in unico senso. Tutti i nodi terminali prendono parte alla trasmissione. Quando un nodo terminale genera un flusso informativo immette quest'ultimo nell'anello avvalendosi del proprio collegamento logico per la trasmissione dati. Il nodo terminale che segue il nodo trasmittente riceve per primo il flusso informativo, lo memorizza, lo analizza per stabilire se è diretto effettivamente a se stesso ed eventualmente, quindi, lo ritrasmette nell'anello.



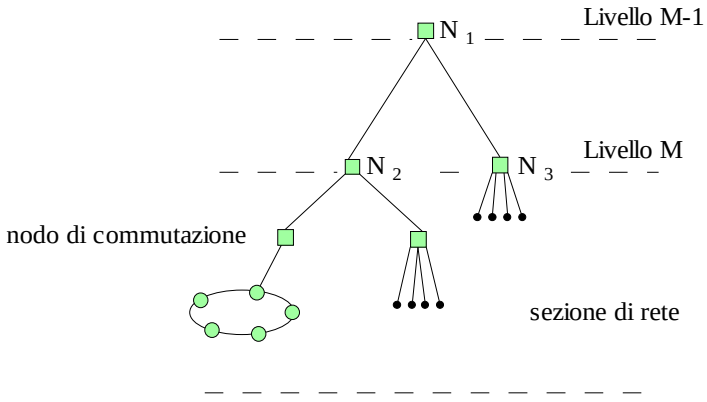
La topologia ad anello chiuso, così come quella a maglia, richiede dei collegamenti fisici piuttosto lunghi, specie quando la distnza tra i nodi terminali della rete è eccessiva. I vantaggi della topologia ad anello risiedono nella possibilità di utilizzo di mezzi trasmissivi unidirezionali dalle elevate prestazioni, essa si addice bene alla costruzione di una estesa rete di telecomunicazione. Tuttavia, la topologia ad anello non minimizza la lunghezza del cavo, inoltre, qualora un nodo terminale della rete subisce un guasto temporaneo oppure non funziona a dovere si assiste alla caduta dell'intera rete. Per quanto riguarda il problema dell'affidabilità si può addottare un accorgimento che prevede l'aggiunta al centro dell'anello di un blocco commutatore detto relay. In questo modo la topologia è ad anello solo a livello a logico ma può essere vista anche a stella a livello fisico. Il centro stella ha poi la capacità di mettere fuori rete qualsiasi nodo guasto (in caso di guasto il relay disinseriche dall'anello il relativo cavo di giunzione).



Un diverso approccio al problema dell'affidabilità dell'anello consiste nell'utilizzazione non più di un unico cavo ma di due o più cavi detti di back-up. Il cavo di riserva entra in funzione quando un nodo terminale è staccato dalla rete oppure non funziona. In caso di caduta di due o più nodi terminali si formano anelli parziali fra loro connessi.



Al tra topologia assai nota è quella ad albero, detta anche di tipo gerarchico. Ogni livello gerarchico della rete può essere costituito a sua volta da una topologia a stella oppure ad anello. Qualora il livello gerarchico fosse costituito da una topologia ad anello può anche capitare che un nodo dell'anello appartenga anche al livello superiore oppure esso può appartenere anche ad altri anelli.

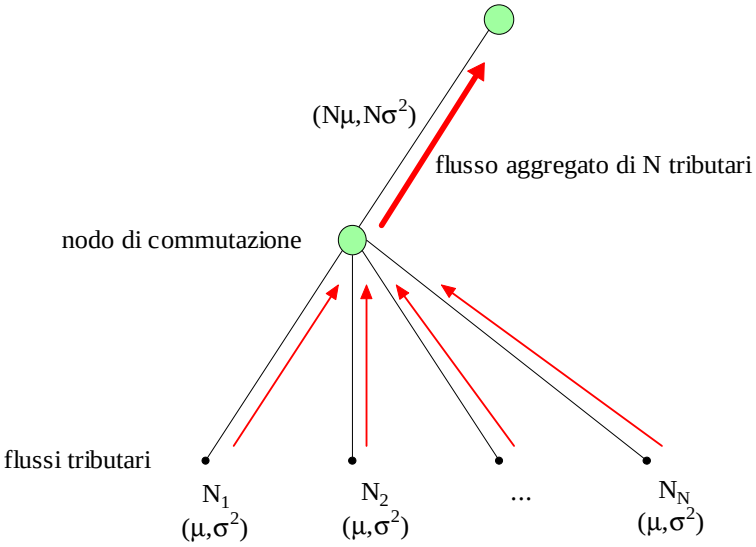


I nodi terminali dislocati sul territorio e che appartengono ad un livello gerarchico dell'albero, sia essi configurati ad anello oppure a stella, che acceddono a flussi informativi della rete attraverso un nodo di rango superiore si dicono appartenere alla medesima sezione di rete. Le sezioni di livello gerarchico superiore sono costituite da nodi in numero via via ridotto, esse trasportano su distanze sempre più elevate (man mano che si sale nella gerarchia) i flussi informativi ottenuti dall'aggregazione dei vari flussi informativi delle varie sezioni di rete. Ogni topologia presenta i suoi vantaggi e svantaggi cosicchè ad ogni problema che essi portano è sempre possibile attuare un accorgimento risolutivo. Ovviamente sarà l'implementazione della soluzione trovata a stabilire la convenienza della soluzione pensata, soprattutto da un punto di vista economico. Per molto tempo il costo dei cavi è stato un ostacolo alla costruzione delle reti, oggi invece la posa dei cavi ed i relativi costi di lavoro si sono talmente abbassati che quasi sempre si preferisce costruire un unico canale fisico dalle elevate qualità piuttosto che realizzare più canali fisici di qualità inferiore (dal canale fisico vengono poi creati più canali logici). Tale scelta ha negli ultimi anni rafforzato la convenienza delle fibre ottiche. Le strutture ad anello ed a stella sono quelle che maggiormente si prestano a collegamenti ad alta velocità grazie al fatto che esse utilizzano collegamenti punto-punto disponendo di ottimi mezzi trasmissivi. Ciò richiede anche un centro stella capace di commutare velocemente i flussi informativi che ad esso coinfluiscono, altrimenti andrebbe vanificata la prestazione della fibra ottica utilizzata per il collegamento punto-punto.

Il rendimento delle sezioni di rete

Man mano che i flussi informativi vengono aggregati, tale processo ha inizio nelle sezioni di rete di rango inferiore e procede in quelle di rango superiore, si assiste ad una riduzione del grado di intermittenza. L'intermittenza è stata definita come il rapporto tra la frequenza di cifra di picco e la frequenza di cifra media: $GI=f_c max/f_c media$.

Quando $GI=1$ la sorgente (il nodo di rete) si dice essere non intermittente. L'aggregazione dei flussi informativi determina l'aggregazione dei gradi di intermittenza e quest'ultimo, man mano che si attraversa la rete dal gasso verso l'altro, tende a ridursi. A tale proposito supponiamo di modellare i flussi tributari mediante una variabile aleatoria gaussiana avente media μ e varianza σ^2, il suo valore massimo è pertanto $\mu + K\sigma^2$ (con K probabilità di sovraccarico).



Quindi GI vale:

$$GI = \frac{f_c \max}{f_c media} = \frac{N\mu + K\sqrt{N\sigma^2}}{N\mu} = 1 + \frac{K\sigma}{\mu\sqrt{N}}$$

Ipotizzando i flussi tributari tra loro incorrelati. K è il valore ottenuto in corrispondenza di una prefissata soglia di probabilità di sovraccarico. Dall'ultima relazione scritta si intuisce come GI diminuisca al crescere di N, numero di tributari o flussi tributari.

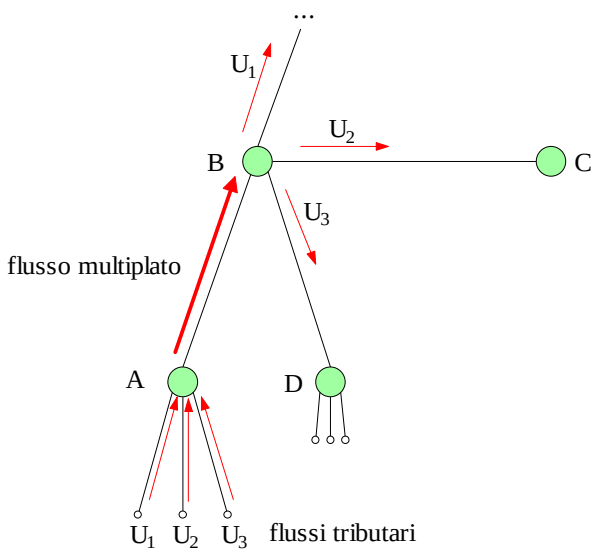
Quando il progettista dimensiona il collegamento per il trasporto del flusso informativo può scegliere per quest'ultimo un valore massimo di frequenza di cifra pari al valore di picco. In questo modo il collegamento risulta idoneo alle entità di rete che richiedono una frequenza di cifra di picco prossima a quella del collegamento. Il rendimento della rete, valutato come il reciproco del GI, può risultare più inefficiente quanto è alto il GI della rete:

$$\text{Rendimento} = \frac{1}{GI} = \frac{N\mu}{N\mu + K\sqrt{N\sigma^2}}$$

Se la rete è ben progettata abbiamo detto che il GI è più basso nelle sezioni di rango superiore della rete, esso al massimo può avvicinarsi o essere uguale all'unità (caso di non intermittenza) ed in tal caso la rete di telecomunicazione sarà sfruttata a dovere.
Se invece il cavo del collegamento fisico è dimensionato in base al valore medio della frequenza di cifra, anzichè quella di picco, il grado di intermittenza si assesta in un intorno dell'unità, la rete sarà in tal caso sfruttata bene ma non reggerà i picchi della frequenza di cifra della rete stessa. Se il progetto è inoltre basato sul valore medio della frequenza di cifra si incrementeranno i ritardi medi di transito.
Una tale scelta mette in crisi i servizi in tempo reale per cui storicamente si è scelto di separare i servizi a tempo reale da quelli più comuni della rete e così la rete stessa è andata sempre più a specializzarsi in una rete per calcolatori. Per questo motivo, quindi, i pacchetti dei flussi informativi che vengono immessi nella rete nascono in un certo senso timbrati ed a seconda della loro caratteristica sono depositati in una coda con ritardo oppure in una coda con perdita.

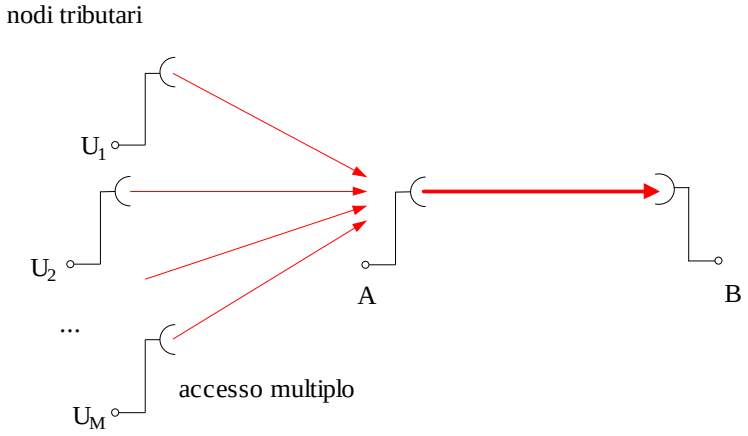
La funzione di multiplazione, schema di multiplazione

Abbiamo finora motivato la necessità di un canale logico tra due nodi della rete, sia essi terminali o interni alla rete. Affinchè sia possibile scambiare l'informazione da un punto ad un altro è allora necessario l'esistenza di almeno un canale fisico. La cascata dei blocchi modulatore-trasmettitore-canale fisico-ricevitore-demodulatore consente poi di generare un canale logico cosicchè i due punti della rete risultano effettivamente collegati logicamente.
Tuttavia lo scenario finora ipotizzato non è l'unico a verificarsi. Infatti, può accadere che in ingresso al nodo A si presintino più flussi informativi che chiameremo flussi tributari. L'esigenza dei flussi tributari è comune e consiste nel raggiungere il nodo B della rete. Successivamente, quindi, ciascun flusso tributario, giunto al nodo B, può prendere percorsi diversi dagli altri flussi tributari. La figura che segue illustra quanto finora è stato detto:



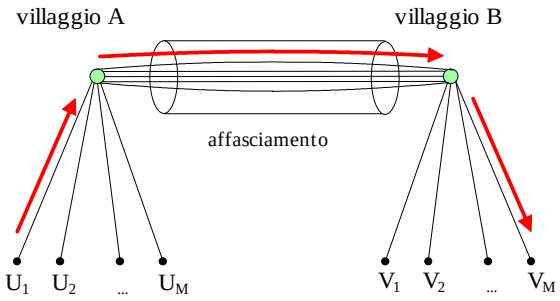
Obiettivo della funzione di multiplazione è la costruzione di un flusso multiplato risultante dall'aggregazione dei flussi tributari in ingresso al nodo A. Presso il nodo B, invece, dal flusso multiplato viene estratto ogni singolo flusso tributario mediante operazione inversa di demultiplazione. Nello scenario qui descritto si è ipotizzato il caso in cui dal canale fisico che collega i nodi A e B venga generato un unico canale logico. E' poi possibile generare, a partire dal canale fisico, M distinti canali

logici e dedicare ciascuno di essi ad ogni flusso tributario. Una tale possibilità è allora valida se gli M canali logici ricavati risultrano di qualità comunque inferiore (se cioè sommati assieme) alla qualità dell'intero canale logico.
Le prestazioni del canale logico e dei dispositivi che lo realizzano dipendono molto dalle qualità del canale fisico adoperato per ogni tratto della rete. I costi connessi alla posa dei cavi (nel caso di propagazione guidata) sono di gran lunga indipendenti dalla tipologia di cavo che viene utilizzato. Tale considerazione vale anche nel caso della propagazione libera, realizzata mediante antenne. Ed allora, per risparmiare sulla struttura della rete, generalmente, si preferisce realizzare un unico canale logico e creare da quest'ultimo M distinti canali logici di qualità inferiore.
La natura dei flussi tributari che si presenta presso il nodo A può essere varia. Infatti, in ingresso al nodo A possono presentarsi flussi tributari di natura numerica o analogica. E' bene precisare che non tutte le tecniche di multiplazione sono adatte ad entrambi i casi. Ci possono poi essere delle modalità particolari di accesso al nodo A. Il caso più ovvio è quello in cui gli M flussi tributari accedono al nodo A mediante M distinti canali logici. Tuttavia, nel caso della propagazione libera, gli M distinti flussi tributari accedono al nodo A utilizzando lo stesso canale logico che fisico).
In tal caso, infatti, si avranno M distinti antenne (ciascuna per ogni flusso tributario) che accedono all'antenna del nodo A. Quest'ultima riceverà, allora, un flusso già multiplato poichè nel flusso informativo ricevuto si fondono i vari flussi tributari. La funzione di multiplazione è in questo caso realizzata dai vari sottonodi tributari che realizzano il flusso multiplato secondo opportune regole di trasmissione. Quando ciò avviene si dice che i flussi tributari realizzano, presso il nodo A, un accesso multiplo.



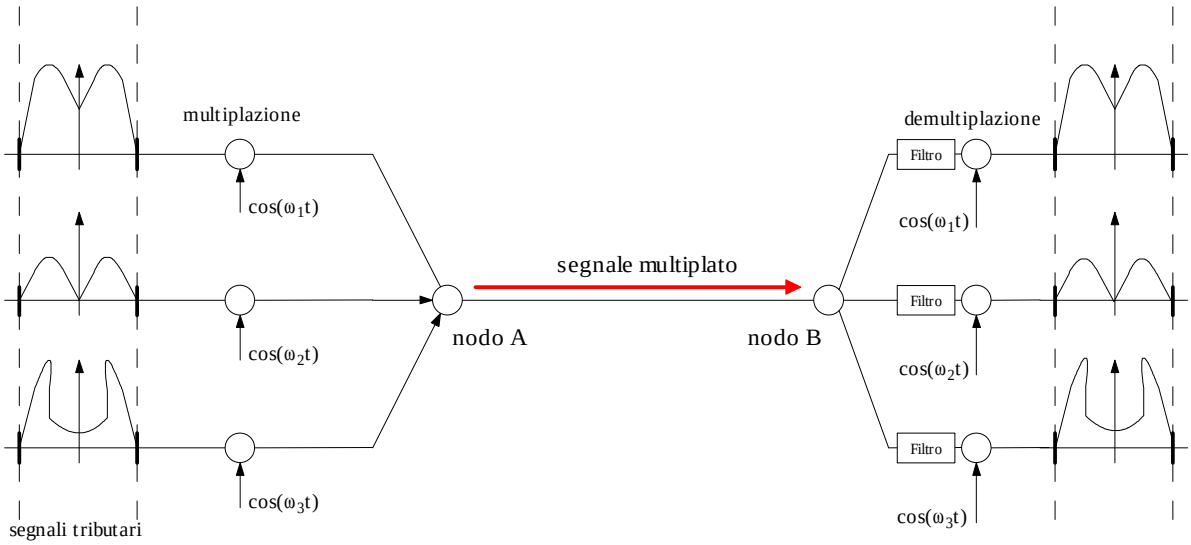
Multiplazione a divisione di spazio (SDM)

Si tratta di una delle prime tecniche di multiplazione adoperata storicamente da Bell in ambito telefonico. Affinchè sia possibile mettere in comunicazione un tributario del villaggio A con uno del villaggio B, Bell dedicava a quest'ultimo un cavo fisico per la comunicazione. La multiplazione a divisione dispazio consiste in questo caso nell'affasciamento dei cavi fisici che si dipartano da A fino a B in un unico tubo. Ad un flusso informativo che richiede un collegamento logico verso B il nodo A può allora assegnare a quest'ultimo un cavo fisico dell'affasciamento.



Multiplazione a divisione di frequenza (FDM per segnali analogici)

La multiplazione a divisione di frequenza FDM si riferisce al caso in cui i flussi tributari sono di natura analogica ed il canale dunque esistente tra A e B è anch'esso analogico. E' opportuno osservare che i flussi analogici qui adoperati possono anche trasportare flussi informativi numerici.
Il flusso multiplato è generato dai singoli flussi tributari mediante operazione di modulazione. Ciascun flusso tributario occuperà secondo una propria caratteristica una banda di interesse [-B,B], dunque mediante modulazione si applica a ciascun flusso tributario una modulazione in frequenza che ha l'effetto di spostare sull'asse delle frequenze i vari flussi tributari che adesso non sono più fra loro sovrapposti. Il flusso multiplato è allora la somma di tutti i segnali tributari modulati in frequenza. Affinchè il canale logico si faccia carico della trasmissione del flusso modulato è poi opportuno che le sue caratteristiche in termini di banda siano tali da contenere il flusso multiplato. In fase di ricezione, presso il nodo B, si effettua l'operazione inversa di demodulazione. Il segnale multiplato viene filtrato nell'intorno della banda di interesse di un segnale tributario e successivamente, mediante operazione inversa di demodulazione, il segnale tributario è riportato in banda base.



Multiplazione a divisione di codice (CDM per segnali numerici)

Questa tecnica di multiplazione si riferisce al caso in cui i flussi informativi tributari sono costituiti da un flusso di cifre binarie. Prima di considerare i due casi più importanti della multiplazione a divisione di codice è opportuno ricordare che ogni trasmettitore riceve in ingresso un flusso di bit e genera in uscita un segnale analogico. Il segnale analogico creato dal trasmettitore contiene l'informazione numerica da trasmettere e viene immesso nel canale logico. Il ricevitore, quindi, riceverà un segnale

analogico simile a quello trasmesso e realizzerà, a partire da quest'ultimo, la conversione del segnale analogico in flusso di bit (l'informazione che si intendeva trasmettere).

Direct sequence

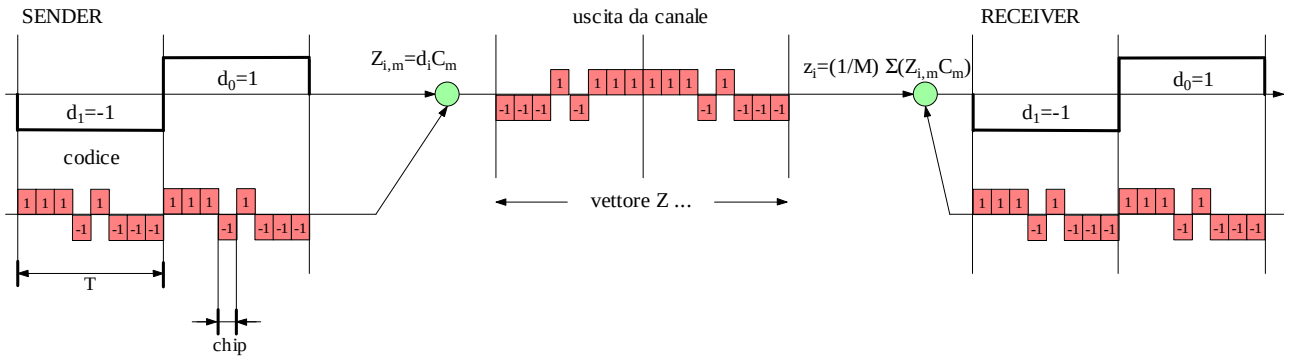
Supponiamo di avere un flusso di NM bit provenienti da N tributari che contemporaneamente inviano i loro flussi informativi presso il nodo trasmettitore. Quest'ultimo dedica a ciascun flusso tributario un intervallo di tempo T necessario a trasmettere M bit. Il meccanismo usato dai tributari per fornire gli M bit può essere vario, ciò ad esempio potrebbe corrispondere al caso in cui dal canale logico che va da A a B siano generati N distinti canali logici (di qualità inferiore alla qualità complessiva del cavo), ognuno per ciascun tributario, ed ognuno di essi capaci di trasmettere 1 bit in T secondi (1/T bps). Altra ipotesi potrebbe ad esempio coincidere con il caso in cui il canale logico è in grado di trasmettere N bit in T secondi (N/T bps). Ad ogni modo, qualunque sia il metodo utilizzato dai flussi tributari, è importante sapere che per ogni flusso tributario viene assegnato un codice. Ogni singolo bit del flusso tributario, quindi, qualora debba essere inviato nella rete, viene codificato mediante il suddetto codice. Infine, l'intervallo di tempo T dedicato a ciascun flusso tributario viene ulteriormente suddiviso in altrettanti intervalli più piccoli detti chip. Ad ogni chip corrisponde un determinato valore del codice di codifica. Il bit da trasmettere è quindi codificato mediante operazione di moltiplicazione del bit in esame con le varie componenti del codice. In altre parole l'i-esimo bit del flusso informativo da trasmettere è dato da $z_{i,m}=C_m S_i$ dove S_i indica il bit da trasmettere e C_m la m-esima componente del codice. Il trasmettitore genera un segnale analogico di durata T (l'intervallo dedicato alla trasmissione di un bit). Indichiamo, allora, con Z il vettore che determina la forma del segnale analogico:

$$\underline{Z} = C_1 S_1 + C_2 S_2 + ... + C_N S_N$$

dove le cifre binarie, per convenienza matematica, sono rappresentato mediante i simboli 1 e -1. C_1 è la cifra binaria proveniente dal primo codice tributario, C_2 dal secondo e così via... Mentre S_1 è il codice abbinato al primo tributario, S_2 quello al secondo e così via... Il codice usato per ogni flusso tributario consiste in una sequenza di *i* bit. I flussi codificati sono, poi, tra loro sommati. Qualora non si verificano interferenze il ricevitore è in grado di ricavare dal flusso multiplato ogni singolo flusso tributario attraverso il seguente calcolo:

$$z_i = \frac{1}{M} \sum_{m=1}^{M} Z_i C_m$$

Un esempio, con unsolo sender, può chiarire il procedimento:



Il mondo è lungi dall'essere ideale ed allora il trasmettitore, così come il ricevitore, lavora in presenza di disturbi ed interferenze ed allora possiamo modellare il vettore ricevuto in ricezione aggiungendo a quest'ultimo una componente vettoriale che modella appunto il disturbo:

$$\underline{Z} = C_1 S_1 + C_2 S_2 + ... + C_N S_N$$
$$\underline{R} = \underline{Z} + \underline{N}$$

R è il vettore ricevuto, N il vettore di variabili aleatorie che modella la presenza dei disturbi agenti sul canale logico e Z è il vettore contenente i flussi multiplati. In tal caso, effettuando l'estrazione dei singoli flussi tributari siamo in presenza di alcune componenti di disturbo, infatti:

$$z_i = \frac{1}{M}\left(\underline{R} \cdot S_i\right)$$
$$z_i = \frac{1}{M}\left(\underline{Z} + \underline{N}\right) \cdot S_i = \frac{1}{M}\left[\left(C_1 S_1 + C_2 S_2 + ... + C_N S_N\right) + \underline{N}\right] \cdot S_i$$
$$z_i = \frac{1}{M} C_i \sum_{i=1}^{N} S_i S_i + \frac{1}{M} C_i \sum_{\substack{i=1 \\ i \neq j}}^{N} S_i S_j + \frac{1}{M} \sum_{i=1}^{N} n_i \cdot S_i$$

Il termine che coinvolge la sommatoria per i≠j prende il nome di interferenza di intercodice ed è uguale a zero se i codici sono tra loro ortonormali. Alcuni esempi di codici ortonormali (C_i C_j = 0) si ottengono dalle matrici di Hadamard. Mentre per rendere piccola la quantità che interessa le componenti dei vettori n_i ed S_i è sufficiente rendere elevata la qualita del canale agendo sul rapporto segnale rumore SNR. Attualmente questa tecnica non viene particolarmente considerata come schema di multiplazione sicchè nei sistemi classici sono altre le tecniche impiegate, tuttavia un impiego significatigo gli viene comunque conferito come ad esempio per la realizzazione dell'accesso multiplo ai sistemi radio cellulari di terza generazione (UMTS).

Frequency hopping

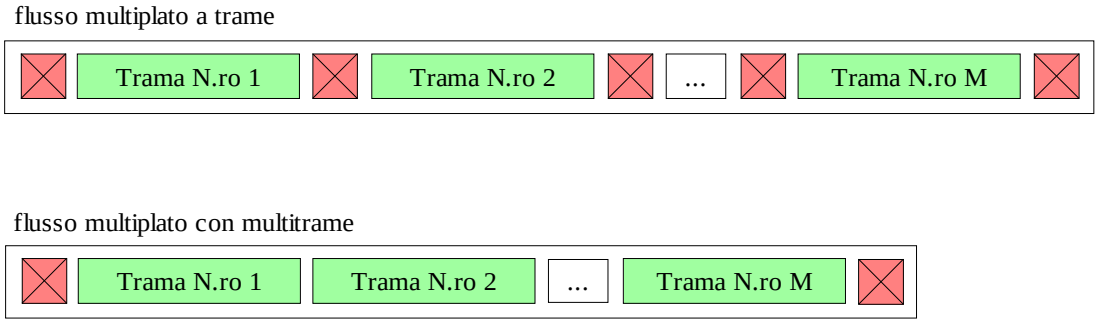
In questo schema il codice, qui detto di hop, determina i salti di frequenza che un flusso tributario deve compiere. Ciascun flusso di bit in ingresso al trasmettitore viene trasformato in un segnale analogico la cui forma dipende fortemente dal contenuto informativo, ciascun segnale analogico viene poi modulato a diverse frequenze per formare il flusso multiplato. Tuttavia in tal caso, a differenza dell'FDM, la banda di frequenza assegnata a ciscun tributario non è sempre la stessa ma cambia in base ad un codice che ne stabilisce i salti. La conoscenza della sequenza di hopping permette al ricevitore di seguire il flusso tributario e di demodulare il segnale.

Il ricevitore può anche demodulare un singolo flusso tributario, per fare ciò è necessario seguire i salti compiuti da un solo flusso tributario. Le modalità con cui cambiano le frequenze del segnale devono garantire la minima sovrapposizione possibile con gli altri tributari. Tuttavia, ogni tanto può accadere che due o più flussi tributari utilizzano contemporaneamente la stessa portante di frequenza generando errori in ricezione.

Tale evento si verifica ad esempio quando il numero di hop assegnati ai flussi tributari superano il limite massimo di hop necessari a garantire l'assenza di interferenza. Anche quando il numero di hop è invece sufficiente a tenere separati i flussi informativi può tuttavia verificarsi una interferenza con un flusso informativo il cui orologio (utile al cambio di frequenza) non è sincronizzato reciprocamente con gli altri.

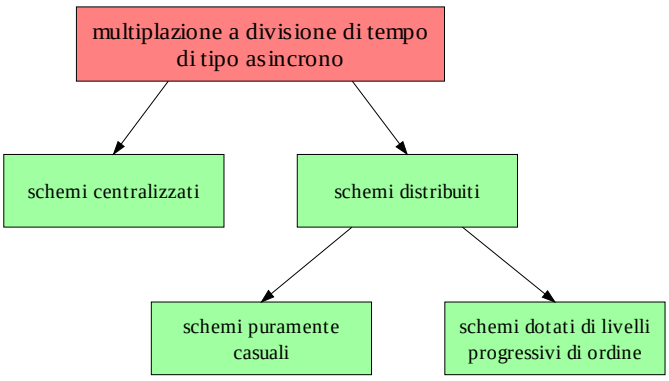
Multiplazione a divisione di tempo (TDM)

Questa tecnica è usata nel caso in cui i flussi tributari sono di natura numerica e quindi binaria. Tuttavia tale tecnica di multiplazione è utilizzata anche in presenza di segnali analogici, in tal caso infatti è possibile vedere il segnale analogico come un segnale a dati campionati e quindi anch'eeso di natura numerica. Quest'ultimo è poi ricostruito presso il nodo di destinazione mediante appositi trasduttori.
Il flusso multiplato, nel caso della multiplazione a divisione di tempo, è costituito da un flusso continuo di bit all'interno del quale prendono posto, secondo un fissato ordine, i bit dei flussi tributari. Per questo motivo il canale logico deve presentare qualità superiori e tali da sostenere la frequenza di cifra del flusso multiplato.
Il flusso multiplato viaggia, dunque, presso il nodo di destinazione. Successivamente i flussi tributari possono prendere direzioni diverse, è allora necessario un operazione inversa di demultiplazione che separa i vari flussi tributari.
Il multiplatore che genera il flusso multiplato può creare quest'ultimo secondo un bit rate costante oppure variabile (frequenza di cifra). Quando il flusso multiplato è a frequenza di cifra costante la multiplazione a divisione di tempo si dice essere sincrona.
All'interno del flusso multiplato si individuano più trame, ogni trama è associata ad un flusso tributario. Ogni trama è lunga N bit consecutivi. Esisterà allora la trama lunga N_1 bit associata al primo flusso tributario, la trama N_2 bit associata al secondo flusso tributario, e così via... Esiste inoltre una trama lunga N_A bit ed associata al canale ed è da quest'ultimo utilizzata per fronteggiare alcuni problemi che in seguito vedremo. Tipicamente ad ogni flusso tributario è concesso aggiungere nella trama otto bit adiacenti. Se però i flussi tributari non hanno la stessa frequenza di cifra alcuni di questi inseriranno nella trama un numero di bit inferiori. Altri tributari, invece, potrebbero addirittura non inserire alcun bit nella trama a loro riservata. In tal caso, per fare in modo che la frequenza di cifra rimanga costante il multiplatore può decidere di riempire le trame vuote. In altri casi, poi, possono esistere flussi tributari che non inseriscono i bit nelle trame a loro assegnate in maniera ordinata, esse si limitano semplicemente ad accodare la propria trama subito dopo quella la precede. Quando ciò accade si dice di generare una supertrama (trame adiacenti).
Se si decide per la tecnica della supertrama il nodo ricevitore non è più in grado di conoscere il bit di inizio e fine trama di ciascun tributario (non ne conosce nemmeno l'ordine con cui le trame si susseguono). Per risolvere tale problema vengono allora usati i bit dedicati al canale per indiciare con dei puntatori i punti in cui iniziano le trame dei flussi tributari. I bit usati per la distinzione delle trame sono talvolta ricordati come bit di allineamento. Essi possono variare ciclicamente per ogni tributario cosicchè questo verrà ad occupare posizioni diverse in ogni multitrama, oppure possono essere sempre gli stessi qualora il flusso tributario si disponga sempre nella medesima posizione all'interno della multitrama.
Se ogni flusso tributario immette a piacimento i propri bit presso il multiplatore di nodo si dice allora di effettuare una multiplazione asincrona. Il flusso multiplato che viene generato può essere, anche in questo caso, a bit rate costante oppure variabile. Se il flusso multiplato è a bit rate costante ed un flusso tributario in un certo istante di tempo non presenta alcun bit presso il multiplatore, affinchè il flusso multiplato continui ad essere a bit rate costante è opportuno inserire dei bit vuoti all'interno della trama. Tali bit di trama vuoti e quindi privi di informazione possono anche essere utilizzati per consentire al ricevitore di identificare l'istante di inizio di ciascun intervallo di tempo riservato a ciascun tributario.



