Capitolo 23

Reti a pacchetto

Alfine è giunto il momento di parlare della *rete delle reti*, ossia di Internet! Il tema è sviluppato con riferimento ai vari strati funzionali che sono coinvolti nella sua operatività, iniziando da una visione di insieme che descrive la concatenazione di indirizzi su cui si basa la trasmissione, per approfondire l'analisi a partire dallo strato di trasporto, giù fino allo strato fisico. Per una visione ancora più ampia sugli aspetti che *sovrastano* lo strato di trasporto, il lettore può far riferimento ad un altro testo dello stesso autore, *Lo strato applicativo di Internet*. Sempre in questo capitolo, sono discussi anche i principi e le pratiche su cui si basa l'ATM, una architettura di rete nata quasi in contemporanea ad Internet, e che pur non avendone eguagliato il successo, rappresenta un caso di scuola per la categoria di reti basate sul paradigma del circuito virtuale. Viceversa, la realizzazione delle reti orientate *alla perdita* ovvero a commutazione di circuito, è rimandata al capitolo 24.

Affermiamo fin da subito che il modello a strati ISO-OSI (pag. 784) è una astrazione concettuale utile per individuare raggruppamenti di funzioni, e serve ottimamente come modello per stimolare l'interoperabilità di apparati di diversi costruttori. D'altra parte, realizzazioni come Internet si sono sviluppate precedentemente alla definizione di tale modello, mentre altre (come ATM) seguono filosofie che solo successivamente sono state incorporate nel modello di riferimento. Pertanto, utilizzeremo le classificazioni ISO-OSI come riferimento culturale e terminologico, mediante il quale analizzare le funzioni delle reti reali.

23.1 La rete Internet

Storia Nel 1964 L. Kleinrock (UCLA) propone un modello di rete non gerarchica e con parti ridondanti, che realizza una modalità di trasferimento senza connessione e senza garanzie di qualità del servizio, rimandando queste ultime ai livelli superiori dell'architettura protocollare. Tale tipologia di servizio è oggi indicata con il termine *best effort*[1]. Nel '69 sono operativi cinque nodi nelle università americane, e nel '72 avviene la prima dimostrazione pubblica di ARPANET, basata su NCP. Nel '73 Kahn e Cerf

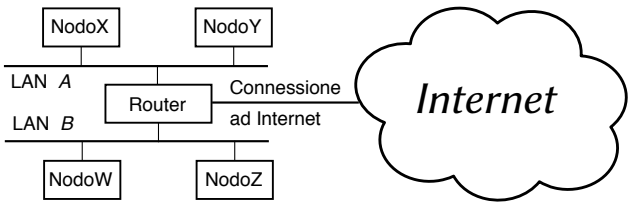
[1] Migliore sforzo, ossia la rete dà il massimo, senza però garantire nulla.

iniziano a definire TCP, da cui viene successivamente separato l'IP per la convenienza di non dover necessariamente aprire sempre una connessione. Fino all'80, il DoD[2] sovvenziona le università per implementare in ambiente UNIX i protocolli, che nel frattempo si vanno arricchendo di servizi, mentre la trasmissione Ethernet (del 1973) è adottata per realizzare LAN.

Nel 1983 il DoD decreta che tutti i calcolatori connessi a ARPANET adottino i protocolli TCP/IP, e separa la rete in due parti: una civile (ARPANET) ed una militare (MILNET). Negli anni seguenti i finanziamenti dalla *National Science Foundation* permettono lo sviluppo di una rete di trasporto a lunga distanza e di reti regionali, che interconnettono LAN di altre università e di enti di ricerca alla rete ARPANET, alla quale si collegano poi anche le comunità scientifiche non americane.

Nel 1990 ARPANET cessa le sue attività, e Barners-Lee (CERN) definisce il WWW, mentre nel '93 Andreessen (NCSA) sviluppa *Mosaic*, il primo *browser* WWW. Dal 1995 L'NSF non finanzia più la rete di interconnessione, ed il traffico inizia ad essere trasportato da operatori privati.

Caratteristiche La parola *Internet* in realtà è composta da due parole, INTER e NET, in quanto le caratteristiche della rete Internet sono quelle di fondere in una unica architettura una infinità di singole reti locali, potenzialmente disomogenee, e permettere la comunicazione tra i computer delle diverse sottoreti.



Ogni nodo della rete è connesso ad una rete locale (LAN[3]), la quale a sua volta è interconnessa ad Internet mediante dei nodi detti *router*[4] che sono collegati ad una o più LAN e ad Internet, e svolgono la funzione di instradare le comunicazioni verso l'esterno. L'instradamento ha luogo in base ad un *indirizzo IP*[5], che individua i singoli nodi in modo univoco su scala mondiale.

Come anticipato a pag. 786, lo strato di rete (o strato IP) realizza un modo di trasferimento a datagramma e non fornisce garanzie sulla qualità di servizio (QoS, QUALITY O SERVICE) in termini di ritardi, errori e pacchetti persi. La situazione è

[2]Department of Defense.

[3]LOCAL AREA NETWORK, ossia *rete locale*. Con questo termine si indica un collegamento che non si estende oltre (approssimativamente) un edificio.

[4]La funzione di conversione di protocollo tra reti disomogenee è detta di *gateway*, mentre l'interconnessione tra reti locali è svolta da dispositivi *bridge* oppure da *ripetitori* se le reti sono omogenee. Con il termine *router* si indica più propriamente il caso in cui il nodo svolge funzioni di instradamento, che tipicamente avviene nello *strato di rete*. Nel caso in cui invece si operi un instradamento a livello dello *strato di collegamento*, ossia nell'ambito di sezioni diverse (collegate da bridge o ripetitori) di una stessa LAN, il dispositivo viene detto detto *switch*. Infine, un *firewall* opera a livello di trasporto, e permette di impostare *regole di controllo* per restringere l'accesso alla rete interna in base all'indirizzo di *sorgente*, al tipo di *protocollo*, e/o a determinati *servizi* .

[5]IP = Internet Protocol.

mitigata dalla strato di trasporto (TCP, Transmission Control Protocol) che offre ai processi applicativi un servizio a circuito virtuale.

I protocolli di Internet sono realizzati in software e sono pubblici; gli utenti stessi e molte sottoreti private contribuiscono significativamente al trasporto, all'indirizzamento, alla commutazione ed alla notifica delle informazioni. Queste sono alcune ragioni fondamentali per cui Internet *non è di nessuno* ed è un patrimonio dell'umanità.

23.1.1 Gli indirizzi

Iniziamo l'argomento discutendo subito la stratificazione degli indirizzi coinvolti in una comunicazione via Internet. Ogni livello funzionale infatti utilizza le proprie convenzioni di indirizzamento, come illustrato nella tabella a fianco. Se a prima vista questa abbondanza di indirizzi può apparire esagerata, è proprio in questo modo che si realizza l'interoperabilità tra ambienti di rete differenti.

Strato	Indirizzo
Applicazione	*protocollo://nodo.dominio.tld*
Trasporto	*socket TCP o porta*
Rete	*indirizzo IP x.y.w.z*
Collegamento	*indirizzo Ethernet a:b:c:d:e:f*

23.1.1.1 IP ed Ethernet

I computer connessi ad Internet (detti *nodi*) sono le sorgenti e le destinazioni dell'informazione, e sono individuati da *un indirizzo IP*, che consiste in un gruppo di 4 byte[6] e che si scrive *x.y.w.z* con ognuna delle 4 variabili separate da punti pari ad un numero tra 0 e 255.

I nodi sono connessi alla rete mediante una interfaccia a volte indicata come MAC (Media Access Control). Prendendo come esempio[7] i nodi connessi ad una LAN Ethernet, l'interfaccia di rete è individuata a sua volta da un *indirizzo Ethernet* composto da 6 byte. Quest'ultimo è unico in tutto il mondo, ed configurato dal costruttore nella scheda di interfaccia. L'indirizzo Ethernet viene però utilizzato solo nell'ambito della LAN di cui il nodo fa parte, ossia dopo che i pacchetti sono stati instradati dai router, per mezzo dell'indirizzo IP, verso la LAN.

23.1.1.2 Sottoreti

Ogni nodo conosce, oltre al proprio indirizzo IP, anche una *maschera di sottorete* composta da una serie di uni seguita da zeri, in numero complessivo di 32 bit, tanti quanti ne sono presenti nell'indirizzo IP. Il termine *maschera* è dovuto all'operazione

[6] Con 4 byte si indirizzano (in linea di principio) $2^{32} = 4.29 \cdot 10^9$ diversi nodi (più di 4 miliardi). E' stato sviluppato il cosiddetto IPv6, che estende l'indirizzo IP a 16 byte, portando la capacità teorica a $3.4 \cdot 10^{38}$ nodi. L'IPv6 prevede inoltre particolari soluzioni di suddivisione dell'indirizzo, allo scopo di coadiuvare le operazioni di *routing*. Per approfondire, vedi ad es. `https://it.wikipedia.org/wiki/IPv6`

[7] Evidentemente esistono molte diverse possibilità di collegamento ad Internet, come via telefono (tramite provider), collegamento satellitare, Frame Relay, linea dedicata, ISDN, ADSL... ma si preferisce svolgere un unico esempio per non appesantire eccessivamente l'esposizione. La consapevolezza delle molteplici alternative consente ad ogni modo di comprendere la necessità di separare gli strati di trasporto e di rete dall'effettiva modalità di trasmissione.

di AND binario (vedi tabella) operata tra la maschera e gli indirizzi IP, per determinare se questi appartengano alla propria stessa LAN oppure risiedano altrove.

Indirizzo IP	Maschera Sottorete	Indirizzo sottorete
151.100.8.33	255.255.255.0	151.100.8.0

Nel caso in cui la sottorete di un nodo Y verso cui il nodo X deve inviare un pacchetto è la stessa su cui è connesso X, allora questi può individuare l'indirizzo Ethernet del destinatario[8] ed inviargli il pacchetto direttamente. In caso contrario, X invierà il pacchetto al proprio *default gateway* verso Internet.

23.1.1.3 Intranet

Alcuni gruppi di indirizzi IP (come quelli 192.168.*w.z* oppure 10.*y.w.z*) non vengono instradati dai router, e possono essere riutilizzati nelle *reti private* di tutto il mondo per realizzare le cosiddette *reti intranet* operanti con gli stessi protocolli ed applicativi che funzionano via Internet.

23.1.1.4 Domain Name Service (DNS)

L'utente di una applicazione Internet in realtà non è a conoscenza degli indirizzi IP dei diversi nodi, ma li identifica per mezzo di nomi simbolici del tipo nodo.dominio.tld, detti anche *indirizzi Internet*. Il processo di risoluzione che individua l'indirizzo IP associato al nome avviene interrogando un particolare nodo, il Domain Name Service (*servizio dei nomi di dominio*). La struttura dei nomi, scandita dai punti, individua una gerarchia di autorità per i diversi campi. Il campo tld è chiamato *dominio di primo livello* (top level domain[9]), mentre il campo dominio in genere è stato registrato da qualche organizzazione che lo giudica rappresentativo della propria offerta informativa. Il campo nodo rappresenta invece una ben determinata macchina, il cui indirizzo Internet completo è nodo.dominio.tld, e che non necessariamente è collegato alla stessa LAN a cui sono connessi gli altri nodi con indirizzo che termina per dominio.tld.

Quando un nodoX generico deve comunicare con nodoY.dominio.tld, interroga il proprio DNS[10] per conoscerne l'IP. Nella rete sono presenti molti DNS, alcuni dei quali detengono informazioni *autorevoli*[11] riguardo ai nodi di uno o più domini, altri (i DNS *radice*, o root) detengono le informazioni relative a quali DNS siano autorevoli per i domini di primo livello, ed altri fanno da tramite tra i primi due ed i *client* che richiedono una risoluzione di indirizzo. Se il DNS di nodoX *non è autorevole per*

[8] Mostriamo in seguito che questo avviene mediante il protocollo ARP.

[9] I top level domain possono essere pari ad un identificativo geografico (.it, .se, .au...) od una delle sigle .com, .org, .net, .mil, .edu, che sono quelle utilizzate quando Internet era solo americana.

[10] Il "proprio" DNS viene configurato per l'host in modo fisso, oppure in modo dinamico dai Service Provider raggiungibili via ADSL, e convenientemente corrisponde ad un nodo situato "vicino" al nodo che lo interroga.

