

Aldo Roveri

RETI DI TELECOMUNICAZIONI

Indice

I	SERVIZI E RETI DI TELECOMUNICAZIONE.....	6
I.1	LA STRUTTURA DI UNA RETE DI TELECOMUNICAZIONE	9
I.1.1	<i>La topologia di una rete</i>	10
I.1.2	<i>Sezione interna</i>	14
I.1.3	<i>Sezione di accesso</i>	18
I.2	SORGENTI DI INFORMAZIONE	21
I.2.1	<i>I segnali emessi</i>	23
I.2.2	<i>Tecniche di compressione di banda</i>	29
I.2.3	<i>Caratteristiche di emissione.....</i>	32
I.3	I SERVIZI E LE POTENZIALITA' DI FORNITURA	36
I.3.1	<i>L'ambiente di comunicazione</i>	36
I.3.2	<i>Potenzialita' coinvolte</i>	40
I.3.3	<i>La tecnica degli attributi.....</i>	41
I.3.4	<i>Modalità di comunicazione</i>	47
I.4	LE ESIGENZE DI COMUNICAZIONE	50
I.4.1	<i>La multimedialita'.....</i>	51
I.4.2	<i>L'ubiquita' e la personalizzazione dei servizi.....</i>	53
I.4.3	<i>La sicurezza e la protezione delle informazioni.....</i>	57
I.4.4	<i>La gestione della rete e dei servizi</i>	58
I.5	EVOLUZIONE DELLE INFRASTRUTTURE	58
I.5.1	<i>Verso l'integrazione di servizi</i>	59
I.5.2	<i>Le reti intelligenti</i>	61
I.5.3	<i>Le reti mobili</i>	64
I.5.4	<i>Comunicazione personale universale.....</i>	65
I.6	CARATTERISTICHE PRESTAZIONALI.....	66
I.6.1	<i>Integrita' informativa</i>	66
I.6.2	<i>Trasparenza temporale.....</i>	67
I.6.3	<i>Modi di connessione.....</i>	69
I.6.4	<i>I servizi nella B-ISDN.....</i>	71
II	LE RISORSE DI RETE.....	74
II.1	RISORSE INDIVISE E RISORSE CONDIVISE.....	75
II.1.1	<i>Il ruolo delle attività di gestione</i>	77
II.1.2	<i>Gestione delle condizioni di contesa.....</i>	78
II.1.3	<i>Gestione degli stati di stallo.....</i>	79
II.2	MODELLI DELL'INTERAZIONE TRA ATTIVITÀ E RISORSE.	80

II.2.1	<i>La popolazione degli utenti potenziali</i>	82
II.2.2	<i>Caratteristiche strutturali di un sistema di servizio</i>	83
II.2.3	<i>La domanda di servizio</i>	85
II.2.4	<i>Funzionalità di controllo</i>	93
II.2.5	<i>Il processo di coda</i>	96
II.2.6	<i>Il concetto di traffico</i>	98
II.3	I FENOMENI DI TRAFFICO IN EQUILIBRIO STATISTICO	102
II.3.1	<i>L'ergodicità del processo di coda</i>	103
II.3.2	<i>Correnti di traffico in equilibrio statistico</i>	107
II.3.3	<i>Caratteristiche prestazionali di un sistema di servizio</i>	111
II.3.4	<i>Code con ingresso e servizio markoviani</i>	115
II.3.5	<i>Il servizio singolo con distribuzione generale e con vacanze</i>	117
II.4	MODELLI PER L'ACCESSO A RISORSE MULTIPLE: LE RETI DI CODE	118
II.4.1	<i>Modello di Jackson</i>	119
II.4.2	<i>Modello di Gordon & Newell</i>	123
II.5	TRAFFICO E PRESTAZIONI NELL'INTERAZIONE TRA ATTIVITÀ E RISORSE	123
II.5.1	<i>Campi di impiego delle risorse indivise e di quelle condivise</i>	123
II.5.2	<i>Risorse virtuali</i>	125
II.5.3	<i>Strategie di assegnazione di risorse condivise</i>	126
II.5.4	<i>Controllo dei sovraccarichi</i>	129
III	ARCHITETTURE E PROTOCOLLI DI COMUNICAZIONE	131
III.1	FUNZIONI DI UN PROCESSO DI COMUNICAZIONE	133
III.1.1	<i>Funzioni orientate al trasferimento</i>	133
III.1.2	<i>Funzioni orientate all'utilizzazione</i>	137
III.2	ARCHITETTURA A STRATI	138
III.2.1	<i>Elementi architetturali</i>	139
III.2.2	<i>Comunicazione tra entità</i>	143
III.2.3	<i>Unità informative</i>	149
III.2.4	<i>Gestione delle connessioni</i>	151
III.2.5	<i>Trasferimento delle informazioni</i>	153
III.2.6	<i>Condizioni di errore</i>	155
III.2.7	<i>Modello astratto del servizio di uno strato</i>	156
III.3	MODELLO OSI	158
III.3.1	<i>Caratteristiche generali del modello</i>	159
III.3.2	<i>Identificazione degli strati componenti</i>	160
III.3.3	<i>Strato di applicazione</i>	161
III.3.4	<i>Strato di presentazione</i>	164
III.3.5	<i>Strato di sessione</i>	164
III.3.6	<i>Strato di trasporto</i>	166
III.3.7	<i>Strato di rete</i>	168
III.3.8	<i>Strato di collegamento</i>	169
III.3.9	<i>Strato fisico</i>	170

IV TECNICHE DI TRASFERIMENTO DELL'INFORMAZIONE172

IV.1 I MODI DI TRASFERIMENTO	173
IV.1.1 Le componenti di un modo di trasferimento	175
IV.1.2 Caratteristiche prestazionali di un modo di trasferimento.....	176
IV.1.3 Esigenza di nuovi modi di trasferimento.....	178
IV.2 GLI SCHEMI DI MULTIPLAZIONE	180
IV.2.1 Modalita' di gestione dell'asse dei tempi	182
IV.2.2 Multiplazione a divisione di tempo sincrona	185
IV.2.3 Multiplazione a divisione di tempo asincrona	188
IV.2.4 Multiplazione con o senza etichetta	191
IV.2.5 Multiplazione ibrida.....	191
IV.3 PRINCIPI DI COMMUTAZIONE.....	192
IV.3.1 Servizi di trasferimento	194
IV.3.2 Modalita' di utilizzazione delle risorse	197
IV.3.3 Assegnazione di risorse fisiche	197
IV.3.4 Assegnazione di risorse logiche.....	200
IV.3.5 Relazioni tra le temporizzazioni negli orologi di nodo.....	202
IV.4 ARCHITETTURE PROTOCOLLARI.....	203
IV.4.1 Protocolli di accesso e di rete interna	204
IV.4.2 Possibili alternative architetturali.....	206
IV.5 ESEMPI DI MODI DI TRASFERIMENTO	210
IV.5.1 Modo di trasferimento a circuito	210
IV.5.2 Modo di trasferimento a pacchetto	212
IV.5.3 Nuovi modi di trasferimento	214
IV.5.4 Modo di trasferimento asincrono.....	217
IV.5.5 Applicazione del LAPD nel Rilegamento di trama	218
IV.6 IL TRATTAMENTO DELLA SEGNALAZIONE	219
IV.6.1 Classificazione dei modi di segnalazione	220
IV.6.2 La segnalazione a canale comune.....	222

V RETI DEDICATE E RETI INTEGRATE.....228

V.1 LE RETI DEDICATE A UN SERVIZIO	228
V.1.1 Rete telefonica.....	228
V.1.2 Reti per dati.....	233
V.2 DEFINIZIONE DELLA ISDN.....	237

VI ELENCO DEGLI ACRONIMI.....239**VII BIBLIOGRAFIA242**

VIII LISTA DELLE FIGURE, TABELLE ED EQUAZIONI	243
VIII.1 FIGURE	243
VIII.2 TABELLE	245
VIII.3 EQUAZIONI	245

PREFAZIONE

Nello sviluppo del capitolo, così come in quello dei capitoli seguenti, si farà riferimento ai risultati dell'attività di normalizzazione svolta in sede internazionale sulla sistemistica per telecomunicazioni. In particolare si citeranno le Raccomandazioni dell'*ITU-T* (nel seguito indicate con l'abbreviazione Racc.) e le Norme *ISO*.