Questa tecnica è particolarmente adatta al caso in cui i flussi tributari hanno un elevato GI. In tal caso è necessario risolvere i possibili problemi di contesa. Se si adotta la modalità orientata al ritardo (coda di attesa per l'accesso alla risorsa), un flusso tributario può essere inoltrato senza ritardo in assenza di contesa oppure può subire un ritardo qualora vi sia contesa fra più flussi tributari. Il ritardo produce effetti negativi sulla frequenza di cifra del flusso tributario. Se la frequenza di cifra risultante è comunque soddisfacente il problema non si pone, qualora il ritardo è invece critico per un applicazione si può allora pensare di compensare il ritardo bufferizzando il flusso di bit in ricezione.
Se invece si adotta la modalità orientata alla perdita (pre-assegnazione per l'accesso alla risorsa) il flusso informativo non subisce ritardi. Tuttavia in caso di contese alcuni bit del flusso informativo potrebbero andare persi.
Per quanto riguarda la pre-assegnazione occorre poi stabilire una regola che permetta di accettare o rifiutare una domanda di assegnazione. Una regola di accettazione ad esempio potrebbe interessare la somma dei ritmi binari di picco degli attuali flussi tributari con quella del flusso tributario richiedente. Qualora tale somma eccede la capacità del canale logico il flusso tributario richiedente viene rifiutato.

Multiplazione a divisione di tempo di tipo asincrono
Quando più flussi tributari, dotati di diversi GI, hanno esigenze di trasmissione verso un nodo della rete essi vi accedono mediante accesso multiplo (nel caso ad esempio della propagazione libera) oppure mediante N distinti canali logici (nel caso ad esempio della propagazione guidata), ciascuno per ogni tributario. I flussi tributari possono non avere poi le stesse esigenze in termini di frequenza di cifra, altri invece possono addirittura non inserire nessun bit nella trama a loro riservata, semplicemente perchè in quell'istante di tempo l'entità di rete che genera il flusso tributario non ha esigenze di trasmissione. In tal caso è opportuna una tecnica di multiplazione di tipo asincrono. Il caso dell'accesso multiplo è complicato dal fatto che ogni tributario non conosce quale sia l'intervallo di tempo da utilizzare per la trasmissione. La multiplazione a divisione di tempo di tipo asincrono ammette diverse realizzazioni, è possibile distinguere due schemi particolari nella quale rientrano tutte le tecniche attualmente in uso, si tratta di: schemi centralizzati e schemi distribuiti. Gli schemi distribuiti a loro volta comprendono gli schemi puramente casuali e gli schemi dotati di livelli progressivi di ordine.



Schemi centralizzati

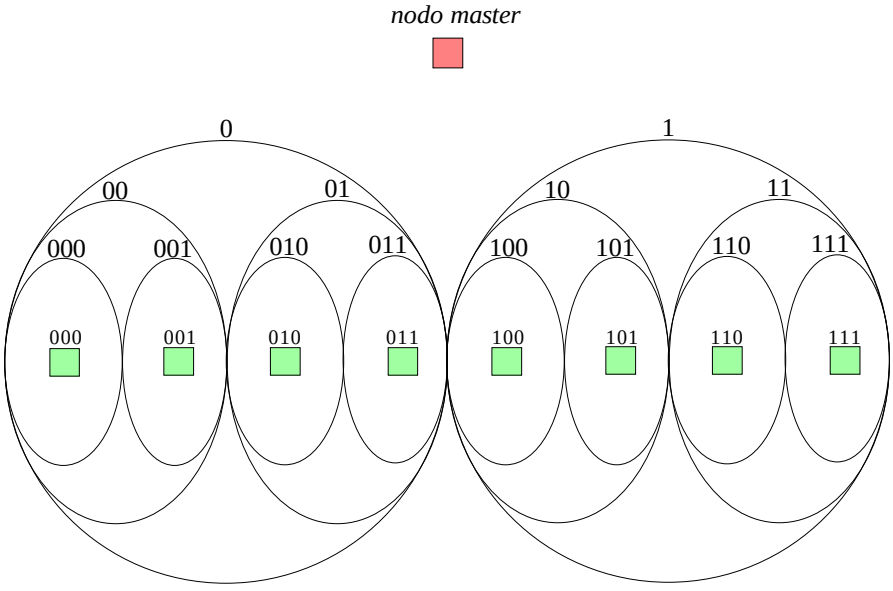
Gli schemi centralizzati sono così chiamati a causa di un nodo centralizzato la cui funzione è essenzialmente quella di raccogliere le reali esigenze di trasmissione che le varie entità di rete ad esso collegate presentano in maniera implicità o esplicita. A differenze delle tecniche casuali gli schemi centralizzati riescono a conseguire una maggiore efficienza nell'utilizzazione delle risorse grazie all'assenza di collisioni (il problema delle collisioni dei flussi informativi è fortemente sentito negli schemi casuali). Tuttavia, il sovraccarico richiesto dal protocollo per evitare le collisioni può rendere tale soluzione inadeguata quando la rete è particolarmente scarica. In tal caso l'entità di rete deve prima manifestare la sua necessità di trasmettere, quindi il nodo centralizzato ne riceve la richiesta e solo dopo l'entità di rete richiedente viene autorizzata dal nodo centralizzato alla trasmissione. Nel caso degli schemi casuali invece, quando la rete è particolarmente scarica (come nel caso di una sola entità di rete avente necessità in trasmissione), l'entità di rete in trasmissione inizia a trasmettere subito sul canale il proprio flusso informativo, senza perdere tempo.
In altre parole, i protocolli ad accesso casuale hanno un throughput pari ad R bit/s, con R capacità di canel, quando nella rete vi è un solo nodo attivo. Quando invece ci sono N nodi attivi il throughput ideale da ottenere mediante protocollo deve essere il più possibile vicino ad R/N bit/s. I protocolli ad accesso casuale, a causa delle frequenti collisioni, non realizzano questa proprietà. Nei protocolli a turno centralizzato si cerca allora di perseguire tale caratteristica. Alcuni importanti schemi centralizzati sono il polling ed il probing (detto anche polling adattativo).

Polling

Il polling è un protocollo a turnazione. Un nodo master sonda a rotazione ciascuna entità di rete ad esso collegato. Il compito del nodo master è quello di verificare se le entità di rete hanno la necessità di trasmettere. Se l'entità di rete risponde in maniera positiva essa, allora, inizierà a trasmettere non appena il nodo master darà a quest'ultima il permesso. Il nodo master, che in fase di polling interroga tutte le entità di rete, ha poi un elenco delle entità interessate a trasmettere. Il primo nodo in trasmissione inizia ad inviare il proprio flusso informativo ed avvisa il nodo master non appena termina la trasmissione dati. Quando ciò si verifica il nodo master avvisa la seconda entità di rete in elenco affinchè questa inizi anch'essa a trasmettere.
Il messaggio che il nodo master invia alle entità di rete viene detto messaggio di polling. Esso contiene al suo interno un campo di indirizzo per il nodo destinatario. Il permesso a trasmettere dati viene invece inviato a tutti i nodi in ascolto ma solo quello che riconosce il proprio indirizzo nel messaggio ne riceve i benefici. Questo protocollo elimina il problema delle collisioni ed ha una efficienza più alta in termini di utilizzo delle risorse, tuttavia presenta l'inconveniente dei tempi di sondaggio che costituiscono inevitabilmente un ritardo. Altro problema assai più serio è la possibilità di un guasto nodo master,

siccome il protocollo fa affidamento ad uno schema centralizzato l'intero canale non può più funzionare.

Probing o polling adattativo
Il probing è una variante del polling e cerca di minimizzare il numero di messaggi che il nodo master invia alle entità di rete. Quando il sistema è particolarmente scarico, ad esempio quando vi è una sola entità di rete interessata alla trasmissione, il nodo master, nel caso del polling, deve comunque interrogare ciascun terminale ad esso collegato. Il probing, invece, tenta di risolvere questo inconveniente partizionando l'insieme dei nodi collegati in tanti sottoinsiemi.
Il nodo master invia in broadcast (quindi verso tutti i terminali) un messaggio di probe. A tale messaggio rispondono solo i nodi interessati alla trasmissione dati (notare che nel messaggio di probe non vi è alcun indirizzo dell'entità di rete). Il nodo master ha poi la capacità di stabilire se al suddetto messaggio di probe giunge almeno una risposta affermativa. Qualora più nodi rispondono in maniera positiva il nodo master comunque stabilisce che esiste l'esigenza di trasmettere dati ed inizia a partizionare l'insieme di nodi a lui collegato fino a scovare il nodo richiedente.
Supponiamo ad esempio che il numero di terminali collegato al nodo master sia una potenza di 2, e quindi $M\ terminali=2^n$. In questo modo ogni nodo terminale è individuato da un unidirizzo lungo n bit. Quando al messaggio di probe segue una risposta affermativa, il nodo master inizia il suo sondaggio a partire dall'indirizzo avente uno o più zeri, se trova il nodo richiedente concede a questo il permesso di trasmettere, altrimente partiziona ulteriormente l'insieme.
Ad esempio, supponiamo che sia $n=3$ e il nodo con indirizzo 010 è l'entità interessata alla trasmissione. Il nodo master inizia il sondaggio a partire dall'insieme di terminali il cui indirizzo inizia con 0, il nodo con indirizzo 010 risponde positivamente. Quindi il nodo master partiziona l'insieme trovato con un messaggio di probe destinato all'insieme di terminali il cui indirizzo inizia per 00, nessuna entità di rete risponde al messaggio. Adesso il nodo master invia il messaggio di probe all'insieme di terminali il cui indirizzo inizia per 01, il nodo con indirizzo 010 risponde positivamente. Infine, il nodo master invia il messaggio di probe al nodo 010 che quindi risponde positivamente ricevendo in seguito il permesso alla trasmissione.

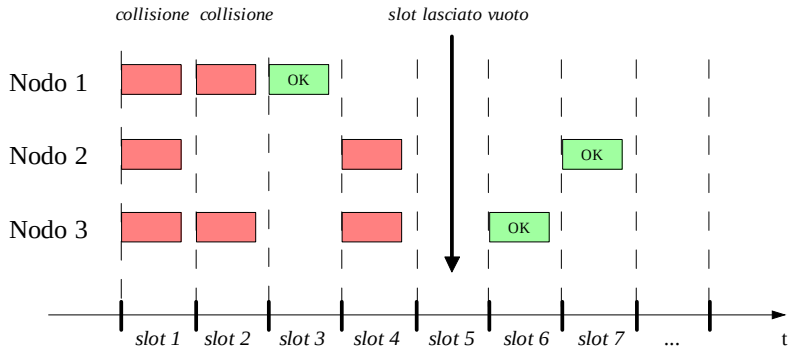


Quando esiste un solo terminale con esigenze di trasmissioni, nel caso del polling sono necessari 2^n messaggi di polling. Nel caso del probing, invece, sono necessari $2n+1$ messaggi di probing (nel caso peggiore, quando cioè il nodo attivo è collocato sull'ultimo indirizzo disponibile, quello con tutti 1).
Il probing è tuttavia inefficiente in caso di carichi pesanti, ad esempio quando tutti gli M terminali hanno dati da trasmettere. In tal caso sono infatti necessari:

$$\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$$

a differenza del polling che ne propone sempre 2^n.

Schemi puramente casuali
Questo schema prevede che un nodo trasmittente, quando ha la necessità di trasmettere dati, invia sul canale comune il flusso informativo. La trasmissione avviene alla massima velocità supportata dal canale. Qualora due o più nodi abbiano la stessa necessità di trasmissione si può verificare la collisione tra i flussi informativi. La ritrasmissione del flusso informativo andato perso non avviene subito ma ciascun nodo sceglie, indipendentemente dagli altri nodi, un intervallo di tempo utile all'attesa.
Il protocolo più semplice, che sfrutta questi principi, è il protocollo p-ALOHA (*pure ALOHA*). I nodi trasmettono in qualunque istante di tempo i propri pacchetti. Siccome il canale è di tipo broadcast ciascun nodo ascolta il flusso trasmesso, se quest'ultimo risulta essere diverso da quello previsto il nodo assume che si sia verificata una collisione e ritrasmette i pacchetto aspettando un intervallo di tempo aleatorio. L'algoritmo di back-off viene utilizzato per fissare l'entità di tale pausa. Successivamente la trasmissione riprende e solo quando si verificano un certo numero di collisioni consecutive l'entità di rete preposta alla trasmissione comunica allo strato superiore l'incapacità di comunicare. Una variante del protocollo p-ALOHA è il protocollo slotted-ALOHA. Assumeremo che tutti i frame consistano esattamente di L bit. Il tempo viene suddiviso in slot, ciascun slot ha la dimensione necessaria alla trasmissione di un frame. Quando un nodo deve trasmettere un frame deve prima attendere l'inizio di un nuovo slot (i nodi trasmettono alla stessa frequenza di cifra). Se non si verifica una collisione, il nodo ha trasmesso con successo il suo frame e non deve effettuare alcuna ritrasmissione. Se invece si verifica una collisione, il nodo rileva la collisione prima del termine dello slot e ritrasmette il suo frame in ciascun slot successivo con probabilità p (ciò si ripete finchè la trasmissione avviene con successo). Indichiamo con p la probabilità di ritrasmissione nel successivo slot e con $1-p$ la probabilità che lo slot successivo venga saltato. Quando nella rete vi è un unico nodo attivo il protocollo slotted-ALOH garantisce la trasmissione alla massima frequenza di cifra. In presenza di collisioni e quindi di altri nodi attivi la frequenza di cifra rimane la stessa ma la velocità di trasmissione si riduce a causa delle ritrasmissioni:



Nell'ipotesi di N nodi con *p* probabilità di successo ed *(1-p)* probabilità di ritrasmissione, la probabilità che un dato slot sia uno slot di successo è quello in cui uno degli N nodi riesce a trasmettere mentre i restanti (N-1) nodi attendono. La probabilità che un nodo trasmetta è dunque *p*, la probabilità che i rimanenti nodi non trasmettano è $(1-p)^{N-1}$. Tale probabilità di successo esprime l'efficienza dello slotted-ALOHA. Per ottenere l'efficienza massima occorre valutare la probabilità p^* che rende massima l'efficienza (L'efficienza massima dello slotted-ALOHA è pari ad *1/e* mentre quella del p-ALOHA è *1/2e*).

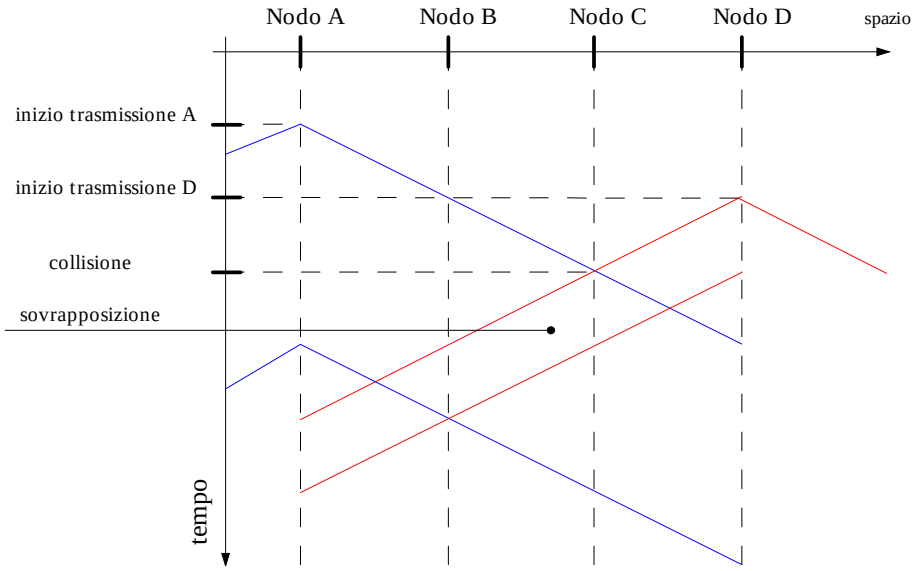
Una variante di tali protocolli è il protocollo ALOHA-*with capture*. Tale protocollo si può presentare sia nella versione p (pure) che nella versione s (slotted). I nodi, divisi in due sottoinsiemi, appartengono ad uno dei due gruppi. Un primo gruppo riguarda tutti i nodi capaci di trasmettere a bassa potenza, un secondo gruppo tutti quei nodi capaci di trasmettere ad elevata potenza. La collisione fra un flusso informativo di un gruppo con quello di un altro gruppo distrugge il pacchetto informativo a bassa potenza facendo sopravvivere quello ad alta potenza che quindi può ugualmente giungere a destinazione (pertanto non è richiesta la ritrasmissione nonostante la sovrapposizione).

Il protocollo CSMA (carrier sesitive multiple access) è capace di ridurre la probabilità del numero di collisioni, esso si basa su due importanti principi:

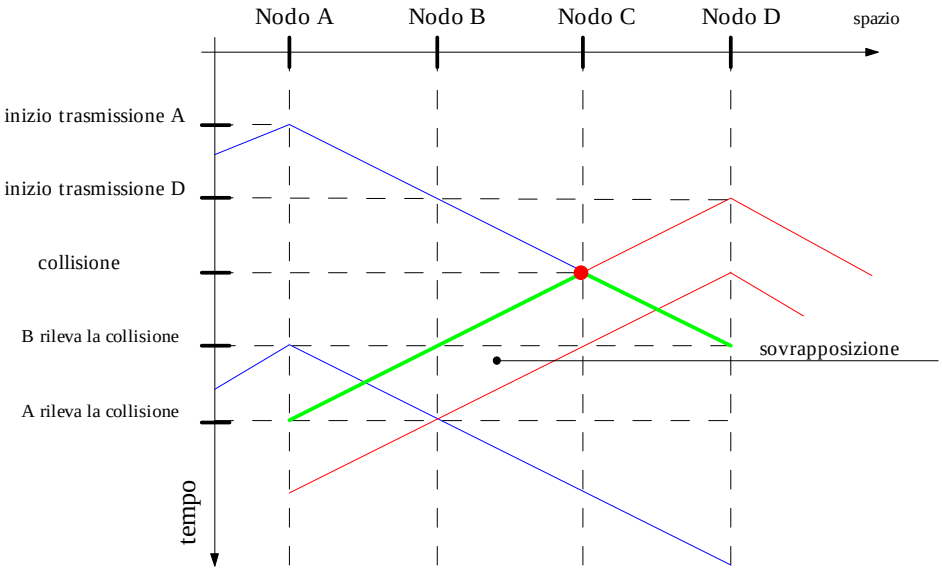
- ascoltare prima di parlare, tale principio si traduce nella rilevazione della portante. Un nodo ascolta il canale prima di trasmettere (carrier detect). Se un frame sta attraversando il canale occupandolo il nodo decide di attendere un tempo aleatorio prima di riascoltare nuovamente il canale nella speranza di trovarlo libero.
- Se qualcuno inizia a parlare contemporaneamente a voi smettete di parlare. Questo principio si traduce nella rilevazione della collisione. Se un nodo trasmette esso ascolta anche il canale. Se quindi rileva una collisione (un altro nodo che interferisce alla trasmissione) il nodo arresta la sua trasmisione (in tal caso infatti è inutile continuare poichè i flussi si sovrappongono e giungono in cattive condizioni a destinazione) e impiega qualche algoritmo per determinare il momento in cui riprendere la trasmissione.

Il protocollo CSMA può essere persistente (se il nodo, trovando il canale occupato, continua a restare in ascolto con lo scopo di trasmettere non appena esso si libera) oppure non persistente (se il nodo, trovando il canale occupato, smette di ascoltarlo e riprova qualche istante di tempo successivo a riascoltarlo). Siccome il CSMA realizza il primo dei due principi prima citati si è pensata ad una sua evoluzione o variante.

Una variante del protocollo CSMA è il protocollo CSMA/CD (CSMA *with collision detect*). Oltre ad ascoltare il canale come già fa il protocollo CSMA per rilevare la portante, un nodo resta in ascolto per un certo intervallo di tempo sufficiente a rilevare eventuali collisioni. Tali collisioni infatti, a causa dei ritardi di propagazione del segnale trasmesso verso gli altri nodi di rete, si possono verificare in fase iniziale di trasmissione. Quando cioè un nodo di rete inizia la sua trasmissione ed il segnale arriva con ritardo ad uno dei nodi a lui adiacenti cosicchè quest'ultimo, ritenendo il canale libero, inizia anch'esso la sua trasmissione. La rilevazione della collisione ha come effetto immediato l'arresto della trasmissione.



Schema CSMA



Schema CSMA/CD

Analisi prestazionale degli schemi casuali

Se indichiamo con G il numero medio di pacchetti trasmessi in un intervallo di tempo lungo T secondi (quindi G/T pacchetti al secondo), con γ probabilità di avere trasmissioni senza collisioni, allora il numero medio di pacchetti trasmessi senza alcuna collisione, che denotiamo con S, è pari a $S=G\gamma$.

Se immaginiamo di modellare il traffico del canale con una variabile di Poisson possiamo allora calcolare la probabilità γ che si abbia una collisione. Affinchè non si verifichino collisioni è allora necessario che nell'intervallo di tempo di durata *t* non arrivino nel canale altri pacchetti. Tale probabilità la possiamo valutare come $p=e^{-\gamma t}$. Dove γ è il numero medio di pacchetti che nell'unità di tempo viene immessa nel canale ed è quindi pari a G/T. Una collisione ha luogo se un altro nodo della rete, diverso da quello considerato, inizia anch'esso una trasmissione durante un intervallo di durata 2T centrato sull'istante di tempo in cui il nodo considerato inizia a trasmettere. Tale intervallo è solitamente detto intervallo di vulnerabilità. Nel caso del p-ALOHA l'intervallo di vulnerabilità è di 2T secondi:

$$\left.\begin{array}{l} S = G\gamma \\ \gamma = e^{-\lambda t} \\ \lambda = \dfrac{G}{T} \end{array}\right\} \rightarrow \gamma = e^{-\frac{G}{T}2T} = e^{-2G} \rightarrow S = Ge^{-2G}$$

Il massimo valore di S si ha per G=0.5 (vale circa 0.184). Nel caso dello s-ALOHA l'intervallo di vulnerabilità è invece di T secondi:

$$\left.\begin{array}{l} S = G\gamma \\ \gamma = e^{-\lambda t} \\ \lambda = \dfrac{G}{T} \end{array}\right\} \rightarrow \gamma = e^{-\frac{G}{T}T} = e^{-G} \rightarrow S = Ge^{-G}$$

Il valore massimo di S si ha per G=1 (vale circa 0.368).

Schemi distribuiti di tipo ordinato

Gli schemi distribuiti di tipo ordinato presentano notevoli vantaggi in termini di affidabilità e di prestazioni nei confronti di quelli casuali ma sono più complicati da attuare e gestire, proprio per questo motivo non hanno ancora raggiunto un certo livello di diffusione. L'imposizione di un ordine gerarchico ai terminali della rete risolve il problema delle collisioni. L'ordine, ad esempio, può essere ricavato anche dall'indirizzo fisico dei nodi. Un piccolo frame detto token viene adoperato per particolari scopi, esso è ad esempio scambiato tra i nodi della rete e solo il nodo che lo possiede ha diritto alla trasmissione. Terminata la trasmissione il nodo che al momento dispone del token deve passare quest'ultimo al successivo nodo. Ogni nodo non deve quindi conocere gli indirizzi di tutti i nodi della rete, per il suo funzionamento è opportuno sapere il solo indirizzo del nodo successivo e quindi del nodo che topologicamente lo segue. Anche se in questo schema non vi è la presenza di alcun nodo master che potrebbe causare la caduta dell'intera rete a causa di un guasto sullo stesso nodo master è comunque possibile una situazione di blocco temporaneo della rete. Esistono infatti delle procedure che tentano di ripristinare un ordine logico qualora ad esempio il token venga perso.
Il token viene trasmesso dal nodo che lo possiebe secondo una modalità broadcast, tutti i nodi quindi lo ricevono. Tuttavia, il nodo che lo ha usato per ultimo ripone nel token stesso l'indirizzo logico del nodo a lui successivo cosicchè solo il nodo che rinosce il proprio indirizzo nel token può effettivamente usarlo.

Il nodo N1 dotato di token ne fa uso trasmettondo il proprio flusso informativo e lo passa quindi al successivo nodo N2. Se dopo un certo istante di tempo il nodo N1 non ascolta alcuna trasmissione in transito sulla rete assume che il token mandato ad N2 sia andato perso e lo ritrasmette. Se dopo un certo numero di volte il nodo N2 non inizia a trasmettere, allora, il nodo N1 chiede in broadcast il successivo nodo di N2. Ad una siffatta richiesta dovrebbe rispondere il nodo N3 che in seguito riceverà da N1 il token (N1 è sempre lì in ascolto, finchè non passa il token). Qualora nessun nodo risopnde al messaggio broadcast mandato da N1 quest'ultimo assume che l'anello logico si è ridotto al solo nodo N1 che quindi userà da solo il canale. Solitamente un nodo che non passa un token viene rimosso dall'anello logico, esso viene quindi scavalcato nella maniera appena vista (messagio broadcast per scoprire il nodo a lui successivo).

La funzione di commutazione

Non tutti i flussi tributari hanno la medesima destinazione cosicchè a partire dal flusso multiplato è necessario demultiplare quest'ultimo ed anlizzare ogni suo flusso tributario affinchè venga stabilito per ciascuno di esso il giusto collegamento di uscita dal nodo. I flussi tributari sono mandati al multiplatore che controlla l'uscita dal nodo. Questa funzione di smistamento dei flussi tributari costituisce la funzione di commutazione. Più in generale la funzione di commutazione viene realizzata per mezzo delle funzioni di attraversamento e di instradamento che a breve vedremo.

La funzione di attraversamento

Per attraversamento si intende il percorso interno al nodo compiuto dal flusso informativo (dopo che esso sia stato demultiplato). Il flusso multiplato in ingresso al nodo viene dunque demultiplato ed ogni singolo tributario (dopo essere stato analizzato) è mandato al multiplatore che comanda un collegamento in uscita dal nodo e che quindi formerà un nuovo flusso multiplato. L'attraversamento del nodo avviene seguendo un percorso interno al nodo e che quindi collega l'ingresso del nodo ad una sua uscita, esso può avvenire secondo due tecniche:

- tecnica della connessione diretta;
- tecnica della connessione ad immagazzinamento e rilancio anche detta store and forward;

La tecnica della connessione diretta è stata la prima ad essere storicamente impiegata e prevede un ritardo di attraversamento pressochè nullo, requisito importante per le applicazioni telefoniche. Infatti, tale tecnica è stata per la prima volta adoperata proprio in ambito telefonico. All'epoca, quando questa tecniva veniva presentata, non esistevano ancora i calcolatori e l'informazione viaggiava nella rete sottoforma di segnale analogico. La memorizzazione del flusso informativo che transitava nel nodo, oltre che inutile era quindi impossibile.
La tecnica di store end forward, invece, si è affermato quasi contemporaneamente all'introduzione dei calcolatori elettronici che consentivano facilmente la memorizzazione di un certo numero di bit del flusso informativo. Quando questa tecnica fu introdotta non si avevano limiti stringenti sui ritardi, i flussi informativi scambiati erano essenzialmente orientati allo scambio dati tra calcolatori elettronici ed inizialmente i ritadi introdotti dallo store and forward sono stati trascurati. Ciò escludeva da questa tecnica l'integrazione di tutti quei servizi che richiedevano un elevato grado di trasparenza temporale.
Nella tecnica store and forward i flussi informativi presenti presso il nodo di rete sottoforma di bit vengono memorizzati in una coda di attesa (la presenza di blocchi di memoria ha permesso appunto di realizzare tale coda) prima di essere trasferiti in uscita. I flussi informativi ammessi in ingresso da un nodo di rete che realizza la commutazione mediante tecnica di store and forward possono avere

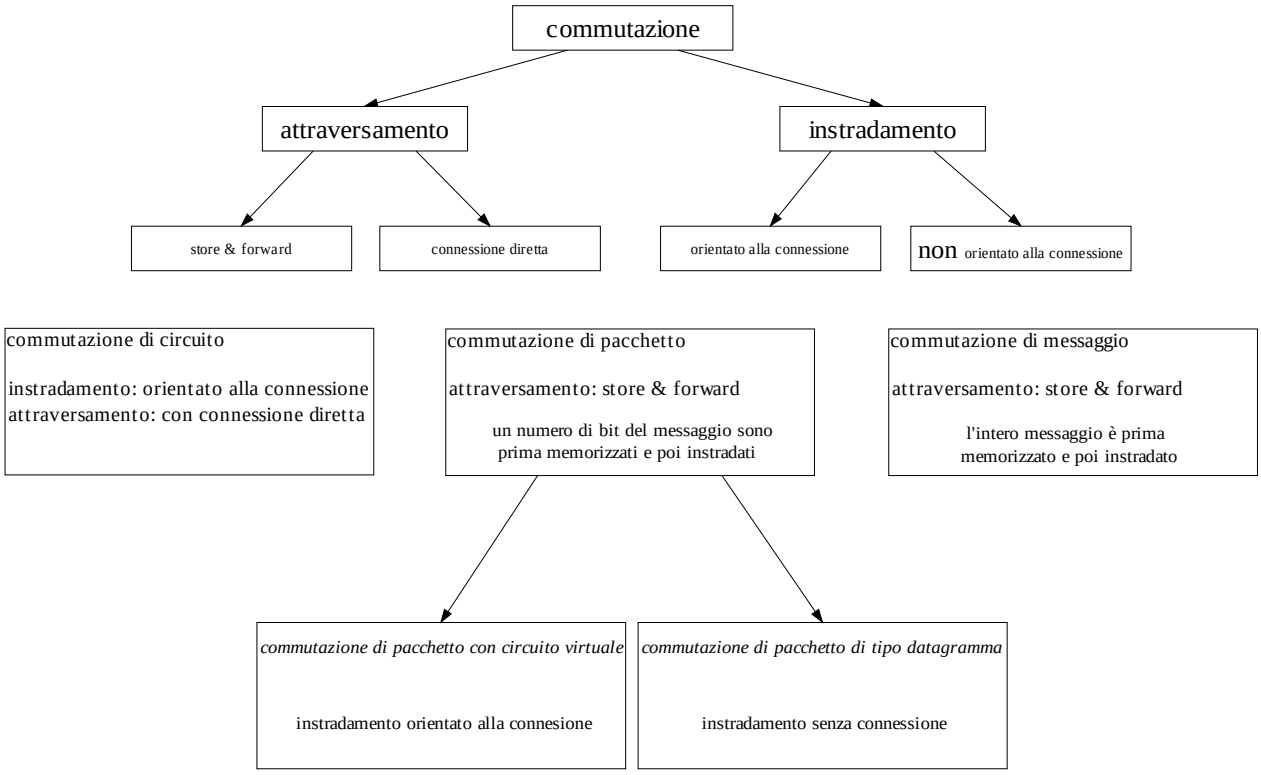
frequenze di cifra variabili. Tuttavia le capacità di memorizzazione del nodo consentono a quest'ultimo di memorizzare un certo numero di bit in transito.
Mentre la tecnica della connessione diretta, non prevedendo alcuna capacità di memorizzazione del nodo, richiede al progettista di dimensionare i flussi in ingresso e quelli di uscita in maniera tale che in ogni istante di tempo il flusso in uscita sia maggiore o uguale alla somma dei valori di picco dei flussi in ingresso. Pertanto si deve assegnare ad ogni flusso tributario a ritmo variabile un servizio di trasferimento con fissate qualità in termini di frequenza di cifra (dimensionata rispetto al valore di picco) e ciò limita fortemente il numero dei tributari che possono accedere al multiplatore del nodo.
Per il motivo appena visto la tecnica della connessione diretta è spesso utilizzata qualora i flussi informativi in ingresso abbiano un ritmo binario costante. Tuttavia è opportuno precisare che non sempre le frequenze di cifre considerate per il dimensionamento della rete siano effettivamente quelle con cui si trasmettono i bit nella rete. Ci possono infatti essere dei lievi scostamenti dovuti all'utilizzo di diversi orologi di nodo che emettono perciò frequenze diverse. E' importante assicurare condizioni di sincronizzazione di rete onde evitare la perdita di allineamento della trama, si distinguino pertanto i seguenti casi:

- condizione di mesocronismo, le frequenze emesse dagli orologi di nodo sono mediamente uguali;
- condizione di plesiocronismo, le frequenze emesse dagli orologi di nodo differiscono entro ristretti margini di tolleranza;
- condizione di sincronismo, un unico orologio di rete comanda la sincronizzazione per tutti i nodi della rete;

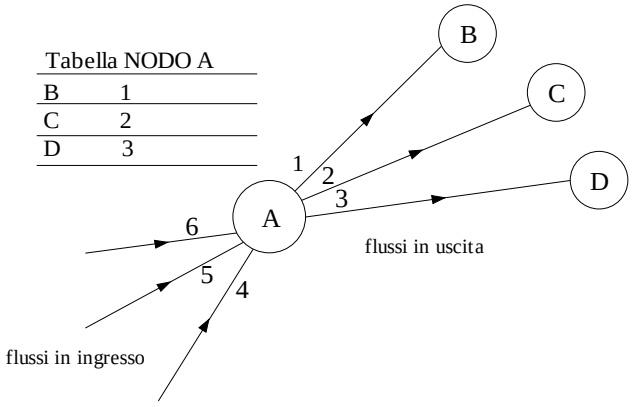
La tecnica di store and forward consente di tollerare le diverse frequenze di cifra avvalendosi di opportuni blocchi di memoria sufficientemente idonei a memorizzare i flussi informativi. Se, inoltre, si dispone di ampie capacità di memorizzazione è possibile adottare anche una temporizzazione asincrona, in tal caso infatti è comunque possibile compensare eventuali scostamenti tra le frequenze di cifra dei vari flussi.