[11] Chi registra il dominio deve disporre necessariamente di un DNS in cui inserire le informazioni sulle corrispondenze tra i nomi dei nodi del proprio dominio ed i loro corrispondenti indirizzi IP. In tal caso quel DNS si dice *autorevole* per il dominio ed è responsabile di diffondere tali informazioni al resto della rete.

nodoY, allora[12] provvede ad inoltrare la richiesta, interrogando prima un DNS radice per individuare chi è autorevole per .tld, quindi interroga questo per trovare chi è autorevole per .dominio.tld, e quindi usa la risposta ottenuta per dirigere la richiesta di risoluzione originaria. Se la cosa può sembrare troppo macchinosa per funzionare bene, è perché la stessa sequenza di operazioni *non deve* essere effettuata sempre: il DNS utilizzato da nodoX riceve infatti, assieme all'IP di nodoY, anche una informazione detta TIME TO LIVE (TTL o *tempo di vita*) che descrive la scadenza della coppia *nome-IP* ottenuta. Genericamente il TTL è di qualche giorno, e fino alla sua scadenza il DNS *ricorda*[13] la corrispondenza, in modo da fornire la propria copia in corrispondenza delle richieste future, e ridurre sensibilmente il traffico legato alla risoluzione degli indirizzi Internet. L'insieme delle risoluzioni apprese è denominata *cache* del DNS[14].

23.1.1.5 Indirizzi TCP

Si è detto che ogni nodo è individuato in Internet mediante il proprio indirizzo IP, ma questo non è sufficiente ad indicare con quale particolare programma (che implementa uno specifico *servizio* come nel caso del DNS) si vuole entrare in comunicazione. I programmi che sono pronti a ricevere connessioni si pongono *in ascolto* su ben determinate *porte* (o *socket*[15]), identificate da numeri[16], e che sono referenziati in modo simbolico (es. *http://*, *ftp://*) dagli applicativi di utente che si rivolgono allo strato di trasporto (il TCP) per stabilire un collegamento con un server presente su di un nodo remoto.

Alcuni servizi rispondono ad indirizzi *ben noti*, fissi per tutti i nodi, in quanto il chiamante deve sapere a priori a quale porta connettersi. Il nodo contattato invece apre con il chiamante una connessione di ritorno su di un numero di porta diverso, che è stato comunicato dal chiamante al momento della richiesta di connessione, e per il quale sempre il chiamante non ha già aperto altre connessioni differenti.

23.1.2 TCP

Discutiamo ora del TCP[17], che offre ai processi applicativi un servizio di trasporto a

[12] In realtà esiste anche una diversa modalità operativa, che consiste nel delegare la ricerca ad un diverso DNS (detto *forwarder*), il quale attua lui i passi descritti appresso, e provvede per proprio conto alla risoluzione, il cui esito è poi comunicato al primo DNS e da questi ad *hostX*. Il vantaggio di tale procedura risiede nella maggiore ricchezza della *cache* (descritta appresso) di un DNS utilizzato intensivamente.

[13] Il DNS ricorda anche le altre corrispondenze ottenute, come il DNS autorevole per *.tld* e per *.dominio.tld*; nel caso infine in cui si sia utilizzato un forwarder, sarà quest'ultimo a mantenere memoria delle corrispondenze per i DNS intermedi.

[14] CACHE è un termine generico, che letteralmente si traduce come *nascondiglio dei viveri*, e che viene adottato ogni volta si debba indicare una memoria che contiene copie di riserva, o di scorta...

[15] *Socket* è un termine che corrisponde alla... presa per l'energia elettrica casalinga, ed in questo contesto ha il significato di una *presa* a cui si "attacca" il processo che richiede la comunicazione. Per l'esattezza, un *socket Internet* è individuato dal numero di porta TCP e dall'indirizzo IP.

[16] Spesso gli indirizzi che identificano i punti di contatto di servizi specifici vengono indicati come SERVICE ACCESS POINT (SAP), anche per situazioni differenti dal caso specifico delle porte del TCP.

[17] TCP = *Transport Control Protocol*.

circuito virtuale, *attaccato* ad una porta[18] di un nodo remoto individuato dall'indirizzo IP. Il suo compito è quello di ricevere dai processi applicativi dei dati, suddividerli in pacchetti, ed inviarli al suo pari che svolge il processo inverso.

23.1.2.1 Il pacchetto TCP

La struttura di un pacchetto TCP è mostrata in figura, e comprende una intestazione composta da 6 gruppi (o più) di 4 byte per un minimo di 192 bit, a cui segue un numero variabile di gruppi di 4 byte di dati, provenienti dagli strati applicativi superiori. Troviamo subito i numeri delle porte a cui si riferisce la connessione, mentre gli indirizzi IP sono aggiunti dallo strato di rete. I numeri *di Sequenza* e di *Riscontro* servono rispettivamente a numerare i bytes dei pacchetti uscenti, ed a notificare l'altro lato del collegamento del numero di sequenza del prossimo byte che si aspetta di ricevere[19], riscontrando implicitamente come correttamente arrivati i pacchetti con numero di sequenza più basso.

Offset (4 bit) codifica il numero di parole da 4 byte dell'intestazione, mentre nei 6 bit *Riservati* non è mai stato scritto nulla. I 6 bit del campo *Controllo* hanno ognuno un nome ed un significato preciso, qualora posti ad uno. Il primo (URG) indica che il campo urgent pointer contiene un valore significativo; ACK indica che si sta usando il Numero di Riscontro; PSH indica un pacchetto urgente che non può rispettare la coda in ricezione; RST segnala un malfunzionamento e impone il reset della connessione; SYN è pari ad uno solo per il primo pacchetto inviato per richiedere di creare una connessione; FIN indica che la sorgente ha esaurito i dati da trasmettere.

I 16 bit di *Finestra* rappresentano il numero di byte che, a partire dal valore espresso dal *Numero di Riscontro*, chi invia il pacchetto è in grado di ricevere, ed il suo utilizzo sarà meglio illustrato tra breve nel contesto del controllo di flusso. Il *Checksum* serve

[18] Il numero di porta costituisce in pratica l'*identificativo di connessione* del circuito virtuale. Nel caso in cui un server debba comunicare con più client, dopo avere accettato la connessione giunta su di una *porta ben nota*, apre con i client diversi canali di ritorno, differenziati dall'uso di porte di risposta differenti.

La lista completa dei servizi standardizzati e degli indirizzi ben noti (*socket*) presso i quali i serventi sono in attesa di richieste di connessione, è presente in tutte le distribuzioni Linux nel file */etc/services*.

[19] Il numero di sequenza si incrementa ad ogni pacchetto di una quantità pari alla sua dimensione in bytes, ed ha lo scopo di permettere le operazioni di controllo di flusso. Il valore iniziale del numero di sequenza e di riscontro è diverso per ogni connessione, e generato in modo pseudo-casuale da entrambe le parti in base ai propri orologi interni, allo scopo di minimizzare i problemi dovuti all'inaffidabilità dello strato di rete (l'IP) che può perdere o ritardare i datagrammi, nel qual caso il TCP trasmittente ri-invia i pacchetti precedenti dopo un time-out. Questo comportamento può determinare l'arrivo al lato ricevente di un pacchetto duplicato, e consegnato addirittura dopo che la connessione tra i due nodi è stata chiusa e riaperta. In tal caso però la nuova connessione adotta un diverso numero di sequenza iniziale, cosicché il pacchetto duplicato e ritardato risulta fuori sequenza, e non viene accettato.

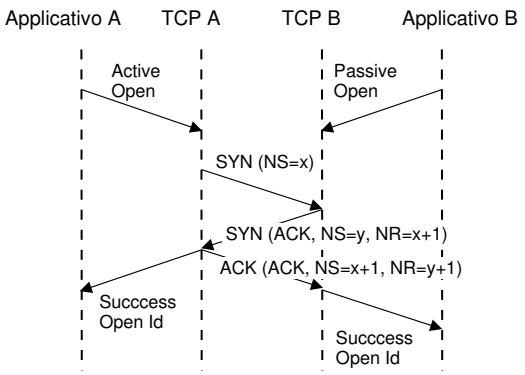
al ricevente per verificare se si sia verificato un errore, il *Puntatore Urgente* contiene il numero di sequenza dell'ultimo byte di una sequenza di dati urgenti, e le *Opzioni* (di lunghezza variabile) sono presenti solo raramente, ed utilizzate a fini di controllo, ad esempio per variare la dimensione della finestra. Infine, il *Riempimento* conclude l'ultima parola da 32 bit.

Uno stesso pacchetto TCP può svolgere funzioni di sola segnalazione, o di sola trasmissione dati, od entrambe.

23.1.2.2 Apertura e chiusura della connessione

Il TCP offre un servizio di di trasporto a circuito virtuale, e prima di inviare dati, deve effettuare un colloquio iniziale con il nodo remoto di destinazione. In particolare, il colloquio ha lo scopo di accertare la disponibilità del destinatario ad accettare la connessione, e permette alle due parti di scambiarsi i rispettivi numeri di sequenza descritti alla nota 19.

L'estremo che viene "chiamato" riveste il ruolo di *server*, e l'altro di *client*. Dato che anche quest'ultimo deve riscontrare il numero di sequenza fornito dal server, occorrono tre pacchetti per terminare il dialogo, che prende il nome di THREE WAY HANDSHAKE[20]. Il diagramma a lato mostra l'evoluzione temporale del colloquio tra un processo applicativo client (A), ed un server (B) che si pone in ascolto, mostrando come al primo SYN che pone $NS_A = x$, ne segua un altro che pone $NS_B = y$, seguito a sua volta dall'ACK di chi ha iniziato[21]. La chiusura può avvenire per diverse cause: o perché è terminato il messaggio, segnalato dal bit FIN, o per situazioni anomale, che il TCP indica con il bit RST.



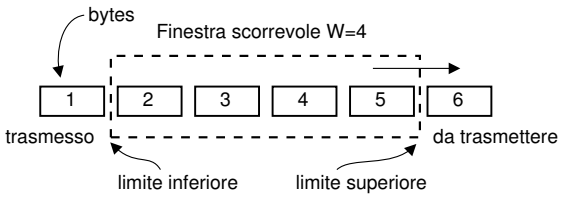
23.1.2.3 Protocollo a finestra

La funzione di controllo di flusso (ossia il dosaggio del ritmo con cui trasmettere i pacchetti) viene attuata dal TCP sfruttando la conoscenza del numero di riscontro NR inviato dal ricevente.

La lunghezza di *Finestra* comunicata con il SYN del ricevente determina la quantità di memoria riservata per i buffer dedicati alla connessione, che viene gestita come una memoria a scorrimento o *finestra scorrevole* (SLIDING WINDOW). Questa memoria è presente per gestire i casi di pacchetti ritardati o fuori sequenza, e contiene i bytes già trasmessi. Il trasmittente (vedi figura) non fa avanzare il limite inferiore finché

[20] HANDSHAKE = *stretta di mano.*

[21] Per ciò che riguarda i valori dei numeri di riscontro NR, questi sono incrementati di 1, perché la *finestra* (descritta nel seguito) inizia dai bytes del prossimo pacchetto, a cui competeranno appunto valori di NS incrementati di uno.



non riceve un riscontro con NR maggiore di tale limite. In questo modo non occorre attendere il riscontro di tutti i bytes, o di tutti i pacchetti (che devono comunque essere di dimensione inferiore alla finestra), ma ci può avvantaggiare trasmettendo l'intero contenuto della finestra.

Una finestra del tutto analoga è utilizzata dal ricevente, allo scopo di ricomporre l'ordine originario dei pacchetti consegnati disordinatamente dallo strato IP di rete. Non appena il ricevente completa un segmento contiguo al limite inferiore, sposta quest'ultimo in avanti di tanti bytes quanti ne è riuscito a leggere in modo contiguo, ed invia un riscontro con NR pari al più basso numero di byte che ancora non è pervenuto[22].

Nel caso in cui sia settato il bit URG[23] significa che si stanno inviando dati urgenti fuori sequenza, e che non devono rispettare il protocollo a finestra, come ad esempio per recapitare un segnale di interrupt relativo ad una sessione Telnet per terminare una applicazione remota.