L'*ITU-T* è uno dei tre settori in cui è stato recentemente (inizio del 1993) riorganizzato l'*ITU* (International Telecommunication Union), che è una agenzia specializzata delle Nazioni Unite (ONU) avente come missione tutto ciò che riguarda i rapporti internazionali per lo sviluppo delle Telecomunicazioni. L'*ITU-T* ha come compito l'attività di normativa in questa area, includendo quanto precedentemente (fino al 1992) svolto dal *CCITT* (Comitato Consultivo Internazionale per la Telegrafia e la Telefonia) e da una parte del *CCIR* (Comitato Consultivo Internazionale per le Radiocomunicazioni).

Le norme che sono prodotte dall'*ITU-T* e che in passato sono emerse dall'attività di *CCITT* e *CCIR* sono chiamate "Raccomandazioni", nel senso che il loro rispetto è volontario per i Paesi aderenti all'*ITU*; tuttavia esse hanno un'influenza determinante sulle attività dei diversi attori sulla scena delle Telecomunicazioni, non solo nell'ambito internazionale, ma anche in quelli regionale e nazionale e costituiscono quindi norme "de-facto".

L'*ISO* (International Organization for Standardization) è anch'esso un organismo delle Nazioni Unite, che è stato creato con l'obiettivo di promuovere lo sviluppo della normativa internazionale per facilitare il commercio di beni e di servizi nel mondo. Nell'ambito di questa missione a larghissimo spettro, l'*ISO* opera anche per lo sviluppo della normativa nell'area delle Telecomunicazioni; ciò normalmente in cooperazione con le attività svolte in seno all'*ITU-T*. Anche le norme *ISO* costituiscono solo raccomandazioni per i Paesi aderenti, che rimangono liberi di rispettarle o meno.

I SERVIZI E RETI DI TELECOMUNICAZIONE

Una comunicazione a distanza avviene solitamente sulla base di un rapporto di domanda e di offerta, che ha come oggetto, un *servizio di*

telecomunicazione. Le parti interessate sono in generale identificabili nei *clienti del servizio*, nel *gestore di rete* e nel *fornitore* di quanto richiesto dai clienti.

La prima di queste parti, e cioè il cliente del servizio, ha l'esigenza specifica di comunicare a distanza con altri clienti, secondo modalità operative e prestazioni di qualità definiti all'interno di un opportuno quadro normativo.

Sono stati definiti vari tipi di servizi di telecomunicazione (ad esempio, i servizi telefonico, telegrafico, telefax, ecc.); altri sono in corso di definizione o lo saranno in un prossimo futuro. Essi si distinguono innanzitutto in base alla *natura dell'informazione* che deve essere trattata (ad esempio, voce, dati o immagini, in unione o in alternativa) e alla particolare *applicazione* che è consentita da questo trattamento (per esempio, una tele-elaborazione, un tele-controllo, una automazione delle procedure d'ufficio, ecc.).

La fruizione di un servizio comporta il richiamo e l'esecuzione, in un ordine prestabilito, di componenti funzionali, che qualificano il servizio nei suoi vari aspetti e che debbono svolgersi rispettando opportune regole. Queste costituiscono la *logica del servizio*.

Un cliente sottoscrive con il fornitore un impegno nel quale sono precisate le forme desiderate di comunicazione a distanza e le modalità di fornitura, ivi compresi gli aspetti di qualità e di costo. In questo quadro contrattuale il cliente può operare lui stesso come *utente* di un servizio scelto tra quelli sottoscritti. In alternativa o in unione, un cliente delega uno o più utenti a fruire del servizio. In generale, un utente di un servizio si identifica con un operatore umano o con una macchina di elaborazione, che agiscono come *sorgenti* e/o come *collettori* di informazione.

Ogni servizio comporta un trasferimento e una utilizzazione di informazione che interessano, di volta in volta, due o più utenti del servizio. Il trasferimento deve avvenire da una sorgente ad almeno un collettore tra loro a distanza. L'utilizzazione deve rispondere in primo luogo alle esigenze applicative degli utenti posti in comunicazione, ma deve anche includere quanto è necessario per completare l'operazione di trasferimento con il fine di rendere possibile una fruizione dell'informazione scambiata. Questo completamento configura lo svolgimento della funzione di *trasporto*.

La seconda parte interessata, e cioè il gestore di rete, deve rendere disponibili le risorse infrastrutturali che sono essenziali per la fornitura di quanto richiesto dai clienti/utenti. L'infrastruttura in parola è una *rete di telecomunicazioni*, che ha il ruolo di piattaforma su cui eseguire il programma contenente la logica di ogni servizio.

Compiti essenziali del gestore di rete sono attivare e mantenere operativi i mezzi tecnici e organizzativi per:

- trasportare l'informazione a distanza secondo quanto richiesto nell'espletamento di ogni servizio;
- controllare e gestire la rete nel suo complesso e nelle sue parti componenti, in modo che il trasporto dell'informazione avvenga entro prefissati obiettivi di qualità e di costo;
- consentire al cliente/utente e al fornitore il controllo e la gestione dei servizi supportati dall'infrastruttura di rete.

L'estensione territoriale di una rete dipende dalle posizioni relative dei suoi utenti. Da questo punto di vista, le reti di telecomunicazioni possono essere sostanzialmente *in area geografica* e *in area locale*. Nel primo caso gli utenti sono distribuiti su un'area molto estesa (una nazione, un continente, l'intero globo terrestre); nel secondo l'area interessata è ristretta a un singolo edificio o a un complesso di insediamenti entro il raggio di qualche chilometro.

Relativamente invece alla gamma dei servizi che possono essere supportati, le reti possono essere *dedicate a un servizio* o *integrate nei servizi*.

Le *reti dedicate a un servizio* sono state concepite e realizzate in passato per la fornitura di un singolo servizio e possono oggi essere utilizzate anche per un insieme ristretto di altri servizi, seppure con limitazioni severe per ciò che concerne la qualità conseguibile. Esempi significativi di reti di questo tipo sono la *rete telefonica* e le *reti per dati*. In particolare, la rete telefonica, che è di gran lunga la più capillare in termini di numero di utenti serviti, è stata in passato, ed è tuttora, oggetto di continui aggiornamenti tecnici con la gradualità imposta dall'enorme mole di investimenti richiesti.

Le *reti integrate nei servizi* sono invece di concezione molto recente e la loro realizzazione è attualmente in atto nei Paesi industrialmente più evoluti, assumendo come punto di partenza il presente stato di maturità conseguito dalla rete telefonica. Il loro obiettivo è rendere possibile la fornitura di una vasta gamma di servizi di telecomunicazione con prestazioni di qualità e di costo decisamente migliori rispetto a quelle ottenibili con le reti dedicate.

La terza e ultima parte interessata è il fornitore del servizio. Questi, che può o meno coincidere con il gestore di rete, deve offrire al cliente quanto richiesto per soddisfare, in modo qualitativamente accettabile, esigenze che in generale sono comunicative ed elaborative. Per questo scopo il fornitore deve anche ottenere dal gestore di rete le risorse infrastrutturali necessarie per espletare i servizi offerti.

Nel quadro di riferimento sopra delineato, il problema generale posto dalla fornitura di un servizio di telecomunicazione può essere così formulato:

data una pluralità di utenti, occorre definire le modalità tecniche e operative per consentire un trasferimento e una utilizzazione di informazione nell'ambito di quanto previsto dal servizio richiesto; i vincoli da rispettare sono una accettabile *qualità di servizio* e un *costo di fornitura* commisurato al beneficio ottenibile.

Per chiarire le implicazioni di questi vincoli, occorre tenere conto che:

- la qualità di un servizio e' definita dal valore assunto da un insieme di parametri prestazionali, che possono variare a seconda della categoria di servizi che si considera;
- nella fornitura di uno specifico servizio e in relazione al conseguimento di limiti accettabili per i valori degli associati parametri prestazionali, la rete deve rendere disponibili opportuni tipi di risorse;
- per ognuno di questi tipi, occorre stabilire la potenzialita' e la quantità di risorse da mettere a disposizione di ogni richiesta di servizio;
- nella scelta progettuale di cui al punto precedente, occorre ovviamente operare un compromesso tra prestazioni desiderate e livelli di costo conseguenti.

Questo capitolo introduce alcuni concetti preliminari per la soluzione del problema sopra formulato. A tale scopo verranno precisate dapprima (par. I.1) quali siano le motivazioni tecnico-economiche che guidano nel definire la struttura di una rete di telecomunicazione. Si passa poi (par. I.2) a considerare le sorgenti di informazione per chiarire quali siano le forme fisiche e logiche secondo cui queste emettono nell'ambito della fruizione di un servizio. Il terzo argomento trattato (par. I.3) riguarda le caratteristiche dei servizi di telecomunicazione secondo i vari tipi di classificazioni che se ne fanno nell'ambito delle attività di normalizzazione in sede internazionale.