La funzione di instradamento
L'instradamento è una funzione decisionale e consiste nello stabilire verso quale collegamento di uscita del nodo vada instradato ogni flusso che si presenta in ingresso al nodo. Ad esempio, nelle vecchie centrali telefoniche un operatore umano realizzava con la sua mente la funzione di instradamento e attuava tale scelta (funzione di attraversamento) mediante connessione fisica del collegamento in ingresso con uno dei collegamenti in uscita (solitamente ciò si traduceva nell'inserimento di una prolunga che si diramava da un punto di ingresso da cui arrivava il flusso tributario di una utenza fino ad un punto di uscita che invece rappresentava la dstinazione richiesta dall'utente).
Per realizzare l'instradamento esistono due tecniche, la prima si dice essere orientata alla connessione mentre la seconda si dice, al contrario, non essere orientata alla connessione. Nella modalità orientata alla connessione, prima di instaurare la connessione logica fra due utenti viene preventivamente determinato un percorso attraverso i nodi della rete, il flusso informativo dovrà poi seguire tale percorso per tutta la durata della connessione. Nella modalità non orientata alla connessione si ammette la possibilità di avere all'interno della rete diversi percorsi che conducono un pacchetto a destinazione e pertanto non vi è alcuno sforzo iniziale, i pacchetti saranno immessi nella rete e oltre a seguire diversi percorsi logici possono anche giungere presso il nodo di destinazione in unordine diverso da quello con cui sono stati trasmessi.
Nell'instradamento orientato alla connessione possono poi verificarsi due eventualità: la sequenza dei canali logici che viene individuata nella fase iniziale, prima cioè di instaurare il collegamento logico,

può essere interamente dedicata alla connessione logica dei due nodi che intendono scambiarsi l'informazione. In tal caso si dice che la sequenza di canali logici individua un canale fisico; altro caso è quello in cui la sequenza dei canali logici individuati nella rete è invece condivisa con altri flussi tributari di altri terminali di rete. In tal caso la sequenza di nodi individuati viene ricordata con il nome di circuito virtuale o circuito logico.
Non vi è dubbio che nel caso di circuito fisico vi è una vera è propria corsia preferenziale per il flusso informativo scambiato da due nodi di rete, il percorso è interamente esclusivo alla comunicazione e offre notevoli garanzie nei tempi di transito. Il circuito virtuale, invece, poichè prevede delle contese di utilizzazione (più tributari potrebbero ad esempio fare richiesta di un canale logico per la trasmissione del proprio flusso informativo), esso introduce pertanto dei ritardi temporali dovuti all'attesa che questi sono tenuti a rispettare prima di impegnare effettivamente le risorse di rete. Nel caso del circuito fisico le contese possono essere solo di pre-assegnazione, pertanto l'entità di nodo che accede alla risorsa dopo la pre-assegnazione è l'unica a sfruttarne le caratteristiche.
Una rete che realizza la funzione di attraversamento mediante connessione diretta e quella di instradamento orientata alla connessione si dice essere una rete a commutazione di circuito.
Una rete che realizza la funzione di attraversamento mediante tecnica di store & forward è invece detta rete a commutazione di pacchetto se un certo numero di bit che compongono un messaggio informativo sono prima memorizzati e successivamente instradati. Una rete che realizza la funzione di attraversamento mediante tenica di store & forward è, al contrario, detta rete a commutazione di messaggio se il messaggio è interamente memorizzato prima di essere instradato.
Ed ancora, una rete si dice essere a commutazione di pacchetto con circuito virtuale se l'instradamento dei flussi informativi avviene secondo la modalità orientata alla connessione. Essa si dice invece essere una rete a commutazione di pacchetto di tipo datagramma se l'instradamento avviene secondo la modalità non orientata alla connessione.



L’instradamento, l’atto decisionale con cui si stabilisce il collegamento di uscita dal nodo,viene attuato da ogni nodo sulla base di opportune tabelle di instradamento. Ogni nodo possiede una tabella che riporta in corrispondenza di ogni riga i nodi di destinazione, su ogni riga il nodo legge il collegamento di uscita da utilizzare per l’instradamento. Un instradamento siffatto viene anche detto dinamico qualora lo stato della rete e quindi dei suoi collegamenti fisici viene seguito aggiornando le tabelle di instradamento con una certa frequenza. Qualora, invece, l’instradamento viene deciso da un solo nodo di rete, per tutti gli altri nodi, esso si dice essere centralizzato. Un instradamento distribuito è poi possibile se i nodi della rete cooperano tra di loro (scambiandosi informazioni) con lo scopo di determinare congiuntamente il percorso che ciascun pacchetto dovrà seguire per giungere a destinazione.



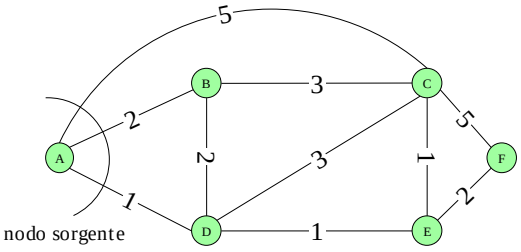
Algoritmi di costruzione di una tabella

L’algoritmo di Djkstra (detto anche LS, link state) usa un informazione globale della rete per stabilire il nodo successivo verso cui instradare i pacchetti informativi, un algortimo decisamente ottimale è invece il distance vector o più semplicemente DV. Si tratta di un algoritmo iterativo, asincrono e distribuito. I due algoritmi appena citati sono quelli maggiormenti adoperati nell’ambito dell’instradamento basato su tabella. Per lo studio degli algoritmi di instradamento modelleremo la rete mediante nodi tra loro collegati. Il collegamento che congiunge due nodi della rete presenterà poi un costo di attraversamento che in seguito analizzeremo meglio e che per ora possiamo intuitivamente legare alla qualità del collegamento fisico.

Algoritmo di Djkstra

L’algoritmo di Djkstra determina i percorsi di costo minimo che partono da un nodo A di sorgente e raggiungono gli altri nodi della rete. Tutto ciò sulla base di informazioni che descrivono l’attuale stato della rete e che costituiscono quindi l’input per l’algoritmo di instradamento. L’algoritmo di Djkstra si basa sullo stato dei link, si tratta di un algoritmo iterativo e ha la proprietà che dopo la K-esima iterazione esso conosce il percorso di minor costo verso K nodi di destinazione.

Per capire meglio il funzionamento dell’algoritmo di Djkstra faremo riferimento ad un esempio di rete generica, i cui collegamenti formano una topologia magliata. L’algoritmo verrà quindi applicato a partire da un nodo iniziale e le successive iterazioni determineranno man mano i cammini avente costo minimo verso ogni nodo della rete.



Nel passo di inizializzazione i percorsi a minor costo sono conosciuti da A, nodo di partenza a cui applicare l'algoritmo, attraverso i vicini a lui collegati direttamente. Nel nostro esempio si tratta dei nodi B, C e D ed avente rispettivamente link con costi di attraversamento di 2, 5 ed 1. I costi verso i nodi al momento non raggiungibili e quindi verso i nodi E ed F sono per adesso messi a ∞.
Nella prima iterazione vengono considerati quei nodi che non ancora sono stati aggiunti al gruppo N, l'insieme di link a minor costo. Nel nostro caso, si vedono i vicini al nodo di partenza A e si aggiunge all'insieme N quello avente minor costo e quindi D(1). Successivamente vengono aggiornati i costi dei link considerando il nodo D appena aggiunto ad N: il costo verso B (sia da A che da D) rimane invariato e vale 2 (in relatà da D è più difficile raggiungere B poichè il costo da pagare da D diviene 3, si preferisce quindi tenere il costo più basso in tabella); il costo verso C, che da A valeva 5 diviene adesso 4 (passando per D); il costo verso E (adesso raggiungibile da A passando attraverso D) è invece 2. Quest'ultimo nodo viene poi aggiunto all'insieme N poichè si tratta del costo al momento più basso trovato in questa iterazione. Le iterazioni continuano fino a toccare tutti i nodi della rete:

N	d(B),p(B)	d(C),p(C)	d(D),p(D)	d(E),p(E)	d(F),p(F)
A	2,A	5,A	**1,A**	∞	∞
AD	2,A	4,D	-	**2,D**	∞
ADE	**2,A**	3,E	-	-	4,E
ADEB	-	**3,E**	-	-	4,E
ADEBC	-	-	-	-	**4,E**
ADEBCF	-	-	-	-	-

(Nella tabella si indica con d(.) la distanza complessiva che intercorre tra un nodo considerato ed il nodo A di partenza mentre con p(.) il precedente nodo)
Il nodo consulta la tabella nel seguente modo, supponendo che un pacchetto abbia la necessità di raggiungere il nodo E, partendo dal nodo A. In corrispondenza della colonna che fa riferimento al nodo E si scorrono le righe fino a trovare quella caratterizzata dalla scritta in grassetto, fatto ciò si legge il valore del nodo p(.). Se tale nodo è l'immediato vicino del nodo A il pacchetto viene inoltrato altrimenti si ripercorre la tabella usando come nodo di destinazione il valore del nodo appena trovato.
Se ad esempio il pacchetto ha la necessità di raggiungere, a partire da A, il nodo F si consulterà la tabella una prima volta: si troverà che il nodo p(F) è E che non è un immediato vicino di A. Quindi si ripercorre la tabella questa volta usando come nodo di destinazione il nodo E, si troverà quindi che p(E) è D. Il pacchetto verrà pertanto spedito al nodo D, quest'ultimo anch'esso dotato di una propria tabella di instradamento ricaverà per il pacchetto in ingresso il link in uscita nelle modalità appena viste.

Algoritmo distance vector
L'agoritmo distance vector (DV) è un algoritmo distribuito, in ogni nodo viene mantenuta in memoria un apposita tabella delle distanze, tale tabella ha una riga per ogni nodo di destinazione ed una colonna

per ogni vicno immediatamente collegato. Supponiamo che un nodo X sia interessato all'instradamento verso Y, passando attraverso il nodo Z. Nella tabella la voce relativa alla distanza viene così valutata

$$D^X(Y,Z) = C(X,Z) + \min_w D^Z(Y,W)$$

Ciascun nodo deve conoscere il costo dei percorsi a minor costo dei suoi vicini.

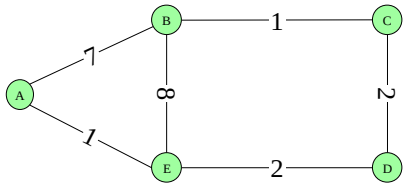
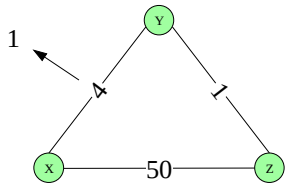


Tabella del NODO E

$D^E()$	A	B	D
A	**1**	14	5
B	7	8	**5**
C	6	9	**4**
D	4	11	**2**

Tale algoritmo può anche essere adoperato in maniera centralizzata (come può avvenire per l'algoritmo di Djkstra) in tal caso esso risulta tuttavia meno efficiente e per questo motivo viene largamente usato in un contesto distribuito. Le righe della tabella intercettano le possibili destinazioni che un pacchetto a partire dal nodo E può richiedere, le colonne rappresentano invece i nodi vicini ad E. Ogni riga della tabella viene completata quantificando il percorso che da E (nodo di partenza) conduce al nodo di destinazione (elemento di riga) passando per un nodo vicino. La tabella di inoltro viene così usata per instradare i pacchetti in uscita verso i vicini che offrono un minor costo di percorrenza. Quando un nodo rileva una variazione del costo del link fra se stesso e un suo vicino esso aggiorna la sua tabella delle distanze e se interviene una variazione nel costo del percorso di minimo costo informa i suoi vicini dell'evento. Sia ad esempio:

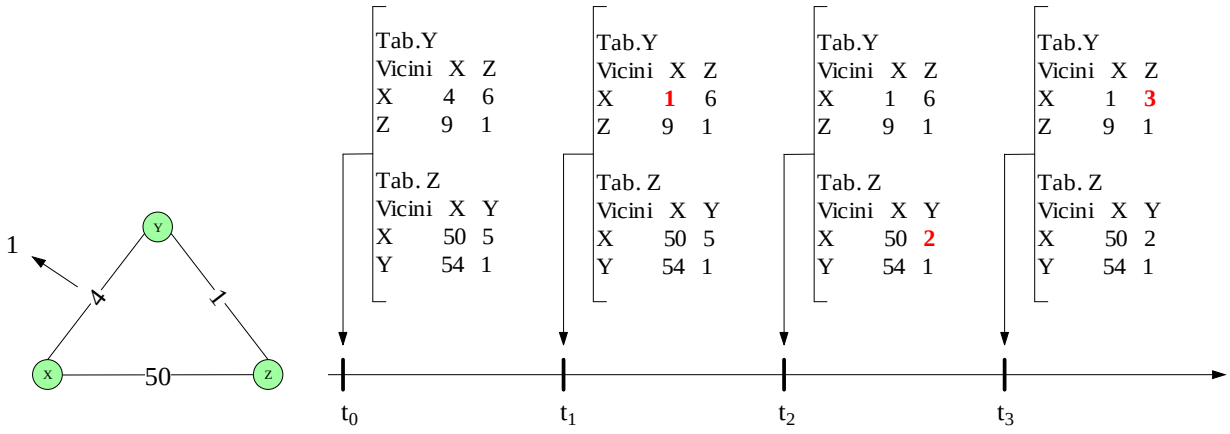


Tab. X		
Vicini	Y	Z
Y	4	51
Z	5	50

Tab.Y		
Vicini	X	Z
X	4	6
Z	9	1

Tab. Z		
Vicini	X	Y
X	50	5
Y	54	1

Supponiamo adesso che il link che collega X ed Y assuma un costo inferiore, passando cioè da 4 ad 1. Analizziamo di seguito gli ingressi nelle tabelle delle distanze di Y e Z verso la destinazione X:



(il collegamento da X ad Y adesso ha costo 1)

Al verificarsi di un cambio del costo di link si ha un periodo di tempo durante il quale i pacchetti rimbalzano continuamente tra i nodi. Risulta cruciale stabilire allora la frequenza di aggiornamento per i link. Se da un lato sembrerebbe corretto inoltrare subito verso i proprio vicini l'informazione di un percorso con costo variato dall'altro si rischia di dedicare una quota delle risorse esclusivmente alle segnalazioni, a scapito quindi della comunicazione dei flussi informativi. Per questo motivo un nodo non comunica immediatamente ogni variazione di un certo costo ma memorizza tale segnalazione e la rende disponibile solo quando questa rimane stabile per uncerto intervallo di tempo (la stessa variazione del costo del link potrebbe infatti variare nuovamente). Di sicuro l'impossibilità di operare di un nodo va comunicata immediatamente a tutti gli altri nodi in maniera tale da evitare una eccessiva perdita dei pacchetti che a questo vengono indirizzati.

Determinazione dei costi dei link

La scelta più semplice per l'assegnazione dei costi ai link della rete potrebbe essere quella che assegna a ciascuno di questi un costo unitario. In tal caso il percoso minimo, trovato applicando uno dei due algoritmi visti, verrebbe ad essere quello che compie il minor numero possibile di salti. Tuttavia ciò non rispecchia la reale situazione della rete, alcuni collegamenti potrebbero infatti essere più performati, altri invece meno caricati di altri, per questo motivo il peso da associare ad link deve solitamente considerare:

- la frequenza di cifra del collegamento fisico, ciò attribuisce un peso più basso al link dotato di maggiore frequenza di cifra (a parità di tutti gli altri parametri);
- il ritardo di transito che mediamente viene riscontrato su di un pacchetto che lo sta attraversando, un ritardo eccessivo potrebbe significare una coda di attesa prossima al traboccamento e quindi un collegamento che termina in un nodo particolarmente affollato;
- in virtù dell'osservazione fatta prima, il numero medio di pacchetti in coda presso un nodo di rete che attendono di essere trasmessi (una misura più rigorosa dovrebbe riguardare non solo il numero di pacchetti in coda ma anche la dimensione media che questi pacchetti hanno e la frequenza di cifra con cui essi sono iniettati nel nodo);
- la lunghezza fisica del cavo, un cavo particolarmente lungo è più soggetto a rumori e interferenze, il segnale poi subisce un attenuazione che è tanto più accentuata quanto più è lunga la dimensione del cavo;

Vengono quindi definite delle funzioni non lineari che in base alle caratteristiche del cavo ed alla loro appartenenza ad uno o più gruppi sopra citati attribuiscono al collegamento un appropriato costo (Ad esempio, costo 1 se il cavo appartiene ad uno solo dei casi sopra in elenco, costo 2 se il cavo appartiene a due dei casi sopra in elenco etc...).

Instradamento gerarchico
L'instradamento da tabella soffre di alcuni problemi legati alle dimensioni delle reti, quest'ultima infatti caratterizza inevitabilmente le dimensioni della tabella di instradamento che quindi possono richiedere maggiori risorse in termini di memorizzazione a ciascun nodo della rete. Inoltre, al crescere dei nodi della rete aumenta anche il carico di lavoro presso ogni nodo (la ricerca di una destinazione all'interno di una tabella potrebbe ad esempio richiedere più tempo). La soluzione al problema appena visto è stata trovata in una forma di instradamento gerarchico.
La rete viene organizzata in cluster, ogni cluster ingloga un certo numero di nodi. L'operazione di aggregazione permette di individuare unnodo di rete mediante l'indirizzo generico *x.y* dove *x* identifica il cluster di appartenenza ed *y* il particolare nodo. L'instradamento gerarchico ha come obiettivo la diminuzione della tabella di instradamento: ogni nodo (in questo caso) mantiene in tabella una riga dedicata a ciascun terminale appartenente al proprio cluster e dedica altrettante righe per ogni cluster della rete. La lettura del primo pezzo dell'indirizzo consente al nodo di stabilire se il pacchetto va destinato ad un nodo appartenente al proprio cluster oppure va instradato verso un cluster esterno alla rete. Il secondo pezzo dell'indirizzo, invece, rintraccia il nodo di destinazione. Quando la rete assume considerevoli dimensioni un cluster può essere ulteriormente suddiviso in sottocluster. La tabella di instradamento, grazie all'instradamento gerarchico, si riduce notevolmente. Tuttavia i pacchetti, adesso, transitano attraverso più nodi (ciò avviene ad esempio quando il terminale di destinazione non si trova nel cluster dipartenza) e ciò introduce dei ritardi che si aggiungono alla trasmissione dei pacchetti.

Instradamento senza tabella
L'instradamento senza tabella ha il pregio di far risparmiare al nodo risorse in termini di memoria, in questo paragrafo verranno citati i metodi più usati attualmente:

Instradamento mediante canale comune
Qualora il canale di accesso al nodo è comune ad alti nodi può risultare comodo risolvere il problema dell'instradamento mediante trasmissione in broadcast (i pacchettisono cioè diretti a tutti i nodi della rete) dei pacchetti.

Instradamento casuale o aleatorio
Viene utilizzato un meccanismo pseudo-aleatorio che sceglie il ramo di uscita dal nodo in maniera tale da garantire con una certa frequenza che il collegamento scelto è effettivamente esatto. Il meccanismo di scelta del collegamento di uscita si basa quindi sulle precedenti operazioni di instradamento, la funzione che stabilisce il collegamento in uscita è allora modellata statisticamente in base alle destinazione dei pacchetti.

Instradamento calcolato
Ogni nodo possiede informazioni sulla struttura della rete e le usa per calcolare mediante formula il ramo di uscita che il nodo deve utilizzare nei confronti di un pacchetto in ingresso. Tale strategia differisce dall'instradamento con tabella in quanto la tabella stessa è sintetizzata tramite la formula adoperata. Non vi è dubbio che adesso la tabella non necessita di essere memorizzata nel nodo ma ciò

comporta un'utilizzo di tale strategia esclusivamente a reti di tipo statico, ossia a reti i cui terminali di appartenenza sono noti a priori.

Instradamento da sorgente
Nel caso di instradamento centralizzato il nodo di rete che ha la necessità di instradare un pacchetto si rivolge ad un nodo server che quindi elencherà al nodo richiedente un elenco di nodi intermedi attraverso cui far passare i pacchetti. Qualora l'instradamento si sviluppa in un contesto non celntralizzato, e quindi distribuito, sarà il nodo sorgente ad occuparsi della ricerca del percorso.
Esso inizia un procedura di path descovering mediante un apposito messaggio PDM (path descovering message) che invia a tutti i nodi della rete a cui è direttamente collegato. Ogni nodo della rete che riceve unmessaggio PDM aggiunge a quest'ultimo (in coda ad un elenco già presente) il proprio messaggio e successivamente manda il messaggio PDM ai nodi ad esso adiacenti Se si capisce che il pacchetto è già transitato per il nodo che riceve il PDM (ad esempio perchè il nodo riconosce nella lista il proprio indirizzo) quest'ultimo viene scartato e non è più inoltrato. Al nodo di destinazione giungono più messaggi PDM, ognuno con un elenco di possibili nodi di transito, spetta pertanto al nodo di destinazione stabilire quale percorso adottare.
In questa fase iniziale di setup o di scoperta del percorso molti bit vengono dedicati alla funzione di instradamento, alcuni di essi infatti sono utilizzati per numerare i percorsi della rete in maniera tale da identificarli e poterli quindi elencare nel PDM. Una soluzione alternativa è quella che invece numera i percorsi interni al nodo che quindi sostituisce al numero del collegamento in ingresso quello di uscita. Tuttavia quando la rete è molto estesa, sia che per la prima che per la seconda soluzione sono comunque necessari un numero di bit per segnalare il percorso. Quindi, per non rovinare l'efficienza ed impiegare al meglio le risorse di rete è consigliabile sfruttare pienamente il percorso trovato. Tale strategia si addice quindi ad una trasmissione massiccia di pacchetti informativi piuttosto che ad una trasmissione occasionale dei flussi informativi.

Instradamento verso destinazione multipla
L'instradamento verso destinazione multipla del pacchetto prevede più destinazione per un singolo flusso informativo ed è talvolta detto trasmissione multicast (più destinatari di uno stesso pacchetto).
Una prima soluzione, assai banale, consente di utilizzare le procedure già esistenti per l'indirizzamento verso una singola destinazione, i pacchetti informativi possono infatti essere indirizzati separatamente. Ad ogni indirizzamento si provvede ad aggiornare il campo indirizzo con il successivo destinatario.
Una successiva evoluzione è quella che prevede di modificare il campo destinato all'indirizzo del destinatario e ammettendo la possibilità di inserire più indirizzi di destinazione. Il campo multi-indirizzo, così viene detto, potrebbe essere troppo esteso in presenza di un numero elevato di destinatari.
Altra tecnica è il flooding o diffusione e consiste nell'inviare un certo messaggio dal nodo sorgente a tutti i nodi conui esso è direttamente collegato (con l'esclusione di quello dal quale il pacchetto è giunto). Si tratta di una tecnica inefficiente ma robusta nei confronti dei possibili malfunzionamenti di alcuni rami e/o nodi della rete. Allo scopo di evitare che un pacchetto inoltrato finisca nuovamente ad un nodo che ha già attraversato, oppure che i pacchetti spediti dalla sorgente finiscano poi per essere confusi con altri pacchetti in attesa, si usa inserire un numero di sequenza per etichettare i pacchetti. Il ricilo dei bit usati nelle numerazione pone comunque un vincolo al numero massimo dei pacchetti che possono essere etichettati e genera ugualmente confusione tra i pacchetti (a causa dei ritardi un pacchetto può ancora vagare nella rete prima di giungere a destinazione ed il suo numero di identificazione potrebbe poi essere stato riutilizzato e quindi assegnato ad un nuovo pacchetto che potrebbe addirittura giungere prima del pacchetto atteso). A tale proposito si inserisce nel pacchetto un campo dati detto campo di age, esso contiene il tempo di vita del pacchetto. Quando il pacchetto viene

trasmesso per la prima volta al campo dati age viene assegnato unopportuno valore. Ad ogni attraversamento esso è decrementato di un unità e quando il valore si esaurisce del tutto il pacchetto non viene più inoltrato.
Se la rete è molto estesa il numero di salti consentiti da nodo a nodo non può essere troppo piccolo, ciò infatti limiterebbe i pacchetti che difficilmente raggiungerebbero i nodi più lontani della rete. Un numero si salti, invece, troppo alta potrebbe tuttavia portare presso i nodi della rete dei pacchetti vecchi (poichè già pervenuti). In generale, il numero di bit per la numerzione dei pacchetti cresce all'aumentare della estensione geografica della rete.

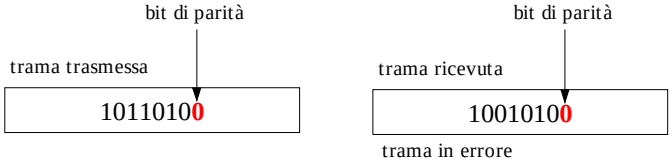
Spanning tree forwarding
Data una rete è possibile determinare su di essa un albero minimo di copertura, lo spanning tree è il più piccolo insieme di rami che forniscono connettività completà a tutti i nodi della rete e non prevede, inoltre, cicli chiusi. Ogni nodo mantiene informazione sui nodi ad esso collegati direttamente e appartenenti allo spanning tree.
I pacchetti broadcast sono distinti da un campo indirizzi particolare (piuttosto che elencare ogni singolo destinatario), generalmente tutti 1. Quando un nodo riceve un pacchetto broadcast inoltra una sua copia su ciascuno dei nodi appartenenti allo spanning tree escludento il nodo da cui il pacchetto è pervenuto. Lo scema tenta di migliorare il meccanismo della diffusione evitando possibili cicli, lo svantaggio tuttavia risiede nel fatto che i collegamenti appartenenti allo spanning tree devono necessariamente essere ricordati. Inoltre, in caso di nuovi collegamenti, l'albero di copertura deve essere aggiornato. La riconfigurazione può avvenire manualmente (un operatore umano aggiorna con una certa frequenza lo spanning tree) oppure tramite un algoritmo che aggiunge o rimuove i nodi che non appatengono più allo spanning tree.

Reverse path forwarding
Premesso che anche in questo caso i pacchetti broadcast sono caratterizzati da un indirizzo speciale, il reverse path forwarding richiede l'esistenza di instradamento punto-punto basato su tabella. Quando un nodo riceve un pacchetto broadcast esso usa la sua tabella di instradamento per detrminare il ramo di uscita che userebbe per inoltrare il pacchetto verso il nodo sorgente, quindi verifica se il pacchetto broadcast arrivi effettivamente attraverso il ramo ricavato. Se la condizione non è soddisfatta il pacchetto è cancellato e non viene inoltrato, se la condizione è invece soddisfatta il nodo invia il pacchetto su tutti i nodi di uscita con l'eccezione di quello da cui il pacchetto proviene.