Controllo di errore Trascorso un certo tempo (detto *timeout*) nell'attesa di un riscontro, il trasmittente ritiene che alcuni pacchetti siano andati persi, e li re-invia[24]. Il valore del *timeout* viene calcolato dinamicamente dal TCP in base alle sue misure di *round-trip delay*[25], ossia del tempo che intercorre in media tra invio di un pacchetto e ricezione del suo riscontro. In questo modo il TCP si adatta alle condizioni di carico della rete ed evita di ri-spedire pacchetti troppo presto o di effettuare attese inutili. In particolare, nel caso di rete congestionata aumenta la frequenza dei pacchetti persi, e valori di *timeout* troppo ridotti potrebbero peggiorare la situazione.

Controllo di flusso Il meccanismo a finestra scorrevole determina, istante per istante, il numero massimo di bytes che possono essere trasmessi verso il destinatario, e pertanto consente al nodo meno veloce di adeguare la velocità di trasmissione alle proprie capacità. La dimensione della finestra può essere variata (su iniziativa del ricevente) nel corso della connessione, in accordo al valore presente nel campo *Finestra* dell'intestazione TCP. Ad esempio, una connessione può iniziare con una dimensione di finestra ridotta, e poi aumentarla nel caso in cui non si verifichino errori, la rete

[22] Il riscontro può viaggiare su di un pacchetto già in "partenza" con un carico utile di dati e destinato al nodo a cui si deve inviare il riscontro. In tal caso quest'ultimo prende il nome di PIGGYBACK (*rimorchio*), o *riscontro rimorchiato*.

[23] In tal caso, il campo *Puntatore Urgente* contiene il numero di sequenza del byte che delimita superiormente i dati che devono essere consegnati urgentemente.

[24] Il mancato invio del riscontro può anche essere causato dal verificarsi di un *cecksum* errato dal lato ricevente, nel qual caso quest'ultimo semplicemente evita di inviare il riscontro, confidando nella ritrasmissione per timeout.

[25] Con licenza poetica: *il ritardo del girotondo*, che qui raffigura un percorso di andata e ritorno senza soste.

sopporti il traffico, ed i nodi abbiano memoria disponibile.

Controllo di congestione Il TCP può usare la sua misura di *round-trip delay* come un indicatore di congestione della rete, e lo scadere di un *timeout* come un segnale del peggioramento della congestione. In tal caso la dimensione della finestra di trasmissione può essere ridotta, riducendo così il carico della rete.

23.1.2.4 UDP

Lo *User Datagram Protocol* è ancora un protocollo di trasporto, che opera senza connessione, e sostituisce il TCP per inviare pacchetti isolati, o serie di pacchetti la cui ritrasmissione (se perduti) sarebbe inutile. Ad esempio, è utilizzato nella trasmissione di dati in tempo reale, oppure per protocolli di interrogazione e controllo come il DNS.

23.1.3 IP

L'Internet Protocol costituisce *l'ossatura* di Internet, realizzandone i servizi di rete ed interfacciando le diverse sottoreti a cui sono connessi i nodi. Le sue principali funzioni sono pertanto l'indirizzamento, l'instradamento e la variazione della dimensione[26] dei pacchetti prodotti dal TCP o da altri protocolli degli strati superiori. Ogni pacchetto è inviato come un messaggio indipendente, in modalità datagramma; la consegna dei datagrammi non è garantita[27], e questi possono essere persi, duplicati o consegnati fuori sequenza.

l'IP riceve dallo strato superiore (il TCP od un altro protocollo) un flusso di byte suddivisi in pacchetti, a cui si aggiunge l'indirizzo IP di destinazione; tale flusso è utilizzato per riempire un proprio buffer di dimensione opportuna, che quando pieno (od al termine del pacchetto ricevuto *dall'alto*) è *incapsulato* aggiungendo una intestazione (*l'header*) che codifica la segnalazione dello strato di rete realizzato dal protocollo IP.

23.1.3.1 Intestazione IP

Codifica le informazioni mostrate nella figura a lato. Il campo *VER* indica quale versione si sta utilizzando, e permette sperimentazioni e miglioramenti senza interrompere il servizio. *HLEN* e *TLEN* indicano rispettivamente la lunghezza dell'header e di tutto il pacchetto, mentre *TOS* codifica un *Type of Service* per differenziare ad esempio la QoS[28] richiesta. L'*identificazione* riporta lo stesso valore

1	5	9	17	20 32
VER	HLEN	TOS	TLEN	
Identificazione			Flags	Frag. Offset
TTL	Protocollo		Checksum	
IP Address Sorgente				
IP Address Destinazione				
Opzioni			Riempimento	

[26] L'IP può trovarsi a dover inoltrare i pacchetti su sottoreti che operano con dimensioni di pacchetto inferiori. Per questo, deve essere in grado di frammentare il pacchetto in più datagrammi, e di ricomporli nell'unità informativa originaria all'altro estremo del collegamento.

[27] Si suppone infatti che le sottoreti a cui sono connessi i nodi non garantiscano affidabilità. Ciò consente di poter usare sottoreti le più generiche (incluse quelle affidabili, ovviamente).

[28] La Qualità del Servizio richiesta per il particolare datagramma può esprimere necessità particolari, come ad esempio il ritardo massimo di consegna. La possibilità di esprimere questa esigenza a livello

per tutti i frammenti di uno stesso datagramma, mentre l'*Offset di frammento* indica la posizione del frammento nel datagramma (con frammenti di dimensione multipla di 8 byte).

Solo 2 dei tre bit di *Flags* sono usati, DF (*Don't Fragment*) per richiedere alla rete di non frammentare il datagramma, e MF (*More Fragments*) per indicare che seguiranno altri frammenti. Il *TTL* (Time To Live) determina la massima permanenza del pacchetto nella rete[29], il *protocollo* indica a chi consegnare il datagramma all'arrivo (ad es. TCP o UDP), e *Checksum* serve per verificare l'assenza di errori nell'header[30].

Gli *Indirizzi IP* di sorgente e destinazione hanno l'evidente funzione di recapitare correttamente il messaggio, mentre il campo *Opzioni* ha una lunghezza variabile, può essere omesso, e consente ad esempio di richiedere il tracciamento della serie di router attraversati.

23.1.3.2 Indirizzamento e Routing

A pagina 798 si è anticipata la relazione che lega la parte iniziale dell'indirizzo IP ad una determinata sottorete, in modo da partizionare i 2^{32} indirizzi su di una gerarchia a due livelli e delegare la consegna all'host finale ad uno o più router responsabili di servire la sottorete[31]. In realtà la gerarchia presenta una ulteriore suddivisione, dettata sia da esigenze amministrative che funzionali.

I bit più significativi dell'indirizzo IP identificano 5 diversi gruppi (o *classi*) di indirizzi, descritti dalla seguente tabella:

Inizio IP addr	Classe	bit rete/nodo	N. reti	N. nodi per rete
0	A	7/24	128	16 777 216
10	B	14/16	16 384	65 536
110	C	21/8	2 097 152	256
1110	D	28 bit di indirizzo multicast per 268 435 456 canali		
11110	E	27 bit per usi futuri e ricerca		

Quando una organizzazione decide di essere presente in Internet, richiede l'assegnazione di un lotto di indirizzi IP ad apposti organismi, i quali attribuiscono all'organizzazione un gruppo di indirizzi di classe A, B o C in base al numero di nodi che l'organizzazione prevede di mettere in rete. Una rete in classe B ad esempio è

IP fa parte dello standard, ma per lunghi anni non se ne è fatto uso. L'avvento delle comunicazioni multimediali ha risvegliato l'interesse per il campo TOS.

[29]Lo scopo del TTL è di evitare che si verifichino fenomeni di loop infinito, nei quali un pacchetto "rimbalza" tra due nodi per problemi di configurazione. Per questo, TTL è inizializzato al massimo numero di nodi che il pacchetto può attraversare, e viene decrementato da ogni nodo che lo riceve (e ritrasmette). Quando TTL arriva a zero, il pacchetto è scartato.

[30]In presenza di un frammento ricevuto con errori nell'header viene scartato tutto il datagramma di cui il frammento fa parte, delegando allo strato superiore le procedure per l'eventuale recupero dell'errore.

[31]Possiamo portare come analogia un indirizzo civico, a cui il postino consegna la corrispondenza, che viene poi smistata ai singoli condomini dal portiere dello stabile. Il servizio postale, così come la rete Internet, non ha interesse di sapere come sono suddivise le sottoreti delle diverse organizzazioni, ed i router instradano i pacchetti IP in base alla parte "rete" dell'indirizzo, delegando ai router della rete di destinazione il completamento dell'instradamento.

individuata da 14 bit (ossia, assieme ai bit di classe, dai primi due bytes dell'indirizzo IP), e quindi esistono $2^{14} = 16384$ diverse reti in classe B, ognuna con una capacità di $2^{16} = 65536$ diversi nodi. Chi è intestatario di un gruppo di indirizzi, provvede ad assegnarli ai singoli nodi della propria sottorete.

23.1.3.3 Subnetting e Supernetting

Osserviamo ora che la maschera di sottorete presentata a pag. 798 *non* coincide con il gruppo di bit che identifica la classe e la rete: infatti, l'insieme di indirizzi 151.100.x.y corrisponde ad una rete in classe B, mentre la maschera di sottorete 255.255.255.0 individua una sottorete in classe C. Praticamente, la rete in classe B è stata ulteriormente suddivisa (*subnettata*) in 256 sottoreti di classe C, permettendo di realizzare un instradamento gerarchico su due livelli nell'ambito dell'organizzazione intestataria della rete in classe B[32]. L'operazione inversa (detta *supernetting*), ossia quella di aggregare più reti di dimensione ridotta in una di dimensione maggiore, ha senso all'interno del router che instrada il traffico verso l'organizzazione intestataria delle sottoreti, in quanto permette di ridurre la dimensione delle tabelle di routing, che contengono così un solo elemento relativo alla super-rete, anziché un elemento per ogni singola sottorete.

23.1.3.4 Classless Interdomain Routing - CIDR

Nella prima metà degli anni '90 apparve evidente che il partizionamento degli indirizzi nelle tre classi A, B e C non era rispondente alle richieste dell'utenza; accadeva infatti che le reti in classe C erano troppo "piccole", mentre quelle in classe B rischiavano di esaurirsi a breve, pur essendo sfruttate molto poco[33]. Per questo motivo, è stata rimossa la suddivisione rigida nelle tre classi, e si è sistematicamente applicato il principio del supernetting. In pratica, si è ridefinita la maschera di sottorete, come una sequenza di *uni* allineata a sinistra, permettendo così di definire reti di dimensione pari a una potenza di due qualsiasi. Come risultato, ora una sottorete è identificata da una coppia indirizzo/maschera del tipo (ad es.) 172.192.0.0/12, che rappresenta tutti 2^{20} indirizzi che vanno da 172.192.0.0 a 172.207.255.255, che hanno i 12 bit più elevati uguali a 101011001100: questa sequenza prende il nome di *prefisso* della rete. In definitiva quindi, la maschera è espressa come il numero di bit più significativi in comune a tutti i nodi della sottorete.

23.1.3.5 Longest Match

Un router decide su che porta instradare un pacchetto IP in base al confronto tra l'indirizzo di destinazione e tutti i prefissi presenti nella tabella di routing, associati ciascuno alla "migliore" porta di uscita verso la sottorete definita dal prefisso. Nel caso in cui si verifichi più di una uguaglianza, si sceglie l'instradamento caratterizzato dal *maggior numero* di bit coincidenti, ossia relativo al prefisso *più lungo*. Infatti, in tal

[32] In questo caso, l'Università di Roma "La Sapienza" è intestataria della rete 151.100.

[33] Ad esempio, organizzazioni con poco più di un migliaio di nodi erano costrette a richiedere una intera classe B con capacità di 65536 nodi.

modo viene preferita la direzione *più specifica* verso la destinazione finale. In assenza di uguaglianze invece, il pacchetto è inoltrato in base ad una *default route*, che tipicamente rimanda la decisione ad un router "gerarchicamente più elevato"[34].

23.1.3.6 Sistemi Autonomi e Border Gateway

Vi sono router collegati direttamente con le LAN, e configurati per instradare correttamente i pacchetti diretti a destinazioni locali. Vi sono poi router collegati solo ad altri router, che *apprendono* gli instradamenti verso le reti locali mediante appositi *protocolli di routing*[35] che consentono ai router di primo tipo di *pubblicizzare* (ADVERTISE) le reti raggiungibili direttamente, ed ai router del secondo tipo di fare altrettanto nei confronti dei loro pari.