Si passa poi (par. I.4) alle attuali esigenze di comunicazione e agli obiettivi che sono in corso di conseguimento. Si parlerà anche (par. I.5), seppure in termini del tutto preliminari, dell'attuale stato evolutivo delle infrastrutture per telecomunicazioni. Si concluderà (par. I.6) con la presentazione di alcune caratteristiche prestazionali richieste nella fornitura di un servizio: lo scopo è individuare gli oneri richiesti a una rete per supportare questo compito.

I.1 La struttura di una rete di telecomunicazione

Una rete di telecomunicazione è un complesso di mezzi che, attraverso le risorse tecniche e operative necessarie, permette ai suoi utenti il trasferimento di

informazione nell'ambito della fruizione del servizio desiderato. Essa è costituita in analogia ad altri tipi di strutture preposte al trasferimento di beni materiali; ad esempio, è strutturalmente simile ad una rete stradale o ferroviaria o ad una rete per il trasporto dell'energia elettrica. L'idea base che guida nella definizione di queste strutture è quella di consentire il trasferimento del bene desiderato in modo economicamente conveniente.

Questo paragrafo è dedicato alla introduzione dei criteri generali che oggi sono costantemente adottati per strutturare una rete di telecomunicazione in area geografica, e cioè, in particolare, per definirne la configurazione geometrica e per individuarne gli apparati componenti.

1.1.1 La topologia di una rete

Supponiamo di dover progettare una rete al servizio di U utenti distribuiti con posizione statica su un'area geograficamente estesa. Il requisito primario è quello di consentire ad ognuno di questi utenti di poter comunicare, individualmente o collettivamente, con ognuno degli altri.

Il punto di partenza di questo progetto è la definizione di una conveniente struttura geometrica della rete, e cioè della sua *topologia*. Elementi componenti di questa sono i *rami* ed i *nodi*. Un ramo, rappresentato graficamente da un segmento di retta o di curva, costituisce l'elemento di connessione di due nodi, mentre un nodo è il punto di estremità di due o più rami. Il significato di queste entità geometriche è tuttavia diverso a seconda che si consideri il trasferimento *logico* o quello *fisico* dell'informazione attraverso la rete.

Dal punto di vista del trasferimento logico, un ramo è il percorso diretto che l'informazione deve seguire per passare da una sua estremità all'altra, mentre un nodo costituisce il mezzo di scambio tra due rami che ad esso fanno capo. L'insieme dei rami e dei nodi con questo significato costituisce il *grafo logico* della rete e descrive, in termini topologici, il modello identificato dalle funzioni di rete aventi natura logica. Questo modello è chiamato *rete logica*.

Se il punto di vista è invece quello del trasferimento fisico, un ramo è la via per il trasferimento dei segnali che supportano l'informazione, mentre un nodo è il punto di trasmissione e/o di ricezione di questi segnali. Nodi e rami con questo ulteriore significato formano il *grafo fisico* della rete. Questo descrive la topologia di un secondo modello di rete, che è chiamato *rete fisica* e che è identificato dalle funzioni di rete aventi finalità trasmissive.

I due modelli di rete ora definiti sono in stretta relazione gerarchica, dato che, come è ovvio, le funzioni di natura logica debbono utilizzare quelle di

natura fisica e queste ultime sono al servizio delle prime. Tuttavia i due grafi logico e fisico, associati a questi modelli, non coincidono in generale, poichè diversi sono i criteri tecnici ed economici che guidano nella definizione dei cammini logici e fisici dell'informazione attraverso una rete.

In un grafo, logico o fisico, due nodi sono *adiacenti*, se esiste un ramo che li connette direttamente. Un *percorso di rete* è una sequenza di nodi adiacenti nel grafo logico, mentre la *lunghezza logica di un percorso di rete* è il numero di rami che lo compongono. Infine la *lunghezza fisica di un ramo* riguarda il grafo fisico ed è l'estensione del mezzo fisico in corrispondenza modellistica con il ramo.

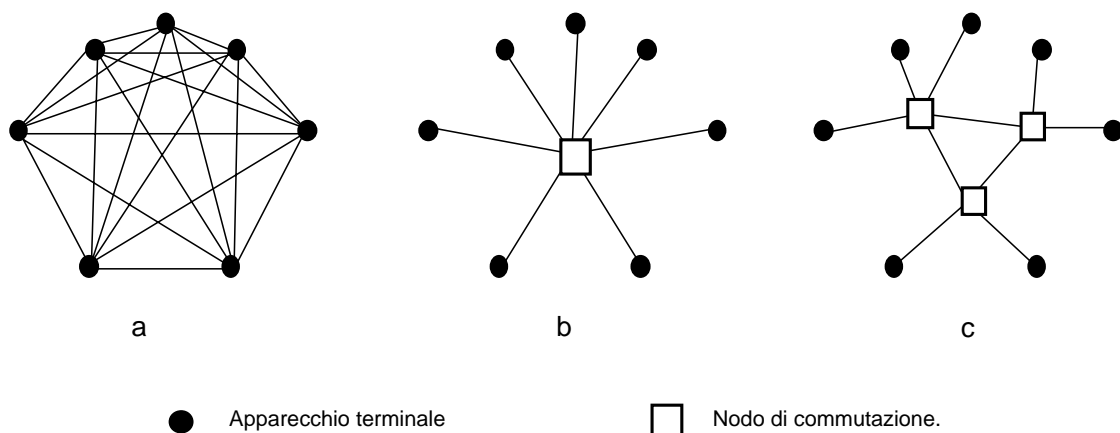


Fig. I-1 - Alternative di grafo di rete.

In Fig. I-1 sono mostrate tre alternative di grafo di rete, in cui si è scelto un piccolo valore di U per semplicità grafica e ove si è supposto che i due grafi logico e fisico siano coincidenti.

La prima alternativa da considerare per interconnettere gli U utenti è mostrata in Fig. I-1a, ove i nodi coincidono con le posizioni dei vari utenti e sono tutti tra loro adiacenti. In tale soluzione quindi ogni utente è connesso direttamente con ognuno degli altri utenti che hanno accesso alla rete. In queste condizioni ogni utente può selezionare, tra gli $U-1$ rami che ad esso fanno capo, il ramo che gli consente di raggiungere un certo altro utente con un percorso di rete avente lunghezza logica unitaria. Inoltre ogni ramo può essere utilizzato solo dai due utenti che sono situati alle sue estremità.

Per valutare la praticabilità economica della soluzione corrispondente al grafo in Fig. I-1a, assumiamo, per semplicità di ragionamento, che il costo di una rete sia determinato in primo luogo da quello dei rami, e cioè dal costo dei mezzi di trasferimento fisico e logico che i rami rappresentano. In queste condizioni, anche prescindendo dalla dipendenza del costo di un ramo dalla sua

lunghezza fisica, è immediato convincersi che la soluzione di Fig. I-1a è impraticabile quando il numero U di utenti supera qualche unità. Basta infatti osservare che il numero di rami necessario in questa soluzione sarebbe uguale a $U(U-1)/2$ e che quindi l'onere economico per una realizzazione crescerebbe con il quadrato del numero degli utenti interconnessi.

Una alternativa più attraente è quella mostrata in Fig. I-1b, ove ogni utente è connesso con un ramo indiviso, chiamato *ramo d'utente*, ad un nodo dislocato, ad esempio, in posizione baricentrica rispetto alla posizione degli utenti. Questo nodo, a differenza di quelli di Fig. I-1a, è rappresentativo di una risorsa condivisa, che deve essere utilizzata ogni qualvolta si voglia stabilire un percorso di rete (avente lunghezza logica uguale a 2) tra due utenti qualunque della rete. La funzione svolta da tale risorsa mette in corrispondenza, quando richiesto e per la durata necessaria, due rami d'utente che ad essa fanno capo, in modo che l'informazione trasportata su uno di questi possa essere inoltrata sull'altro verso la sua destinazione. Il nodo in questione, considerato appartenente al grafo logico, è quindi rappresentativo di un'apparecchiatura di scambio e, per la funzione che questa è preposta a svolgere, è chiamato *nodo di commutazione*.