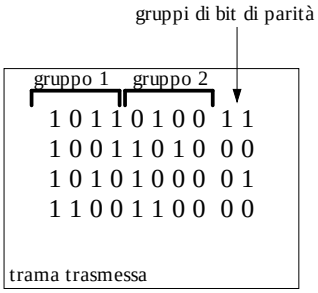
Il controllo di errore
La trasmissione di un flusso informativo da sorgente a destinazione non avviene mai inmaniera del tutto ideale: i bit giunti presso il nodo di destinazione possono infatti contenere errori e l'applicazione che in quel momento sta richiedendo il trasferimento dati potrebbe non funzionare correttamente.
La connessione dati è quindi priva di errori ed è considerata accettabile da un'applicazione solo quando la probabilità di errore assume valori bassi, in realtà alcune applicazioni possono anche pretendere una comunicazione perfetta dei pacchetti informativi. Gli strati superiori cercano allora di migliorare la qualità del canale abbassandone la probabilità di errore mediante apposite funzioni di controllo di errore. Il principio che sta alla base del controllo di errore prevede l'aggiunta al pacchetto di un insieme di bit preposti al controllo dell'errore. Il pacchetto così modificato e talvolta chiamato trama viaggia nella rete con il suo carico aggiuntivo. In questo modo l'entità di strato superiore alla pari con quella di sorgente può effettuare uncontrollo di errore sul pacchetto pervenuto che sarà inoltrato verso lo strato superiore solo se la funzione di controllo di errore ne accerta l'integrità. In alcuni casi oltre al controllo di errore si può avere anche la correzzione dell'errore individuato. Gli approcci che qui seguiremo

possono essere raggruppati sotto due grandi famiglie dicontrollo di errore: approccio ARQ ed approccio FEC.
Secondo l'approccio ARQ (automatic repeat request) l'entità alla pari di quella di sorgente ha il compito di rilevare l'errore giunto a destinazione e di richiedere la trasmissione del pacchetto qualora ciò dovesse verificarsi. Quando invece la funzione di controllo di errore convalida la trama pervenuta essa stessa viene, allora, privata dei bit di errore che l'entità sorgente gli aveva aggiunto ed invia il messaggio all'entità di strato superiore. Alcune importanti considerazioni vanno considerate qualora si decida di adottare l'approccio ARQ: non vi è dubbio che la sua efficienza permetterà di consegnare allo strato superiore un servizio più affidabile, tuttavia per attuare il meccanismo di ritrasmissione dei pacchetti andati corrotti è necessario disporre di un canale logico bidirezionale in cui convogliare da un lato i pacchetti e dall'altro le richieste di ritrasmissioni. La ritrasmissione ovviamente causa un rallentamento della frequenza di cifra ed aggiunge ulteriori tempi di attesa.
Nell'approccio FEC (forward error correction) alla rilevazione dell'errore segue la sua correzione che quindi ripristina il pacchetto. Non sempre è possibile attuare la correzione, tale processo deve poi garantire una bassa probabilità di sbagliare la correzione. L'approccio FEC non richiede un canale logico bidirezionale (canale diritorno), tuttavia maggiori complicazioni sono presenti in fase di progettazione dei dispositivi rivolti alla correzione, essi dovranno poi avere maggiori capacità elaborative facendo crescere il costo realizzativo per ogni nodo.
Infine, un diverso approccio può combinare i benefici portati da entrambi gli schemi citati: i meccanismi di correzione potrebbero ad esempio correggere gli errori più comuni mentre quelli più rari potrebbero invece essere risolti con la ritrasmissione del pacchetto.

Automatic repeat request
Prima di affrontare lo schema ARQ per la rilevazione degli errori è opportuno analizzare alcuni dei più importanti metodi che sono appunto utilizzati per la rilezione di uno o più errori all'interno di una trama. Affinchè sia possibile determinare la presenza di errori in un pacchetto di m bit è necessario ammettere in trasmissione tutte le possibili combinazioni e quindi 2^m pacchetti. Il nodo in ricezione riceve allora pacchetti lunghi n bit ed aggiunge a questi r bit formando così un pacchetto di $m=n+r$ bit.
Gli r bit aggiuntivi dipendono dagli n bit trasmessi secondo una regola che permetterà in seguito di discriminare i pacchetti ricevuti. Inoltre, i bit aggiuntivi devono essere tali da ridurre la probabilità di errore sugli stessi bit aggiuntivi.
Un primo metodo per la rilevazione degli errori prevede l'aggiunta di un solo bit, con tale bit è comunque possibile effettuare operazioni di rilevazione di errore/i. Tale procedura è nota come controllo di parità (oppure disparità). Al pacchetto in trasmissione è concesso un numero pari (oppure dispari) di bit 1 (oppure di bit 0). Il bit aggiunto serve appunto a garantire che i bit 1 (oppure quelli 0) siano in numero pari (oppure dispari). Tuttavia se si dovesse verificare più di un errore si potrebbe non rilevare l'evento, più in particolare, mediante schema con bit di parità non è possibile rilevare unnumero pari di errori nel pacchetto. Dato il suo basso overhead aggiunto al pacchetto questa tecnica è comunque usata laddove la probabilità di errore del canale è assai bassa, quasi trascurabile.



Una regole più articolata della precedente prevede l'aggiunta di *r* bit, in altre parole la trama o il messaggio è suddiviso in *r* sottogruppi e ciascun bit *r* comanda il controllo di parità per ogni sottogruppo. In questo modo è possibile localizzare il sottogruppo in errore.



Se poi si immaggina di disporre il messaggio trasmesso secondo una struttura a matrice in cui ogni trama del messaggio occupa una riga è posibile adottare un interessante schema che in alcuni casi può addirittura individuare il bit in errore che può così essere corretto. Ogni trama termina con il proprio bit di parità, l'ultima riga si ottiene invece conteggiando la parità in base alle colonne delle trame del messaggio. Se si verifica un errore si troverà, presso il nodo di destinazione, un errore di parità su di una riga ed un errore di parità su di una colonna, l'intersezione individua il bit in errore. Se si verificano più errori ma su righe e colonne diverse è ancora possibile individuare il bit compromesso. Se inve si verificano più errori sulla stessa riga o colonna non è più possibile individuare l'errore.
Detto *L* il numero di caratteri del pacchetto, *C* il numero di bit usati per un carattere si ha che $n=CL$ indica il numero di bit che compone il pacchetto. Quindi $r=C+L+1$ sono i bit aggiuntivi. La frazione di bit che porta informazione rispetto al totale del pacchetto è:

$$\frac{CL}{CL+C+L+1}=\frac{CL}{(C+1)(L+1)}$$

Il metodo più spesso utilizzato è leggermente più sofisticato di quelli appena visti. Esso è detto metodo di controllo della ridondanza ciclica o più berevemente CRC (cyclic redundary check). I codici CRC sono anche conosciuti come codici polinomiali poichè è possibile interpretare la stringa di bit che deve essere spedita come un polinomio i cui coefficienti sono i valori binari 1 e 0. La sequenza di bit da spedire è quindi vista come un numero in notazione binaria, tale numero va diviso per un opportuno divisore (precedentemente concordato) che è anch'esso visto secondo la notazione binaria. Il resto di tale operazione viene aggiunto alla sequenza dei bit da spedire per agevolare in seguito la funzione di controllo dell'errore. In ricezione vengono allora estratti dalla trama i primi *n* bit, si effettua su di essi la divisione e si verifica che il resto dell'operazione coincida con i restanti *r* bit del pacchetto. Se il resto ottenuto dall'operazione di divisione è diverso da quello ricevuto viene rilevata la presenza di errore.
Poichè può risultare difficile realizzare i dispositivi per la classica divisione tra numeri in notazione binaria, si realizza una divisione basata su regole simili ma diverse da quelle classiche. Tale procedura può essere descritta considerando la divisione di due polinomi con coefficienti binari ed operando cone le regole dell'aritmetica modulo 2. Il polinomio associato ad una stringa di *m* bit $a_{m-1},a_{m-2},...,a_1,a_0$ è pari a $a_0+a_1x+a_2x^2+...+a_{m-2}x^{m-2}+a_{m-1}x^{m-1}$. Il sender ed il receiver si accordano sul divisore G da usare per le operazioni di divisione:

$$P_1(x)x^r = Q(x)D(x) + R(x)$$

Un esempio:

$$\left.\begin{array}{l} d = 6bit \\ P_1(x) = 100110 \\ \\ r = 3bir \\ x^r = x^3 \end{array}\right\} \rightarrow P_1(x)x^r = x^3 \cdot \left(x^5 + x^2 + x\right) = x^8 + x^5 + x^4$$

Il divisore concordato sia 1011, $D(x)=x^3+x+1$, pertanto:

$$Q(x) = P_1(x)x^r : D(x)$$

$$x^8 + x^5 + x^4 \quad : \underline{x^3 + x + 1}$$

$$\underline{x^8 + x^6 + x^5} \qquad x^5 + x^3 + 1$$

$$x^6 + x^4$$

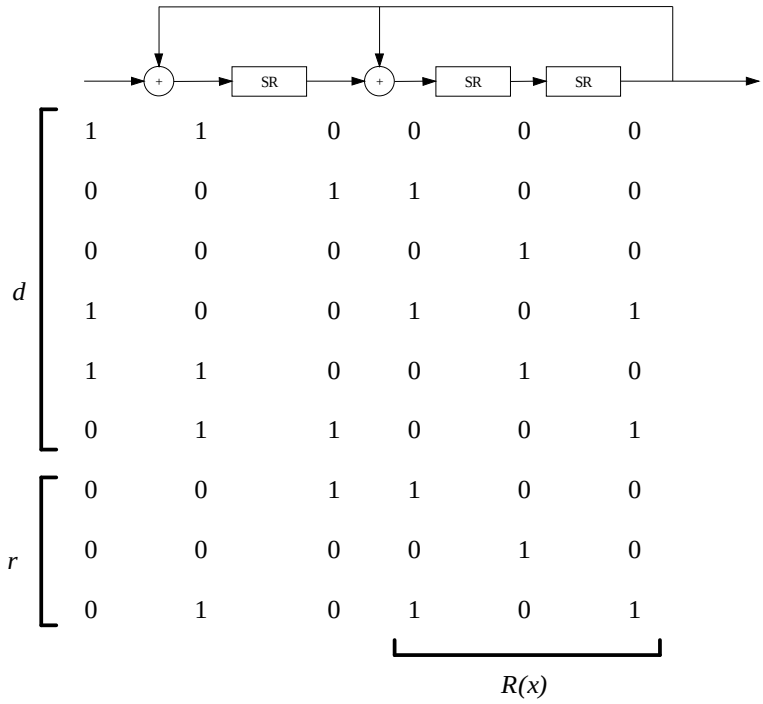
$$\underline{x^6 + x^4 + x^3}$$

$$x^3$$

$$\underline{x^3 + x + 1}$$

$$x + 1$$

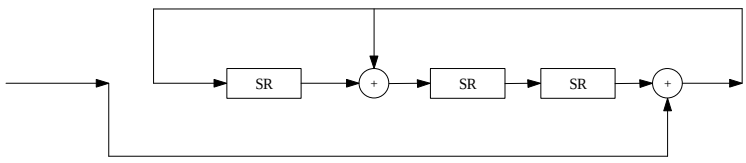
Dunque $R(x)=x+1$ che corrisponde ad 011 mentre $Q(x)=x^5+x^3+1$. Il motivo che spinge all'adozione di tale procedura è dovuta alla semplice realizzazione, mediante registri a scorrimento, di opportune strutture logiche capaci di valutare con efficacia il resto della divisione $R(x)$.



d	1	1	0	0	0	0
	0	0	1	1	0	0
	0	0	0	0	1	0
	1	0	0	1	0	1
	1	1	0	0	1	0
	0	1	1	0	0	1
r	0	0	1	1	0	0
	0	0	0	0	1	0
	0	1	0	1	0	1

R(x)

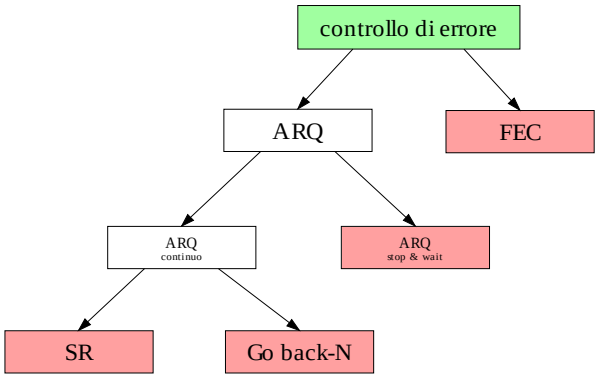
Il pacchetto da spedire è considerato a partire dalla cifra più significativa, terminati i bit di ingresso vengono dati altri *r* bit 0 e dopo l'ultimo clock i registri a scorrimento conterranno il valore di *R(x)*. Osservare come questo schema permetta parallelamente di fare la verifica del resto, infatti terminati *d* bit di ingresso è possibile fornire gli *r* bit del resto che sono allegati al pacchetto. Se non si sono verificati errori in trasmissione i registri a scorrimento conterranno tutti il valore 0. Uno schema alternativo è il seguente:



1	1	0	1	0	0
0	0	1	1	1	0
0	1	0	1	1	1
1	0	1	1	1	1
1	0	0	0	1	1
0	1	0	1	0	1

R(x)

In questo schema, con gli stessi dispositivi, non sono più necessari gli *r* bit. Il dispositivo, quindi calcola *R(x)* con un numero inferiore diclock.
Due importanti tipologie di schemi ARQ sono gli schemi ARQ continuos e l'ARQ stop & wait. Per la trattazione di tali schemi si assumerà che un singolo trasmettitore invia informazioni ad un singolo ricevitore, il canale ipotizzato sarà di tipo bidirezionale (il ricevitore dispone in questo modo di un canale per inviare messaggi di conferma diretti al trasmettitore), i bit di conferma contengono anch'essi bit di controllo di errore, i pacchetti ricevuti e ritenuti errati sono scartati dal ricevitore.



Schema stop & wait
Il trasmettitore, nel caso in cui non si verificano errori, invia verso il ricevitore un pacchetto alla volta. Prima di inviare il successivo pacchetto esso attende l'arrivo di un messaggio di conferma che il ricevitore gliinvia. Il successivo pacchetto viene trasmesso solo dopo che il messaggio di conferma è pervenuto. Se il ricevitore, effettuando i propri controlli sul pacchetto, stabilisce che il pacchetto appena giunto è in errore non manda alcuna conferma al trasmettitore. Dall'altro lato, il trasmettitore attende un intervallo di timeout prima di ritrasmettere nuovamente il pacchetto andato perso o corrotto. Appare quindi importante dimensionare correttamente l'intervallo di timeout T_{out}. Il trasmettitore ne stima un determinato valore considerando i tempi di propagazione dell'onda elettromagnetica di andata e di ritorno (2τ), considerando i tempi di trasmissione per il messaggio di conferma T_A ed il tempo di processing T_P che il ricevitore impiega per elaborare il pacchetto. Quindi, il minimo valore di timeout è $2\tau+T_A+T_P$. Tuttavia a causa della variabilità del tempo di processing che varia in funzione del carico presso l'entità in ricezione, il tempo di timeout è spesso maggiorato rispetto al valore minimo trovato.
I messaggi di conferma che il ricevitore manda al trasmettitore è dotato di bit diparità per la rilevazione dell'errore. Qualora un messaggio di conferma giunge al trasmettitore in maniera errata quest'ultimo si comporterà come se nessun messaggi o di conferma sia pervenuto, attenderà lo scadere dell'intervallo di timeout e ritrasmetterà il pacchetto.
Ed allora, siccome il ricevitore vedendosi mandare un pacchetto che però è la nuova copia del precedente pacchetto non confermato potrebbe confondere quest'ultimo con il successivo pacchetto. Per scongiurare questo tipo di problema è sufficiente un unico bit di numerazione cosicchè quando il ricevitore riceve un pacchetto avente lo stesso bit di numerazione dell'ultimo pacchetto ricevuto assume che il proprio messaggio diconferma, precedentemente inviato, non sia ancora giunto al trasmettitore oppure sia andato perso. Esso pertanto provvederà a mandare una nuova copia del messggio di conferma.
Tuttavia possono ancora verificarsi delle ambiguità questa volta nei confronti dei messaggi di conferma. Può infatti accadere che, a causa dei ritardi di transito della rete, un messaggio di conferma giunga con eccessivo ritardo presso il trasmettitore che vedendo scadere l'intervallo di timeout provvede, come da protocollo, a ritrasmettere una nuova copia dell'ultimo pacchetto inviato. Il

ricevitore, grazie alla numerazione dei pacchetti, saprà riconoscere la copia del pacchetto e manderà una nuova copia del messaggio di conferma. Il trasmettitore, che nel frattempo ha provveduto ad inviare una nuova copia, si vede recapitare il vecchio messaggio di conferma, esso arresterà l'attuale ritrasmissione in corso e passa al successivo pacchetto in elenco. Adesso nella rete stanno viaggiando il pacchetto successivo all'utlimo pacchetto confermato ed il messaggio di conferma del vecchio pacchetto che era stato ritrasmesso. Quando al trasmettitore giungerà il vecchio messaggio di conferma si confonderà quest'ultimo con il messaggio diconferma dell'ultimo pacchetto inviato. Per questo motivo anche i messaggi di conferma, così come i pacchetti, vengono numerati (la sequenza che deve avvenire è: pacchetto 1, messaggio conferma 1, pacchetto2, messaggio conferma 2, etc...).

ARQ continui

Lo schema ARQ continuo è stato introdotto per ovviare al problema dei tempi morti del ricevitore e del trasmettitore nello scema ARQ stop & wait. Infatti, in tale schema i pacchetti sono inviati in successione anche se il messaggio di conferma non è stato ancora ricevuto dal trasmettitore. Per questo motivo, il trasmettitore ed il ricevitore funzionano in maniera continua. Lo schema ARQ continuo richiede però un canale full duplex (lo schema ARQ stop & wait necessita di un canale half duplex).

Quando si verifica un errore nella ricezione di un pacchetto è possibile seguire due diversi tipi di approccio. Un primo approccio è rivolto alla soluzione SR, selctive repeat. Quando un ricevitore riceve un pacchetto contenente un errore quest'ultimo richiede al trasmettitore la sua ritrasmissione. Ciò presuppone la numerazione dei pacchetti affinchè il ricevitore possa indicare al trasmettitore il pacchetto da ritrasmettere. Inoltre, quando il trasmettittore non vede arrivare un messaggio di conferma per un pacchetto precedentemente inviato esso attende (così come nell'ARQ stop & wait) lo scadere del tempo di timeout e ne effettua in maniera automatica la ritrasmissione. Siccome sono inviati più pacchetti contemporaneamente sono allora necessari più bit per la numerazione, non è poi garantito l'ordine con cui i pacchetti arrivano che è quindi sparso. Se dunque l'entità di strato superiore necessita dei pacchetti ordinati e opportuno memorizzare questi in un apposito buffer di memoria e consegnare i pacchetti una volta ordinati. La capacità di memoria è cruciale se si vuole limitare, presso il nodo ricevente, i problemi introdotti dai ritardi di propagazione della rete.

Uno schema diverso è lo schema go back-N in cui il ricevitore insiste nel voler ricevere i pacchetti secondo un ordine che è quello corretto ed invia al trasmettitore messaggi di conferma oppure messaggi di mancata ricezione (segnali di ACK e di NACK) dei pacchetti di informazione. Il ricevitore manda un segnale di NACK anche quando riceve un pacchetto corretto che però non era atteso (il ricevitore insiste sull'ordine dei pacchetti). Il trasmettitore, invece, invia i pacchetti consecutivi entro una certa finestra di pacchetti e quando riceve un messaggio di NACK (il NACK è numerato da un indice associato al pacchetto a cui esso si riferisce) riprende la trasmissione dal pacchetto non pervenuto. Un eventuale messaggio di ACK (anch'esso numerato) ha poi l'effetto di conferma nei conferma del pacchetto a cui esso si riferisce e di tutti i pacchetti che esso precede.

Un elemento importante nei protocolli go back-N è il numero di bit da destinare al conteggio dei pacchetti che quindi fissa la dimensione della finestra. I numeri destinati al conteggio non vanno riutilizzati durante il periodo di tempo in cui il relativo messaggio di conferma è atteso:

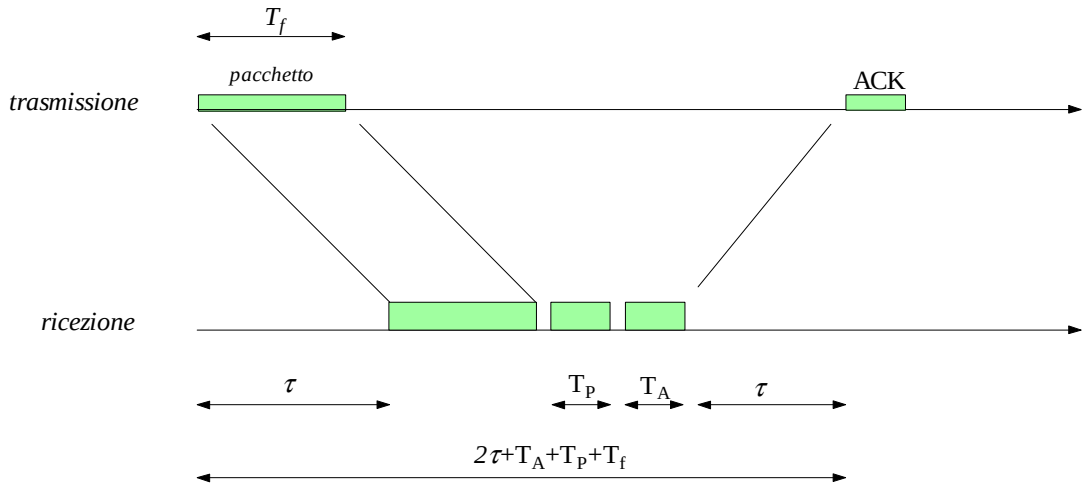
$$2^b > \frac{T_{out}}{T_f}$$

b è il numero di bit destinati al conteggio. Il numero di bit stabilisce la dimensione della finestra, essa è influenzata dal meccanismo che gestisce il controllo del flusso, ma ciò sarà argomento del successivo capitolo. Per adesso è sufficiente sapere che la velocità di trasmissione del trasmettitore potrebbe essere

tale da non consentire al ricevitore di processare i pacchetti che ad esso giungono. Il ricevitore si comporta allora nel seguente modo: anche se ad esso perviene un pacchetto che segue la sequenza logica in atto quest'ultima ritarda l'invio del messaggio di conferma cosicchè il nodo trasmittente effettuerà una ritrasmissione del pacchetto.
Definita una finestra fatta da W pacchetti, il trasmettitore inietta nella rete al massimo W pacchetti successivi e senza attenderne la conferma. Qualora giunga un messaggio di conferma, ad esempio del primo pacchetto della sequenza logica, il trasmettitore incrementa di un unità la finestra dei pacchetti. Se giunge la conferma per il terzo pacchetto in ordine logico, essendo confermati anche quelli precedenti (e quindi il primo ed il secondo), la finestra è incrementata di tre pacchetti cosicchè il trasmettitore procede nella trasmissione del flusso. Osservare che nel caso in cui la finestra ha esattamente W=1 (un solo pacchetto), il protocollo go back-N si riduce ad un ARQ continuo di tipo stop & wait. Affinchè non vi siano pacchetti duplicati occorre che il numero di bit sia tale che:

$$2^b > \min\left(W, \frac{T_{out}}{T_f}\right)$$

Analisi delle prestazioni dello schema ARQ stop & wait



L'efficienza in assenza di errori è:

$$\eta = \frac{T_f}{T_F} = \frac{T_f}{\tau + T_A + T_P + T_f}$$

dove T_f è il tempo necessario alla trasmissione, T_F è il tempo di fornitura del servizio allo strato sottostante, τ è il ritardo di propagazione, T_P è il tempo di processing, T_A è il tempo utile alla trasmissione del messaggio di ACK. Consideriamo adesso il caso in cui il canale presenti degli errori. E' possibile che un certo pacchetto e la conferma relativa siano stati ricevuti senza errori (primo caso). E' possibile che uncerto pacchetto o la sua conferma relativa sia stato ricevuto in errore, ritrasmesso dopo un tempo $T=T_f+T_{out}$ e poi ricevuto correttamente (secondo caso). Ed amcora, è possibile che un certo pacchetto o la sua relativa conferma sia stato ricevuto in errore, ritrasmesso dopo un tempo

$T=T_f+T_{out,}$ ricevuto ancora in errore, ritrasmesso una seconda volta dopo un tempo T e poi infine ricevuto correttamente come la sua conferma relativa.
Assumendo che P_e sia la probabilità che un pacchetto sia in errore ed assunta l'indipendenza dell'evento errore su un pachetto dall'evento errore su un altro pacchetto, sipuò ottenere il tempo medio $T_{f,m}$ richiesto per la trasmissione in maniera tale da poter poi stimare l'efficienza dello schema.
Il primo caso (pacchetto e conferma ricevuti correttamente) si presenta con probabilità $1-P_e$ e richiede un tempo di trasmissione T_f. Il secondo caso, primo tentativo fallito e secondo riuscito, avviene con probabilità $P_e(1-P_e)$ e richiede un tempo di trasmissione di $T+T_f$. Il terzo caso sipresenta con probabilità $P_e^2(1-P_e)$ e richiede un tempo di trasmissione di $2T+T_f$. In generale, l'i-esimo caso si presenta con probabilità $P_e^{i-1}(1-P_e)$ e richiede un tempo di trasmissione di $(i-1)T+T_f$. Il tempo medio $T_{f,m}$ richiesto per la trasmissione vale:

$$T_{f,m}=\sum_{i=1}^{+\infty}\left[(i-1)T+T_f\right]P_e^{\,i-1}(1-P_e)$$
$$T_{f,m}=\sum_{i=0}^{+\infty}\left[iT+T_f\right]P_e^{\,i}(1-P_e)$$
$$T_{f,m}=(1-P_e)\left[T\sum_{i=0}^{+\infty}iP_e^{\,i}+T_f\sum_{i=0}^{+\infty}P_e^{\,i}\right]$$
$$T_{f,m}=(1-P_e)\left[TP_e\sum_{i=0}^{+\infty}iP_e^{\,i-1}+T_f\sum_{i=0}^{+\infty}P_e^{\,i}\right]$$
$$T_{f,m}=(1-P_e)\left[TP_e\frac{1}{(1-P_e)^2}+T_f\frac{1}{1-P_e}\right]$$
$$T_{f,m}=T\frac{P_e}{1-P_e}+T_f\approx P_eT+T_f$$

Si sono adoperate le relazioni notevoli:

$$\sum_{i=0}^{+\infty}x^i=\frac{1}{1-x}$$
$$\sum_{i=0}^{+\infty}ix^{i-1}=\frac{1}{(1-x)^2}$$

Con $T=T_f+T_{out}$:

$$T_{f,m}=T\frac{P_e}{1-P_e}+T_f=(T_f+T_{out})\frac{P_e}{1-P_e}+T_f=T_f\left(1+\frac{P_e}{1-P_e}\right)+T_{out}\frac{P_e}{1-P_e}=T_f\left(\frac{P_e}{1-P_e}\right)+T_{out}\frac{P_e}{1-P_e}$$
$$T_{f,m}=\frac{T_f+T_{out}+P_e}{1-P_e}$$

L'efficienza η è valutata dal rapporto fra il tempo di trasmissione ed il tempo T_t in cui il canale è occupato:

$$\eta = \frac{T_f}{T_t} = \frac{T_f}{\frac{T_f + P_e T_{out}}{1 - P_e} + 2\tau + T_A + T_P}$$

Analisi delle prestazioni degli schemi continui

Nel caso di assenza di errori nel canale valutiamo l'efficienza η, se *W* è la finestra di pacchetti e *(W-1)Tf* è il tempo di trasmissione di tali pacchetti, nell'ipotesi in cui $2\tau+T_A+T_P \sim 2\tau$ sia il tempo necessario a ricevere un messaggio di conferma per ogni pacchetto, si distinguono due casi:

$$\eta = \begin{cases} 1 \quad se \quad (W-1)T_f > 2\tau \\ \dfrac{T_f + (W-1)T_f}{T_f + 2\tau} \quad se \quad (W-1)T_f > 2\tau \end{cases}$$

In caso di errori il pacchetto è trasmesso N_m volte (numero medio delle ritrasmissioni):

$$\eta = \begin{cases} \dfrac{1}{N_m} \quad se \quad (W-1)T_f > 2\tau \\ \dfrac{1}{N_m}\dfrac{T_f + (W-1)T_f}{T_f + 2\tau} \quad se \quad (W-1)T_f > 2\tau \end{cases}$$

Forward error correction

La capacità del ricevitore dirilevare e correggere gli errori è conusciuta come correzione degli errori in avanto o più semplicemente FEC. La tecnica FEC prevede la trasmissione del pacchetto una sola volta, il ricevitore effettua quindi una rilevazione e correzione dell'errore. Se da un lato l'efficienza η in tal caso diventa unitaria occorre subito precisare che il numero di bit da destinare alla funzione di controllo dell'errore aumenta in quanto tali bit consentono oltre alla rilevazione anche la correzione dell'errore.

In uno schema FEC bisogna per prima cosa fissare il numero di bit t massimi ammissibili in errore cosicchè se nel bacchetto ci sono T bit in erore tali che T<t allora il pacchetto in uscita dal correttore è sicuramente ripristinato. Più in generale, detta P_0 la probabilità che un pacchetto controllato sia in errore (probabilità che di sicuro va minimizzata), il numero di bit t è tale che la quantità:

$$P_0 = \sum_{k=t+1}^{m} p^K (1-p)^{m-k}$$

sia la più piccola possibile (p è la probabilità che un singolo bit sia in errore). Notare che P_o diminuisce all'aumentare di t e quindi dei bit impiegati al controllo di errore. Lo schema di rilevazione dell'errore che adotta un singolo bit, bit di parità, difficilmente si presta all'individuazione precisa del bit in errore. Esso pertanto consente la sola rilevazione ma impedisce la correzione.

Uno schema che invece si presta alla correzione dell'errore è il controllo di parità verticale e longitudinale. In tale schema infatti, qualora sia presente un errore sul pacchetto, risultano due errori

nei meccanismi di controllo della parità che quindi individuano mediante sistema a coordinate il bite in errore. Individuata quindi la posizione esatta del bit in errore all'interno del pacchetto risulta poi immediata la correzione. Tuttavia lo schema di controllo della parità verticale e longitudinale non è idoneo qualora nel pacchetto siano presenti due o più errori (il bit di parità infatti ne indica uno solo per riga o colonna).
Per ovviare a questo problema e consentire dunque la correzione di più errori si considerano valide, in fase di correzione, 2_n stringhe piuttosto che tutte le possibili 2^m stringhe. Le stringhe ammissibili sono progettate in maniera tale che tra esse intercorra una certa distanza detta di hamming cosicchè in fase di correzione il pacchetto converga verso la stringa ammissibile più vicina che quindi risulterà essere quella probabilmente inviata. Se la distanza fra le stringhe ammissibili è di $2d+1$ è allora possibile correggere d errori.
Per ridurre i bit r aggiuntivi richiesti ed il numero dioperazioni necessarie per realizzare la funzione di correzione dell'errore è possibile ricorrere al metodo della ridondanza ciclica CRC. Occorre allora fissare un numero di bit che costituirà la parte informativa del messaggio e stabilire quindi il numero di bit massimi in errore. Fatto ciò si costruisce una tabella in cui si riporta in corrispondenza del divisore utilizzato nel calcolo del resto i possibili valori che questo può assumere. Ad ogni valore del resto corrisponde una particolare sindrome di errore che quindi può essere corretta. Quindi, il resto conseguente alla operazione di divisione consente di risalire all'errore e può individuare i bit da correggere (ad esempio: resto $R(x)$=101, bit 0 e 7 in errore, etc...). Non si ha sempre bisogno della tabella per la rilevazione del tipo di errore. Infatti, considerata la struttura a shift register, vista quando si è analizzato lo schema a blocchi che realizza che realizza il controllo CRC, si può fornire come ingreso (dopo aver dato d bit di ingresso) altri bit (anzichè i bit 0). Si otterrà un'opportuno resto che coincide proprio con quello che si verrebbe a determinare qualora un bit adiacente all'ultimo ingresso sia in errore (primo clock). Se invece si forrniscono due ulteriori ingressi e quindi altridue colpi di clock, il resto assume la forma tipica nel caso in cui i bit in errore sono stavolta adiacenti al penultimo etc... Si può allora memorizzare la stringa che rappresenta il resto in corrispondenza di un singolo errore verificatosi ad esempio su di un bit e realizzare, mediante funzione logica AND, il controllo (e successivamente la correzione) con l'attuale resto contenuto nei registri a scorrimento.