L'insieme di sottoreti (e router, nodi e DNS) gestite da una stessa organizzazione prende il nome di *Autonomous System* (AS), e nel suo ambito sono attivi protocolli di routing detti *Interior Gateway Protocols* (IGP), che distribuiscono le informazioni di raggiungibilità interna. Alcuni router di uno stesso AS svolgono il ruolo di *Border Gateway* (BG), e comunicano con i BG di altri AS mediante appositi *Exterior Gateway Protocols* (EGP), pubblicizzando all'esterno le proprie sottoreti, apprendendo dagli altri BG la raggiungibilità delle sottoreti esterne, e distribuendo tali informazioni ai router interni. Un compito particolare dell'EGP, è quello di attuare qualche politica nei confronti del *traffico di transito* tra due AS diversi dall'AS di cui il BG fa parte: in tal caso, il protocollo prende il nome di *Border Gateway Protocol* (BGP).

L'applicazione del CDIR comporta, per ogni scambio di informazioni di routing, la necessità di aggregare o disaggregare i prefissi di sottorete, in modo da mantenere al minimo la dimensione delle tabelle di instradamento.

23.1.3.7 Multicast

Tornando all'esame della tabella di pag. 804, in cui la classe E costituisce evidentemente una "riserva" di indirizzi per poter effettuare sperimentazioni, la classe D individua invece dei canali *multicast*[36]. Quando un nodo decide di aderire ad un canale multicast, invia un messaggio[37] in tal senso al proprio router più vicino, che a sua volta si occupa di informare gli altri router. Questi ultimi provvederanno quindi, qualora osservino transitare un pacchetto avente come destinazione un canale multicast, ad instradarlo verso l'host aderente. In presenza di più nodi nella stessa sottorete in ascolto dello stesso canale, solo una copia dei pacchetti attraverserà il router: il traffico multicast[38]

[34] Sebbene la topologia di Internet possa essere qualunque, nella pratica esistono dei *carrier* internazionali che svolgono la funzione di *backbone* (spina dorsale) della rete, interconnettendo tra loro i continenti e le nazioni.

[35] Vedi ad es. `https://didattica-2000.archived.uniroma2.it/rt/deposito/rt04-12.pdf`

[36] Il termine *multicast* è ispirato alle trasmissioni *broadcast* effettuate dalle emittenti radio televisive.

[37] Mediante il protocollo IGMP (*Internet Group Management Protocol*) che opera sopra lo strato IP, ma (a differenza del TCP) fa uso di datagrammi non riscontrati, similmente all'UDP ed all'ICMP.

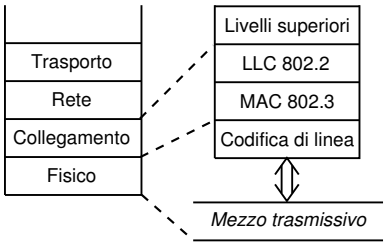
[38] Data l'impossibilità a stabilire un controllo di flusso con tutti i destinatari, il traffico multicast viaggia all'interno di pacchetti UDP.

evita infatti di aprire una connessione dedicata per ogni destinatario, ma si suddivide via via nella rete solo quando i destinatari sono raggiungibili da vie diverse.

23.1.4 Ethernet

Ci occupiamo qui di un caso particolare di realizzazione dei primi due livelli del modello ISO-OSI. Come anticipato a pag. 797, molti nodi di Internet sono univocamente individuati da un indirizzo (Ethernet) di 6 byte che, sebbene sia unico al mondo, viene usato solamente nell'ambito della LAN a cui il nodo è connesso, in quanto la distribuzione mondiale degli indirizzi Ethernet è casuale[39]: se infatti questi fossero usati come indirizzi a livello di rete, le tabelle di instradamento dovrebbero essere a conoscenza di *tutti* i nodi esistenti[40]. Puntualizziamo inoltre che un nodo di Internet può essere connesso alla rete anche in svariati altri metodi come mediante modem telefonico, rete cellulare, WiFi: qui ci limitiamo a descrivere il caso delle LAN Ethernet, peraltro particolarmente diffuso.

Ethernet individua un particolare tipo di pacchetto dati, adottato inizialmente dalla Xerox, adatto ad incapsulare dati provenienti da protocolli diversi. Successivamente, il formato è stato standardizzato dall'IEEE, e per ciò che ci interessa le specifiche sono quelle identificate dalle sigle 802.2 (Logical Link Control, LLC) e 802.3 (Carrier Sense Multiple Access - Collision Detect, CSMA/CD).



La figura mostra il legame tra queste due sigle e gli strati del modello; lo strato MAC in cui si realizza il CSMA/CD individua il Media Access Control. Il mezzo trasmissivo è un cavo, coassiale o coppia simmetrica, sul quale sono collegati tutti nodi della LAN, che si *contendono* il mezzo trasmissivo, in quanto vi può trasmettere solo un nodo per volta. Inoltre, tutti i nodi sono in ascolto sullo stesso mezzo per ricevere i pacchetti a loro destinati, riconoscibili per la presenza del proprio indirizzo Ethernet nel campo destinazione. Un pacchetto Ethernet può inoltre riportare un indirizzo di destinazione particolare, detto di *Broadcast*, che obbliga *tutti* i nodi presenti alla ricezione del pacchetto.

23.1.4.1 Address Resolution Protocol - ARP

Quando un pacchetto IP giunge ad un router, e l'indirizzo IP indica che il destinatario è connesso ad una delle LAN direttamente raggiungibili dal router[41], questo invia su quella LAN un pacchetto *broadcast*, su cui viaggia una richiesta ARP (ADDRESS RESOLUTION

[39] E rappresenta quindi ciò che viene detto uno *spazio di indirizzi piatto* (FLAT ADDRESS SPACE).

[40] Al contrario, il partizionamento dell'indirizzo IP in rete+nodo permette di utilizzare tabelle di routing di dimensioni gestibili.

[41] Ad ogni porta del router è associata una coppia sottorete/maschera (vedi pag. 798) che descrive l'insieme degli indirizzi direttamente connessi alla porta. La verifica di raggiungibilità (o *adiacenza*) è attuata mettendo in AND l'IP di destinazione con le maschere, e confrontando il risultato con quello dell'AND tra le maschere e gli indirizzi delle sottoreti collegate.

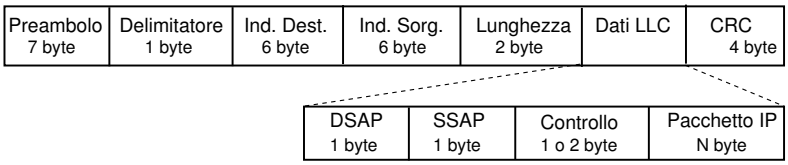


Figura 23.1: Formato di un pacchetto (o *trama*) Ethernet

PROTOCOL), allo scopo di individuare l'indirizzo Ethernet del nodo a cui è assegnato l'indirizzo IP di destinazione del pacchetto arrivato al router. Se tale nodo è presente ed operativo, riconosce che la richiesta è diretta a lui, ed invia un pacchetto di risposta comunicando il proprio indirizzo Ethernet, che viene memorizzato dal router in una apposita tabella[42].

Operazioni simili sono svolte da ognuno dei nodi della LAN, ogni volta che debbano inviare un pacchetto ad un altro nodo direttamente connesso alla stessa rete locale. Se al contrario l'IP di destinazione non fa parte della stessa LAN, il pacchetto è inviato al *default gateway*.

23.1.4.2 Formato di pacchetto

Il pacchetto Ethernet è generato dall'LLC e dal MAC, ognuno dei quali incapsula il pacchetto IP con le proprie informazioni di protocollo, con il risultato finale mostrato in fig. 23.1 . In testa troviamo 7 byte di *preambolo*, necessario a permettere la sincronizzazione dell'orologio del ricevente con quello in trasmissione; dato che la sincronizzazione richiede un tempo non noto a priori, un byte di *flag* segnala l'inizio del pacchetto. Troviamo quindi gli *indirizzi Ethernet* di sorgente e destinazione, due byte che indicano la *lunghezza* della restante parte del pacchetto, e quindi l'incapsulamento dei dati prodotti dall'LLC. In fondo, sono presenti 4 byte che realizzano il *controllo di errore*.

L'LLC da parte sua inserisce (in testa al pacchetto IP) due indirizzi *SAP* (SERVICE ACCESS POINT) di sorgente e destinazione, da utilizzare per indicare il codice che identifica il tipo di rete e/o protocollo del pacchetto incapsulato (ad es., IP od ARP). Nel campo di *controllo* possono essere anche ospitati numeri di sequenza, per i casi che lo possano richiedere, ed infine troviamo il pacchetto IP originario.

D'altra parte, per ovviare al numero limitato di possibili incapsulamenti esprimibili utilizzando solo gli 8 bit dei campi SAP, è stata introdotta una estensione all'LLC denominata SNAP (*Subnetwork Access Protocol*)[43] che pone i campi DSAP, SSAP e controllo pari a `0xAAAA03`, a cui aggiunge altri 5 bytes, dei quali i primi tre sono denominati OUI (*Organizationally Unique Identifier*) che, se posti tutti a zero, stabiliscono che i due

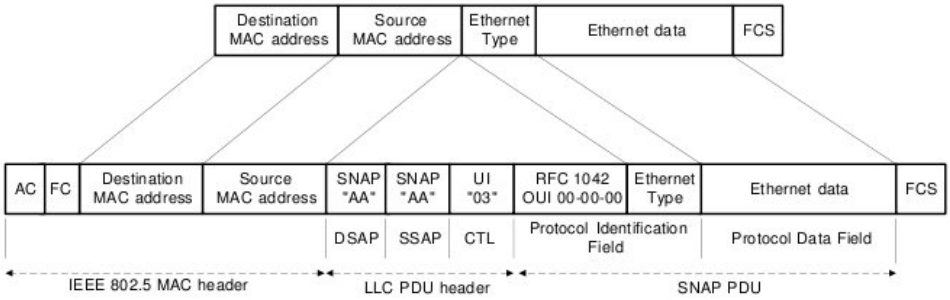
[42] Dato che i nodi possono essere spostati, possono cambiare scheda di rete e possono cambiare indirizzo IP assegnatogli, la corrispondenza IP-Ethernet è tutt'altro che duratura, ed ogni riga della tabella ARP indica anche quando si sia appresa la corrispondenza, in modo da poter stabilire una scadenza, ed effettuare nuovamente la richiesta per verificare se sono intervenuti cambiamenti topologici.

Se il nodo ha cambiato IP, ma non il nome, sarà il TTL del DNS (mantenuto aggiornato per il dominio del nodo) a provocare il rinnovo della richiesta dell'indirizzo.

[43] `http://en.wikipedia.org/wiki/Subnetwork_Access_Protocol`

byte seguenti (indicati come *protocol ID*) debbano essere interpretati come un codice *Ethertype*[44], lo stesso usato nel formato *Ethernet II* discusso appresso, permettendo quindi di specificare finalmente il protocollo incapsulato.

Infine, viene molto frequentemente usato un formato di trama ancora diverso, detto *Ethernet II* o DIX[45], che corrisponde a quello definito inizialmente prima che l'IEEE emettesse gli standard della serie 802, e che usa i 16 bit del campo *lunghezza* per indicare direttamente l'*Ethertype* della SDU incapsulata, ed omette i campi DSAP, SSAP e di controllo. In tal caso, il campo lunghezza rappresenta un numero più grande di 0x0600, maggiore della massima lunghezza prevista, e ciò fa si che venga interpretato come codice Ethertype, e che se sono incapsulati pacchetti IP, vale 0x0800. La figura seguente, tratta dal documento dell'IEEE, illustra la corrispondenza tra i campi del formato SNAP e DIX.