Da un punto di vista economico, la soluzione di Fig. I-1b richiede, oltre all'onere connesso al nodo di commutazione, un costo che è legato agli U rami componenti e che, quindi, cresce linearmente con il numero degli utenti interconnessi.

Per valutare però la praticabilità di questa soluzione, che è quella tipica di una *rete a stella*, occorre tenere conto della dipendenza dei costi di ogni ramo dalla corrispondente lunghezza fisica. Poichè tale dipendenza comporta in generale un costo crescente al crescere di questa lunghezza, la soluzione di rete a stella è utilizzabile quando la dislocazione degli utenti è tale da consentire il contenimento delle lunghezze fisiche di ramo entro qualche chilometro.

Una soluzione alternativa, che può essere impiegata senza i limiti della rete a stella, è quella illustrata in Fig. I-1c. In questo caso si suddivide l'utenza, in base alla propria dislocazione, in insiemi di postazioni che siano fisicamente entro il raggio massimo di qualche chilometro e si interconnettono gli utenti appartenenti a ognuno di questi insiemi con una struttura a stella. Ogni insieme di utenti fa quindi capo ad un proprio nodo di commutazione. Per consentire la connettività tra utenti che fanno capo a nodi di commutazione diversi, occorre allora provvedere a connettere i vari nodi della rete con rami specifici, chiamati *rami di giunzione*.

I nodi di Fig. I-1c, considerati appartenenti al grafo logico, sono rappresentativi di apparecchiature che debbono svolgere, oltre alla funzione già descritta nel caso del nodo di Fig. I-1b, anche l'ulteriore funzione di mettere in corrispondenza, quando richiesto e per la durata necessaria, rami d'utente con rami di giunzione ovvero coppie di rami di giunzione.

Si osserva che questa terza alternativa comporta l'impiego di ulteriori risorse condivise rispetto alla soluzione di Fig. I-1b. Infatti, oltre ai nodi di commutazione, anche i rami di giunzione costituiscono una risorsa che deve essere utilizzata ogni qualvolta si voglia stabilire un percorso tra due utenti qualsiasi facenti capo a nodi di commutazione diversi.

Tali rami possono quindi essere utilizzati con maggiore frequenza media rispetto ad un qualsiasi ramo d'utente: infatti ad un ramo di giunzione possono fare riferimento le richieste di comunicazione presentate da tutti gli utenti della rete, mentre un ramo d'utente deve soddisfare le richieste del solo utente che ad esso fa capo.

Le considerazioni sopra svolte hanno ovviamente carattere del tutto preliminare rispetto al problema complesso della determinazione della topologia di una rete di telecomunicazioni. Esse tuttavia consentono di concludere che l'obiettivo di convenienza economica di una rete può essere conseguito con l'applicazione di un criterio generale che può essere così espresso:

è conveniente mettere a disposizione degli utenti risorse condivise, cioè mezzi che possono essere utilizzati, contemporaneamente o in tempi successivi, per costituire, secondo le esigenze dell'utenza, una pluralità di percorsi di rete; è altrettanto conveniente contenere l'onere delle risorse indivise, quali, ad esempio, i rami d'utente.

L'applicazione di questo criterio ha un'immediata conseguenza sulla topologia delle reti in area geografica. In queste reti, infatti, è sempre possibile distinguere due sezioni principali: la *sezione di accesso* e la *sezione interna*.

La sezione interna comprende gli apparati di commutazione e di trasmissione, che consentono di porre in corrispondenza clienti comunque dislocati. La sezione di accesso comprende invece i mezzi con i quali ogni cliente è in relazione con la rete per utilizzarne i servizi. Ulteriori particolari su queste due partizioni di rete sono forniti in § I.1.2 e I.1.3.

Circa le possibilità di accesso a una rete di telecomunicazione si possono distinguere due condizioni altrettanto significative, che debbono essere affrontate con differenti mezzi tecnici per la realizzazione della sezione di accesso e che, almeno finora, hanno condotto a distinguere due tipi di reti.

La prima di queste condizioni si verifica quando i servizi supportati dalla rete sono accessibili solo da parte di utenti che, ogniqualvolta desiderino comunicare, siano in posizione statica o che, pur in movimento, rimangano in un intorno relativamente ristretto di un sito di riferimento (abitazione, ambiente di lavoro, ecc.). Si parla in tal caso di rete per applicazioni fisse o, in breve, di *rete fissa*.

Se invece l'accesso è consentito ad utenti che sono in movimento senza limitazioni alle loro possibilità di deambulazione (a piedi o su veicoli), si tratta di un ambiente di comunicazione per applicazioni mobili o, in breve, di una *rete mobile*. L'estensione territoriale di una infrastruttura siffatta (*regione di copertura*) è quella che può essere percorsa dai suoi utenti conservando la possibilità di comunicare.

I.1.2 Sezione interna

Nella sezione interna di una rete fissa o mobile, i rami e i nodi del grafo logico rappresentano le *linee di giunzione* e gli *apparati di commutazione*, rispettivamente. Invece, un ramo e i due nodi alle sue estremità nel grafo fisico rappresentano le parti componenti di un *sistema di trasmissione*, in cui gli *apparati terminali* sono collocati nei siti dei nodi e quelli di *linea* (con le annesse tratte di mezzo trasmissivo) sono schematizzati dai rami.

Si è già puntualizzato che i due grafi logico e fisico non coincidono in generale. La motivazione tecnica di questa distinzione risiede nella diversità dei criteri di dimensionamento seguiti per determinare la capacità di trasferimento dei rami dei due tipi di grafi.

Esempio I.1-1 - In Fig. I-2 è mostrato un esempio di grafo logico (Fig. I-2a) e di grafo fisico (Fig. I-2c), relativi alla sezione interna di una rete di telecomunicazioni. Il grafo fisico risulta dalla scelta di percorso dei mezzi trasmissivi quale è rappresentata in Fig. I-2b. I quattro nodi *A*, *B*, *C* e *D* del grafo logico sono connessi, a due a due, da rami specifici come può essere giustificato dalle relazioni di traffico esistenti tra questi nodi. Questa connessione diretta significa, ad esempio, che l'informazione, avente origine in un apparecchio terminale facente capo all'apparato di commutazione rappresentato dal nodo *A* e avente come destinazione un apparecchio terminale connesso all'apparato di commutazione rappresentato dal nodo *B*, può essere trasportata seguendo il percorso di rete diretto che congiunge *A* con *B*.

Dal punto di vista fisico, consideriamo, ad esempio, il segnale che supporta il trasferimento dell'informazione da *A* a *B*. In questo caso la modularità delle capacità trasmissive può indurre, come in Fig. I-2c, a trasferire questo segnale dal nodo *A'* a quello *D'* e da quest'ultimo al nodo *B'*. La conseguente struttura del grafo fisico di Fig. I-2c ci dice che il transito attraverso il nodo *D'* ha significato solo da un punto di vista trasmissivo e non coinvolge quindi l'apparato di commutazione rappresentato dal nodo *D* nel grafo logico.

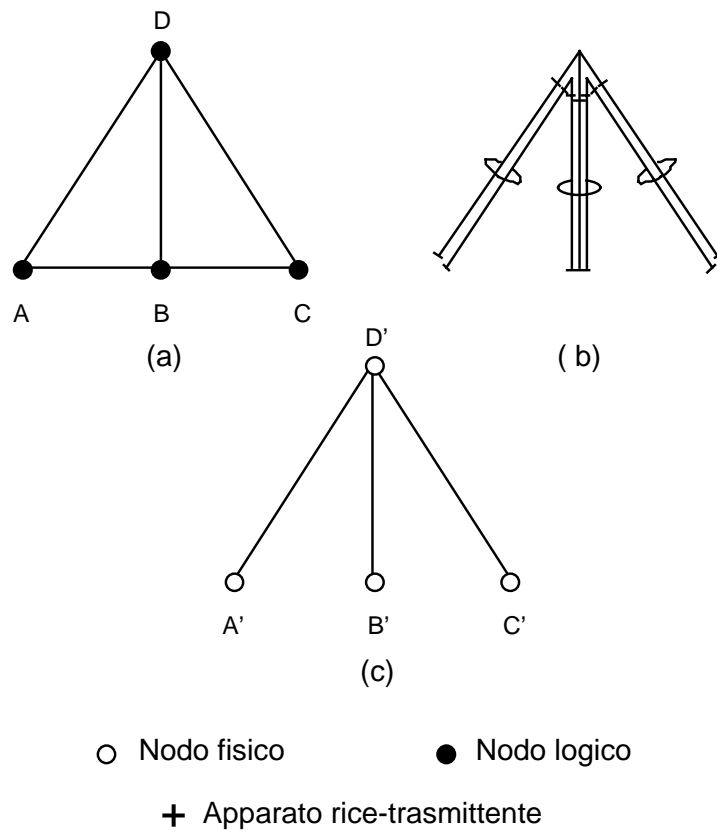


Fig. I-2 - Esempio di grafo logico (a) e di grafo fisico (c), relativi alla sezione interna di una rete di telecomunicazione; il grafo fisico risulta dalla scelta di percorso dei mezzi trasmissivi quale e' rappresentata in (b).