Il controllo del flusso e della congestione

Reti con attraversamento diretto

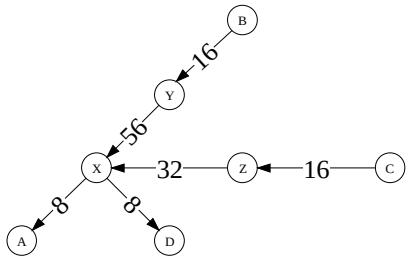
Storicamente parlando, le reti con attraversamento diretto si sono affermate nei periodi in cui non erano ancora disponibili risorse di memorie presso i nodi di rete. Solo quando tali risorse sono state rese disponibili esse hanno allora arricchito le capacità dei nodi che tuttavia ne fanno un uso limitato per contenere i tempi di attraversamento dei flussi informativi. Le reti con attraversamento diretto vanno incontro alla congestione (blocco della rete a causa di una situazione di stallo che si è verificata) quando la somma dei flussi in ingresso ad un nodo supera la frequenza di uscita ammissibile in uscita. Infatti, nell'eventualità che ciò si verifichi le memorie di nodo vanno incontro ad una lenta o rapida saturazione (dipende di quanto si oltrepassa la frequenza di cifra ammissibile in uscita) con l'impossibilità, quindi, di accettare nuovi flussi in ingresso. Per questo motivo, una rete con attraversamento diretto richiede necessariamente un controllo del flusso. Il controllo del flusso, realizzato presso ogni nodo di rete, favorisce il controllo della congestione. Un metodo per realizzare il controllo del flusso consiste ad esempo nell'imporre una riduzione dei flussi di rete in ingresso in maniera tale da evitare la congestione della rete. Tuttavia è anche opportuno non imporre ai flussi tributari un eccessivo abbassamento della frequenza di cifra per non ridurre troppo il rendimento della rete. La soluzione, per reti con attraversamento diretto, è dunque il controllo delle frequenze di cifra dei flussi informativi in ingresso.

Reti con attraversamento stor & forward

Le reti a commutazione di pacchetto sono efficienti perchè condividono meglio tra i vari utenti della rete le risorse disponibili. Ad esempio, un determinato collegamento fisico di una rete potrebbe essere usato per condurre dati tra diverse coppie di utenti. La memoria dei nodi e le capacità elaborative di questi sono risorse condivise e possono essere usate per soddisfare i requisiti di comunicazione di diverse coppie di utenti. In questo scenario appena descritto è proprio la condivisione delle risorse che può portare ai problemi di congestione visti prima.

Si presenta una situazione di congestione della rete quando la somma delle richieste avanzate dai vari flussi tributari per accedere alle risorse supera le capacità che questa offre per soddisfare le richieste avanzate. L'aumento delle risorse non è una buona soluzione poichè esse potrebbero scongiurare le congestioni ma rimarrebbero largamente inutilizzate quando alla rete si rivolgono poche utenze. Occorre perciò trattare come possibile l'evento congestione e cercarne di minimizzarne gli effetti. Questo richiede un'attenta progettaione della rete tale che la probabilità che essa si manifesti sia la piccola possibile. Data la natura imprevedibile del traffico telematico e della distribuzione delle diverse domande degli utenti, risulta necessario prevedere delle procedure che consentano alla rete di uscire da uno stato dicongestione in cui è purtroppo possibile entrare. Osserveremo per prima cosa la necessità di tali funzioni di controllo del flusso e della congestione, quindi, nei successivi paragrafi, si osserveranno i protocolli che realizzano gli adeguati controlli.

Supponiamo di avere una rete non controllata, il traffico può quindi accedere alla rete a tutte le risorse che queste mette a disposizione e senza alcuna limitazione alla frequenza di cifra. Ogni nodo è autorizzato ad iniettare il flusso informativo nella rete. La rete che stiamo immaginando adotta poi una strategia di controllo dell'errore secondo la modalità hop by hop. Pertanto in essa si manifesta un primo inconveniente che di seguito descriviamo: un nodo invia un pacchetto al nodo adiacente e ne attende la conferma. La rete che invece si trova soppiantata dalle richieste ha esaurito le risorse cosicchè il pacchetto, anche se giunto senza errori viene perso. Dall'altro lato, il nodo trasmettitore è in attesa della conferma e tiene anch'esso impegnato le proprie risorse nei confronti del pacchetto in oggetto. Allo scadere dell'intervallo di timeout essor ritrasmetterà il pacchetto. Possono nascere, fondamentalmente, due tipi di problemi in una rete non controllata: la degradazione del ritmo binario e la non equità, ossia un diverso trattamento dei flussi tributari. Per capire meglio questi problemi si considera la rete in figura:



Nell'istante di tempo in cui vogliamo analizzare la rete troviamo in essa le seguenti richiesti: una richiesta di trasferimento dati dal nodo B al nodo A con frequenza di cifra F1; una richiesta ditrasferimento dato dal nodo C al nodo D con frequenza di cifra F2. La prima richiesta segue il percorso di nodi BYXA mentre la seconda richiesta segue il percorso CZXD (il nodo X è quindi condiviso da entrambe le richieste). Due le eventualità che consideriamo:

1) F1=7 Kbit/s ed F2=0, non costituisce nessuna situazione di congestione poichè la richiesta da B ad A è supportata dalle risorse esistenti nella rete (ovviamente tale scenario si può verificare a ruoli invertiti).
2) F1=8+X Kbit/s ed F2=0, i pacchetti sono adesso immessi nella rete ad una frequenza di cifra che è superiore a quella offerta dal collegamento XA (che ne sopporta al massimo 8 Kbit/s). Al nodo X pervengono, dunque, delle richieste iningresso che superano quelle offerte in uscita dal nodo! Con il passare del tempo la memoria del nodo si riempirà e quando ciò si verifica i pacchetti che arrivano dal collegamento YX inizieranno ad essere scartati (per indisponibilità di memoria), non arriveranno nemmeno le conferme ai pacchetti ed allora il nodo Y impegnerà le proprie risorse per conservare una copia dei pacchetti mandati (dovendone effettuare la ritrasmissione ogni timeout). In definitiva, anche il nodo Y andrà incontro ad una congestione poichè il nodo B continuerà ad inviare i pacchetti e questi, inizialmente colmeranno la memoria di Y e successivamente inizieranno ad essere scartati.

Per il punto 2): siccome la frequenza di cifra richiesta è di 8+X Kbit/s (essendo il collegamento XA a frequenza di 8 Kbit/s) ad X i pacchetti giungono con tale frequenza di cifra e sono inviati ad A alla massima frequenza di cifra che il collegamento XA offre. Il nodo X, pertanto, si riempe di X pacchetti al secondo. Quando tutta la memoria del nodo X è occupata esso continua a incamerare pacchett ad un ritmo di 8+X pacchetti al secondo, 8 pacchetti al secondo sono inviati al nodo A mentre X pacchetti al secondo sono scartati. Un istante immediatamente successivo, il nodo Y (che da B ha ricevuto altri pacchetti) esso avrà da trasmettere gli 8+X pacchetti più gli X pacchetti che non hanno ricevuto conferma, quindi 8+2X pacchetti al secondo. Di tali pacchetti, ancora una volta, 8 saranno accodati e quindi spediti (8 pacchetti al secondo presi da una coda e trasmessi fanno posto ad 8 nuovi pacchetti). Al nodo Y servirà posto in memoria per 2X pacchetti, quelli non confermati. Ciò si ripete ogni secondo finchè il ritmo binario sul collegamento YX non raggiunge il massimo valore che il collegamento può sopportare (56 Kbit/s nel nostro esempio). Quando anche la memoria di Y si è saturata il problema si sposta sul nodo B e sul collegamento BY, non passerà molto tempo ed alla fine anche B satura tutta la sua memoria. La situazione che si è delineata prende il nome di deadlock, tutte le risorse sono occupate dai pacchetti, ma nessun pacchetto viaggia nella rete! Ciò è stato causato, come anticipato ad inizio paragrafo, da un aliquota di pacchetti in eccesso che non era supportata da un collegamento (il collegamento XA offre 8 Kbit/s in ingresso al nodo X vi giungevano invece 8+X Kbit/s). Il problema della congestione nell'esempio considerato può essere affrontato in due modi:

1) fornire alla rete risorse a sufficienza in modo che il carico offerto alla rete non superi la capacità delle risorse della rete;
2) limitare, nel nostro caso, il ritmo binario di B ad 8 Kbit/s.

La soluzione 1 richiede di dimensionare il massimo carico che la rete può sopportare e può essere una soluzione valida solo se tale carico massimo si avvicina ai picchi della rete. La soluzione 2 è invece quella più ragionevole e necessita di opportuni protocolli per essere attuata.

Analizziamo un nuovo caso, sempre riferendoci alla rete prima considerata. In questa circostanza supponiamo F1=7 Kbit/s ed F2=7 Kbit/s. Entrambi i flussi sono sopportati dalle frequenze di cifra dei collegamenti logici e non determinano una condizione di congestione. Nell'ipotesi in cui il flusso da B ad A aumenti la frequenza dicifra fino a 8+X Kbit/s è ragionevole pensare che per questo si prospetti la stessa situazione analizzata prima mentre per il flusso che va da C a D (essendo questo al di sotto delle frequenze di cifra dei collegamenti logici) non vi siano problemi. In realtà entrambi i flussi atraversando il nodo X e quindi condividono le risorse di memoria presso tale nodo. Dall'esempio visto

in precedenza possiamo già dire che la memoria di X andrà incontro ad una saturazione (gli 8+X Kbit/s eccedono gli 8 Kbit/s che X propone in uscita). Ciò basta per innescare il meccanismo di prima: i pacchetti da Z ad X inizieranno ad essere scartati (a causa dell'esaurimento di memoria del nodo X) con conseguente traboccamento della memoria del nodo Z prima, e del nodo C dopo. Non appena il nodo X riceve una conferma dal nodo A o dal nodo D può liberare spazio in memoria ed ospitare nuovi pacchetti, ma che direzione è favorità? Di sicuro la frequenza di cifra del collegamento YX (56Kbit/s) è più grande di quella ZX (32 Kbit/s), pertanto è più probabile che X accolga i pacchetti provenienti con maggiore frequenza da Y (la memoria di X è quindi maggiormente occupata da pacchetti di Y che tra l'altro è il nodo che ha causato il pasticcio poichè ha richiesto un trasferimento ad 8+X Kbit/s !). Una tale situazione ha riservato al nodo C un diverso trattamento poichè esso è stato penalizzato molto più di B sebbene fosse B la causa della congestione. Per risolvere il problema è posibile usare le soluzioni 1 e 2 ma in realtà una terza soluzione è possibile. Ciò consiste nel riservare una porzione sufficiente dimemoria, presso il nodo , al traffico che è destinato a D. In questo modo il traffico di rete che da C va a D non è soggetto ai problemi appena considerati. Tuttavia, riservare delle risorse per una attività risulta essere una scelta contraria alla strategia generale di assegnazione a domanda che comporta l'aumento dell'efficienza delle reti a commutazione di pacchetto. In altri termini, se F2 varia con intermittenza la memoria riservata presso il nodo X non può essere utilizzata da una diversa attività. Risulta quindi necessario un compromesso tra efficienza ed equità.

Classificazione delle procedure

Le funzioni di controllo del flusso e della congestione sono state introdotte per assicurare che una sorgente (troppo veloce) non inondi una destinazione con un traffico superiore a quello che questa può gestire. Esse possono essere attuate secondo la modalità hop-by-hop oppure secondo la modalità end-to-end a seconda che la funzione sia assegnata ad uno strato superiore o inferiore a quella che svolge la funzione di instradamento.

Le procedure per il controllo del flusso e della congestione possono essere di tipo reattivo se è prevista nello schema la retroazione con alcune variabili di rilievo che simboleggiano l'attuale stato della rete. Tali variabili sono scambiate tramite messaggi per cui questi schemi sono efficienti solo se il tempo neccessario allo scambio di tali messaggi è più piccolo della durata della congestione. Questi schemi mettono in atto tutta una serie di azioni che sono prese dalla rete per scongiurare la congestione.

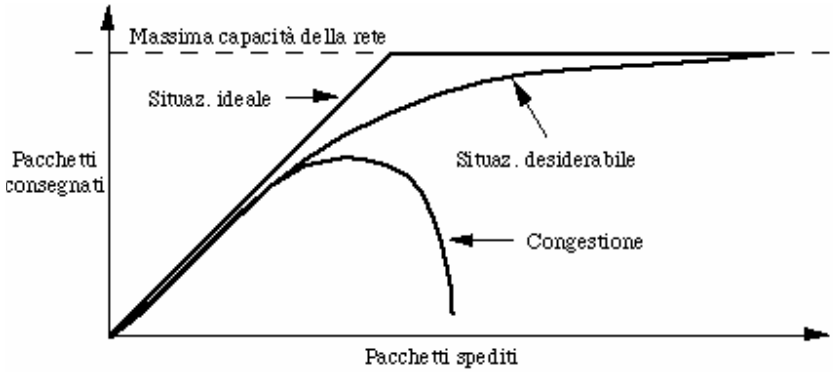
Esiste poi una strategia di controllo preventivo che tentano di evitare la formazione di una congestione. Se i ritardi di rete non sono significativi, la retroazione piò essere utilizzata in una procedura preventiva; in caso contrario, la prevenzione della congestione può essere realizzata sulla base di uno schema a ciclo aperto (senza retroazione) che quindi cerca di anticipare lo stato della rete ed alloca di conseguenza le risorse della rete.

Schemi a finestra

Lo scopo principale ci uno schema a finestra è quello di adattare il ritmo binario del trasmettitore al ritmo binario del ricevitore, esso realizza per questo motivo il controllo della congestione. La funzione principale degli schemi a finestra è quella di limitare, quando opportuno, il flusso dipacchetti inviati dalla sorgente. Questo ha il duplice effetto di limitare il ritmo a cui i pacchetti giungono al nodo di destinazione (controllo del flusso) ed il carico offerto alla rete di transito (controllo della congestione). Gli schemi a finestra realizzano tutti e due i tipi di controllo. Il numero di pacchetti in transito nella rete e non ancora riscontrati stabiliscono la dimensione della finestra. Ogni nodo memorizza in una variabile, una sorta di contatore, la dimensione della finestra che aumenta di una unità per ogni pacchetto trasmesso e viene ridotta di una unità per ogni pacchetto confermato. Quando la finestra è satura (quando il contatore della finestra è pari alla dimensione della finestra) esso scarta tutti i pacchetti richiesti al di fuori della finestra.

Controllo del flusso
Nell'ipotesi in cui il nodo di transito verso sorgente e destinazione sia uno solo è possibile assistere a due particolari scenari. Il nodo sorgente A che invia verso destinazione B fa transitare quindi i suoi pacchetti verso X, nodo intermedio. Il nodo A attende risposta di conferma da X, ma questo invia tale conferma solo dopo che essa è prima mandata da B (X quindi la propaga ad A). Appena giunta la conferma, il nodo A invia un nuovo pacchetto verso B e transitando sempre in X. In X però lo spazio necessario a memorizzare il pacchetto di A già c'era in precedenza (quando cioè B ha confermato ad X il pacchetto di A). In questo scenario i pacchetti occupano la memoria di A poichè attendono di essere trasmessi. Un secondo scenario è invece quello che prevede l'invio del messaggio di conferma quando il pacchetto è ricevuto da X e non da da B. Questa soluzione, invece, occupa la memoria di X.
Un approccio consiste nello scindere i due messaggi che i nodi si scambiano, ossia quelli relativi al messaggio di conferma e quelli relativi alla possibilità di ospitare in coda un altro messaggio. La forma più semplice di questo approccio è l'uso del messaggio RNR, receive not ready. Esso indica al trasmettitore la corretta ricezione di un pacchetto ma avverte quest'ultimo sullo stato di memoria del nodo intermedio che al momento non può ospitare successivi pacchetti.
Un diverso schema per la regolazione del flusso è quello che fa uso di crediti per la trasmissione: un messaggio di conferma è trasmesso quando un pacchetto è ricevuto ed un pacchetto che smboleggia un credito è usato per indicare al trasmettitore la capacità ad ospitare più pacchetti.

Controllo della congestione
Quando troppi pacchetti sono presenti in una parte della retet, si verifica una congestione che degrada le prestazioni. Ciò dipende dal fatto che, quando un router non riesce a gestire tutti i pacchetti che gli arrivano, comincia a perderli, e ciò causa delle ritrasmissioni che aggravano ancor più la congestione.



La congestione in un router può derivare da diversi fattori:

- troppi pochi buffer nel router;
- processore troppo lento nel router;
- linea di trasmissione troppo lenta (si allunga la coda nel router di partenza).

Inoltre, la congestione in un router tende a propagarsi ai suoi vicini che gli inviano dati. Infatti, quando tale router è costretto a scartare i pacchetti che riceve non li conferma più, e quindi i router che li hanno spediti devono mantenerli nei propri buffer, aggravando così anche la propria situazione.
Il controllo della congestione è un problema globale di tutta la rete, ed è ben diverso dal problema del controllo di flusso nei livelli data link, network (nel caso dei servizi connection oriented) e trasporto, che invece riguarda una singola connessione sorgente-destinazione.

Ci sono due approcci al problema della congestione:

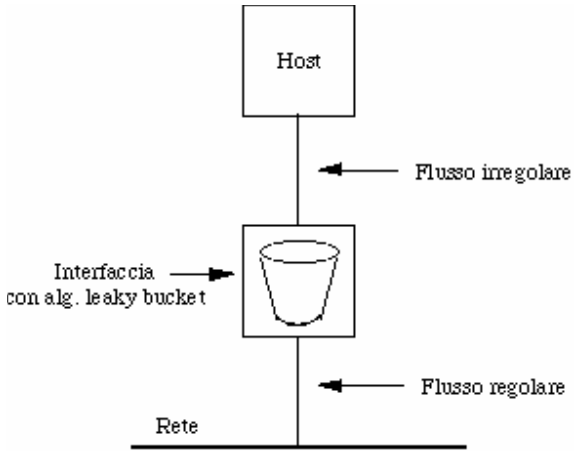
- open loop (senza controreazione);
- closed loop (con controreazione);

Il primo cerca di impostare le cose in modo che la congestione non si verifichi, ma poi non effettua azioni correttive. Il secondo tiene sott'occhio la situazione della rete, intraprendendo le azioni opportune quando necessario.

Nell'approccio traffic shaping, di tipo open loop, l'idea è di forzare la trasmissione dei pacchetti a un ritmo piuttosto regolare, onde limitare la possibilità di congestioni. Verdemo tre tecniche per implementare il traffic shaping:

- leaky bucket;
- token bucket;
- flow specification;

Algoritmo Leaky bucket (secchio che perde)
L'idea è semplice, e trova un analogia reale in un secchio che viene riempito da un rubinetto (che può essere continuamente manovrato in modo da risultare più o meno aperto) e riversa l'acqua che contiene attraverso un forellino sul fondo, a ritmo costante. Se viene immessa troppa acqua, essa fuoriesce dal bordo superiore del secchio e si perde. Sull'host si realizza (nell'interfaccia di rete o in software) un leaky bucket, che è autorizzato a riversare sulla rete pacchetti con un fissato data rate (diciamo bps) e che mantiene, nei suoi buffer, quelli accodati per la trasmissione. Se l'host genera più pacchetti di quelli che possono essere contenuti nei buffer, essi si perdono.



Algoritmo token bucket (secchio di gettoni)
E' una tecnica per consentire un grado di irregolarità controllato anche nel flusso che esce sulla rete. Essenzialmente, si accumula un credito trasmissivo con un certo data rate (fino ad un massimo consentito) quando non si trasmette nulla. Quando poi c'è da trasmettere, lo si fa sfruttando tutto il credito disponibile per trasmettere, fino all'esaurimento di tale credito, alla massima velocità consentita dalla linea. Il secchio contiene dei token, che si creano con una cadenza prefissata (ad esempio uno ogni millisecondo) fino a che il loro numero raggiunge un valore M prefissato, che corrisponde all'aver riempito il secchio di token. Per poter trasmettere un pacchetto (o una certa quantità di byte), deve

essere disponibile un token. Se ci sono k token nel secchiello e h>k pacchetti da trasmettere, i primi k sono trasmessi subito (al data rate consentito dalla linea) e gli altri devono aspettare dei nuovi token.
Dunque, potenzialmente dei burst di M pacchetti possono essere trasmessi in un colpo solo, fermo restando che mediamente non si riesce a trasmettere ad una velocità più alta di quella di generazione dei token. Un'altra differenza col leaky bucket è che i pacchetti non vengono mai scartati (il secchio contiene token, non pacchetti). Se necessario, si avverte il livello superiore, produttore dei dati, di fermarsi per un pò. Questi due algoritmi possono essere usati per regolare il traffico host-router e router-router; in quest'ultimo caso però, se il router sorgente è costretto a fermarsi invece di inviare dati e non ha spazio di buffer a sufficienza, questi possono perdersi.

Flow specification
Il traffic shaping è molto efficace se tutti (sorgente, subnet e destinazione) si accordano in merito. Un modo di ottenere tale accordo consiste nello specificare: le caratteristiche del traffico che si vuole inviare (data rate, grado di burstiness, ecc.); la qualità del servizio (ritardo massimo, frazione di pacchetti che si può perdere, ecc.).
Tale accordo si chiama flow specification e consiste in una struttura dati che descrive le grandezze in questione. Sorgente, subnet e destinatario si accordano di conseguenza per la trasmissione. Questo accordo viene preso prima di trasmettere, e può essere fatto sia in subnet connesse (e allora si riferisce al circuito virtuale) che in subnet non connesse (e allora si riferisce alla sequenza di pacchetti che sarà trasmessa).

Choke packet
In questo approccio, di tipo closed loop, è previsto che un router tenga d'occhio il grado di utilizzo delle sue linee di uscita. Il router misura, per ciascuna linea, l'utilizzo istantaneo U e accumula, entro una media esponenziale M, la storia passata:

$$M_{nuovo}=a\ M_{vecchio} + (1 - a)U$$

dove il parametro a (compreso fra 0 ed 1) è il peso dato alla storia passata; (1-a) è il peso dato all'informazione più recente. Quando, per una delle linee in uscita, M si avvicina a una soglia di pericolo prefissata, il router esamina i pacchetti in ingresso per vedere se sono destinati alla linea d'uscita che è in allarme. In caso affermativo, invia all'host di origine del pacchetto un choke packet (to choke significa soffocare) per avvertirlo di diminuire il flusso. Quando l'host sorgente riceve il choke packet diminuisce il flusso (tipicamente lo dimezza) e ignora i successivi choke packet per un tempo prefissato, perché tipicamente ne arriveranno molti in sequenza. Trascorso tale tempo prefissato, l'host si rimette in attesa di altri choke packet. Se ne arrivano altri, riduce ancora il flusso. Altrimenti, aumenta di nuovo il flusso.

Principi e modelli di riferimento: il modello OSI

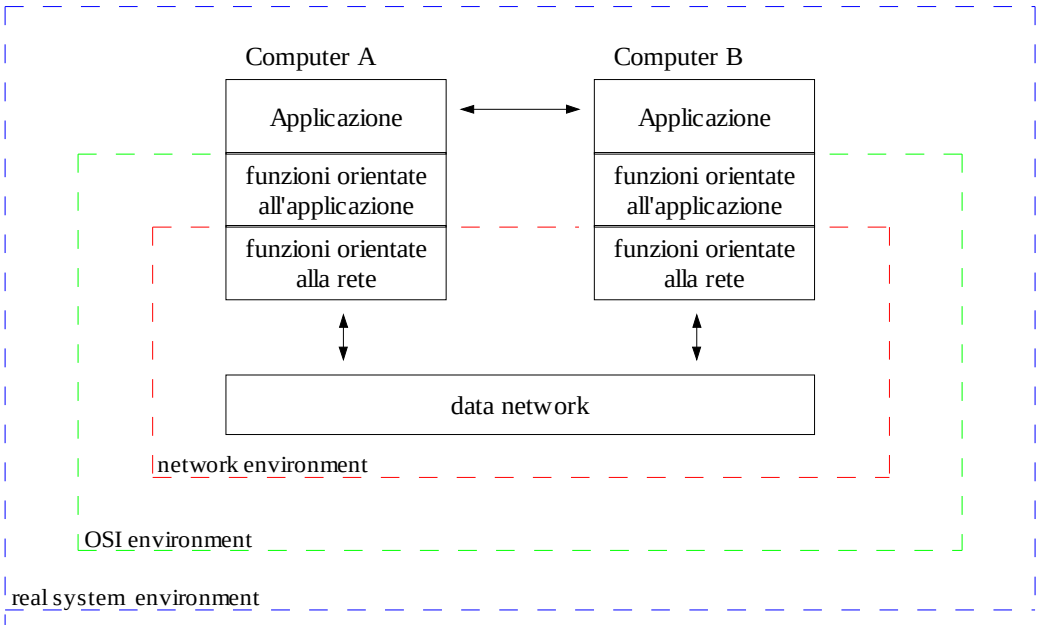
Negli anni 70’ esistevano sul mercato diverse aziende produttrici di hardware e di software che dedicavano i loro sforzi verso la realizzazione di dispositivi orientati alla comunicazione. Ogni dispositivo garantiva tuttavia la comunicazione solo con i dispositivi della stessa casa produttrice. In altre parole il software dedicato alla gestione delle interfacce di rete era proprietario e ciò non consentiva l’interfacciamento verso dispositivi diversi ostacolando, dunque, l’espansione della rete.
Verso la fine degli anni 70’ iniziò un processo di standardizzazione, iniziato dalla ISO (international standardization organization). L’OSI è un modello di riferimento aperto a tutti i produttori e garantisce, qualora siano rispettate un insieme di regole, la comunicazione fra processi applicativi residenti su computer di case produttrici diverse. Il modello OSI non definisce il funzionamento delle singole

funzioni necessarie ad instaurare una comunicazione fra due o più terminali cosicchè ad ogni singolo produttore è comunque lasciata la possibilità di trovare un algortimo più efficiente.
Il modello OSI si presenta strutturato in sette livelli, ciascuno dei quali implementa distinte funzioni di rete e fornisce servizi ai livelli adiacenti. La stratificazione consente di suddividere una complessa funzione come è la comunicazione in tante sottofunzioni più semplici. Ogni livello si comporta come se la comunicazione avvenisse con il livello paritario presente presso l'entità coinvolta nel processo di scambio informativo. Spesse volte infatti si dice che i livelli paritari sono tra loro virtualmente connessi. I sette livelli del modello OSI sono poi suddivisi in due categorie, in base alle funzioni che essi svolgono:

- network depended function;
- application oriented function;

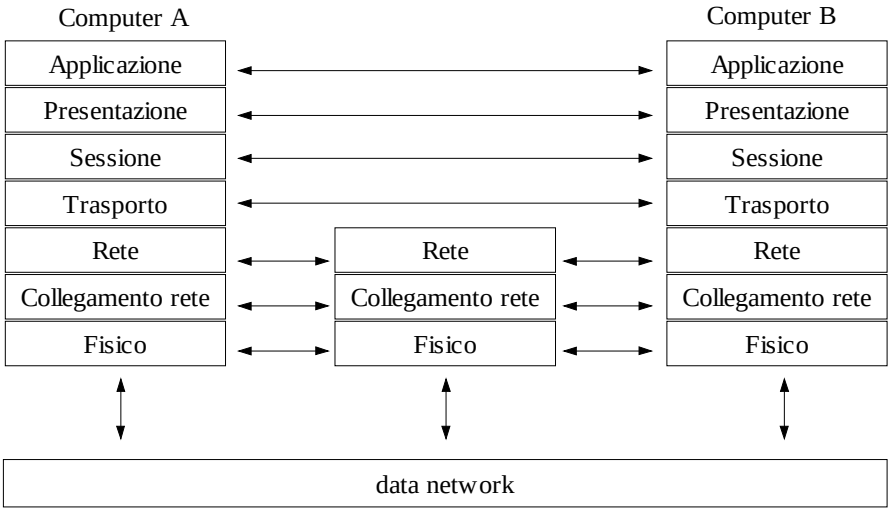
questo porta alla definizione di tre ambienti operativi:

- network environment;
- OSI environment;
- real system environment;



I principi che hanno portato alla stratificazione e che ne regolano la scelta sono i seguenti:

- l'aggiunta di un livello è necessaria solo se è richiesto un altro livello di astrazione;
- ogni livello deve essere caratterizzato da determinate funzioni;
- la funzione assegnata a ciascuno strato deve essere scelta considerando i soli protocolli standard e quindi riconosciuti a livello internazionale;
- la definizione di uno strato deve minimizzare il flusso scambiato alle interfacce;
- il numero dei livelli logici deve essere sufficientemente alto in maniera tale da non mischiare assieme, nello stesso livello, funzioni che non hanno nulla in comune;



Questa figura ci permette di apprezzare un interessante particolare: i primi tre livelli dell'architettura OSI (livelli fisico, collegamento dati e rete) fanno parte del network environment e pertanto sono replicati presso i nodi intermedi della rete che connettono l'host A all'host B.
Quando un utente invia un messaggio quest'ultimo attraversa tutti gli strati, dal livello di applicazione fino a quello fisico. Il messaggio originario si sposta nella pila protocollare attraverso le interfacce che costituiscono quindi delle porte di accesso presso ogni strato. Ogni strato aggiunge al messaggio tutte le informazioni necessarie affinchè esso possa essere poi elaborato dal livello adiacente o dal livello paritario. Quando il messaggio giunge presso l'entità di nodo in ricezione quest'ultimo percorre la pila protocollare dal basso verso l'alto, in tal caso ogni strato elabora il messaggio ricevuto prelevando le informazioni aggiunte dal livello paritario e mandando verso l'alto il messaggio, fino al livello di applicazione.
Il livello fisico si preoccupa della gestione del mezzo trasmissivo e quindi del canale logico (inteso come la cascata di trasmettitore, collegamento fisico e ricevitore) su cui avviene lo scambio di informazioni. Il livello fisico è poi responsabile dell'attivazione, disattivazione e mantenimento del collegamento fra trasmettitore e ricevitore. Il messaggio giunge attraverso l'interfaccia del livello fisico ed è qui convertito in un flusso continuo di bit, successivamente viene trasmesso. Esso definisce le modalità di connessione tra il cavo e la scheda di rete e di conseguenza regola anche le caratteristiche dei mezzi fisici da adoperare. Sono ad esempio definite (dallo standard OSI) le seguenti caratteristiche:

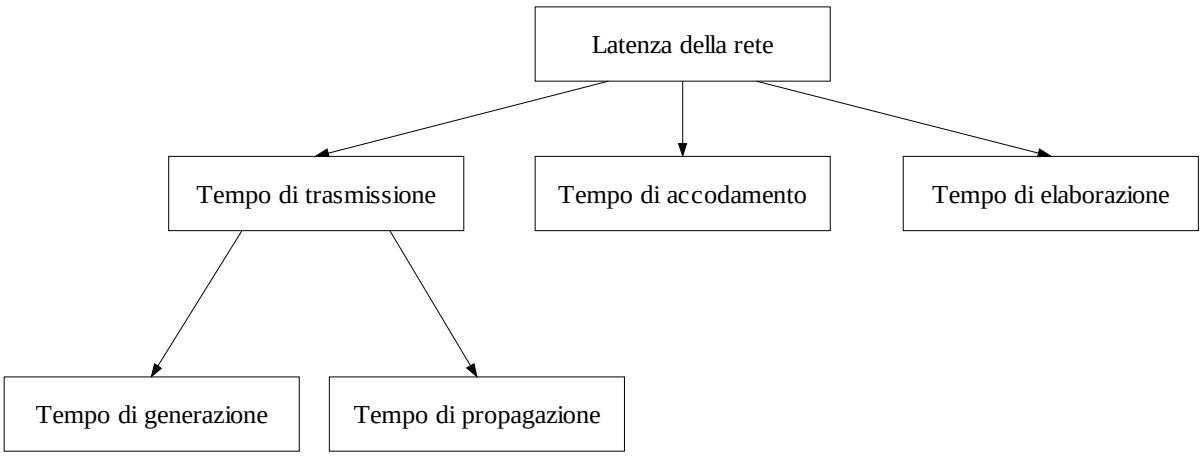
- caratteristiche fisice, come il numero di piedini di un connettore;
- caratteristiche elettriche, come i valori di tensione per i livelli logici adottati per la codifica;
- caratteristiche funzionali, come il significato dei pin di un componente della scehda di rete;

Il secondo strato, dal basso verso l'alto, è lo strato data-link o collegamento dati. Esso assume già un certo livello di importanza e poggia sul livello sottostante, il livello fisico visto in precedenza, cercando di creare un collegamento privo di errori in termini di bit trasmessi. A questo livello diventa molto importante la topologia della rete. Il data-link si preoccupa allora dell'indirizzo fisico ed organizza i suoi dati in una struttura logica detta frame (oppure trama) cosicchè possa essere garantita la ricezione presso il dispositivo di rete che è in questo modo identificato. Il livello di collegamento dati prevede inoltre la possibilità che la linea di comunicazione possa alterare il contenuto di un frame a causa dei disturbi elettromagnetici. Esso pertanto supporta dei meccanismi di ritrasmissione dei frame andati

persi e/o corrotti. Il problema della numerazione dei pacchetti è qui fortemente sentito, a causa delle ritrasmissioni, la numerazione dei pacchetti, risove il problema dei pacchetti duplicati.
Una importante funzione affidata allo strato collegamento dati è quella che cerca di evitare che un trasmettitore troppo veloce metta in difficoltà un ricevitore relativamente più lento. Occorre pertanto inserire un meccanismo di regolazione del traffico che consenta al trasmettitore di sapere quanto spazio di memoria buffer è andato utilizzato presso il nodo ricevitore. Il livello collegamento dati è stato pensato in presenza di collegamenti punto-punto che connettono tra loro vari nodi. Qualora, invece, il canale di accesso al mezzo fisico è comune si definisce, allora, un sottostrato detto livello MAC (medium access controll).
Lo strato di rete svolge un importante funzione che è quella di routing. Lo strato sottostante e quindi il livello data-link mette a disposizione un canale di comunicazione affidabile per la trasmissione dati. Lo strato di rete quindi deve adesso risolvere il problema di trovare un percorso idoneo attraverso cui far transitare i pacchetti informativi verso il nodo di destinazione. Per questo motivo lo strato di rete si occupa della traduzione del nome logico dell'host destinatario. L'instradamento è quindi realizzato attraverso la consultazione di apposite tabelle, esse possono essere statiche oppure dinamiche. Sempre nello strato di rete si trovano implementate le funzioni per il controllo della congestione (troppi pacchetti nella rete contemporaneamente).
Il quarto strato è lo strato di trasporto, si tratta di un cosiddetto punto end-to-end, cioè non è presente nei nodi intermedi della rete e si deve far carico di tutte le problematiche relative al trasferimento dei dati dalla sorgente alla destinazione (i soli due punti che nel processo di scambio informativo sono dotati di tale strato). Ad esempio, se la tecnica di trasferimento è di tipo datagram, quando si dovrà ricostruire il messaggio, nel nodo di destinazione, bisognerà effettuare opportuni controlli per ripristinare il messaggio originale andato corrotto. In questo modo, al livello superiore, si fornisce un servizio di comunicazione affidabile. Pertanto le funzioni svolte da tale livello sono quelle della frammentazione del messaggio e del controllo di errore.
Molti terminali sono in grado di aprire diverse connessioni di trasposrto dati verso altri terminali ed allora ogni pacchetto di tale strato possiede una informazione aggiuntiva che indica appunto la connessione di appartenenza. Qualora un'applicazione richieda un elevato throughput possono allora essere aperte più connessioni di rete. Lo strato di trasporto, infine, offre allo stratto superiore (livello di sessione) un tipo di servizio che viene caratterizzato dai cosiddetti parametri di QOS (quality of service). Il modello OSI ne prevede 5 classi, numerati in maniera sequenziale da 0 a 4.
Nonostante gli sforzi messi in atto nello strato di trasporto per risolvere le problematiche che si incontrano per realizzare un trasporto affidabile, i servizi finora messi a disposizione non permettono ancora a due processi apllicativi di colloquiare. Essi infatti potrebbero intraprendere un azione di comunicazione in contemporanea con l'inevitabile sovrapposizione del flusso informativo. Ed allora il livello di sessione permette di organizzare e sincronizzare il dialogo delle entità di livello superiore. E' necessario un sistema a token che permetta la trasmissione solo all'entità che lo possiede. Sono ovviamente necessari schemi di assegnazione del token.
Il livello di presentazione si preoccupa di preparare le informazioni ricevute dal livello superiore (livello applicazione) in un formato adatto alla trasmissione, ciò avviene effettuando delle conversioni secondo gli standard attualmente riconosciuti: ASCII ed EBCD per gestire file di testo, standard GIF (graphic interchanging format), JPEG (joint photographics eperts group) e TIFF (tagged image file format) per la rappresentazione di immagini, standard MPEG (motion picture eperts group) per la codifica dei flussi video, standard MIDI (musical instrumental digital interface) per l'audio digitale etc...
Il livello più alto della pila protocollare del modello OSI è il livello di applicazione, data la sua posizione all'interno del modello, quest'utlimo non offre servizi ad ulteriori livelli ma interagisce in modo diretto con le applicazioni usate dall'utente. Al fine di soddisfare i servizi richiesti dall'utente

esso verifica innanzittutto se vi sono sufficienti risorse per stabilire un processo di comunicazione tra due sistemi. Lo strato di apllicazione è fornito di due interfaccie, una diretta e l'altra indiretta. Attraverso l'interfaccia diretta con il livello sottostante lo strato di applicazione serve le applicazioni tipiche di una rete, quali ad esempio lo scambio di e-mail, il trasferimento dei file, l'accesso ai database, l'accesso ai siti web, la gestione remota di applicazioni distribuiti e l'emulazione di terminali. Nel controllo dei sistemi remoti, per non incorrere in problemi dovuti ai diversi tipi di terminali esistenti, viene definito come standard il terminale VT100.
L'interfaccia indiretta e un redirector vengono invece forniti per le apllicazioni stand-alone in un ambiente LAN. Il compito del livello di applicazione è quello di virtualizzare l'ambiente OSI fornendo alle applicazioni utenti i mezzi per accedere a tale ambiente.
Il modello OSI ha il grande vantaggio di distinguere in modo concettualemte molto chiaro tra servizio di uno strato (insieme di funzioni di strato), interfaccia di uno strato (come si richiedono i servizi ad uno strato) e protocolli di uno strato (in che modo le entità di uno strato svolgono le funzioni). Le porte di accesso ad uno strato sono chiamate SAP, service access point. In esse transitano i messaggi che gli strati solitamente si scambiano e che si dicono talvolta IDU (information data unit). L'IDU è diviso in due blocchi, l'ICI (interface controllo information) che contiene la richiesta di un servizio al livello adiacente e l'SDU (service data unit) che invece contiene i dati veri veri e propri che non vengono per questo motivo modificati. Ad ogni livello viene aggiunto un PCI (protocol control unit) che assieme all'SDU forma la PDU (protocol data unit) che rappresenta l'informazione ricevuta dal livello paritario.

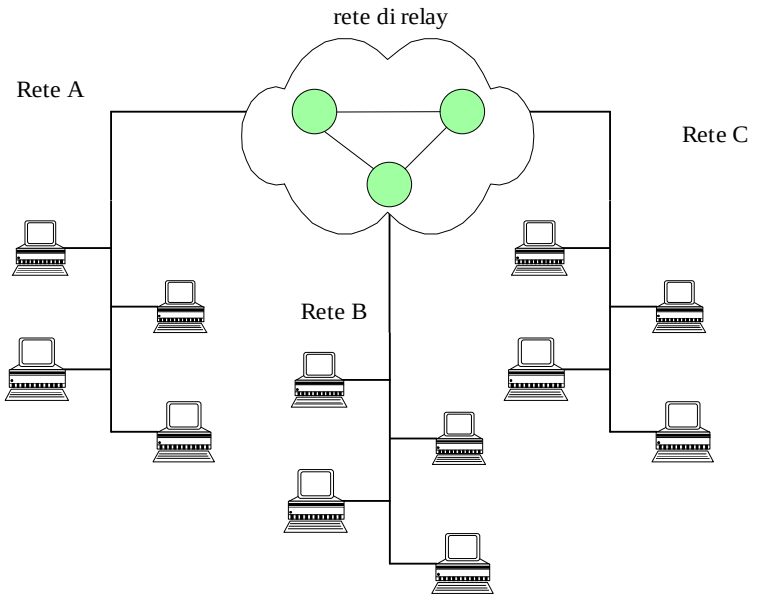
Latenza della rete
Uno dei parametri di qualità del servizio offerto da una rete di telecomunicazioni è il ritardo di transito del flusso informativo inteso come il tempo in cui un blocco di dati permane nella rete. E' compito del progettista stimare il tempo di ritardo a cui un pacchetto va incontro, magari modellando la rete attraverso complicate funzioni.
In generale la latenza della rete si compone di tre componenti: il tempo di trasmissione, il tempo di accodamento ed il ritardo di elaborazione. Tali componenti sono da intendersi nei valori medi poichè variano da nodo a nodo ed ognuno di questi è poi caratterizzato da un diverso stato in termini di capacità elaborative, capacità di memoria e capacità trasmissive.
Il ritardo di trasmissione è a sua volta composto da due componenti, una di queste è il ritardo di generazione, ossia il tempo necassario ad inviare nella rete un blocco di N bit alla frequenza di R bit/s, dunque pari ad N/R e talvolta detto tempo di pacchettizzazione. L'altra componente del ritardo di trasmissione è invece il ritardo dovuto alla propagazione del segnale elettromagnetico.
Il ritardo dovuto all'accodamento è dovuto essenzialmente alla condivisione delle risorse di nodo. Infatti, due o più applicazioni che presentano una richiesta alle risorse di nodo vanno incontro a dei tempi di attesi dovuti al fatto che l'entità che per prima richiede una risorsa è anche la prima ad usufruirne delle capacità. Le richieste che invece pervengono successivamente presso il nodo di rete finiscono in una coda di attesa. Un nodo di rete interessato da maggior traffico introduce sui pacchetti un maggiore ritardo dovuto ad una coda di attesa più affollata.
Altra componente della latenza della rete è il ritardo dovuto all'elaborazione di ogni singolo pacchetto. Tale elaborazione consente l'instradamento poichè da un'analisi del pacchetto il nodo stabilisce la porta di uscita verso cui inviare il pacchetto. Tuttavia tempi ben più grandi sono necessari per stabilire se il pacchetto ricevuto sia effettivamente integro. In caso contrario saranno necessarie un certo numero di ritrasmissioni finchè il pacchetto giunge integro presso il nodo di destinazione. La variabilità dei ritardi della latenza della rete deve infine considerare la possibilità che diversi pacchetti possono intraprendere percorsi diversi, essi quindi sono soggetti non tutti agli stessi ritardi.



Interconnesione tra reti

Quando il modello OSI fu proposto ai costruttori di dispositivi e software esistevano già delle reti che non essendo state realizzate sulla base di accordi presi erano pertanto strutturate, in termini architetturali, diversamente da come lo standard prevede. In questo scenario, gli utenti di una rete non potevano allora scambiare flussi informativi verso gli utenti di un'altra rete.

Questo problema di interconnessione tra reti esistenti viene risolto utilizzando appositi dispositivi detti relay che si aggiungono alle reti esistenti senza richiedere quindi alcuna modifica hardware alla rete e consegnano agli utenti una visione della rete unitaria.



Nel caso di due reti distinte, l'interconnessione si realizza mediante un relay connesso direttamente ad entrambi le reti. Qualora invece l'interconnessione riguarda più didue reti si introduce tra di loro una rette di frame relay, ogni rete è collegata ad un frame relay il quale scambia flussi informativi con le altre reti mediante l'individuazione di un cammino ottimale verso il frame relay di destinazione.

Un primo problema da risolvere nella gestione dell'interconnessione mediante frame relay è la collocazione del frame relay nei confronti delle due reti da connettere. E' cioè necessario individuare

nella rete A e nella rete B I nodi a cui connettere il relay. Si tratta di una scelta che va presa anche considerando la topologia della rete nonchè I flussi informativi che la attraversano.
Altro problema da risolvere è la numerazione dei nodi: affinchè il relay sia in grado di rintracciare un nodo di rete in maniera univoca, è necessario etichettare i nodi delle due reti in maniera tale da scongiurare eventuali ambiguità. Siccome il frame relay, nel caso di interconnessione fra due reti, deve connettere due reti diverse, esso sarà dotato di una coppia di pile protocollari. Se la rete A si connette al frame relay alla porta di ingresso 1 e se la rete B si connette al frame relay alla porta di ingresso 2, essondo le due reti (ad esempio) organizzate diversamente, esse implicheranno presso il nodo frame relay la presenza di due pile protoccollari: una di queste sarà dedicata alla rete A e l'altra sarà invece dedicata alla rete B.

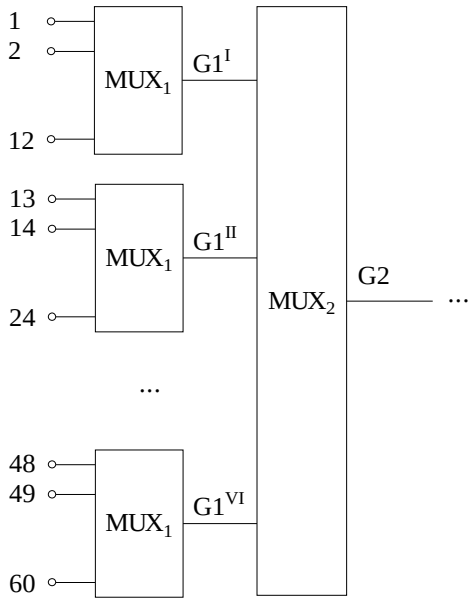
Siccome la possibilità di interconnettere due reti tra di loro è molto varia non esistono delle regole generali che valgono in tutti i casi di interconnesione. Il caso più banale di interconnesione è ad esempio costituito dalla connessione delle due reti mediante un semplice cavo.
Il caso di interconnessione più complicato è invece quello che ricorre all'uso di dispositivi più elaborati detti gateway. Essi tengono in considerazione tutte le differenze che esistono tra le architetture protocollare delle reti che devono essere connesse. Talvolta in ciascuna rete da interconnettere vengono aggiunti più di un gateway, ciò è dovuto alla estensione geografica della rete da interconnettere: in questo modo si prevedono cioè più punti di interconnessione.
Con l'introduzione di relay e gateway si risolve il problema dell'interconnessione tra reti, tuttavia l'interconnessione fra due reti ha da sempre generato competizioni fra le reti connesse qualora allo strato di applicazione e quindi presso l'utente finale essa stessa si presenta con caratteristiche diverse.
Nel caso incui le due reti dispongano dello stesso protocollo nello strato di applicazione l'interconnessione è allora trasparente agli utenti che ne fanno uso.
Qualora, invece, le due reti adottano differenti protocolli nello strato di applicazione, l'interconnessione porta alle competizioni prima citate. Se un numero comunque cospicuo di utenti preferisce entrami le due reti allora il gateway mette in atto l'interconnessione fornendo agli utenti delle due reti la possibilità di dialogare fra loro. Qualora, invece, una delle due reti è maggiormente più seguita dell'altra (in termini di utenti soddisfatti), allora la rete con meno utenti è solitamente destinata a scomparire nel tempo. Oppure, con il passare degli anni essa si guadagnerà la fornitura di un particolare servizio (lasciando gli altri servizi all'altra rete) qualora il servizio offerto rimane comunque migliore della rete antagonista.
Un altra situazione di interesse storico è il caso in cui duu reti forniscono servizi differenti ed entrambi hanno un seguito nei numerosi utenti. In questo scenario le due reti continueranno ad esistere, in maniera tra loro separate ma interconnesse, rimanendo a lungo tra loro in equilibrio. La scoperta, poi, e l'evoluzione tecnologica può in ogni caso rompere questo equilibrio quando una delle due reti si avvale di nuove tecnologie capaci di fornire entrambi le applicazioni. Gli utenti allora inizieranno a scegliere per una o per l'altra rete.
Quest'ultima situazione si è verificata nel perioo in cui le reti TCP/IP e le reti geografiche dati realizzavano un servizio orientato allo scambio dei dati telematici mentre le reti telefoniche offrivano il servizio della conversazione in tempo reale. L'evoluzione della rete telefonica e della rete geografica dati in una nuova rete ATM che forniva entrambi i servizi sembrava ormai destinata a soppiantare le reti TCP/IP. Tuttavia le reti ATM sono state standardizzate prima del diffondersi di un importante apllicazione, il web browsing. Le reti TCP/IP (a caccia di nuove tecnologie e applicazioni da inglobare) hanno da subito accolto l'applicazione killer del web browsing e ciò ha consentito la sopravvivenza delle reti TCP/IP. Dall'altro lato, invece, le reti ATM non potendo inglobare il web browsing (ciò richiederebbe una modifica significativa alla struttura di ATM) hanno continuato ad offrire i servizi di scambio dati e conversazione in tempo reale. Quando le reti TCP/IP, lentamente, hanno trovato il modo

di fornire anche esse il servizio della telefonia in tempo reale, la situazione di equilibrio prima descritta ha cominciato a cedere a favore di TCP/IP che si è inserita man mano nelle comunicazioni sulle dorsali di reti, laddove ATM è stata a lungo utilizzata.

Standard di rete per flussi non intermittenti

Gerarchia FDM

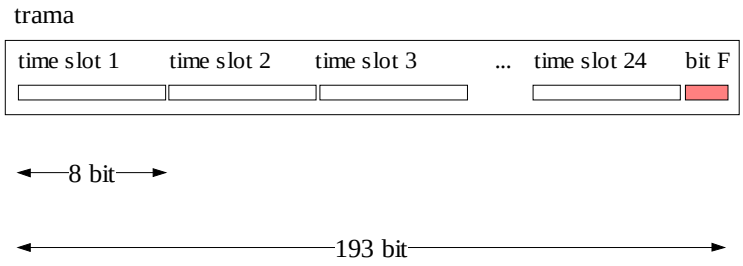
Nella gerarchia FDM, i cui multiplatori adottano una multiplazione a divisione di frequenza, il flusso informativo si presenta sotto forma di segnale analogico con banda lorda di 4KHz (segnale vocale analogico). I flussi tributari sono accorpati nei vari livelli per formare, man mano che si sale nella gerarchia, un flusso multiplato che aggrega in se più tributari. Il primo livello della gerarchia si occupa di multiplare 12 canali telefonici collocandoli nella banda 60-108 KHz e generando in questo modo un segnale multiplato che è solitamente detto gruppo primario o G1. Il secondo livello si occupa, invece, della multiplazione di 5 gruppi primari per formare un gruppo secondario G2 che ingloba quindi 60 canali telefonici.
Per la formazione dei livelli gerarchici superiori si possono presentare due altrenative. Una prima alternativa è la multiplazione di 15 gruppi secondari per la formazione di un gruppo terziario G3 al cui interno trovano posto 900 canali telefonici (soluzione adottata in Italia). Una diversa strategia prevede invece la multiplazione prima di 5 gruppi secondari per formare un gruppo terziario G3 da 300 canali telefonici e successivamete, dalla multiplazione di 3 gruppi terziari, si origina un segnale multiplato di 900 canali telefonici (soluzione adottata in America). I mezzi trasmissivi adoperati sono le coppie simmetriche, le coppie coassiali nelle due versioni 1.2/4.4 mm (coassialino) e 2.6/9.5 mm (coassiale), nonchè i ponti radio.



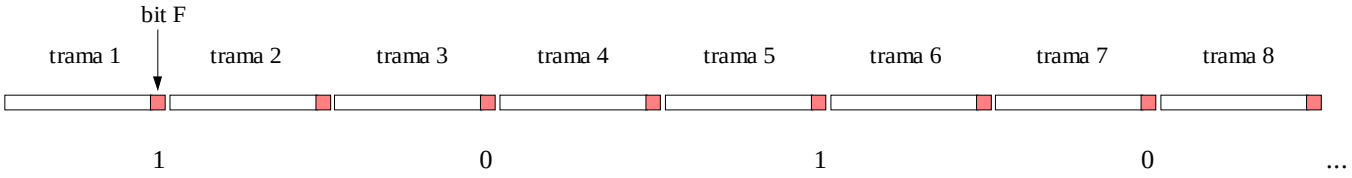
Gerarchia plesiocrona

Nella gerarchia plesiocrona il flusso tributario si riferisce a flussi numerici con frequenza di cifra costante e pari a 64Kbit/s (segnali campionati con tecnica PCM). Il segnale è campionato 8000 volte in un secondo, con frequenza quindi di 8KHz. Ogni campione è rappresentato con stringhe di 8 bit. Il flusso numerico ottenuto dal campionamento ha così una frequenza di cifra pari a 64 Kbps.

Nel sistema T1 sono multiplati 24 time slot che formano così una trama. Ogni time slot è lungo 8 bit e fa riferimento quindi ad un tributario, un ulteriore bit F è poi aggiunto in fase di multiplazione per realizzare l'allineamento.



In un secondo si presentano 8000 trame ed essendo la trama lunga 193 bit si ha un flusso multiplato che ha una frequenza di cifra di (8000·193) 1.544Mbps. Tale struttura di trama 1 (T1) è anche detta DS1 (digital signaling 1). Per una corretta demultiplazione è richiesto di identificare l'inizio di ciascun time slot (che fa riferimento ad un tributario) e di ciascuna trama, per questo motivo sono aggiunti dei bit di allineamento. Per realizzare l'allineamento nel sistema T1 si utilizza il bit F aggiunto, più in particolare: i bit F delle trame dispari hanno forma 101010 e permettono l'allineamento di trama, i bit F delle trame pari hanno invece forma 001110 e realizzano l'allinemaneto di multitrama.



In seguito è stata introdotta una modifica del sistema T1 (extended framinf format), che prevede il raggruppamento di 24 trame consecutive per formare una supertrama, qui i bit F delle trame dispari sono usati per le segnalazioni mentre quelli nelle trame pari non multipli di 4 sono usate per il controllo dell'errore, i bit F delle trame pari che sono invece multipli di 4 realizzano l'allineamento di supertrama.
A partire dal flusso T1 (1.544Mbps), mediante multiplatore M12, si genera il flusso T2 mediante multiplazione di 4 flussi T1. Il multiplatore di tale livello aggiunge 17 bit: (193·4)+17=789 bit (essendoci 8000 trame al secondo) il segnale T2 ha quindi una frequenza di cifra di 6.312Mbps.
Il multiplatore M23 accorpa invece 7 segnali T2 (del livello gerarchico sottostante) ed aggiunge 69 bit: (789·7)+69=5592 bit (essendoci 8000 trame al secondo) il segnale T3 ha una frequenza di cifra di 44.736 Mbps.
Infine, il multiplatore M34 accorpa 6 segnali T3 ed aggiunge 720 bit: (5592·6)+720=34272 bit (essendoci 8000 trame al secondo) il segnale T4 ha una frequenza di cifra di 274.176 Mbps.
La gerarchia plesiocrona si è diffusa per prima in America e successivamente in Europa. I sistemi anche se tra loro incompatibili coesistono grazie ad appositi convertitori.
Il primo livello della gerarchia plesiocrona, seconda la normativa europea, ha una frequenza di cifra standardizzata di 2.048 Mbps. Anche le frequenze di cifra del secondo, terzo e quarto livello sono standardizzate: il secondo livello multipla 4 tributari di primo livello (8.448 Mbps), il terzo livello affascia 4 tributari di secondo livello (34.368 Mbps), il quarto livello affascia altrettanti tributari di livello inferiore (139.264 Mbps) (lo schema è quindi 4-4-4). In Europa si affasciano 32 time slot per la

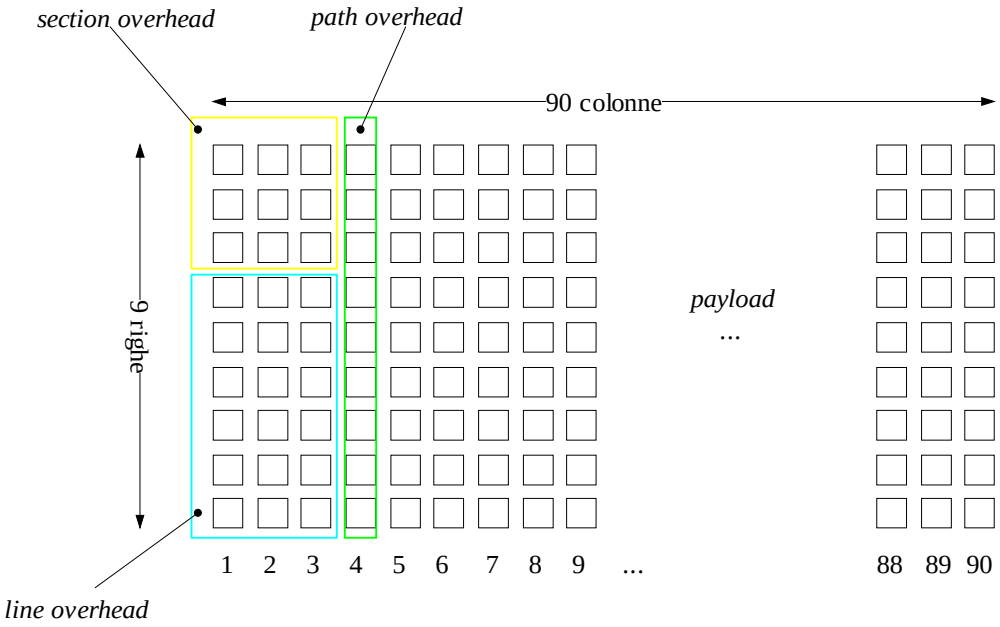
formazione della trama contro i 24 usati nello schema americano. Di questi 32 time slot, 2 sono interamente dedicati alla sincronizzazione (prima trama) ed alla segnalazione (trama numero 17). Quindi complessivamente si hanno a disposizione 30 time slot per ogni tributario, un time slot è lungo 8 bit: 32·8=256 bit (essendoci 8000 trame al secondo) la frequenza di cifra del segnale T1 europeo è di 2.048 Mbps. La ricerca per l'instradamento viene effettuata sezione per sezone, la centrale locale instrada il flusso verso una prima centrale di transito e poi si prosegue l'instradamento finchè non si giunge a destinazione. La commutazione avviene su ciruito fisico preventivamente scovato e per instaurare la connessione il chiamante invia informazioni al chiamato.

Gerarchia numerica sincrona

La gerarchia SDH (synchronous digital hierarchy) e la gerarchia SONET (synchronous optical network) sono gerarchie di multiplazione definite negli anni 80' e che vengono adoperate in corrispondenza di collegamenti fisici costituiti da fibre ottiche. I flussi sono multiplati da entrambe le gerarchie in maniera sincrona. Nella gerarchia plesiocrona è emerso un importante punto critico: è richiesta l'operazione di demultiplazione completa per l'accesso anche al singolo tributario.
Mentre nei sistemi asincroni, per ottenere flussi ad elevate frequenze di cifra, il sistema deve multiplare man mano i segnali secondo una opportuna gerarchia, nei sistemi SDH e SONET la multiplazione è realizzata in un singolo passo grazie alla sincronia della rete stessa. L'estrazione di un solo tributario, poi, non richiede la demultiplazione dell'intero segnale multiplato.
SONET/SDH ha un numero di vantaggi considerevoli ed è adottato come standard mondiale nella trasmissione ad alto bit-rate. Esso è stato sviluppato a partire dagli anni 80' con una serie di studi svolti agli AT&T Bell labs e poi in tutti i laboratori europei e giapponesi. Il progetto di rete da cui prende spunto si chiamava METROBUS. Il clock viene distribuito, utilizzando la rete satellitare GPS, alle synchronous supply unit poste nei nodi principali della rete e da qui, tramite operazioni di slaves ai nodi secondari. Gli orologi di riferimento sono orologi al cesio, idrogeno o rubidio ed hanno un accuratezza da 10^{-11} a 10^{-13} (in realtà a causa di alcuni fenomeni l'accuratezza complessiva è inferiore a quella originaria ma è comunque buona).
La multiplazione SONET/SDH è basata sul concetto di organizzare in uno spazio complesso le informazioni da trasportare dalla rete (payload) e le informazioni OAMP (operation, administration, maintenance e provisioning). SONET/SDH è organizzato secondo frame di lunghezza temporale di 125 μs (ovvero 8000 caratteri al secondo). Se quindi inserisco un carattere ogni trama, 8000 volte al secondo, sento un segnale audio da 4 Kbyte.
La trama STS-1 (quella di SONET) è organizzata in una struttura avente 90 colonne e 9 righe, ogni cella è di 8 bit, in totale ci sono: (90·9·8)=6480 bit trasmessi in 125 μs per un flusso avente frequenza di cifra di 51.84 Mbps. La trama STM-1 (quella di SDH) utilizza invece 3 piani sovrapposti della suddetta struttura., la frequenza di cifra (adesso triplicata) è quindi di 155.52Mbps. La struttura si presenta suddivisa in 4 aree principali: il section overhead, il line overhead, il path overhead ed il payload. Quest'ultimo occupa solo una parte dell'intero frame, più precisamente 86·9 byte.
Il frame SONET viene riempito mappando su di esso i payloads client attraverso un organizzazione a tributario chiamata virtual tributaries o virtula containers (VT/VC). I VT sono a loro volta divisi in 7 higher level che possono quindi ospitare diveri lower VT, più esattamente:

- 4 VT 1.5 ciascuno da 27 bytes = 108 bytes;
- 3 VT 2 ciascuno da 36 bytes = 108 bytes;
- 2 VT 3 ciascuno da 54 bytes = 108 bytes;
- 1 VT 6 ciascuno da 108 bytes = 108 bytes;

Quindi ogni VT contiene 108 bytes organizzati secondo 12 colonne di 9 righe, 108 bytes. Ora, 12 colonne per 7 tributari fanno 84 colonne, 3 in meno delle 87 destinate al payload. Di queste 3, una come abbiamo visto serve come overhead del payload stesso e 2 sono tenute di riserva per eventuali problemi di sincronizzazione dei payload.