23.1.4.3 Collisione

Come anticipato, il mezzo trasmissivo è in comune con tutti i nodi, e dunque si è studiata una particolare soluzione il cui nome CSMA/CD indica che l'*Accesso Multiplo* avviene in due fasi: prima di trasmettere, si ascolta se non vi sia già qualcuno che trasmette (CARRIER SENSE), e durante la trasmissione, si verifica che nessun altro stia trasmettendo contemporaneamente (COLLISION DETECT). Pertanto, ogni nodo che debba trasmettere si pone prima in ascolto, e se osserva che già vi sono trasmissioni in corso, attende un tempo casuale e riprova. Quando trova il mezzo "libero", inizia a trasmettere, ma contemporaneamente verifica che nessun altro inizi a sua volta la trasmissione: questo fatto può accadere, in virtù del tempo di propagazione[46] non nullo, e determina un periodo (detto di *contesa*, e che dipende dalla massima lunghezza del cavo) entro il quale un nodo può erroneamente credere che nessun altro stia trasmettendo.

[44] `http://en.wikipedia.org/wiki/EtherType`

[45] Vedi `http://en.wikipedia.org/wiki/Ethernet_II_framing`. La sigla DIX deriva dalle iniziali delle aziende che l'hanno definito, ossia DEC, Intel and Xerox

[46] Su di un cavo coassiale *tick* da 50 Ω, la velocità di propagazione risulta di $231 \cdot 10^6$ metri/secondo. Su di una lunghezza di 500 metri, occorrono 2.16 μsec perché un segnale si propaghi da un estremo all'altro. Dato che è permesso congiungere fino a 5 segmenti di rete per mezzo di ripetitori, e che anch'essi introducono un ritardo, si è stabilito che la minima lunghezza di un pacchetto Ethernet debba essere di 64 byte, che alla velocità di trasmissione di 10 Mbit/sec corrisponde ad una durata di 54.4 μsec, garantendo così che se si è verificata una collisione, le due parti in causa possano accorgersene.

Qualora sia rilevata una contesa, i due nodi smettono di trasmettere, e riprovano solo dopo una attesa di durata casuale.

23.1.4.4 Trasmissione

Il segnale relativo al pacchetto Ethernet viene trasmesso adottando una codifica di linea di tipo Manchester differenziale (§ 15.2.1). La configurazione con tutti i nodi collegati su di uno stesso cavo è detta *a bus*, e sono state coniate apposite sigle per identificare il tipo di connessione, come ad esempio 10BASE5 e 10BASE2, relative al collegamento di banda base a 10 Mbps, su cavo *tick* e *thin*[47], con estensione massima 500 e 200 metri[48].

23.1.5 Fast e Gigabit Ethernet

Mentre si proponeva ATM come una soluzione idonea per quasi tutti gli ambiti, la tecnologia Ethernet ha incrementato la velocità trasmissiva di un fattore pari a mille, e si propone sempre più come soluzione generalizzata.

23.1.5.1 Fast Ethernet

Nel 1995 è stato definito lo standard IEEE 802.3u detto *Fast Ethernet*, che eleva la velocità di trasmissione a 100 Mbps ed impiega due diversi cavi UTP[49] per le due direzioni di trasmissione, rendendo eventualmente la comunicazione *full-duplex*[50]. In quest'ambito sono definiti i sistemi 10BASET e 100BASET, relativi all'uso del cavo UTP anziché di un coassiale, e prevedono una topologia *a stella* per la LAN, realizzata utilizzando una unità centrale (detta HUB o *mozzo di ruota*) da cui si dipartono tanti cavi, ognuno che collega un unico nodo. Nel caso di un HUB economico, questo svolge solo le funzioni di ripetitore (ritrasmette tutto su tutte le sue porte) e dunque le collisioni possono ancora verificasi.

23.1.5.2 LAN Switch

D'altra parte, i dispositivi detti BRIGDE o LAN SWITCH *apprendono* dai pacchetti in transito gli indirizzi ethernet dei nodi collegati alle porte, ed evitano di ritrasmettere i pacchetti sulle porte dove *non si trova* il destinatario. Dato che gran parte del traffico è inviato verso il *gateway* della LAN, lo SWITCH apprende in fretta su che porta questo si trovi, cosicché tutti i pacchetti destinati all'esterno non sono ritrasmessi sugli altri rami della LAN, ed il traffico tra i nodi connessi allo SWITCH non si propaga al resto della LAN.

La lunghezza massima dei collegamenti è ora ridotta a 100 metri, per il motivo che un pacchetto di dimensione minima di 64 byte trasmesso a 100 Mbps, impiega un tempo che è $\frac{1}{10}$ di quello relativo alla velocità di 10 Mbps, e quindi per consentire la

[47] TICK = *duro* (grosso), THIN = *sottile*. Ci si riferisce al diametro del cavo.

[48] Le sigle indicano infatti la velocità, se in banda base o meno, e la lunghezza della tratta.

[49] UNSHIELDED TWISTED PAIR (UTP), ossia la coppia ritorta non schermata.

[50] La trasmissione *full-duplex* si instaura quando entrambe le interfacce agli estremi ne sono capaci. Una interfaccia *half-duplex* deve invece gestire situazioni *interne* di collisione, quando un pacchetto uscente da un nodo si scontra con uno entrante.

detezione di collisione, si è dovuta ridurre di pari misura la massima distanza tra nodi trasmittenti.

23.1.5.3 Dominio di broadcast e VLAN

Anche se i dispositivi BRIDGE e SWITCH evitano di trasmettere traffico verso le porte diverse da quella di destinazione, alcuni pacchetti devono comunque essere ritrasmessi in tutte le direzioni: si tratta del traffico *broadcast*, diretto verso un ben preciso insieme di indirizzi ethernet, ed usato per funzioni di coordinamento tra i nodi della LAN, come ad esempio l'*esplora risorse di rete*. I traffico broadcast non esce dalla LAN, arrestandosi al router di livello IP; una eccessiva presenza di traffico broadcast può però pregiudicare l'efficienza sia della LAN che dei suoi nodi, oltre che produrre problemi di sicurezza; per questo si è sviluppata la possibilità di assegnare le porte di uno switch a diversi *domini di broadcast*, detti *LAN virtuali* (VLAN), che non scambiano traffico, realizzando di fatto molteplici LAN con uno stesso cablaggio. Per interconnettere le LAN, occorre attraversare un dispositivo router.

23.1.5.4 Gigabit Ethernet

Nel giugno 1998 viene standardizzato l'IEEE 802.3z, che porta ad 1 Gbps la velocità di trasmissione delle trame Ethernet, rimpiazzando lo strato di codifica di linea dell'802.3 con i due strati inferiori dell'ANSI X3T11 *Fiber Channel*[51]. In questo modo, si mantiene la compatibilità con gli strati LLC e MAC di Ethernet, mentre la trasmissione avviene su fibra ottica o su cavo in accordo alla tabella seguente.

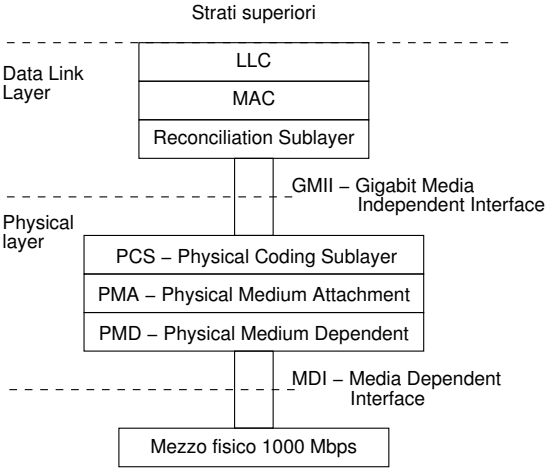
media	distanza	mezzo	sorgente
1000BASE-SX	300 m	f.o. multimodo ϕ 62.5 μm	laser 850 nm
	550 m	f.o. multimodo ϕ 50 μm	laser 850 nm
1000BASE-LX	550 m	f.o. multimodo ϕ 50 o 62.5 μm	laser 1300 nm
	3000 m	f.o. monomodo ϕ 9 μm	laser 1300 nm
1000BASE-CX	25 m	cavo STP (*shielded twisted pair*)	
1000BASE-T	25-100 m	4 coppie di cavo UTP	

23.1.5.5 Packet bursting

Dato che ora la velocità di trasmissione è 10 volte quella del fast Ethernet, la compatibilità con il MAC CSMA/CD richiederebbe di ridurre la massima lunghezza del collegamento a 10 metri. Al contrario, è stata aumentata la durata minima di una trama portandola a 512 byte, in modo da aumentare la durata della trasmissione e garantire la detezione di collisione. In effetti, il MAC ethernet continua a produrre pacchetti di durata minima 64 byte, e questi sono riempiti (*padded*) fino a 512 byte con una *carrier extension* di simboli speciali. Questa operazione è particolarmente inefficiente se i pacchetti da 64 byte sono frequenti; in tal caso si attua allora il *packet bursting* che, esauriti i 512 byte minimi realizzati come indicato, accoda gli ulteriori pacchetti nello stesso burst trasmissivo, fino ad una lunghezza di 1500 byte.

[51] Vedi ad es. `https://it.wikipedia.org/wiki/Fibre_Channel`

23.1.5.6 Architettura di Gigabit Ethernet



La figura a lato mostra la pila protocollare per Gigabit Ethernet. La GMII permette di usare lo strato MAC con qualunque strato fisico, ed opera sia in full-duplex che in half-duplex, alle velocità di 10, 100 e 1000 Mbps, mediante due percorsi dati (Tx e Rx) da 8 bit, più due segnali di strato per indicare presenza di portante e detezione di collisione, che sono mappati dal RS nelle primitive riconosciute dallo strato MAC preesistente.

Lo strato fisico è suddiviso in tre sottolivelli. Il PCS fornisce una interfaccia uniforme al RS per tutti i media. Provvede alla conversione 8B/10B tipica del *Fiber Channel*, che rappresenta gruppi di 8 bit mediante *code group* da 10 bit, alcuni dei quali rappresentano i simboli, ed altri sono codici di controllo, come quelli usati per la *carrier extension*. Il PCS genera inoltre le indicazioni sulla portante e sulla collisione, e gestisce la auto-negoziazione sulla velocità di trasmissione e sulla bidirezionalità del media.

Il PMA provvede alla conversione parallelo-serie e viceversa, mentre il PMD definisce l'MDI, ossia la segnalazione di strato fisico necessaria ai diversi media, così come il tipo di connettore.

23.1.5.7 Ripetitore full-duplex e controllo di flusso

Se tutte le porte di un ripetitore sono di tipo full-duplex, allora non può più verificarsi contesa di accesso al mezzo; semmai la contesa avviene all'interno del ripetitore, che (non essendo un SWITCH) copia tutte le trame in ingresso (debitamente bufferizzate in apposite code) in tutte le code associate alle porte di uscita. Pertanto, la lunghezza massima dei collegamenti non è più dettata dalla necessità di rilevare collisioni, ma dalle caratteristiche del mezzo trasmissivo. D'altra parte, possono verificarsi situazioni di *flooding* delle code di ingresso; il comitato IEEE 802.3x ha quindi definito un meccanismo di controllo di flusso, che mette in grado i ripetitori (e gli switch) di richiedere ai nodi connessi la sospensione temporanea della trasmissione.

23.1.5.8 10 Gigabit Ethernet

Nel 2002 viene definito lo standard IEEE 802.3ae, che stabilisce le modalità operative di un collegamento Ethernet operante solo in full duplex su fibra ottica. Lo standard prevede di interoperare con la trasmissione SONET/SDH.

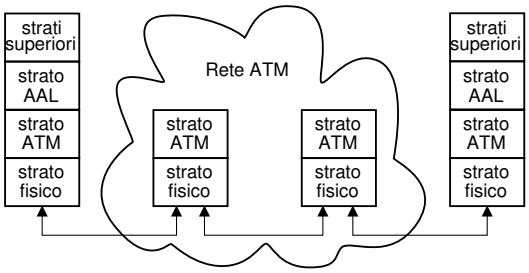
23.2 ATM

La sigla ATM sta per *Asynchronous Transfer Mode*, ed identifica una particolare rete progettata per trasportare indifferentemente traffico di diversa natura, sia di tipo

dati che real-time[52], che per questo motivo è indicata anche come B-ISDN[53]. Il suo funzionamento si basa sul principio della *commutazione di cella* (CELL SWITCHING), dove per cella si intende un pacchetto di lunghezza fissa di 53 byte. I primi 5 byte delle celle contengono un identificativo di connessione, ed il loro instradamento avviene mediante dei circuiti virtuali. La commutazione delle celle tra i nodi di rete ha luogo in maniera particolarmente efficiente, e questa è una delle caratteristiche più rilevanti dell'ATM.