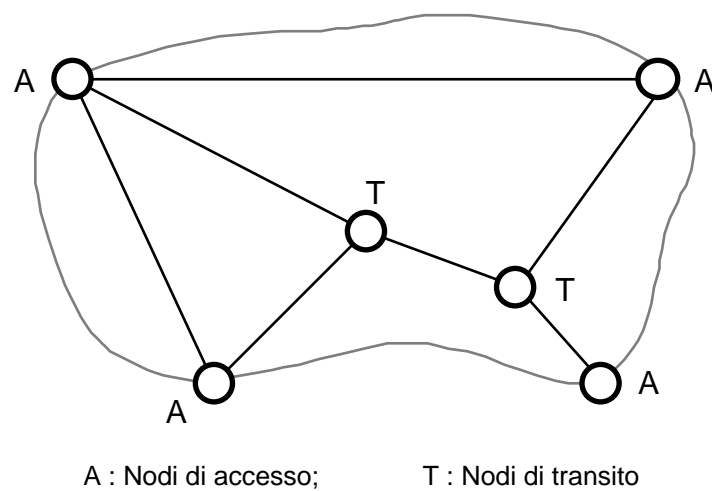


Fig. I-3 - Grafo logico relativo alla sezione interna di una rete di telecomunicazione: distinzione tra nodi di accesso (A) e di transito (T).

Nel grafo logico della sezione interna sono distinguibili due tipi di nodi (Fig. I-3), e cioè i *nodi di accesso* e i *nodi di transito*. I nodi di accesso, che sono collocati sui *bordi* della sezione interna, sono legati direttamente con gli utenti della rete (nel senso, ad esempio, che ogni utente di rete fissa fa capo a uno specifico nodo di questo tipo) e sono interconnessi con percorsi diretti o indiretti. I nodi di transito, che fanno parte del *nucleo* della sezione interna, non hanno invece relazioni dirette con la sezione di accesso e costituiscono il mezzo di scambio per formare percorsi indiretti tra due nodi d'accesso.

Gli apparati di commutazione sono gli organi della sezione interna preposti a svolgere la funzione che consente di utilizzare in modo flessibile le risorse di una rete di telecomunicazione (*funzione di commutazione*): ciò al fine di fornire agli utenti della rete, a seguito di loro richieste, un trasferimento di informazione secondo le esigenze di uno specifico servizio. Mediante tale funzione è infatti possibile porre in comunicazione utenti, variabili di volta in volta, in base alle richieste che vengono ad essa presentate. Lo scopo della funzione di commutazione viene conseguito con mezzi utilizzabili, contemporaneamente o in tempi successivi, per rispondere ad altre richieste.

Gli apparati di commutazione, quando operano in modo automatico, e cioè senza l'intervento determinante di operatori umani, sono chiamati *autocommutatori*. Le loro funzioni variano in relazione ai tipi di nodi del grafo logico della rete che li rappresentano. In particolare, i nodi di accesso e quelli di transito sono in corrispondenza con gli *autocommutatori locali* e con quelli *di transito*, rispettivamente.

Un autocommutatore è azionato da richieste di servizio che gli pervengono da ognuno degli apparati appartenenti all'insieme di rete logica con cui è in relazione diretta. Tale azione si manifesta con richieste di servizio che sono inoltrate verso specifici apparati dello stesso insieme di rete. Per consentire queste interazioni, nella periferia di un autocommutatore sono previsti *punti di accesso logico*, che sono denominati *terminazioni di commutatore* e che si distinguono in *terminazioni di ingresso* o *di uscita*, a seconda della loro specializzazione a trattare richieste di servizio, rispettivamente, entranti nell'autocommutatore o uscenti da questo.

Oltre a questi compiti di natura logica, la periferia di un autocommutatore svolge anche funzioni di apparato terminale (trasmittente o ricevente) per tutti i sistemi di trasmissione, che sono utilizzati nella sezione di accesso o in quella interna e che ad esso fanno capo. In questo ulteriore ruolo, che riguarda le funzionalità della rete fisica, un autocommutatore presenta anche *punti di accesso fisico*, che sono denominati *porte*.

Alle *porte di ingresso e di uscita* fanno capo i canali trasmissivi con verso di trasmissione, rispettivamente, entrante nell'autocommutatore e uscente da questo. Le porte di ingresso e di uscita coincidono fisicamente quando lo stesso mezzo fisico è supporto di un canale trasmissivo bidirezionale (*linea a due fili*); sono invece distinte quando le due direzioni di trasmissione sono supportate da canali trasmissivi unidirezionali (*linea a quattro fili*). In tutti i casi ogni porta di un autocommutatore è associata a una specifica terminazione e costituisce con questa l'interfaccia fisica e logica tra l'autocommutatore e la rete.

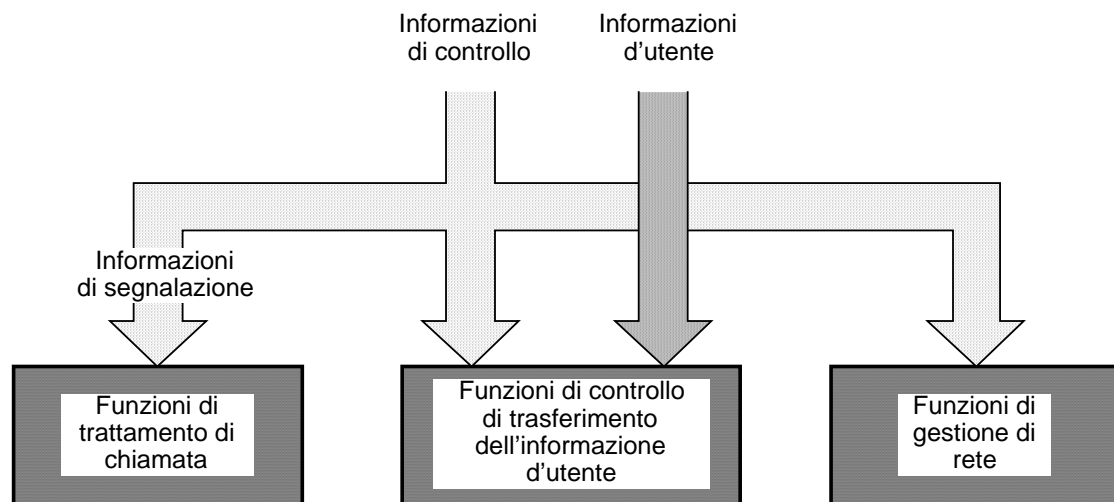


Fig. I-4 - Principali utilizzazioni delle risorse di elaborazione in un autocommutatore.

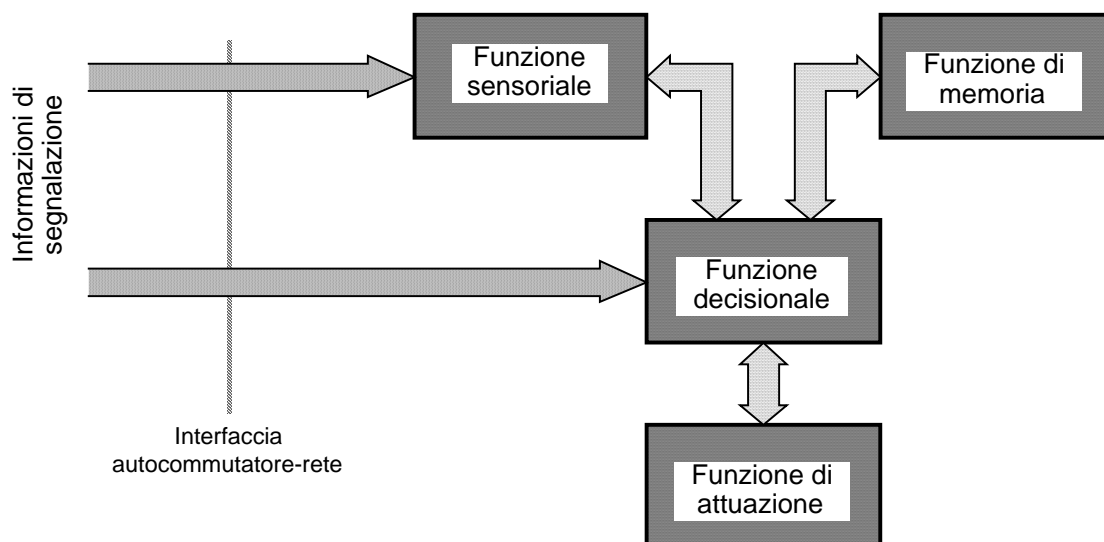


Fig. I-5 - Relazione tra le funzioni di trattamento di chiamata e le informazioni di segnalazione

I.1.3 Sezione di accesso

Nella sezione di accesso di una rete fissa vanno distinti, per ogni cliente, i mezzi di comunicazione, di cui questo dispone nel proprio ambiente di abitazione o di lavoro (*postazione di utente*), e quelli preposti alla connessione (*allacciamento di utente*) tra tale postazione e il pertinente nodo di accesso (Fig. I-6).