Le operazioni di aggregazione e preparazione del traffico da inserire nella trama SONET sono eseguite con opportune regole dagli apparati service adapter, sviluppati dalle aziende che preparano le intefacce SONET. Un altro concetto importante nell'organizzazione di SONET/SDH è rappresentatao dal pointer. Può succedere infatti che per motivi di protocollo o di efficienza di trasporto il payload non sia dentro il singolo frame di SONET. In questo caso esso inizia da un punto qualsiasi del payload per continuare poi nel frame successivo. Il pointer è un numero contenuto nel line overhead che indica a che punto (numero di cella) del payload cercare l'inizio del payload stesso. L'impiego del pointer è molto importante anche perchè permette di ospitare a bordo di SONET payload floating, cioè fluttuanti rispetto alla trama stessa. Le aree riservate all'interno della trama rispecchiano la configurazione di SONET che presenta alla stessa maniera 3 tipi di componenti.
Il path terminal equipment effettua la multiplazione in cui il payload viene mappato. Il line terminal equipment è un hub che provvede anch'esso alla multiplazione dei segnali SONET. I section terminal equipment sono dei rigeneratori di linea che eseguono le funzioni di allineamento e controllo di errore.

Reti di calcolatori

Quando i calcolatori elettronici fecero la loro comparsa sul mercato ad essi veniva prevalentemente assegnato il compito di elaborare i dati secondo una modalità molto semplice, essi infatti furono inizialmente concepiti per operare in maniera isolata. Lo scambio dei dati fra due calcolatori (qualora fosse stato necessario) avveniva tramite supporti di memorizzazione e ciò implicava lo spostamento di personale qualificato, oppure della persona interessata, presso l'edificio in cui era collocato il calcolatore. I primi supporti di memorizzazione, oltre ad essere delicati, non erano particolarmente capienti. La ricerca verso supporti di memorizzazione più capienti e dalle dimensioni ridotte ha solo posticipato nel tempo l'esigenza ormai sentita da più utenti, ossia lo scambio di informazioni con altri calcolatori di altri utenti.

L’informazione, così come è avvenuto in altri settori, viene affidata al campo elettromagnetico e ciò costituisce in molti casi un notevole miglioramento poichè riduce i tempi utili al trasposrto dell’informazione stessa (nel caso in cui non è necessario lo spostamento fisico di persone). La realizzazione di una rete di telecomunicazioni per calcolatori costituisce pertanto una moderna soluzione al problema dello scambio dell’informazione. Ethernet è il modello di rete locale più diffuso che consente di collegare diversi calcolatori distribuiti in area locale (sia essa un edificio oppure un complesso di edifici appartenenti allo stesso comprensorio).
Questo sistema favorisce quindi lo scambio diretto di dati in formato elettronico tra tre o più calcolatori senza ricorrere al passaggio di supporti di memorizzazione. Infatti, il numero di calcolatori coinvolti nella connessione deve essere almeno tre poiché se i calcolatori fossero soltanto due non si potrebbe più parlare di rete, ma bensì di collegamento diretto da punto a punto, come ad esempio quello che si crea quando si usano particolari tipi di cavo seriale o parallelo per trasferire dati da un calcolatore ad un notebook.
L’informazione è memorizzata presso i calcolatori sotto forma numerica, tuttavia ogni calcolatore adotta un proprio metodo per la rappresentazione ed elaborazione delle informazioni. Alcuni calcolatori ad esempio, disponendo di maggiori capacità elaborative e di memorizzazione, sono specializzati in particolari funzioni e pertanto si rivolgono a precisi scopi di utilizzazione, altri invece si dicono di tipo general purpose poichè più generici e semplici nell’architettura interna. Dunque i nodi terminali della rete possono essere diversi tra loro ma ad ogni modo dotati di capacità elaborative e memorizzative che ne caratterizzano inevitabilmente le prestazioni. Le stesse funzioni elaborative sono inoltre molto complesse e richiedono un certo grado di competenza. Non vi è alcun dubbio che i calcolatori, a differenze di altri terminali di altre reti (ad esempio terminali telefonici, televisivi etc...), offrono maggiori potenzialità al punto tale da attrarre sempre più utenti.
Ethernet definisce un area locale (detta LAN, Local Area Network) in cui è consentito colloquiare liberamente con qualsiasi calcolatore collegato e di trasmettere la stessa informazione contemporaneamente a tutte le macchine in ascolto (modalità broadcasting). Essa non costituisce la migliore tecnologia per lo scambio dati ma nel corso degli anni si è affermata sempre più al punto da innescare un economia di scala che ha conenstito di realizzare interfacce di rete sempre più economiche.

La sua storia ha inizio nei primi anni ‘70 presso il Palo Alto Research Center (PARC), il laboratorio di ricerca di Xerox, per opera di Robert Metcalfe e David Bloggs. Il lavoro iniziò intorno al 1972, ma la sua prima definizione pubblica risale a un articolo pubblicato nel 1976 con la firma dei due inventori. Il nome, ideato e registrato da Xerox, suggerisce l'idea dell'etere, cioè di quella sostanza immateriale che in passato si supponeva pervadesse tutta l'aria e consentisse il propagarsi della luce. Così intesa, verrebbe la tentazione di pensare a una rete che usa onde radio elettromagnetiche per la distribuzione delle informazioni, quando invece è sempre necessario un cavo in rame oppure in fibra ottica per convogliare i segnali. Xerox, come è accaduto successivamente per altre invenzioni sviluppate nei suoi laboratori californiani, non ebbe l'intraprendenza di trasformarla immediatamente in un prodotto commerciale e dobbiamo aspettare il dicembre del 1980 per averne la prima versione utilizzabile, dovuta all'iniziativa congiunta di Xerox, Digital Equipment e Intel. Nel 1982 lo standard iniziale fu sostituito dalla versione 2.0, detta anche Ethernet II oppure DIX (Digital, Intel, Xerox) che costituisce ancora oggi uno standard di riferimento per numerosi impianti. Il passaggio finale fu affidarne la standardizzazione a un ente al di sopra delle parti. Considerando le potenzialità di diffusione mondiale, Ethernet non poteva restare affidata nelle mani di tre società private. Tutti gli altri produttori non avrebbero investito in una tecnologia che sfuggisse al loro controllo. Il ruolo di arbitro fu affidato all'Institute of Electrical and Electronics Engineers (IEEE), un ente statunitense con sede a New York che riunisce scienziati, ingegneri e studenti e che nella prima metà degli anni ‘80 creò un

comitato, identificato dal numero 802, il cui compito è di codificare tutti i tipi primari di rete locale, incluso naturalmente Ethernet. La sua prima formulazione ufficiale risale al 1983 con la pubblicazione del documento IEEE 802.3 in cui si definiscono le specifiche elettriche e fisiche per una rete Ethernet a 10 Mbit/s su cavo coassiale. Successivamente il documento è stato perfezionato a più riprese, cominciando dal 1985 con la definizione del metodo di accesso e proseguendo, poi, con l'aggiunta di versioni capaci di funzionare anche su cavi di tipo differente e a velocità diverse (10 Mbit/s, 100Mbit/s, 1 Gbit/s e 10Gbit/s).

Una LAN Ethernet può avere una topologia a bus comune (è questa la soluzione più usata) oppure una topologia a stella e può funzionare su cavo coassiale, su doppino telefonico o fibra ottica. Il cavo viene steso in maniera tale da collegare le stazioni di lavoro, ognuna di queste riceve tutto quello che attraversa la rete, la possibilità di trasmettere è data ad una sola stazione. L'informazione viaggia nella rete organizzata in una struttura detta trama al cui interno prendono posto, oltre ai dati che costituiscoo l'effettivo payload, l'indirizzo sorgente e quello di destinazione. In questo modo, un nodo terminale che riceve una trama ha la possibilità di accogliere il messaggio qualora questo sia effettivamente indirizzato a lui, oppure può scartare quest'ultimo in caso contrario. Il meccanismo appena descritto facilità poi l'inserimento di nuovi nodi terminali nella rete, essi infatti ricevono tutti i pacchetti in transito nella rete ed acquisiscono la facoltà di trasmettere quando questa è libera.
Esistono diversi standard Ethernet, tutti raggruppati sotto la sigla 802.x e che si differenziano tra di loro per le velocità raggiunte in fase di trasmissione. Tuttavia, ogni standard mantiene intatta la struttura organizzativa dei dati favorendo quindi l'interazione anche con i nodi terminali più lenti.
Ogni scheda di rete dispone di un indirizzo permanente espresso in numeri esadecimali e lungo 48 bit. I primi 24 bit di tale indirizzo indicano il costruttore dell'interfaccia, i restanti 24 bit sono invece assegnati dal costruttore a ciascuna interfaccia prodotta. Pertanto, qualunque sia la topologia fisica della rete e qualunque sia la velocità dell'interfaccia di rete, la tecnica trasmissiva rimane invariata e consiste nel trasmettere un segnale prossimo ad un onda quadra che oscilla tra due valori di tensione, uno positivo ed uno negativo cosicchè ad ogni transizione da negativo a positivo e viceversa è possibile indicare l'occorrenza di una cifra binaria e quindi di 1 oppure 0. Il sistema appena descritto prende il nome di codifica Manchester ed ha il notevole vantaggio di non confondere i valori logici poichè per essi si misurano le inversioni di polarità anzichè le ampiezze degli impulsi facilmente sensibili a disturbi.
Le sigle adottate dalle versioni di Ethernet forniscono informazioni su come funziona la rete, ad esempio la versione di Ethernet 10BASE-5 indica che la rete funziona in trasmissione a 10Mbit/s e può sopportare un segmento di rete lungo 500 metri. Il termine BASE è invece l'abbreviazione del termine banda base (baseband) ed indica una trasmissione in banda base, secondo codifica Manchester, con una portante di 20MHz. Ogni stazione stende un cavo per la connessione che va dalla scheda al cavo di rete. Una primissima soluzione adottata per questo tipo di allacciamento furono le prese a vampiro. Queste da un lato si agganciavano comodamente alla scheda di rete mediante un connettore che lo fissava ad una slitta successivamente avvitata su se stessa, mentre all'estremo opposto un secondo connettore si agganciava al cavo di rete perforandone la calza esterna e toccando quindi l'anima interna del cavo.
Al troncone di cavo che quindi si distende nell'edificio e collega i nodi terminali va poi aggiunto un tappo di terminazione, tipicamente un resistore che scarica qualsiasi segnale in arrivo affinché non si rifletta all'indietro e non vada a collidere con gli altri impulsi trasmessi.
In una successiva versione di Ethernet, la 10BASE-2 (quindi 10Mbit/s in trasmissione e 200 metri di cavo per il bus comune) fu introdotto un diverso connettore per favorire l'allacciamento a nuovi terminali, si tratta del connettore a T. Una estremità del connettore a T era collegata direttamente all'interfaccia di rete, il cavo in ingresso e quello in uscita dalla stazione erano poi allacciati al

connettore a T. La rete non è composta da un singolo spezzone di cavo, ma da tanti cavi concatenati. Il punto in cui lo spezzone incontra il successivo coincide con il punto in cui un nodo si collega alla rete. Se da un lato questa soluzione economica abbatte di molto i costi iniziali è opportuno ricordare una importante vulnerabilità di Ethernet: qualunque interruzione provocata nei pressi di una stazione terminale provoca la caduta dell'intera rete. Esistono così dei particolare tester che applicati ad una rete guasta ne indicano il punto di rottura, più precisamente essi forniscono una stima della distanza espressa in metri a partire dal punto di applicazione della sonda fino al punto di rottura. Siccome il cavo attraversa l'edificio curvandosi in più punti è allora opportuno conoscere con esatteza la lunghezza dei singoli tratti che compongono la rete in maniera tale da ricavarne approsimativamente il luogo che ospita la stazione interessata dal guasto.
Le recenti versioni di Ethernet si sono concentrate su altri mezzi trasmissivi che oggi giorno risultano particolarmente di moda: il doppino telefonico (10BASE-T) e la fibra ottica (10BASE-FP).
In queste ultime due soluzioni la topologia elettrica è a bus mentre la topologia fisica è a stella. Questo significa che nello stendere i cavi all'interno dell'edificio si segue una topologia a stella: tutte le connessioni di un certo gruppo confluiscono in un singolo punto dove vengono collegate a un concentratore, detto hub di rete. Il concentratore funziona anche da ripetitore (esso rigenera il segnale che si attenua man mano che questo percorre la rete), tuttavia fa in modo che un raggio di questa stella sia elettricamente il prolungamento dell'altro e quindi operi come se fosse un troncone di cavo coassiale ininterrotto. Il fatto di portare tutte le connessioni verso un singolo punto, oltre all'economicità del doppino, comporta due vantaggi importanti: è possibile allestire in anticipo diverse prese in punti uniformemente distribuiti nel locale, senza doverle attivare tutte immediatamente (basta non collegare al concentratore quei rami che sono temporaneamente inattivi); inoltre qualsiasi ramo difettoso viene automaticamente escluso senza influire sul funzionamento del resto della rete. Gli hub di rete in commercio dispongono di un numero minimo di porte (solitamente almeno 4 per i modelli economici) mentre alcuni modelli sono modulari, è sempre possibile cioè aggiungere nuove porte al dispositivo. Il numero di porte del concentratore di rete dipende dall'estensione della rete che si vuole allestire. E' tuttavia preferibile dotare il concentratore di rete di un numero di porte maggiore di quelle necessarie in maniera tale da garantire la connesione ai futuri nodi terminali che si aggiungeranno.
Supponiamo che l'adattore Ethernet A con indirizzo fisico AA-AA-AA-AA-AA-AA abbia la necessità di inviare dati all'adattatore Ethernet B con indirizzo fisico BB-BB-BB-BB-BB-BB. In una rete Ethernet non esiste un nodo master che regola i diritti di trasmissione, tutti i nodi sono autodisciplinati. Ogni nodo, infatti, si astiene dalla trasmissione qualora un altro nodo della rete impegna il canale in una conversazione. Dunque, la prima operazione che l'interfaccia A deve intraprendere è l'ascolto del canale. Se il canale è libero da altre conversazioni la trasmissione dell'informazione verso l'interfaccia B può avere inizio, in caso contrario l'interfaccia A dovrà necessariamente attenderne la fine. La tecnica di trasmissione usata da Ethernet è nota con la sigla CSMA/CD (carrier sense multiple acces with collisione detect). La prima operazione che l'interfaccia svolge è dunque la rilevazione della portante (carrier sense). Tuttavia, siccome l'accesso è multiplo, cioè comune a tutti i nodi della rete, la scelta di iniziare una trasmissione può avvenire in concomitanza con un altro nodo terminale (magari più distante dall'interfaccia A che per questo motivo non si accorge dell'inizio della trasmissione), quando ciò avviene si dice che si verifica una collisione. I flussi informativi si sommano e perdono quindi il loro significato.
Se non esistesse nessun dispositivo in grado di rilevare una collisione, l'interfaccia del nodo terminale A e quella che precedentemente ha iniziato a trasmettere in concomitanza con questa continuerebbero a trasmettere nella rete i bit che compongono i rispettivi messaggi nella totale convinzione che questi giungano ai nodi di destinazione privi di errore.
Per tale motivo si è deciso di dotare l'interfaccia di rete di un interessante circuito il cui compito è appunto la rilevazione delle collisioni (collison detect). Il circuito è inoltre abbastanza semplice: in caso

di collisione i segnali elettrici delle due stazioni oltre a mescolarsi finiscono anche per sommarsi, per questo motivo la tensione risultante che circola nella rete supera i valori di tensione assegnati ai livelli logici. Quando la collisione viene rilevata dall'interfaccia di rete, entrambe le stazioni coinvolte nella collisione non interrompono immediatamente la trasmissione. Esse continuano a inviare bit fino al raggiungimento di 64 byte. Ciò favorisce la rilevazione della collisione anche alle altre interfacce di rete che in questo modo si accorgono dell'attuale stato della rete.
L'algoritmo di backoff fornisce poi il tempo di attesa alla ritrasmissione, ogni nodo coinvolto nella collisione ha il proprio timer cosicchè essi non sceglieranno lo stesso tempo di attesa. Anche a seguito di questo accorgimento possono ancora verificarsi delle collisioni, questa volta con uno dei nodi che riprende ad esempio la trasmissione ed uno invece che aspettava che il canale si liberasse. Ad ogni nuova collisione l'intervallo di attesa viene incrementato, dopo un certo numero di collisioni consecutive l'entità preposta alla trasmissione comunica agli strati superiori la sua incapacità nel portare a termine la comunicazione. L'algoritmo di backoff fissa il tempo di attesa pari a $2r\tau$ dove τ è il massimo ritardo che intercorre fra tutte le possibili coppie di nodi terminali ed r è un numero intero uniformamente distribuito fra 0 e 2min(k,10) con k numero di tentativi di ritrasmissione:

Prima collisione	$(k=0, p=0.5)(k=1, p=0.5)$
Seconda collisione	$(k=0, p=1/4)(k=1, p=1/4)(k=2, p=1/4)(k=3, p=1/4)$
Terza collisione	$(k=0, p=1/8)(k=1, p=1/8)(k=2, p=1/8)(k=3, p=1/8)$ $(k=4, p=1/8)(k=5, p=1/8)(k=6, p=1/8)(k=7, p=1/8)$

...

Nella realtà le collisioni sono più frequenti di quello che a prima vista potrebbe sembrare. Infatti, oltre al caso fortuito visto prima di due stazioni che trasmettono esattamente nello stesso momento, esistono anche altri casi in cui due o più macchine cercano di prendere possesso della linea con la convinzione che sia libera, quando questa in realtà non lo è e c'è già qualcun altro che ha cominciato a trasmettere.
Per capire come questo possa accadere occorre analizzare i tempi impiegati nella trasmissione: alla velocità di 10 Mbit/s ci vogliono 100 nanosecondi per inviare un singolo bit. La trasmissione non è dunque istantanea e si verifica quello che in termini tecnici si chiama ritardo di propagazione.
Ci vuole circa un nanosecondo per percorrere 30 centimetri e, prima che il secondo bit sia uscito dalla scheda di rete che sta trasmettendo, il primo bit ha circa trenta metri di vantaggio. Le reti Ethernet hanno lunghezze di centinaia di metri perciò può benissimo accadere che una seconda stazione, diciamo a 90 metri di distanza dalla prima, ascolti la linea nel momento in cui la prima ha iniziato a trasmettere e la trovi comunque libera, visto che il primo bit non è ancora arrivato fino a lei. In tal caso la seconda stazione inizierebbe la propria trasmissione e quasi subito si troverebbe coinvolta in una collisione. Anzi, anche una terza stazione, ancora più distante, potrebbe partire nel frattempo e provocare un vero e proprio tamponamento a catena. Questo ci fa capire per quale motivo, al crescere del numero di stazioni presenti sulla rete, aumenti anche il numero di collisioni e ci spiega anche perché una rete Ethernet non possa superare una certa lunghezza. Il problema viene ulteriormente complicato dal fatto che, mentre la seconda e la terza stazione si accorgono della collisione quasi immediatamente, la prima non se ne rende conto fino a quando il segnale di collisione rimbalza indietro lungo la rete e ritorna fino a lei.
Quindi si aggiungono ulteriori tempi morti perché, come abbiamo visto prima, bisogna continuare a trasmettere almeno 64 Byte anche in caso di collisione, così da far proseguire la collisione abbastanza a

lungo da consentire a tutte le stazioni coinvolte di accorgersene. La quantità di byte da trasmettere è legata al tempo che il segnale elettrico impiega per completare un viaggio di andata e ritorno (round trip time) sull'intera rete. Per l'Ethernet a 10 Mbp/s le specifiche dicono che, qualunque sia il tipo di cavo utilizzato, un singolo bit non deve impiegare più di 50 microsecondi per coprire l'intera lunghezza della rete nei due sensi, il che equivale a trasmettere 500 bit, cioè 62,5 Byte, arrotondati a 64.
Da questi parametri di partenza deriva una serie di vincoli di lunghezza del cavo, di numero massimo delle stazioni per tratta di cavo e di numero massimo di ripetitori. Per estendere il limite della rete oltre il valore di 50 microsecondi, per l'andata e ritorno, è necessario creare una seconda rete e collegarla alla prima attraverso un dispositivo di rete chiamato bridge che memorizza ogni messaggio in arrivo da una rete e lo ritrasmette alla rete successiva solo se è destinato a questa, oppure lo scarta se si tratta di un messaggio che deve rimanere all'interno della prima rete.
Il bridge è il precursore di un altro dispositivo, lo switch. Le trame sono allora analizzate dal bridge che basandosi sull'indirizzo di destinazione del messaggio stabilisce se esso è indirizzato ad uno dei nodi terminali di una rete collegato ad una delle porte di ingresso/uscita del bridge. A tale proposito il bridge si serve di una apposita tabella, ad ogni nodo terminale collegato al dispositivo corrisponde una riga che lo associa alla porta di ingresso/uscita impegnata. Il compito di uno bridge è poi quello di bloccare il traffico di una rete affinchè questo non inondi l'altro segmento collegato. Quando il bridge riceve una trama su una delle proprie porte o interfacce, ha diverse alternative.

Se l'indirizzo di destinazione appartiene a una macchina che si trova sullo stesso segmento da cui la trama arriva, il bridge scarta la trama poichè questa è destinata a rimanere all'interno della rete dove troverà quindi il destinatario;
Se l'indirizzo di destinazione appartiene a una macchina che si trova su di un altro segmento da cui arriva la trama, il bridge analizza la tabella d'inoltro per ricavare il segmento in uscita;
Se l'indirizzo di destinazione non compare nella tabella di inoltro la trama viene spedita a tutti segmenti di rete collegati al bridge, con la sola ecezione di quello da cui è arrivata;

Lo switch, essendo l'evoluzione dello bridge, realizza le medesime funzioni in hardware anzichè in software. Questo non rallenta il transito dei pacchetti che possono passare alla massima velocità (wire speed). Nella realtà un rallentamento esiste sempre ed è caratterizzato dalle modalità di funzionamento dello stesso switch. Una prima tecnica di switching va sotto il nome di store and forward: ogni trama in arrivo è memorizzata in un buffer di memoria dove viene analizzata per intraprendere una successiva azione di instradamento. La trama, prima di essere ritrasmessa verso destinazione, deve arrivare per intero, e solitamente prima che ciò avvenga si procede ad accertarne la correttezza mediante funzione di controllo di errore (molti dispositivi di rete realizzano facilmente in hardware il controllo di ridondanza ciclica CRC sul pacchetto ricevuto).
Una seconda tecnica è invece il cut through, non appena una parte del pacchetto giunge ad uno switch questo ne analizza immediatamente l'indirizzo di destinazione e lo rilancia in uscita ancora prima che l'intero pacchetto è giunto per intero. Nel frame Ethernet uno bridge oppure uno switch trova le informazioni necessarie all'instradamento del messaggio. Il frame Ethernet si compone di 6 campi, in essi troviamo:

- Un preambolo di 8 byte ha il compito di svegliare il dispositivo di rete: i primi 7 byte hanno valore 10101010, l'ultimo byte è invece 10101011. I primi byte sveliano l'adattatore e sono utili alla sincronizzazione degli orologi tra sender e receiver. L'ultimo byte indica invece che il successivo gruppo di byte indica l'indirizzo di destinazione;
- Un campo per l'indirizzo di destinazione di 6 byte;
- Un campo per l'indirizzo del mittente 6 byte;

- Un campo di 2 byte per indicare il protocollo di strato superiore;
- Un campo dati variabile da un minimo di 40 ad un massimo di 1500 byte;
- Un campo di 4 byte destinato al controllo di errore (controllo CRC);

Reti locali wireless
Le reti locali wireless sono definite nello standard 802.11 e solo di recente hanno fatto la loro comparsa dispositivi in grado di metterne in atto i principi. In realtà lo standard 802.11 raccoglie diverse tecnologie, di queste ci piace citare le più importanti:

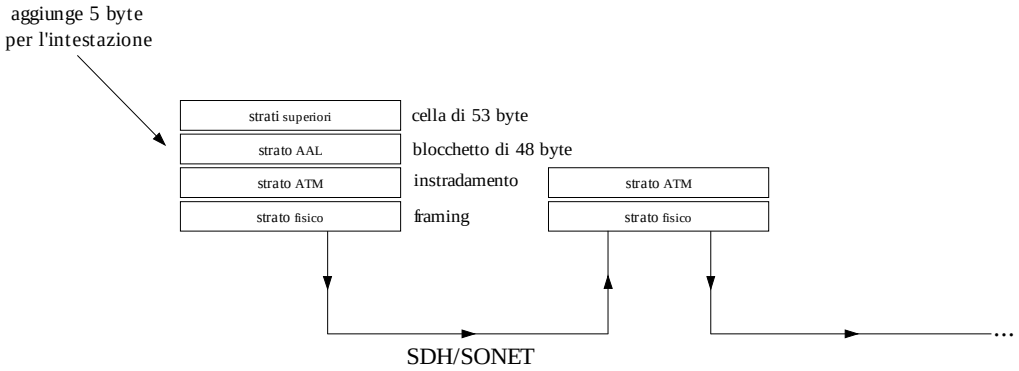
Standard IEEE 802.11b: opera tra i 2.4 ed i 2.483 Ghz della vecchia banda ISM prima destinata per scopi ingegneristici, scientifici e medici e adesso libera. La massima velocità raggiungibile da questo standard è di 11 Mbps in trasmissione. Ogni paese adotta un diverso valore per la massima potenza trasmissibile, in Europa ad esempio tale valore è di 100mW mentre in America è di 1W. Ciò incide molto sulla superficie coperta dal segnale;
Standard IEEE 802.11a: opera a 5GHz in una banda più ampia di quella dello 802.11b e permette di raggiungere una velocità di trasmissione pari a 54 Mbps. Attualmente è assai poco diffusa, se non negli USA, e gli apparati non sono spesso compatibili con l'802.11b, limitando in tal modo la sua diffusione, nonostante una velocità di trasmissione assai maggiore;
Standard IEEE 802.11g: è il nuovo standard che estende le caratteristiche di trasmissione dello standard 802.11b, portando la velocità di trasmissione a 54 Mbps, pur operando a 2,4GHz e garantendo compatibilità con gli apparati di questo standard.

La rete locale è suddivisa in celle dette BSS (basic service set), ogni cella è dotata quindi di un punto di accesso AP (access point). All'AP si rivolgono tutti i dispositivi collegati alla rete senza fili. Ogni terminale che intende interfacciarsi con la rete dispone di un NIC (network interface connection). Il dispositivo NIC opera da interfaccia tra il dispositivo mobile e la rete a radiofrequenza svolgendo le funzioni di accesso alla rete, organizzazione dei dati in pacchetti e di modulazione/demodulazione delle sequenze di bit.
Poichè una cella BSS può risultare insufficiente a coprire l'intera area locale si realizza allora una rete di celle BSS e di AP. Gli AP sono poi tra loro collegati per formare un impianto di distribuzione. Quest'ultimo può essere realizzato mediante rete Ethernet oppure tramite un altra rete wireless. Nello standard 802.11 viene adottato un protocollo di accesso multipllo detto CSMA/CA (carrier sense multiple with collision avoid) che evita le collisioni piuttosto che rilevarle semplicemente. La stazione che ha dati da trasmettere resta in ascolto del mezzo trasmissivo e se questo è occupato attende ritardando la trasmissione. Se il canale è invece libero per un tempo sufficiente alla trasmissione esso allora viene impiegato dalla stazione in attesa.
Siccome è possibile che anche un altra stazione possa trasmettere contemporaneamente, il ricevitore controlla il pacchetto appena ricevuto e trasmette al mittente un messaggio di conferma. Il messaggio di conferma indica al trasmettitore che non si è verificata alcuna collisione ed informa tutti gli altri nodi terminali della rete sul tempo necessario al completamento dell'attuale trasmissione in corso. Se il trasmettitore non riceve il messaggio di conferma ripete la trasmissione un certo numero di volte e solo dopo alcuni tentativi consecutivi non andati a buon fine comunica all'entità di strato superiore la sua incapacità di svolgere il compito assegnatogli.
Questo sistema soffre tuttavia di un problema noto come il problema del terminale nascosto. Per comprenderlo supponiamo di avere un area locale partzionata da tre celle BSS, ognuna di queste è quindi dotata di un AP alla quale si rivolgono i terminali che si trovano in prossimità delle aree coperte. Supponiamo di chiamare le tre area locali coperte con le lettere A,B e C ed i terminali che si trovano nelle suddette aree con le lettere a,b e c. Le tre celle BSS condividono inoltre alcune aree di copertura, i

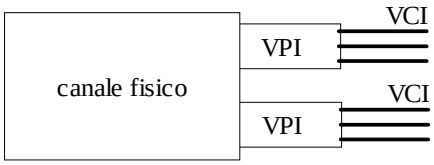
terminali b e c si trovano rispettivamente nella zona di copertura delle celle B e C. Il terminale A invece si trova in un area locale coperta dalla cella BBS A e dalla cella C.
Il terminale b trovando il canale disponibile inizia una trasmissione dati diretta al terminale c e si rivolge al suo AP di copertura. Quest'ultimo osserva la trama ed inoltra il messaggio alla cella BSS C che copre la zona in cui si trova il terminale c. Il messaggio è tuttavia ricevuto oltre che dal terminale c anche dal terminale a che si trova in una zona a doppia copertura. Per scongiurare quindi che il terminale nascosto a intraprenda anch'esso un azione di trasmissione, il terminale c invia al terminale b un messaggio di avvenuta ricezione del messaggo. Tale messaggio viene per primo rivolto all'AP dell'area locale C ed è pertanto ricevuto anche dal terminale nascosto a. Il terminale c indica quindi nel messaggio di ack verso il destinatario il tempo che questo impiegherà per portare a termine la trasmissione. Tale informazione è così percepita anche dai terminali nascosti (nel nostro caso da a) che in qusto modo evitano la trasmissione aspettando che il canale si liberi.

Le reti ATM

ATM, asynchronous transfer mode, indica il funzionamento di una rete numerica magliata in cui l'informazione è trasmessa in blocchi di lunghezza fissa detti celle, multilati con la tecnica della multiplazione a divisione di tempo. L'attributo "asincrono" si riferisce alla modalità di allocazione della capacità trasmissimiva del mezzo e non alla tecnica di trasmissione a livello fisico che invece è sincrona.
L'architettura di una rete ATM è organizzata secondo una struttura a quattro strati: strato fisico, strato ATM, strato AAL (adaptation layer) e strati superiori. Gli strati superiori forniscono l'informazione da trasferire allo strato sottostante AAL in blocchetti di 48 byte. Lo strato AAL aggiunge ad ogni blocchetto 5 byte per l'intestazione formando la cosiddetta cella che è passata, quindi, allo strato sottostante ATM.
Allo strato fisico compete la funzione di framing, mentre allo stato ATM è affidata la funzione di instradamento che avviene secondo una modalità orientata alla connessione. Allo strato ATM è poi assegnata la funzione di controllo dell'errore basato sui soli 5 byte di intestazione. Una funzione di errore sui byte di carico è poi svolta dagli strati superiori, in questo modo viene ridotto il carico per ciascun nodo di transito che svolgeranno così le sole funzioni di instradamento.