23.2.1 Architettura ATM

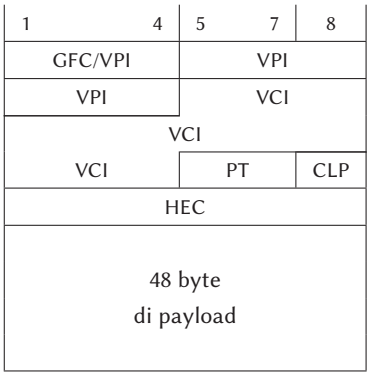
La rete ATM viene anche definita come una *Overlay Network*, in quanto operativamente si *sovrappone* ai livelli inferiori di una rete esterna. Dal canto suo, ATM è strutturata sui tre strati funzionali di adattamento (AAL), di commutazione ATM, e fisico. Mentre i nodi ai bordi della rete devono realizzare tutti e tre gli strati, i nodi interni svolgono solo le funzioni attuate da quelli inferiori. La tabella 23.1 riporta le principali funzioni svolte dai tre strati, e pone in evidenza come in uno stesso strato siano identificabili diverse sotto-funzioni.



23.2.2 Strato fisico

Il mezzo primario di trasmissione (con cui è in contatto il sotto-strato PM) per ATM è la fibra ottica, in accordo alla struttura di trama dell'SDH/SONET, per la quale sono state standardizzate le velocità di 1.5 e 2 Mbps (DS1/E1), 155 Mbps (OC3) e 622 Mbps (OC12c). La velocità di 155 Mbps è disponibile anche su FIBRE CHANNEL, e su cavo ritorto, mentre la velocità di 100 Mbps è disponibile su FDDI. Infine, sono previste anche velocità di interconnessione di 139, 52, 45, 34 e 25 Mbps.

In funzione del mezzo trasmissivo, può variare la *struttura di trama*[54] (mostrata in figura) in cui



Formato della cella ATM

[52]Per traffico real-time si intende sia quello telefonico, sia più in generale quello di natura multimediale.

[53]Siamo alla fine degli anni '80, e la definizione *Integrated Service Data Network* (ISDN) si riferisce ad una rete in grado di permettere, oltre al normale trasporto dei dati, anche servizi di rete. La rete ISDN era però limitata ad una velocità massima (presso l'utente) di 2 Mbps, e per questo venne chiamata *narrow-band* ISDN (N-ISDN). A questa, avrebbe fatto seguito la *broad-band* ISDN (B-ISDN) che ha poi dato luogo alla definizione dell'ATM.

[54]Sono definite due tipi di *interfaccia utente-network* (UNI): quella SDH/SONET, in cui le celle sono inserite nel *payload* della trama SDH, e quella CELL-BASED, che prevede un flusso continuo di celle. Mentre nel primo caso il bit rate lordo comprende *l'overhead* di trama, nel secondo comprende la presenza di celle di tipo *Operation and Maintenance* (OAM).

strato	sotto-strato	funzioni
ATM Adaptation Layer (AAL)	• Convergenza (CS)	Definisce il servizio offerto agli strati superiori
	• Segmentazione e Riassemblaggio (SAR)	Suddivide i dati in modo compatibile con la dimensione di cella, e li ricostruisce in ricezione
ATM layer		Multiplazione e demultiplazione delle celle Traslazione delle etichette VPI/VCI Generazione/estrazione dell'HEADER della cella Gestione del controllo di flusso GFC
Physical Layer (PL)	• Convergenza di trasmissione (TC)	Delimitazione delle celle Inserimento celle IDLE per adattamento velocità Generazione e verifica dell'HEC (controllo di errore) Generazione della trama di trasmissione
	• Mezzo Fisico (PM)	Temporizzazione e sincronizzazione Gestione del mezzo

Tabella 23.1: Stratificazione delle funzioni in una rete ATM

vanno inserite le celle. Il quinto byte della intestazione di cella contiene l'*Header Error Code* (HEC) calcolato sui 4 byte precedenti, che viene usato in ricezione per rivelare due errori e correggerne uno[55]. Nel caso in cui la sorgente produca dati a velocità inferiore a quella del collegamento sono inserite celle aggiuntive di tipo IDLE, rimosse al ricevitore[56]. Infine, la *delimitazione delle celle* è attuata in ricezione in base alla correlazione tra i primi quattro byte dell'header, ed il campo HEC dello stesso.

23.2.3 Strato ATM

Mentre lo strato fisico si occupa di trasmettere e ricevere celle, lo strato ATM si occupa di elaborarle. Nei nodi *di frontiera* le celle sono multiplate e demultiplate, mentre *dentro la rete* sono commutate tra gli ingressi e le uscite.

Nei primi quattro byte dell'header di cella trova posto *l'etichetta* necessaria a realizzare il trasferimento a circuito virtuale; questa etichetta è suddivisa in due campi, il *Virtual Path Identifier* (VPI) ed il *Virtual Channel Identifier* (VCI)[57].

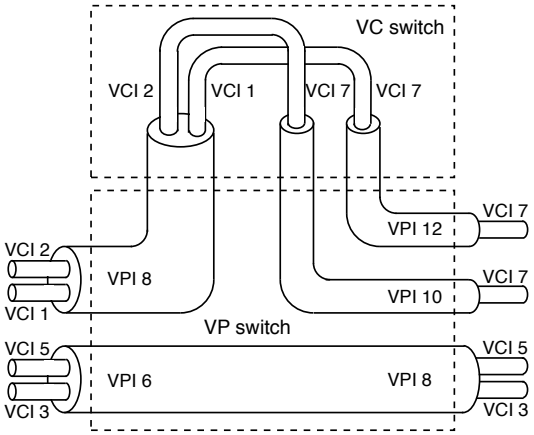
Il motivo della suddivisione risiede nella possibilità di raggruppare logicamente diversi circuiti virtuali che condividono lo stesso percorso nella rete. Nei collegamenti di cui è composto il percorso comune, viene usato uno stesso VPI per tutte le celle, mentre le diverse connessioni su quel percorso sono identificate mediante diversi VCI.

[55] Nel primo caso la cella viene scartata, mentre nel secondo inoltrata correttamente. La presenza di più di due errori, provoca un errato inoltro della cella.

[56] Le celle IDLE sono riconoscibili in base ad una particolare configurazione dei primi 4 byte dell'header, così come avviene per le celle OAM, nonché per altri tipi particolari di cella, che trasportano la segnalazione degli strati superiori.

[57] Mentre per VCI sono riservati 16 bit, per VPI si usano 12 bit all'interno della rete, e 8 bit ai suoi bordi, riservando 4 bit indicati come *Generic Flow Control* (GFC) per regolare il flusso delle sorgenti.

L'instradamento congiunto delle celle con uguale VPI è effettuato nei nodi (VP *switch*), che si occupano solo[58] di scambiare il VPI delle celle, e di porle sulla porta di uscita corretta, come indicato dalle tabelle di instradamento.



La sequenza dei nodi attraversati dall'instradamento è indicata come *Virtual Path Connection* (VPC), è composta da zero o più VP switch, ed è delimitata tra due nodi (VC o VP/VC switch[59]) che elaborano anche i VCI. La sequenza dei VC switch che elaborano i VCI, e che si estende tra due nodi che terminano lo strato di adattamento, è indicata invece con il termine *Virtual Channel Connection* (VCC)[60] e comprende uno o più VPC, coincidendo spesso[61] con il percorso tra ingresso ed uscita[62] della rete ATM.

La creazione delle tabelle di instradamento può essere di tipo manuale, dando luogo ad una *Permanent Virtual Connection* (PVC), oppure può essere il risultato di una richiesta estemporanea, dando luogo ad una *Switched Virtual Connection* (SVC)[63]; l'oggetto della richiesta può essere una VCC od una VPC, ed in questo secondo caso la VPC verrà usata per tutte le VCC future tra i due nodi.

23.2.4 Classi di traffico e Qualità del Servizio (QoS)

Nella fase di *setup* sono attuate delle verifiche dette *Connession Admission Control* (CAC) per assicurarsi che la nuova connessione non degradi le prestazioni di quelle già in corso, nel qual caso la chiamata è rifiutata. La sua accettazione determina invece la stipula tra utente e rete di un *Traffic Contract* a cui la sorgente si deve attenere. Nel corso della trasmissione, i nodi ATM verificano che le caratteristiche del traffico in transito

[58] Questa semplificazione del lavoro di instradamento, quando confrontata con quello relativo ad una rete IP, è all'origine della vocazione *fast switching* della rete ATM. Per di più, permette la realizzazione *hardware* dei circuiti di commutazione. D'altra parte, mentre per IP l'instradamento avviene al momento della trasmissione, in ATM avviene durante il *set-up* della connessione, quando le tabelle di instradamento sono inizializzate.

[59] Nel caso in cui venga invece scambiato solo il VCI, si ottiene uno switch VC puro.

[60] La rete ATM assicura la consegna delle celle di una stessa VCC nello stesso ordine con cui sono state trasmesse, mentre non assicura l'ordinamento per le celle di una stessa VPC.

[61] Può accadere infatti di incontrare uno switch VC puro, in cui è scambiato solo il VCI, ed al quale fanno capo due diverse VCC.

[62] I nodi di ingresso ed uscita sono indicati come *ingress* ed *egress* nella terminologia ATM.

[63] Nella richiesta di una SVC, l'utente invia i messaggi di *setup* su di una particolare (*well known*) coppia VPI/VCI=0/5. In generale, le prime 32 VCI di ogni VPI sono riservate per propositi di controllo. In queste, sono contenuti dei messaggi di segnalazione che aderiscono alle specifiche Q.2931, che fanno parte di *User Network Interface* (UNI) 3.1, e che sono un adattamento di Q.931 per N-ISDN. Le specifiche UNI 4.0 prevedono la negoziazione della QoS, e la capacità di richiedere una SVC per una VPC.

nelle VCC siano conformi al rispettivo contratto, svolgendo un *Usage Parameter Control* (UPC) detto anche *policing*[64]. Prima di proseguire, forniamo però alcune definizioni.

Come anticipato, ATM si è sviluppata per trasportare diversi tipi di traffico, classificabili come segue, nei termini dei parametri indicati di seguito:

- *Constant Bit Rate* (CBR) identifica il traffico real-time come la voce[65] ed il video non codificato;
- *Variable Bit Rate* (VBR) può essere di tipo real time (es. video MPEG) oppure no, ed allora può tollerare variazioni di ritardo (CDV) ma non l'eccessiva perdita di dati (CLR);
- *Available Bit Rate* (ABR) tenta di sfruttare al meglio la banda disponibile. Il contratto prevede la fornitura di un MCR da parte della rete, e le sorgenti sono in grado di rispondere ad una indicazione di congestione, riducendo di conseguenza l'attività;
- *Unspecified Bit Rate* (UBR) condivide la banda rimanente con ABR, ma non gli è riconosciuto un MBR, né è previsto nessun controllo di congestione. Le celle in eccesso sono scartate. Idonea per trasmissioni insensibili a ritardi elevati, e che dispongono di meccanismi di controllo di flusso indipendenti[66].

Le classi di traffico sono descrivibili mediante i parametri

- *Peak Cell Rate* (PCR) applicabile a tutte le classi, ma è l'unico parametro per CBR;
- *Sustainable Cell Rate* (SCR) assieme ai tre seguenti, descrive le caratteristiche di VBR: velocità comprese tra SCR e PCR sono non-conformi, se di durata maggiore di MBS;
- *Minimum Cell Rate* (MCR) caratterizza la garanzia di banda offerta alla classe ABR;
- *Maximum Burst Size* (MBS) descrive la durata dei picchi di traffico per sorgenti VBR.

Il contratto di traffico, mentre impegna la sorgente a rispettare i parametri di traffico dichiarati, vincola la rete alla realizzazione di una *Quality of Service* (QoS), rappresentata dalle grandezze (tra le altre)

- *Cell Tansfer Delay* (CTD) assieme alla seguente, è molto importante per la classe CBR;
- *Cell Delay Variation* (CDV) rappresenta la variabilità nella consegna delle celle, dannosa per le applicazioni real-time. La presenza di una CDV elevata può inoltre provocare fenomeni di momentanea congestione all'interno della rete, e può

[64] Letteralmente: POLIZIOTTAMENTO. Il controllo può anche essere effettuato su di una intera VPC.