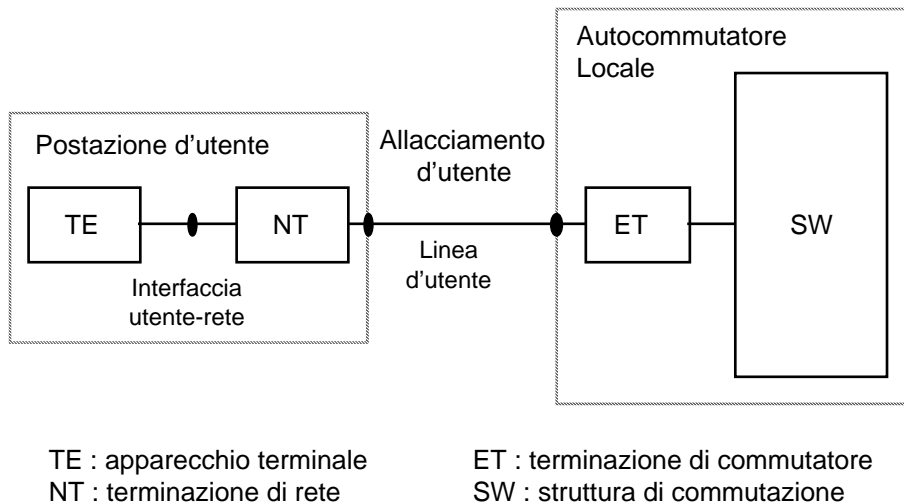


Fig. I-6 - Esempio di sezione di accesso con linea di utente individuale.

Presso la postazione di utente sono disponibili uno o più *apparecchi terminali*, ognuno dei quali costituisce il mezzo con cui un utente della rete usufruisce di uno o più servizi di telecomunicazione. Un apparecchio terminale è in grado di svolgere funzioni più o meno complesse in relazione al servizio richiesto ed è connesso alla rete tramite le potenzialità fisiche e logiche dell'allacciamento di utente.

Per sfruttare tali potenzialità, vengono utilizzate funzioni attivate dall'utente e dalla rete. Le prime risiedono nell'apparecchio terminale, mentre le seconde sono presenti in una apparecchiatura, che fa parte della postazione di utente e che è denominata *terminazione di rete*. L'interfaccia tra apparecchio terminale e terminazione di rete è detta *di accesso* o, con termine equivalente, *interfaccia utente-rete*.

Un'interfaccia utente-rete può presentare varie configurazioni, in base alla relazione tra apparecchio terminale e terminazione di rete. Tali configurazioni sono *punto-punto* o *punto-multipunto*, a seconda che una singola terminazione di rete venga posta in corrispondenza con uno o con più apparecchi terminali.

Per ciò che riguarda poi l'allacciamento di utente, si possono avere varie soluzioni. La più semplice di queste prevede, per il legamento di una postazione di utente con il pertinente nodo di accesso, una connessione (*linea di utente individuale*), che è ad uso esclusivo degli utenti operanti nell'ambito di tale postazione. Soluzioni alternative sono basate, ad esempio, su *linee d'utente multipunto* o su apparecchiature di *concentrazione/multiplazione*. Queste hanno l'obiettivo comune di contenere, nei limiti del possibile e nel rispetto di una accettabile qualità di servizio, gli oneri economici connessi alla sezione di accesso, che costituiscono, in generale, una quota parte largamente prevalente del costo complessivo di una rete, in termini di oneri sia per la prima installazione, sia per l'esercizio e la manutenzione.

Una linea d'utente individuale connette una terminazione di rete con una terminazione dell'autocommutatore locale (Fig. I-6) e, come tale, è rappresentata, nel grafo logico della rete, da un ramo che fa capo, da un lato, al punto individuato dalla postazione di utente e, dall'altro, al nodo di accesso.

Qualunque soluzione venga impiegata per l'allacciamento di utente in una rete fissa, il mezzo trasmissivo impiegato per connettere la postazione di utente con il pertinente nodo di accesso può essere uno qualunque tra quelli resi disponibili dalle attuali tecnologie (mezzo in rame, in fibra ottica o su portante radio) e può essere scelto in base a considerazioni economiche e a seconda della capacità di trasferimento che si desidera conseguire.

Esempio I.1-2 - Per esemplificare i concetti e le definizioni precedentemente introdotte, consideriamo il caso di una rete dedicata per dati, appartenente alla categoria delle reti fisse e nota anche come *rete di calcolatori*. Reti di questo tipo sono strutture capaci di collegare un insieme di apparati per il trattamento dell'informazione; esse quindi consentono di condividere e di distribuire le risorse di elaborazione disponibili in vari sistemi più o meno distanti tra loro.

Le attuali reti per dati presentano dimensioni e configurazioni molto varie: si passa dalla semplice struttura comprendente un calcolatore e pochi terminali a strutture complesse che interconnettono tra loro migliaia di terminali e decine di processori con differenti potenzialità di elaborazione. Tali strutture sono in grado di fornire una vasta gamma di servizi con prevalenti finalità di tele-elaborazione.

Uno schema generale di rete per dati, che non corrisponde a nessun caso reale, ma che riassume alcune tra le caratteristiche topologiche proprie di una rete di calcolatori, è mostrato in Fig. I-7.

Rinviamo al seguito per un chiarimento sui vari elementi componenti ivi rappresentati, è possibile identificare in questa figura le postazioni d'utente, la sezione di accesso e quella interna.

Le postazioni d'utente comprendono varie risorse di elaborazione e di trasferimento collocate in prossimità degli utenti della rete. Tra tali risorse si distinguono in primo luogo gli apparecchi terminali, che possono presentarsi in varie configurazioni, e cioè come

apparecchiature *singole* o *multiple*. In questo secondo caso si possono avere configurazioni a *grappolo* (cluster), o a distribuzione *multipunto*, o con interconnessione in rete (*reti in area locale*).

In ogni caso si parla di *apparecchi terminali per dati* ovvero, in sigla, di *DTE* (Data Terminal Equipment), con riferimento agli apparecchi terminali che accedono alla rete attraverso un'interfaccia fisicamente individuabile.