La tecnica dell'instradamento orientata alla connessione permette di ridurre i ritardi di transito in ciascun nodo, ciascuna cella contiene nel proprio header l'informazione necessaria al percorso preventivamente individuato (ciò favorisce l'instradamento). Il percorso virtuale è identificato da una stringa di bit che è aggiunta nell'header ed è detta VPI. Ogni percorso virtuale è caratterizzato da una proprio frequenza di cifra e da un certo numero di canali virtuali che può ospitare. Ogni canale virtuale è anch'esso identificato da una stringa di bit detta VCI.



Strato fisico

Lo strato fisico prevede molte varianti sia per quanto riguarda la frequenza di cifra che per il mezzo fisico e l'eventuale gerarchia trasmissiva usata. Allo stato attuale degli standard si possono avere interfacce funzionanti con frequenze di cifra comprese fra 2 e 622 Mbit/s, utilizzanti come mezzo fisico fibra ottica, cavo coassiale o doppino telefonico.

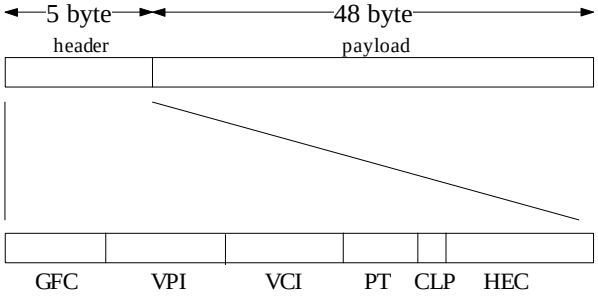
Allo strato fisico compete la funzione di framing, le celle vengono trasmesse dallo strato fisico avvalendosi della organizzazione offerta dal mezzo trasmissivo. Nella maggior parte dei casi si tratta di SDH o SONET. Le celle possono essere trasmesse secondo due modalità:

- modalità basata su SDH, si utilizzano le strutture informative e quindi le funzioni di controllo definite nell'ambito dei sistemi SDH;
- modalità basata su celle, non si fa ricorso ad una struttura specifica della trama che viene così definita caso per caso.

Un esempio della modalità basata su SDH è la trasmissione di flussi di 155.52 Mbps e 622.68 Mbps. Il flusso ATM è allora messo in un contenitore C4 la cui capacità non è però un multiplo intero della lunghezza della cella. Pertanto, le celle non occupano posizioni fisse e possono anche trovarsi spezzate in due contenitori consecutivi di tipo C4. La modalità basata su celle prevede che ogni 26 celle consecutive venga emessa una cella libera in modo da avere un flusso coincidente con quello che il contenitore C4 è in grado di trasportare. Anche il trasporto di flussi a 1.544 e 2.048 Mbps viene previsto qualora si usino i sistemi trasmissivi della gerarchia plesiocrona. Quando ciò avviene si hanno a disposizione 30 celle (2 sono destinate alla segnalazione ed al controllo di errore) e può accadere che ogni singola cella non sia sufficientemente grande da ospitare l'intera cella ATM che può quindi trovarsi a cavallo di due celle consecutive

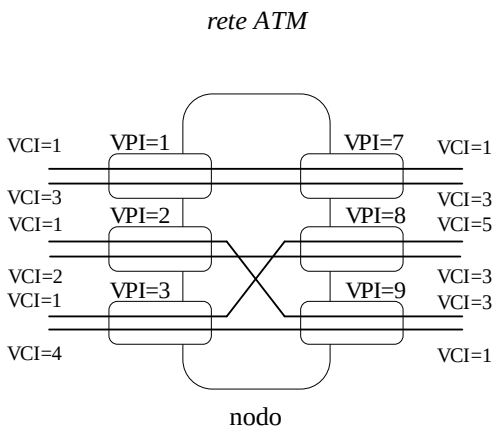
Strato ATM

Lo strato ATM svolge la funzione di instradamento delle celle , esso si basa sui primi 3 bite e mezzo dell'intestazione in cui ogni cella viene timbrata con opportuni valori di VPI e VCI. Altri bit nell'header consentono la funzione di controllo di errore dell'intestzione. La cella ha dimensioni fisse di 53 byte suddivisi in 5 byte di intestazione e 48 byte di payload.



I campi assumono i seguenti significati:

- GFC, generic flow control (4(bit), viene utilizzato per effettuare operazioni di controllo del flusso;
- VPI e VCI, virtual path/virtal channel identifier (24 bit), consente di individuare la connesione virtuale. 16 bit definiscono il canale della particolare connessione che è invece identificato da 8 bit;
- PT, payload type (3 bit), serve a discriminare l'informazione trasportata;
- CLP, cell loss priority (1 bit), serve alla rete per sapere se il pacchetto può essere scartato in caso di necessità (al verificarsi ad esempio di una congestione);
- HEC, header error control (8 bit), è usato dallo strato fisico per la rilevazione/correzione di errore nell'header della cella. Per il payload non è prevista alcuna forma di controllo che quindi vanno realizzate in modalità end-to-end dagli strati superiori.



Quando tutti i canali di un circuito virtuale sono commutati verso la stessa direzione in uscita (vedi nodo di rete ATM in figura, da VPI=1 a VPI=7 con VCI=1 e VCI=3), l'instradamento è realizzato dal nodo aggiornando semplicemente il campo di VPI nell'intestazione. Qualora, invece, un solo canale è commutato verso una diversa direzione di uscita, l'instradamento si realizza aggiornando congiuntamente il VPI ed il VCI.

Strato AAL

Allo strato AAL viene affidata la funzione di controllo di errore sull'intero carico della cella. Notare infatti che un nodo intermedio si compone dei soli strati fisico ed ATM. Lo strato AAL è il primo strato ad essere dunque presente su entrambi i nodi di sorgente e destinazione. Pertanto la funzione di controllo dell'errore non può che essere assegnata allo strato AAL in maniera tale da prevedere quindi un controllo dell'errore secondo la modalità end-to-end. Ogni funzione di strato AAL viene implementata da quattro diversi protocolli, ognuno di questi adatti ad un particolare flusso informativo. Si hanno i seguenti protocolli:

- AAL1, a tale protocollo competono le funzioni di gestione e temporizzazione del flusso trasportato. Per conseguire i propri obiettivi il protocollo sfrutta il primo dei 48 byte del pacchetto (solitamente destinati al payload) per numerare ciclicamente 8 pacchetti consecutivi (ciò richiede quindi l'uso di 3 bit). Per controllare l'errore all'interno dello stesso byte usa invece 4 bit (sempre ricavati dal primo byte) ed infine l'ultimo bit è usato per fornire un segnale

di temporizzazione (bit CSI). I bit CSI dei pacchetti pari forniscono la temporizzazione (101010...) mentre la sequenza dei bit CSI nei pacchetti dispari indica lo scostamento dalla frequenza del segnale;

- AAL2, si occupa della multiplazione di diversi flussi sulla stessa connessione ATM. Siccome vincola la lunghezza massima dei pacchetti in ingresso che al più possono essere lunghi 64 byte offre tempi ritardo molto contenuti e si addice al trasferimento di flussi con spiccate esigenze temporali;
- AAL3/4, solge anch'esso le funzioni di multiplazione di più flussi sulla stessa connessione ATM ma a differenza di AAL2 non limita lalunghezza massima dei pacchetti ad un numero basso di byte che in tal caso è di 65535 byte;
- AAL5, è stato l'ultimo protocollo ad essere definito e per questo motivo ha portato con se numerose semplificazioni che nei precedenti protocolli AAL erano invece assenti. Questa evoluzione è potuta avvenire solo quando il canale, congiuntamente agli altri protocolli che lo gestiscono, ha iniziato a offrire una probabilità di errore sempre più bassa

Strati superiori

Gli strati superiori della rete hanno come compito la contrattazione del servizio con l'utente. I servizi basati su ATM realizzano il trasferimento da un estremo all'altro e ciò presuppone la conversione dell'informazione nella cella ATM, l'instradamento della stessa ed infine la sua consegna. Tutto viene offerto dallo strato ATM che consegna, dunque, agli strati superiori una visione globale delle capacità offerte dal sistema. Un nodo ATM, detto anche hub ATM, raccoglie in ingresso i flussi informativi che possono essere sia di natura ATM oppure di altri sistemi. Per il secondo caso al nodo viene anche fornito un adattatore ATM dei flussi. Ciò è essenziale per instradare in seguito i flussi ATM.

Questa potenzialità proiettava ATM verso il successo: immaginate una rete che ospita anche flussi di altre reti, tutti con il passare del tempo sarebbero migrati verso ATM. Finora le reti ATM con strato fisico SDH/SONET sono state ampiamente utilizzate solo allo strato di trasporto nella realizzazione delle dorsali di rete. Il suo successo è stato clamorosamente stoppato da un errore non prevedibile. Infatti, nel momento in cui ATM veniva standardizzata, non era ancora stato inventato il protocollo http che di lì a poco dalla sua comparsa inizio a caratterizzare fortemente i flussi che iniziarono ad essere legati al web. Il web browsing, applicazione killer, del protocollo http non era contemplato da ATM ed è di difficile integrazione per essa stessa. ATM funziona svolge la sua funzione di instradamento nella modalità orientata alla connessione e questo non si addice ad http poichè richiederebbe, per aprire ad esempio una pagina web, di aprire e chudere diverse connessioni (una per ogni pagina di server diverso). Aggiungiamo poi che tutto ciò va fatto in tempi ristretti ed ecco che ATM si rileva inefficiente per http.

Le reti geografiche wireless

I collegamenti elettromagnetici su cui spesso si fa viaggiare l'informazione possono essere caratterizzati da una propagazione guidata oppure da una propagazione libera. Sui primi canali elettromagnetici, quelli cioè con propagazione guidata, si sono costruite le prime reti di telecomunicazioni. La rete telefonica e la rete di calcolatori costituiscono due esempi di rete fortemente legati all'uso del cavo nei collegamenti fisici. Con l'avvento delle tecnologie wireless i canali con propagazione libera hanno iniziato ad inserirsi man mano nelle attuali reti esistenti, in alcuni casi poi la realizzazione di un collegamento wireless è addirittura più economica. Basti pensare a quei luoghi laddove la posa dei cavi è impensabili: località di montagna, edifici storici etc... Ciò non scoraggia il progettista che, qualora le antenne usate per il collegamento wireless vengono poste in visibilità, può fare affidamento su un canale con caratteristiche superiori a canali con propagazione guidata. Quello

appena visto non è l'unico vantaggio offerto da un collegamento wireless che adesso offre un importante possibilità: quando un collegamento wireless coinvolge un terminale della rete lo libera dal consueto cavo fisso e gli consente dunque la mobilità.
Per quanto riguarda il discorso della posa dei cavi e della economicità della soluzione wireless va precisato che quest'ultima è stata importante almeno fintanto che i costi per la fibbra ottica non si sono abbassati fino a scendere a prezzi ragionevoli. Il vantaggio della mobilità, invece, era offerto solo a quei terminali capaci di autoalimentarsi (con batterie o comunque senza il vincolo che il cavo dell'alimentazione elettrica imponeva loro) e che avevano poi caratteristiche di ingombro e di peso abbastanza contenute (nessuno si sarebbe portato dietro un armadio se pur mobile). I terminali che rispondeva a questi requisiti non erano molti e per giunta questi avevano un prezzo troppo elevato per innescare politiche di economie di scale. Il costo di un singolo terminale era poi giustificato dal fatto che erano ammessi solo un certo numero di terminali funzionanti in contemporanea. Tutto ciò era causato dall'indisponibilità di una soluzione efficiente per la realizzazione dell'accesso multiplo a divisione di spazio nella banda radio. Possiamo individuare due periodi storici di rilievo per le reti wireless, un primo periodo è quello precedente all'introduzione di tecniche efficienti per la multiplazione a divisione di spazio ed un secondo periodo è invece quello successivo.
Per il primo periodo storico, anche in base alle considerazioni fatte, i collegamenti wireless erano rivolti verso quei terminali dotati di poco ingombro e peso ridotto. Rientravano in questa categoria, al momento, le sole radio. Le stazioni radio erano tra loro collegate da ponti radio, essi venivano dislocati su posti di montagna per raggiungere la condizione di visibilità fra le antenne e costituivano la rete di distribuzione. Ad ogni nodo veniva affidata la copertura di una certa zona, il nodo riceveva dalla rete di distribuzione il segnale e lo replicava, secondo modalità broadcat (verso tutti i terminali). Quando l'area ottenuta era insufficiente a coprie il territorio nazionale si provvedeva ad inserire nella rete di distribuzione un ulteriore nodo. Il problema qui è essenzialmente quello di non tollerare che la stessa banda di frequenza venisse utilizzata nelle aree adiacenti onde evitare una sovrapposizione del segnale. Pertanto, ogni nodo doveva sì distribuire il segnale ma su frequenze diverse. Ciò allontanava tra loro, quindi, tutti quei nodi che adottavano la stessa frequenza. Una stessa area può essere coperta da due nodi solo se questi usano una diversa frequenze (trasmettono in bande diverse).
Durante il secondo periodo le reti esistenti continuano ad esistere, tuttavia sono adesso disponibili tecniche di multiplazione a divisione di spazio che si concentrano su aree più ristrette, talvolta anche rivolte al solo ambito urbano. I centri urbani sono poi tra loro collegati mediante fibra ottica. Queste tecniche cellulari, così sono ricordate, riaprono le prospettive offerte dalla mobilità. Inizia anche ad essere usata, nella sezione di distribuzione, la tecnica numerica che ha il pregio di ammettere che la stessa zona venga coperta da più nodi.
Le reti più importanti costruite nel primo periodo sono: le reti radiofoniche e la rete televisiva terrestre.
Le reti radiofoniche si affidano in unprimo momento alla banda di frequenza AM che offre una portante di alcune decine di KHz. In realtà l'uso della parola rete è improprio: a causa delle proprietà di tale banda il segnale trasmesso può irradiare una vasta area per cui in realtà non vi è una rete vera e propria che invece nasce quando le trasmissioni radiofoniche si affidano alla banda FM (88-108MHz).
La banda AM, multiplata secondo tecnica di multiplazione a divisione di frequenza, offriva pochi canali per lo più intercontinentali. E così il desidorio di disporre di un numero maggiore di trasmissioni vocali ha motivato l'introduzione della banda FM. Quest'ultima offriva una copertura più ridotta al punto che più enti radiofonici che intendevano effettuare una copertura nazionale dovevano prevedere più nodi per formare così una rete di distribuzione. Con l'avvento dell'FM il termine rete ha iniziato ad essere usato propriamente.
La rete televisiva terrestre neccessita, invece, di più trasmettitori per la copertura di una singola zona. La banda occupata dal segnale video è di circa 5-7 MHz e richiede una frequenza portante di almeno una cinquantina di MHz. Le bande di frequenza individuate a tale scopo sono le VHF e UHF. Anche in

questo contesto prende forma una opportuna rete di distribuzione per garantire al segnale video una copertura nazionale. E possibile pensare alla rete di accesso come ad una rete distinta da quella di distribuzione. Il trasmettitore trasmette sulla rete di accesso i diversi segnali video che le diverse reti di distrubuzione gli inviano. In questo scenario, più interessante, si manifesta l'esigenza di un relay che favorisca l'interconnesione delle diverse reti di distribuzione. Il relay, in ambito di reti televisive, è il nodo a cui accedono i terminali, esso è anche detto ripetitore.
Ad ogni terminale deve essere inviato un segnale che contine N_a segnali analogici, per questo motivo è importante dimensionare l'area di copertura di modo che $N_a << N_t$ con N_t numero di terminali per sfruttare meglio le risorse). Questa scelta comporta la fornitura di un servizio audio visivo privo del controllo di presentazione (nel controllo di presentazione ogni tributario sceglie i flussi analogici che vuole vedere). Infatti, affidare il controllo di presentazione ad N_t terminali richiede di trasmettere dal nodo N_t flussi informativi differenti (un bel lavoro se gli utenti sono molti). Osservare che la sola rete telefonica al momento è capace di garantire $N_a=N_t$. Se si vule offrire il controllo di presentazione occorre anzitutto abbassare le frequenze di cifra dei flussi trasmessi su richiesta e ridurre il numero di utenti che vi possono accedere.

Le reti TCP/IP

Un primo modello di riferimento per reti di calcolatori e antenato dell'internet odierna fu ARPANET: la prima rete di calcolatori a commutazione di pacchetto. ARPANET fu un progetto sponsorizzato dal DoD (US dipartment of defense) e sviluppato presso varie università americane. Gli obiettivi che si volevano raggiungere eano la possibilità di interconnettere più reti e la robustezza intesa come la capacità di funzionare anche in caso di guasti ai nodi intermedi. La ricerca permise la realizzazione di protocolli in grado di garantire le caratteristiche dette prima e ciò diede inizio a quello cge oggi si chiama modello TCP/IP. Nel modello TCP/IP sono definite quattro diverse stratificazioni: strato fisico, strato di rete, strato di trasporto e strato di applicazione. Ad ogni strato si presentano diversi protocolli, ognuno di essi offre un determinato servizio.
Nello strato più elevato, strato di applicazione, si trovano i protocolli adoperati dalle applicazioni utente. Tali protocolli si distinguono in protocolli di supporto e protocolli utente. Tra i protocolli utente più popolari ci sono:

- FTP, file transfer protocol, per il trasferimento di file;
- SMTP, simple mail transfer protocol, per il trasferimento di posta semplice;
- HTTP, hyper tet transfer protocol, definisce le modalità concui vengono scambiate le pagine html dal server we al web browser dell'utente;

i protocolli di supporto includono i protocolli per la gestione della rete:

- SNMP, simple network management protocol;
- DNS, domain name system, utile alla risoluzione degli indirizzi dal formato numerico a quello alfabetico;

Tutti protocolli definiscono le modalità con cui avvengono lo scambio dei messaggi nonchè il formato del messaggio stesso. Lo strato di trasporto, come nel modello OSI, fornisce servizi di comunicazione end-to-end alle applicazioni. I due protocolli più importanti sono TCP e UDP. Il protocollo TCP (trasmission control protocol) offre un servizio orientato alla connessione e un servizio di trasferimento affidabile dei dati:

- servizio orientato alla connessione: ancora prima che i messaggi da scambiare fra client e server iniziano a fluire, il protocollo TCP prevede una procedura di handshake che allerta il client ed il server preparandoli ad un arrivo massiccio di pacchetti. Quando la procedura di handshake è avvenuta con successo si dice che esiste una connessione (full-duplex) tra i processi client e server (tramite socket). Al termine della comunicazione la connessione viene chiusa.
- Servizio affidabile: i processi che si affidano a TCP possono confidare in una consegna precisa ed ordinata di tuttii pacchetti;

TCP prevede anche un meccanismo di controllo della congestione che strozza un processo quando la rete è congestionata (cosa non buona per le applicazioni in tempo reale). Infine, TCP non garantisce nulla sui tempi di consegna e sulla velocità di spedizione.
Il protocollo UDP (user datagram protocol) non effettua il controllo della congestione e il riassemblaggio dei pacchetti. Esso, inoltre, non prevede alcuna procedura di handshake per cui un applicazione che deve inviare dati lo fa senza preoccuparsi di trovare disponibile l'altra applicazione remota. Si tratta per questo motivo di un protocollo senza fronzoli che come TCP non offre tra l'altro garanzia sui tempi di ritardo.
Il successo delle reti TCP/IP è dovuto all'efficacia del protocollo dello strato di rete, il protocollo IP (internet protocol). Esso fornisce servizi di strato di rete realizzati con l'ausilio di un altro protocollo, l'ICMP (internet control message protocol) per lo scambio di messaggi orientati al controllo della rete.
Al di sotto dello strato di rete non sono definiti ulteriori protocolli, nello strato fisico viene ipotizzata la capacità di passare l'informazione da un punto della rete ad un altro. L'interconnessione fra nodi terminali avviene tramite nodi intermedi che non hanno così la necessità di sviluppare i livelli più alti del modello di riferimento.

Il protocollo IP
Il protocollo IP, assieme a TCP, costituisce il nucleo della pila protocollare e fornisce agli strati superiori un servizio inaffidabile e senza connessione:

- servizio inaffidabile: non è garantita la consegna e le modalità di tale consegna del pacchetto. I pacchetti possono essere persi, duplicati o ritardati dalla rete e ciò si verifica senza che alcun messaggio avvisi le entità di strato superiore;
- servizio senza connessione: non esiste alcuna nozione di connessione attiva ed i pacchetti nella rete possono seguire differenti percorsi e andare incontro a differenti ritardi. Non è inoltre previsto un numero di sequenza per i pacchetti;

In base a quanto detto finora circa i servizzi offerti da IP si descrive solitamente il servizio offerto da IP come servizio di tipo best effort" volendo indicare che il protocollo IP fa il massimo sforzo per portare a termine la consegna dei pacchetti rispettandone l'ordine di spedizione. Ovviamente non vengono offerte garanzie sui tempi di consegna oltre che sull'effettiva consegna. Lo strato di rete ha tre componenti:

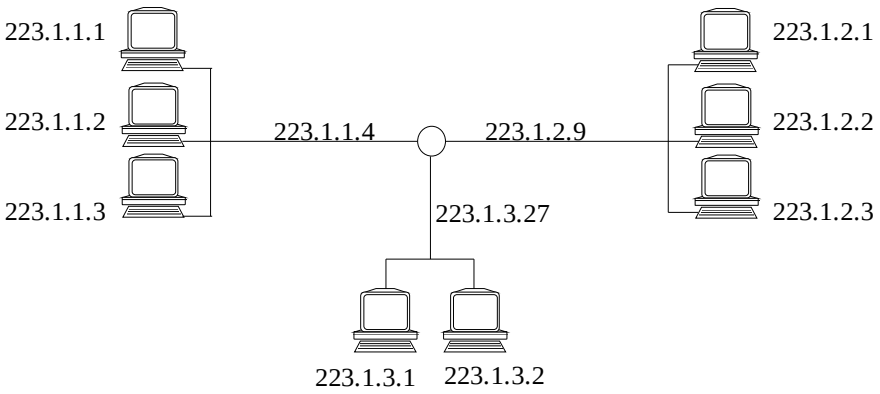
- il protocollo IP che definisce l'indirizzamento dello strato di rete, specifica i campi del datagramma e le azioni che i router ed i terminali devono intraprendere. La versione più usata del protocollo IP è la IPv4, è in corso tutt'ora una transizione che dovrebbe introdurre una successiva versione di IP, la IPv6;
- la seconda componente dello strato di rete determina l'instradamento dei pacchetti attraverso la rete e si poggia sui protocolli RIP, OSPF BGP;

- la terza componente dello strato di rete richiede la possibilità di registrare errori nei datagrammi con lo scopo di segnalarli all'entità trasmittente o ricevente (protocollo ICMP);

Le modalità di instradamento ed il particolare indirizzamento adoperato da IP sono due punti cruciali per l'efficienza del protocollo. Quando un host ha dei dati da spedire invia quest'ultimi su dei link in uscita dall'interfaccia di rete. Il link è poi collegato ad un router che a differenza dell'host è dotato di più interfacce (almeno una in ingresso ed una in uscita) e generalmente di più link in uscita. Il protocollo IP richiede allora che ciascuna interfaccia abbia un indirizzo tale da permetterne l'identificazione. Nell'attuale versione, l'IPv4, gli indirizzi sono specificati mediante stringhe di 32 bit, solitamente scritti nella notazione decimale puntata. I 32 bit sono cioè suddivisi in 4 gruppi da 8 bit ciascuno, ogni gruppo può allora indicare (nella notazione decimale) un numero appartenente all'insieme {0,1,2,...,255} che è separato dal successivo gruppo mediante un punto. Con l'utilizzo di 32 bit si possono allora indirizzare 2^{32} possibili indirizzi IP:

$$193.32.216.9 \rightarrow 11000001.00100000.11011000.00001001$$

Tuttavia gli indirizzi IP non possono essere scelti a caso, una parte dell'indirizzo IP sarà determinato dalla rete a cui il terminale si collega, ad esempio:



In base alla figura osserviamo quanto segue: i tre terminali in alto a sinistra condividono i primi 24 bit dell'indirizzo che identifica quindi la rete solitamente segnata come 223.1.1.0/24 e dove la notazione /24 è detta maschera di rete ed i bit che ne particolarizzano i valori si dicono essere il prefisso della rete. Ed allora la rete 223.1.1.0/24 consiste di tre interfacce host (223.1.1.1, 223.1.1.2, 223.1.1.3) e di una interfaccia router (223.1.1.4). Qualsiasi host che si aggiunge alla suddetta rete dovrà avere un prefisso di rete pari a 223.1.1.xxx/24. Sempre osservando la figura, possiamo adesso dire che in essa si presentano complessivamente tre reti: 223.1.1.0/24, 223.1.2.0/24 e 223.1.3.0/24. E' possibile vedere ciascuna rete che collega le interfacce di rete come un segmento ethernet. Tuttavia, per IP sono segmenti di rete non solo i segmenti ethernet che collegano gli host ad un router ma lo sono anche i segmenti che collegamenti i vari router di una rete al punto che nella figura meritano anch'essi un indirizzo. Originariamente l'architettura IP definiva quattro classi di indirizzo, una quinta classe è poi riservata per usi futuri:

- classe A, i primi 8 bit identificano la rete e gli ultimi 24 bit identificano le interfacce collegate (il primo bit vale 0);

- classe B, i primi 16 bit identificano la rete e gli ultimi 16 bit identificano le interfacce di rete collegate (i primi due bit valgono 10);
- classe C, i primi 24 bit identificano la rete e gli ultimi 8 bit identificano le interfacce collegate (i primi tre bit valgono 110);
- classe D, i primi bit valgono 1110 ed i rimanenti 28 bit identificano gli indirizzi con destinazione multipla;
- classe E, i primi bit valgono 1111, gli indirizzi di tale classe sono riservati ad usi futuri;

Dalle classi così definite scaturiscono i numeri dei terminali collegabili ad una rete: nella classe A sono disponibili $2^{(8-1)}$=128 reti (il primo bit è fisso a 0) con la possibilità di indirizzare 16.777.216 (2^{24}) host. Nella classe B sono disponibili $2^{(16-2)}$=16.384 reti (i primi 2 bit sono fissi a 10) con la possibilità di indirizzare 65536 indirizzi host. Nella classe C sono disponibili $2^{(24-3)}$=2.007.152 reti (i primi tre bit sono fissati a 110) con la possibilità di identificare 2^8=256 host.
Tale gerarchia ha nel corso degli anni esaurito gli indirizzi di rete, i maggiori problemi venivano dalle reti di classi B e C che per la loro dimensione si presentavano o troppo grandi o troppo piccoli.
Fu così che nel 1993 l'IETF standardizzò il CIDR (class inter domain routing) in cui la parte dell'indirizzo di rete poteva essere lunga a piacere (qualsiasi numero di bit) invece che essere di 8,16 oppure 24 bit. Una rete con indirizzamento CIDR ha una notazione decimale del tipo a.b.c.d/x in cui x indica il numero di bit che nei 32 bit complessivi costituisce l'indirizzo di rete.
Ad esempio, una società con 2000 host può farsi assegnare un indirizzo di rete di classe B che indirizza 65536 host con uno spreco di 65536-2000=63536 host oppure può adottare la notazione CIDR a.b.c.d/21 in cui i primi 21 bit identificano la rete ed i restanti 11 bit permettono di indirizzare 2^{11}=2048 host con uno spreco di 2048-2000=48 indirizzi host. Inoltre, la società può suddividere ulteriormente gli 11 bit di indirizzo host per creare le proprie reti interne mediante procedura di subnetting. Il numero di sottoreti X che è possibile creare stabilisce il numero di bit necessari, pari a 2^X-2 bit. Utilizzati X bit degli 8 messi a disposizione dell'ottetto ne rimangono Y che quindi fissano il numero di host rappresentabili pari a 2^Y-2. Ad esempio, dalla rete con IP 192.168.5.0 si vogliono ricavare due sottoreti, occorrono pertanto 2^2-2=2 bit (X=2) che saranno presi da un ottetto di bit. Quindi Y=6, è allora possibile indirizzare, in ciascuna sottorete 2^6-2=62 host (si sottrae 2 perchè da tutti gli indirizzi possibili si tolgono quelli broadcast ed unicast). Quindi definiamo le due subnet mask per le reti:

1) 11000000.10101000.00000101.01000000 $\rightarrow$ 192.168.5.64

2) 111000000.10101000.00000101.10000000 $\rightarrow$ 192.168.5.128

Per la rete 1) il primo indirizzo valido sarà 11000000.10101000.00000101.01000001 (192.168.5.65), mentre l'ultimo indirizzo valido sarà 11000000.10101000.00000101.01111110 (192.168.5.126), l'indirizzo broadcast per tale rete (tutti 1) sarà 11000000.10101000.00000101.01111111 (192.168.5.127).
Per la rete 2) il primo indirizzo valido sarà 11000000.10101000.00000101.10000001 (192.168.5.129), mentre l'ultimo indirizzo valido sarà 11000000.10101000.00000101.10111110 (192.168.5.190), l'indirizzo broadcast per tale rete (tutti 1) sarà 11000000.10101000.00000101.100111111 (192.168.5.191).
Anche dopo l'introduzione del CIDR gli indirizzi IP erano pochi cosicchè iniziarono ad essere trattati come risorse limitate. L'ISP che ne fa richiesta per questo motivo fornisce una documentazione che ne giustifica le pretese. L'attribuzione di un blocco di indirizzi IP è comunque soggetta a restrizioni

temporali. L'utente residenziale che si collega ad Internet chiede all'ISP un indirizzo IP, tale richiesta avviene mediante l'uso del protocollo DHCP. Un altra possibilità per la configurazione degli è l'assegnamento manuale. Il DHCP permette ad un host di ottenere un indirizzo IP in maniera automatica, il DHCP completa la richiesta dell'utente anche con altre informazioni come ad esempio la fornitura di un indirizzo server DNS. Il DHCP è talvolta detto protocollo pug and play, collega e funziona. Si tratta di una soluzione molto di moda tra gli ISP residenziali che solitamente non dispongono di un IP per tutti gli utenti che invece si collegano mediante un indirizzo IP temporaneo.
Ogni qualvolta che un host si collega, l'ISP aggiorna la lista di IP disponibili, al termine della connessione l'IP usato dall'host è rimesso nella lista degli indirizzi disponibili. Il DHCP favorisce la portabilità dell'host (cosa ben diversa della mobilità) che può così collegarsi in diversi posti con diversi indirizzi IP. Ad ogni rete è assegnato un server DHCP per la configurazione automatica dell'IP. Il terminale che intende collegarsi manda un messaggio di scoperta DHCP, siccome non conosce con precisione l'indirizzo del server DHCP invia il messaggio di richiesto nella modalità broadcast all'indirizzo 255.255.255.255. Nel datagramma appena inviato l'host mette a 0.0.0.0 il relativo campo dell'indirizzo sorgente.
Il server DHCP riceve la richiesta e risponde con l'offerta di un indirizzo IP, la maschera di rete ed il tempo di affitto dell'indirizzo assegnato (è una sorta di contrattazione). L'host deve rispondere all'offerta di un server DHCP (ci possono essere più server DHCP e quindi più offerte contemporanee) con un messaggio di ocnferma che dovrà sostanzialmente ripetere nel datagramma l'IP ricevuto assieme agl'altri parametri. Ciò è indispensabile poichè nella rete potrebbero esserci più server DHCP ed allora il messaggio ACK-DHCP ha l'effetto immediato di convalidare la contrattazione verso un server DHCP e di annullare tutte le altre offerte nel frattempo pervenute (cosicchè gli altri server DHCP possono considerare libero l'IP offerto ma rifiutato dall'host).