[65] La classe CBR si presta bene a trasportare traffico telefonico PCM. In questo caso, può trasportare solo gli intervalli temporali realmente occupati.

[66] La classe UBR è particolarmente adatta al trasporto di traffico IP, in quanto questo è un protocollo senza connessione, e gli strati superiori (ad es. il TCP) sono in grado di gestire correttamente un servizio di collegamento con perdita di dati.

essere ridotta adottando degli *shaper*[67], che riducono la variabilità di ritardo a spese un aumento di CTD;

- *Cell Loss Ratio* (CLR) rappresenta il tasso di scarto di celle del collegamento.

Nel caso in cui il policing rilevi che una connessione viola le condizioni contrattuali[68] può intraprendere svariate azioni, tentando di non scartare immediatamente la cella, ma provvede comunque a segnalare l'anomalia, ponendo pari ad uno il bit *Cell Loss Priority* (CLP) dell'header. Ciò fa si che la cella divenga *scartabile*[69] in caso di congestione in altri nodi. Un ulteriore campo dell'header, il *Payload Type* (PT), può infine ospitare una *segnalazione in avanti*, che manifesta il fatto che la cella in questione ha subito congestione.

23.2.5 Indirizzamento

I nodi di una rete ATM sono identificati da un indirizzo di 20 byte, di diverso significato nei casi di reti private o pubbliche, come indicato dal primo byte (AFI). Nel primo caso, detto *formato NSAP*[70], il DCC o l'ICD sono assegnati dall'ISO, e l'indirizzo del nodo è disposto nei 10 byte indicati come *High-Order Domain Specific Part* (HO-DSP). I sei byte dell'*End System Identifier* (ESI) sono forniti dal dispositivo connesso ai bordi della rete, e coincidono con il suo indirizzo *Ethernet*: in tal modo la rete comunica un prefisso che identifica il nodo di ingresso, ed il dispositivo lo associa al proprio ESI per forgiare il proprio indirizzo completo. Infine, il byte SEL può essere usato per multiplare più entità presso il terminale, ed è ignorato dalla rete.

Rete Privata				
AFI	ICD/DCC	HO-DSP	ESI	SEL
Rete Pubblica				
AFI	E.164	HO-DSP	ESI	SEL

Nel caso di rete pubblica, il campo HO-DSP è ristretto a 4 byte, e gli 8 byte di E.164 contengono un indirizzo appartenente alla numerazione telefonica mondiale.

23.2.6 Strato di adattamento

Come mostrato in tab. 23.1, l'AAL è suddiviso in due componenti, *Segmenting and Reassembly* (SAR) e *Convergence Sublayer* (CS); le funzioni di quest'ultimo sono ulteriormente

[67] Un *sagomatore* è composto in prima approssimazione da un buffer di memoria, il cui ritmo di svuotamento *non è mai* superiore ad un valore costante.

[68] Ad esempio, una CBR supera il proprio PCR, od una VBR oltrepassa il PCR per più tempo di MBS, oppure il traffico generato da una UBR non può essere instradato per l'esaurimento della banda.

[69] Alcune classi di traffico pongono CLP=1 già in partenza, sia per una capacità indipendente di risolvere situazioni di perdita di dati, sia per la diversa natura dei dati che possono inviare, come ad esempio una codifica di segnale in cui alcuni dati posso essere interpolati, mentre altri no. Al contrario, alcune sorgenti confidano molto nel rispetto del proprio CLP=0, come ad esempio nel caso in cui queste inviino pacchetti di dati ben più grandi delle celle ATM, e che sono di conseguenza frammentati in molte unità, ed in presenza di una sola cella mancante, devono ritrasmettere l'intero pacchetto. In quest'ultimo caso, sono state elaborate strategie di *scarto precoce* (EARLY DISCARD) di tutte le celle di un pacchetto, per il quale si è già verificato lo scarto di una cella componente.

[70] Il formato NSAP si ispira al *Network Service Access Point* dell'OSI, e se ne differenzia per aver fuso i campi *Routing Domain* e *Area* in un solo campo HO-DSP, per il quale si è adottata una gerarchia di instradamento basata su di un prefisso mobile, in modo simile al CDIR dell'IP.

A	B	C	D
servizio isocrono		ritardo variabile consentito	
bit rate costante	bit rate variabile		
con connessione			senza connessione
AAL 1	AAL 2	AAL 3/4 o 5	AAL 3/4 o 5

Figura 23.2: Classi di servizio della rete ATM

ripartite tra una *Common Part* (CPCS) ed un *Service Specific* cs (SSCS).

Il compito di AAL è quello di generare i 48 byte del payload per le celle ATM a partire dalle SDU ricevute, e di ricomporre queste ultime in ricezione, a partire dal risultato della loro demultiplazione operata (in base alle etichette VPI/VCI) dallo strato ATM ricevente. Mentre il SAR si interfaccia con lo strato ATM, il CS interagisce con i protocolli superiori, e le esatte operazioni svolte dipendono dalla natura del traffico trasportato: la fig. 23.2 mostra quattro diverse situazioni.

La classe A è un classico caso CBR, ed in tal caso si adotta un AAL di tipo 1 in cui lo strato CS è assente, ed il SAR utilizza il primo dei 48 byte di cella per inserire informazioni di controllo sull'ordine di consegna, ed è di ausilio al recupero della temporizzazione di sorgente presso la destinazione.

La classe B (AAL 2) individua sorgenti multimediali a pacchetto, mentre per la C (AAL 3/4 o 5) siamo più tipicamente in presenza di una connessione dati a circuito virtuale. In questa categoria rientra il trasporto di collegamenti X.25 e *frame relay*, sia di tipo ABR che UBR. Lo stesso tipo di AAL (3/4 o 5) è infine usato anche per la classe D, in cui rientra pienamente il trasporto di traffico IP su ATM.

Quando il CS di AAL 3/4 riceve una SDU (di dimensione massima $2^{16} - 1$) dagli strati superiori, la allinea ad un multiplo di 32 byte, e vi aggiunge 32 byte in testa ed in coda con informazioni di lunghezza e di controllo di integrità. La CS-PDU risultante è passata al SAR, che la suddivide in blocchi di 44 byte, a cui ne aggiunge 2 in testa e due in coda[71], e completa così la serie di 48 byte da passare allo strato ATM. Al contrario, il SAR dell'AAL 5 suddivide la CS-PDU in blocchi da 48 byte e non aggiunge informazioni[72], demandando il riconoscimento dell'ultima cella di una stessa CS-PDU ad un bit del campo PT presente nell'header di cella ATM. D'altra parte, la lunghezza della CS-PDU dell'AAL 5 è multipla di 48 byte, aggiungendone un numero appropriato, oltre ai 64 byte di intestazione (ora posta in coda), in cui ora sono presenti anche 8 bit di informazione da utente ad utente.

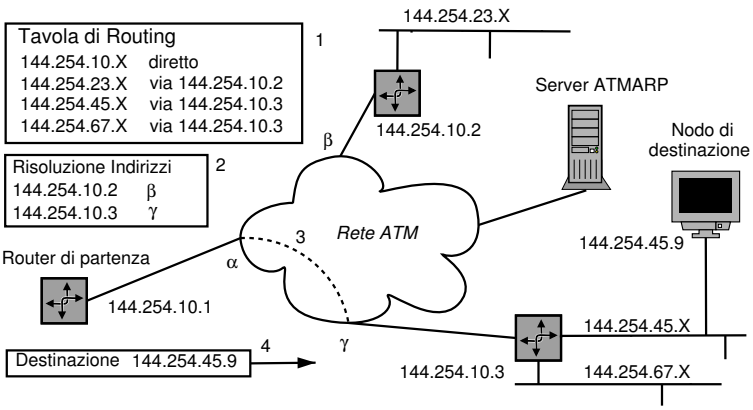
[71] Questi ultimi 4 byte contengono l'indicazione (2 bit) se si tratti della prima, ultima od intermedia cella di una stessa CS-PDU, la lunghezza dei dati validi se è l'ultima (6 bit), un numero di sequenza (4 bit), un controllo di errore (10 bit), ed una etichetta (10 bit) che rende possibile interallacciare temporalmente le celle di diverse CS-PDU.

[72] In questo modo si risparmiano 4 byte ogni 48. Ora però è indispensabile che le celle arrivino in sequenza, e non è più possibile alternare diverse CS-PDU.

23.2.7 IP su ATM classico

Allo stesso tempo in cui si diffonde l'uso di ATM tra gli operatori di TLC, il TCP/IP emerge come *lo* standard comune per l'interconnessione tra elaboratori. Sebbene il TCP/IP si appoggi ad *Ethernet* in area locale, per i collegamenti a lunga distanza[73] l'ATM presenta indubbi vantaggi come la disponibilità di banda su richiesta, la coesistenza con il traffico di tipo diverso, l'elevata efficienza della commutazione, e la possibilità di raggiungere diverse destinazioni. Una prima soluzione, subito scartata, fu quella nota come *peer model*, in cui i nodi ATM possiedono un indirizzo IP, ed usano i protocolli di routing IP. ATM risulta così *appaiata* alla rete IP, ma ciò complica la realizzazione dei nodi ATM, ed il metodo non si generalizza per protocolli diversi da IP.

L'alternativa seguita, detta *overlay model*, vede ATM come uno stato di collegamento su cui opera l'IP, che si comporta come se si trovasse su di una LAN. In particolare, solo i nodi di frontiera tra IP ed ATM prendono un doppio indirizzo, ed individuano una *Logical Subnet* (LIS) definita da uno stesso prefisso IP ed una stessa maschera di sottorete. Con riferimento alla figura che segue, quando il router di partenza vuole contattare il nodo di destinazione, trova (1) prima l'IP del router di destinazione, e quindi invia una richiesta ARP al server ATMARP presente nella LIS[74], che risponde comunicando l'indirizzo γ, il quale è così risolto (2). A questo punto si può instaurare una VCC con B mediante la segnalazione ATM (3), ed effettuare la comunicazione (4). Una tale soluzione è nota come *VC multiplexing*, ed i dati sono incapsulati direttamente nella CPCS-PDU di AAL5. In ricezione, l'etichetta VPI/VCI è usata per consegnare il pacchetto al protocollo di strato superiore che ha realizzato la connessione ATM. D'altra parte, questa elaborazione deve avvenire a *diretto contatto* con AAL5, e ciò preclude la possibilità di interlavoro con nodi esterni alla rete ATM.



Nel caso in cui sia antieconomico creare un gran numero di VC, o se si dispone unicamente di un PVC[75], il pacchetto IP viene incapsulato in un header LLC IEEE 802.2

[73] Quando la distanza tra i nodi oltrepassa dimensioni di un edificio, si parla di *Campus Network* o di *Wide Area Network* (WAN), ed a volte è usato il termine *Metropolitan Area Network* (MAN) per estensione cittadine. Per estensioni ancora maggiori si parla di *reti in area geografica*.

[74] Tutti i nodi della LIS hanno configurato manualmente l'indirizzo ATM del server ATMARP.

[75] Un VC permanente collega solamente una coppia di nodi, ed in tal caso è possibile anche fare a meno del server ATMARP, in quanto un PVC è configurato manualmente. Nei fatti, questo è l'uso più diffuso del trasporto IP over ATM, ed è tipicamente utilizzato per collegare sedi distanti di una stesso sistema

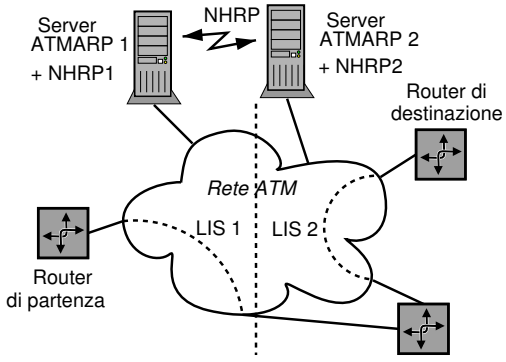
prima di essere consegnato all'AAL5. In tal modo, il router ricevente esamina l'header LLC del pacchetto ricevuto dal nodo ATM di *egress,* per consegnare il pacchetto al protocollo appropriato, realizzando così un *trasporto multiprotocollo* su ATM.