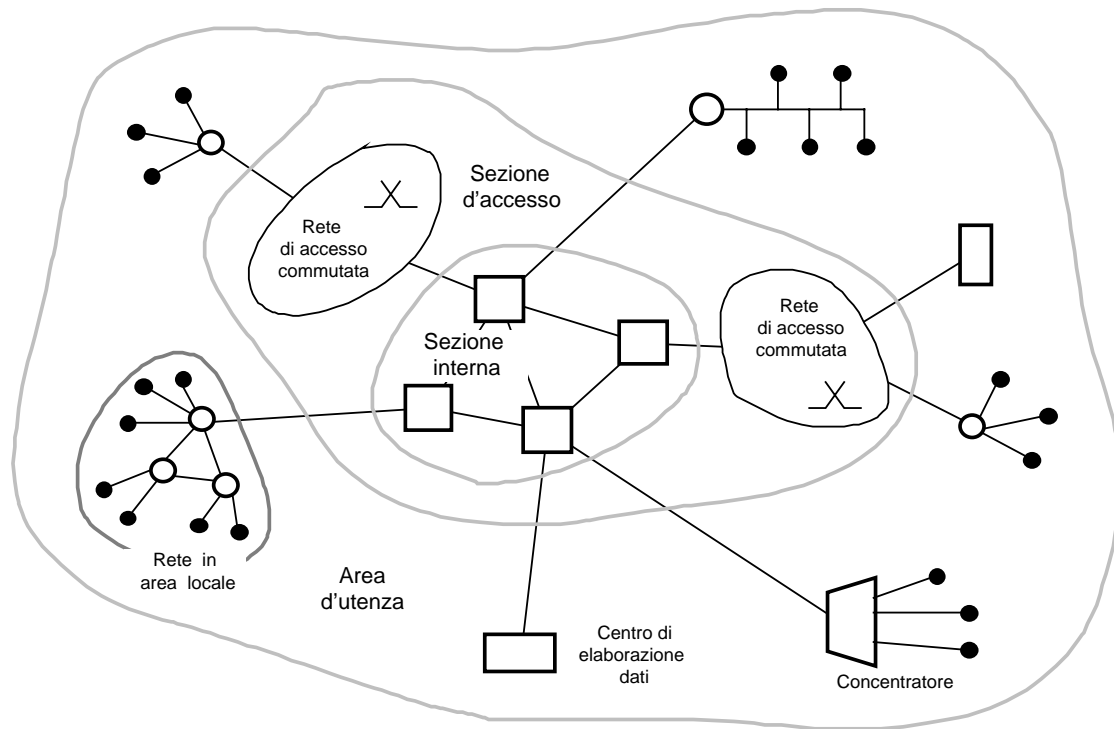


Fig. I-7 - Schema generale di una rete per dati.

La sezione interna di rete comprende l'insieme delle risorse di commutazione e di trasmissione che consentono il trasferimento a distanza dell'informazione d'utente.

Infine la sezione di accesso consente di connettere le postazioni d'utente con la sezione interna di rete. Le sue funzionalità sono di norma messe a disposizione dalle ordinarie reti pubbliche per telecomunicazione (ad esempio dalla rete telefonica pubblica) che eseguono le funzioni di connessione tra i DTE e i punti di accesso della rete per dati. Tali punti di accesso sono la sede delle cosiddette *apparecchiature di terminazione del circuito-dati*, che sono indicate comunemente con la sigla *DCE* (Data Circuit terminating Equipment); esse sono le terminazioni di rete.

La Fig. I-8 mostra il modello base di una catena di trasferimento per dati con configurazione di tipo *punto-punto*, e cioè in presenza di una sorgente che desidera trasferire informazione ad un singolo collettore. In tale schema si distinguono, oltre ad una risorsa di trasferimento messa a disposizione dalla rete per dati, anche due unità DTE di estremità e due unità DCE di interfaccia.

I DTE possono corrispondere, ad esempio, ad un terminale semplice che desidera instaurare un colloquio con un altro DTE, o anche ad un centro di elaborazione capace di gestire contemporaneamente molti colloqui con altrettanti DTE.

In ogni caso in un DTE si possono individuare due sezioni che compiono funzioni

distinte. La prima di queste è rivolta all'*elaborazione dell'informazione*, mentre la seconda controlla il *trasferimento* dei dati attraverso la rete.

Per quanto riguarda poi la risorsa di trasferimento, per la sua realizzazione esistono varie alternative, che saranno oggetto del cap. IV.

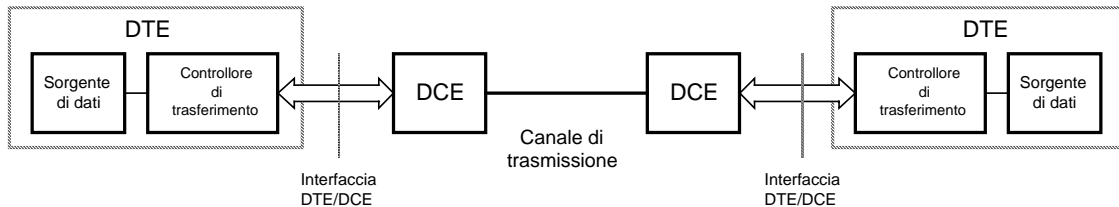


Fig. I-8 - Modello base di un sistema di comunicazione di dati.

Tutto ciò vale nel caso di accesso a una rete fissa. Se invece la rete è mobile, vanno sottolineati alcuni elementi di diversità. In primo luogo la postazione di utente è mobile, e quindi tali sono anche l'apparecchio terminale (singolo o multiplo) e la terminazione di rete che a tale postazione sono associati.

Le conseguenze di questa diversità comportano altri elementi di distinzione; precisamente:

- l'allacciamento d'utente può essere realizzato solo con mezzi trasmissivi di tipo radio, a cui quindi devono essere adattate le funzioni trasmissive alle due estremità della connessione (lato utente e lato rete);
- nella sezione di accesso di una rete mobile, occorre prevedere un insieme di funzioni di natura logica che sono peculiari a questo particolare ambiente di comunicazione (cfr. § I.5.3).

I.2 Sorgenti di informazione

Consideriamo il flusso informativo di utente emesso da una sorgente di informazione nell'ambito di un servizio di telecomunicazione. Al riguardo, è nota la distinzione tra una *sorgente analogica* e una *numerica* ed è altresì noto che il flusso emesso da una sorgente analogica può essere trasformato in forma numerica attraverso un'operazione di conversione analogico-numerica (*conversione A/N*).

Poiché gli attuali orientamenti della tecnica per effettuare un trasferimento o una memorizzazione di informazione rendono preferibili trattamenti in forma numerica, nel seguito faremo unico riferimento a sorgenti che emettono in questa forma. Ciò equivale ad assumere che l'emissione sia costituita da una sequenza di simboli, appartenenti a un insieme di cardinalità limitata. Questa

ipotesi vale ovviamente nel caso di una sorgente numerica, come ad esempio in quello di una *sorgente di dati*; ma vale anche se la sorgente e' analogica, come si verifica per informazioni sonore (*sorgente audio*) o visive (*sorgente video*), e se il flusso informativo emesso e' il risultato di una conversione A/N.

Sorgenti di dati

Una sorgente di dati emette, in corrispondenza dei suoi simboli, un insieme di cifre binarie. La legge di corrispondenza e' quella definita da un codice, mentre l'operazione di codifica e' quella che realizza materialmente questa corrispondenza. I simboli da codificare sono caratteri o combinazioni di questi.

L'insieme dei caratteri costituisce un alfabeto e comprende, in primo luogo, i caratteri grafici, e cioe' le dieci cifre della numerazione decimale, le ventisei lettere utilizzate nella scrittura alfabetica internazionale, segni operatori o di punteggiatura. Sono anche inclusi caratteri di controllo, che sono destinati a facilitare il trasferimento dell'informazione emessa dalla sorgente, a comandare l'avvio o l'arresto delle apparecchiature terminali periferiche e, infine, a organizzare la presentazione dell'informazione tramite stampanti o organi di visualizzazione su schermo.

La rappresentazione binaria di un carattere e' chiamata parola di codice. Se n e' il numero delle cifre binarie componenti ogni parola di codice, si dice che il codice e' a n momenti. In questo caso il numero di caratteri rappresentabili e' al massimo uguale a 2^n . Tutti i codici per comunicazioni di dati utilizzano valori di n compresi tra 5 e 8.

Tra i codici a 5 o 6 momenti si citano l'IA2 (alfabeto internazionale n° 2, anche noto come codice Baudot, a 5 momenti) e il BCD (codice decimale codificato binario, a 6 momenti), che, tuttavia, hanno oggi applicazioni marginali nel campo delle comunicazioni di dati.

Per cio' che riguarda invece i codici a 7 o 8 momenti, gli esempi piu' significativi sono dati dall'IA5 (alfabeto internazionale n° 5 o codice ASCII, a 7 momenti) o dall'EBCDIC (Extended Binary Coded Decimal Interchange Code, a 8 momenti).

Le cifre binarie che compongono una parola di codice e che sono quindi in corrispondenza a un carattere possono essere strutturate in serie o in parallelo. Nel primo caso le singole cifre binarie risultanti dal processo di codifica costituiscono una sequenza predisposta per una emissione della sorgente su una singola uscita; si parla allora di sequenza di dati in serie (o seriali). Nel secondo caso le n cifre binarie di ogni parola di codice sono predisposte per una emissione in parallelo su altrettante uscite; dato che ciascun carattere e' rappresentato da un vettore a n componenti, una sequenza di caratteri e' posta in corrispondenza a una sequenza di vettori; si tratta allora di una sequenza di dati in parallelo.