23.2.8 LANE, NHRP e MPOA

Discutiamo qui brevemente ulteriori possibilità di utilizzo di ATM come trasporto IP, ma a cui verosimilmente sarà preferito l'MPLS.

Mentre l'approccio classico aggiunge un substrato tra IP ed AAL5, per così dire *esterno* alla rete ATM, l'approccio LANE (LAN *Emulation*) ne aggiunge uno *esterno* alla rete IP, che *crede* di avere a che fare con una LAN ethernet. In questo caso anziché una LIS, si definisce una *Emulated LAN* (ELAN), il cui esatto funzionamento prevede diversi passaggi[76].

Sia nel caso classico che in quello LANE, se due router IP sono su due LAN (LIS o ELAN) differenti (con prefissi differenti) la comunicazione tra i due deve necessariamente attraversare un terzo router IP, anche se esiste un collegamento diretto tra i primi due, tutto interno alla rete ATM. La situazione è illustrata nella figura seguente, per il caso classico. Come possiamo notare, i router di partenza e di destinazione potrebbero dialogare direttamente tramite la rete ATM, diminuendo il carico di traffico della stessa, e risparmiando al router intermedio il compito di riassemblare e disassemblare i pacchetti IP in transito, oltre a riclassificarli ai fini del routing. Se i server ATMARP delle due LIS possono scambiarsi le proprie informazioni, il router di partenza può arrivare a conoscere l'indirizzo ATM di quello di destinazione, e creare un collegamento diretto. Lo scambio delle corrispondenze *<ind. IP; ind. ATM>* avviene per mezzo del *Next Hop Resolution Protocol* (NHRP) tra entità indicate come *NHRP Server* (NHS), che possono appartenere ognuno a più LIS, e che instaurano tra di loro un meccanismo



autonomo, eliminando la necessità di sviluppare in proprio un impianto di TLC tra le sedi.

[76] La emulazione di una LAN da parte della rete ATM è possibile dopo aver definito per ogni ELAN un *LAN Emulation Server (LES) a cui ogni LAN Emulation Client* (LEC) si rivolge per conoscere l'indirizzo ATM di un altro LEC, a partire da suo indirizzo MAC (la traduzione da IP a MAC è già avvenuta tramite ARP a livello IP). In una ELAN deve inoltre essere presente un dispositivo *Broadcast and Unknown Server* (BUS) che diffonde a tutti i LEC i pacchetti broadcast Ethernet (come ad es. le richieste ARP), e che viene usato dai LEC che devono inviare un broadcast. Infine, occorre un *LAN Emulation Configuration Server* (LECS) che conosce, per ogni ELAN della rete ATM, l'elenco dei LEC, del LES e del BUS.

All'accensione di un LEC, questo contatta il LECS (conoscendone l'indirizzo ATM, oppure su di una VCC ben nota, o tramite segnalazione ATM) per apprendere gli indirizzi ATM del proprio LES e del BUS. Quindi, registra presso il LES la corrispondenza tra i propri indirizzi MAC ed ATM. Quando un LEC desidera inviare dati ad un altro LEC, dopo averne risolto l'indirizzo ATM interrogando il LES, incapsula le trame IP con un header LLC IEEE 802.2 proprio come nel caso classico.

di *passa-parola*[77], per rispondere alle interrogazioni che ricevono. L'applicazione di un meccanismo in parte simile, porta nel caso delle ELAN alla definizione del *Multi Protocol over ATM* (MPOA[78]).

23.2.9 MPLS

Il *Multi Protocol Label Switching* (MPLS) è un metodo di realizzare una trasmissione a circuito virtuale su reti IP, la cui architettura è descritta nella RFC 3031 dell'IETF, e che verrà esposto meglio in una prossima edizione. Qui illustriamo i legami che MPLS presenta con ATM.

Lo sviluppo di MPLS ha origine dalle iniziative industriali tese a realizzare router Internet economici di prestazioni elevate, e capaci di gestire la banda in modo appropriato. Lo IETF ha ricevuto il compito di armonizzare in una architettura standardizzata i diversi approcci, basati sul principio di inoltrare i pacchetti in base ad una etichetta (LABEL) impostata dal primo router della rete, proprio come avviene in ATM. Dato che erano già disponibili i dispositivi hardware per realizzare i nodi di switching ATM, i primi prototipi hanno semplicemente utilizzato tali switch sotto il diretto controllo di un router IP, collegato ad altri simili tramite la rete ATM. L'MPLS è tuttavia più generale, sia verso l'alto (è *multiprotocollo* in quanto si applica oltre che ad IP, a qualunque altro strato di rete) che verso il basso (funziona indifferentemente dall'implementazione dello strato di collegamento, sia ATM, *ethernet* od altro).

La *label* apposta dal primo MPLS *Router* (LSR) dipende dalla destinazione IP del pacchetto; diverse destinazioni possono coincidere con una sola *Forwarding Equivalence Class* (FEC)[79], identificata da una singola *label*. Tutti i pacchetti di una stessa FEC sono inoltrati verso il medesimo *next hop*, indicato dalla tabella di routing, indicizzata dalla *label*[80]. Nella stessa tabella, si trova anche la nuova *label* da assegnare al pacchetto, prima di consegnarlo all'LSR seguente. In tutti i LSR successivi, il pacchetto non è riclassificato, ma solo inoltrato verso il *next hop* con una nuova *label* come ordinato dalla tabella di routing. Pertanto, è il primo LSR a decidere tutto il tragitto, ed i pacchetti

[77] I NHS risiedono su dispositivi che sono anche router IP, e che quindi mantengono aggiornate le tabelle di instradamento che indicano il prossimo salto (*next hop*) verso destinazioni IP. Le richieste di risoluzione ATMARP per un certo indirizzo IP sono instradate mediante queste stesse tabelle, giungendo di salto in salto fino al router-NHS appartenente alla stessa LIS dell'IP di destinazione, che conosce la risposta. Quest'ultima ripercorre all'indietro il percorso fatto dalla richiesta, fino alla sorgente. I router attraversati dal *passa parola, ricordano* (per un pò) le risposte trasportate, riducendo il traffico NHRP.

[78] Il metodo si basa su di un meccanismo indicato come *flow detection*, attuato dal ruoter IP-ATM prossimo alla sorgente, che è in grado di accorgersi di traffico non sporadico diretto verso una medesima destinazione. Questo router impersona allora un *MPOA Client* (MPC), ed interroga un *MPOA server* (MPS) per conoscere l'indirizzo ATM della destinazione, in modo da creare un collegamento diretto. Ogni MPS serve una o più ELAN, e gli MPS comunicano tra loro mediante il NHRP.

L'MPOA realizza la separazione tra il calcolo dell'instradamento e l'inoltro dei dati. A differenza di un ruoter tradizionale, che svolge entrambi i compiti, l'MPC svolge solo l'inoltro verso l'indirizzo ATM di destinazione, mentre quest'ultimo è fornito dall'MPS, che si comporta quindi come un *route server*.

[79] Nel routing IP tradizionale, una FEC coincide con l'instradamento individuato dal *longest match*.

[80] Nel routing IP convenzionale, per ogni router, la tabella di routing deve essere esaminata per intero per ogni pacchetto, alla ricerca del *longest match* tra le regole presenti.

di una stessa FEC seguono tutti lo stesso *Label Switched Path* (LSP). In tal modo gli switch possono essere più semplici, si possono stabilire instradamenti diversi per una stessa destinazione[81] in base al punto di ingresso, così come le FEC posso essere rese dipendenti non solo dalla destinazione, ma anche da altri parametri, come la classe di servizio richiesta.

L'associazione tra *label* e FEC (ossia il *next hop* per i pacchetti con quella *label*) è stabilita dal LSR di *destinazione*[82], e cioè un LSR indica agli LSR dai quali *si aspetta di ricevere* traffico, quale *label* usare in corrispondenza delle FEC per le quali conosce l'instradamento. Dato che la conoscenza di un istradamento è anche il prerequisito sulla cui base sono annunciate le informazioni di routing *hop-by-hop* in Internet, il *Label Distribution Protocol* (LDP) può essere vantaggiosamente associato ai protocolli di distribuzione delle informazioni di routing già esistenti (es. BGP). Le associazioni tra FEC e *label* si propagano dunque fino ai nodi di ingresso, realizzando un reticolo di "alberi" di LSP, costituiti dagli LSP definiti da una stessa FEC, e che convergono verso uno stesso *egress* a partire da diversi *ingress*. Nel nodo in cui più LSP si riuniscono, è possibile effettuare il *label merging* assegnando la stessa *label* ai pacchetti uscenti, riducendo così la dimensione delle tabelle di routing.

L'etichetta *label* su cui si basa l'MPLS può *genericamente* consistere in un incapsulamento della PDU dello strato di rete, prima che questa sia passata allo strato di collegamento. Quando i LSR sono realizzati mediante switch ATM, la *label* è efficacemente realizzata usando la coppia VPI/VCI, realizzando i LSP come delle VCC. In questo caso però, sorgono problemi nel caso in cui si debba effettuare il *merge* di più LSP relative ad una stessa FEC, che passano da uno stesso ATM-LSR. Infatti, se un nodo adottasse in uscita una stessa *label*-VCC per differenti VCC entranti, le celle in cui sono segmentati i pacchetti IP, ed ora con uguale *label*-VCC, si alternerebbero, rendendo impossibile il riassemblaggio dei pacchetti. Per questo motivo, MPLS può operare anche con LSR che non permettono il *merging*, e che possono quindi essere utilizzati assieme ad altri che ne sono capaci; in tal caso, L'LSR non-merging *non è* notificato automaticamente delle associazioni FEC-*label*, ma gli viene comunicata una (diversa) label ogni volta che ne chiede una (da associare ad una FEC), usando così più *label* del necessario. Una alternativa è quella di codificare la FEC mediante il solo VPI, ed usare il VCI per indicare il nodo di partenza. In questo modo, il *merging* è per così dire *automatico*, senza problemi di alternanza temporale delle celle di diversi pacchetti IP, ed il metodo può essere applicato se è possibile coordinare l'assegnazione dei VCI tra sorgenti diverse, e se il numero delle *label* non oltrepassa la capacità di indirizzamento.

L'esposizione svolta è volutamente semplificata, e trascura per comodità alcune importanti caratteristiche di MPLS.

[81] Il routing IP tradizionale opera su di una base *hop-by-hop*, e per questo non può tenere conto della provenienza. Quando due pacchetti per una medesima destinazione passano da uno stesso router, proseguono per lo stesso percorso.

[82] Infatti, è la *label* del pacchetto *ricevuto* che determina il *next hop*, e quindi è quest'ultimo a definire la semantica della *label* presso i propri vicini.

L'opera

Trasmissione dei Segnali e Sistemi di Telecomunicazione

è il risultato di un progetto ventennale di cultura libera, aggiornato di continuo ed evolutosi fino alla forma attuale. La sua disponibilità pubblica è regolata dalle norme di licenza Creative Commons

Attribuzione - Non commerciale - Condividi allo stesso modo

`https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it`

e tutte le risorse relative al testo sono accessibili presso

`https://teoriadeisegnali.it/libro/`

Puoi contribuire al suo successo promuovendone la diffusione e supportarne lo sviluppo attraverso una donazione, in buona parte devoluta ai progetti *open source*[1] che ne hanno resa possibile realizzazione e divulgazione. Ai donatori viene accordato un accesso *vitalizio* al formato PDF *navigabile* di tutte le edizioni presenti *e future*.

[1]
- Lyx - `http://www.lyx.org/`
- LaTeX- `https://www.latex-project.org/`
- TeX Users Group - `https://tug.org/`
- Inkscape - `http://www.inkscape.org/`
- Gnuplot - `http://www.gnuplot.info/`
- Octave - `http://www.gnu.org/software/octave/`
- Geany - `https://www.geany.org/`
- Linux - `https://www.linux.it/`
- Free Software Foundation - `https://shop.fsf.org/`
- GNOME Foundation - `https://www.gnome.org/`
- Mozilla Foundation - `https://www.mozilla.org/it/`
- Wikipedia - `https://it.wikipedia.org`
- Internet Archive - `https://archive.org/about/`
- Creative Commons - `https://creativecommons.it/chapterIT/`
- WordPress - `https://it.wordpress.org/`
- Phplist - `https://www.phplist.org/`