Nel seguito del paragrafo, verranno dapprima (§ I.2.1) presentate le modalita' di generazione del segnale che sostiene l'informazione emessa da una sorgente numerica o da una analogica su cui si opera una conversione A/N. Successivamente (§ I.2.2), si forniranno alcuni esempi significativi delle tecniche di compressione di banda. Infine (§ I.2.3), verranno definite le caratteristiche di emissione di una sorgente e, sulla base di queste, si

effettueranno alcune prime classificazioni dei servizi di telecomunicazioni ad essa associati.

1.2.1 I segnali emessi

Da un punto di vista fisico, l'emissione di una sequenza di simboli o delle parole binarie in corrispondenza con questi e' sostenuta da un *segnale numerico*, e cioe' da un segnale costituito da una successione di segnali elementari (*elementi di segnale*), che si susseguono nel tempo e a ciascuno dei quali e' associata una porzione dell'informazione emessa dalla sorgente (ad esempio, una cifra binaria). Questa associazione e' attuata facendo assumere a una grandezza caratteristica di ogni elemento di segnale (ad esempio, alla sua ampiezza, durata, ecc.) valori discreti che sono in corrispondenza con l'informazione da rappresentare e che sono individuati sull'asse dei tempi da un *istante significativo*.

Un segnale numerico e' quindi caratterizzato da una duplice discretizzazione: quella dell'asse dei tempi, che e' costituita dalla successione degli istanti significativi, e quella dei valori della grandezza caratteristica che si e' scelta per rappresentare l'informazione.

Conseguentemente, elemento fondamentale per la generazione di un segnale numerico o per l'estrazione da questo dell'informazione sostenuta e' la disponibilita' di un dispositivo che scandisce nei due casi i tempi di generazione e di estrazione, rispettivamente. Tale dispositivo e' un *orologio*, che definisce la *temporizzazione* del segnale.

Le sorgenti audio e l'associato segnale numerico

L'informazione emessa da una sorgente audio (informazione audio) e' associata all'intensita' di suono, generata, ad esempio, da un parlatore (informazione fonica) o da una fonte musicale. Le variazioni nel tempo di tale intensita' (ovvero della corrispondente grandezza elettrica dopo una trasduzione acusto-elettrica) sono rappresentate da un segnale che e' analogico e a tempo-continuo: cioe' le sue ampiezze variano in un continuo e la relativa base dei tempi coincide con l'asse dei numeri reali positivi \mathbb{R}^+ . Sia quindi $x(t)$, $t \in \mathbb{R}^+$ questo segnale.

L'estrazione da $x(t)$, $t \in \mathbb{R}^+$ di un insieme di simboli avente cardinalita' limitata richiede due operazioni di discretizzazione: una, effettuata nel dominio del tempo, e' un campionamento; l'altra e' svolta nel dominio delle ampiezze.

Il campionamento trasforma $x(t)$, $t \in \mathbb{R}^+$ in un segnale ancora analogico, ma a tempo-discreto $x(t_k)$, $k \in I$ (ove I e' l'insieme dei numeri interi), che e' rappresentato dai valori istantanei assunti da $x(t)$, $t \in \mathbb{R}^+$ in una preassegnata sequenza di istanti temporali $\{t_k, k \in I\}$ (istanti di campionamento). Nelle applicazioni piu' usuali tale sequenza e' scelta in modo tale che il campionamento sia uniforme, e cioe' in modo che l'intervallo temporale tra due

successivi istanti di campionamento sia costante; in questo caso, detto intervallo e' il periodo di campionamento T_c , il suo inverso e' la frequenza di campionamento F_c e gli istanti di campionamento formano la sequenza $\{kT_c, k \in I\}$.

Al segnale analogico e a tempo-discreto $x(kT_c)$, $k \in I$, che si ottiene da un campionamento uniforme, possono essere associate concettualmente:

- una sequenza analogica $\{x_k, k \in I\}$, in cui si prescinde da ogni riferimento temporale e il cui elemento k -esimo e' definito da $x_k = x(kT_c)$;
- una base dei tempi, su cui hanno significativita' solo gli istanti di campionamento e in cui l'istante k -esimo $t = kT_c$ e' in corrispondenza biunivoca con l'elemento k -esimo della sequenza $\{x_k, k \in I\}$.

Come e' noto, affinche' un'operazione di campionamento uniforme sia reversibile, e cioe' affinche' dal segnale risultante $x(kT_c)$, $k \in I$ sia possibile ricostruire senza alterazioni il segnale originario $x(t)$, $t \in \mathbb{R}^+$, e' necessario che quest'ultimo sia strettamente limitato nella banda $|f| \leq F_M$ contigua all'origine dell'asse delle frequenze e che la frequenza di campionamento F_c sia scelta di valore non inferiore a $2F_M$.

La discretizzazione dell'asse dei tempi e' premessa essenziale per il vero e proprio processo di conversione A/N, che consiste nella trasformazione della sequenza $\{x_k, k \in I\}$ in quella quantizzata $\{x_{qk}, k \in I\}$; gli elementi di questa sono rappresentati da un numero di cifre finito appartenente a un alfabeto di opportuna dimensionalita' e sono in corrispondenza biunivoca con gli istanti di campionamento.

Dalla sequenza $\{x_{qk}, k \in I\}$ con l'associata base dei tempi, dopo un suo trasferimento o una sua memorizzazione che non introducano alterazioni, si puo' riottenere una versione quantizzata $x_q(t)$, $t \in \mathbb{R}^+$ del segnale originario $x(t)$, $t \in \mathbb{R}^+$ (segnale quantizzato). Cio' e' possibile attraverso:

- una operazione di conversione numerico-analogica (conversione N/A), che trasforma la sequenza $\{x_{qk}, k \in I\}$ in un segnale a tempo-discreto $x_q(kT_c)$, $k \in I$ in forma quantizzata, e cioe' avente ampiezze che possono assumere solo un numero finito di valori possibili; cio' e' possibile utilizzando la base dei tempi associata a $\{x_{qk}, k \in I\}$ e ponendo $x_q(kT_c) = x_{qk}$.
- una operazione di interpolazione, che trasforma il segnale $x_q(kT_c)$, $k \in I$ nella sua versione a tempo-continuo $x_q(t)$, $t \in \mathbb{R}^+$ l'interpolazione e' l'operazione inversa a quella di campionamento e non introduce alterazioni se sono rispettate le condizioni per la reversibilita' del processo di campionamento.

Le operazioni di conversione A/N e N/A, che consentono di trasformare la sequenza analogica $\{x_k, k \in I\}$ nel segnale quantizzato a tempo-discreto $x_q(kT_c)$, $k \in I$, costituiscono il processo di quantizzazione. Prescindendo dalle modalita' con cui viene attuata, la quantizzazione e' un'operazione non-lineare istantanea, che si articola in due fasi temporalmente distinte:

- la prima e' un processo di misura e mira a determinare, in ogni istante di campionamento kT_c e in relazione a una prestabilita suddivisione dell'asse delle ampiezze in un numero finito di intervalli quantici, il particolare intervallo entro cui cade il valore x_k da quantizzare; l'informazione estratta in questa fase e' espressa dal numero d'ordine dell'intervallo quantico individuato; questo numero viene normalmente rappresentato con parole di codice espresse in cifre binarie;
- la seconda consiste in un processo di restituzione e mira a generare, ad ogni istante in corrispondenza biunivoca con quello di campionamento, una ampiezza istantanea $x_q(kT_c)$, avente valore dipendente dal particolare intervallo quantico individuato nella fase di misura.

La quantizzazione può essere uniforme o non-uniforme a seconda che gli intervalli quantici siano o non siano di uguale larghezza. In ogni caso il risultato di questa operazione introduce nel segnale restituito una componente spuria, che è chiamata rumore di quantizzazione.

Nel caso dei segnali telefonici, per assicurare un adeguato rapporto segnale/rumore di quantizzazione in tutta la dinamica dei livelli di potenza media che possono essere assunti dal segnale originario $x(t)$, $t \in \mathbb{R}^+$ e per garantire al tempo stesso un contenimento del flusso binario da trasferire, si adotta una quantizzazione non-uniforme, in cui gli intervalli quantici sono di larghezza via via crescente all'aumentare delle ampiezze da quantizzare. Il legame tra larghezze degli intervalli quantici e ampiezze del segnale telefonico è chiamato legge di compressione.

Le sorgenti video e il segnale di immagine numerico

Nel caso di una sorgente video, l'informazione emessa è costituita da immagini, e cioè dalle proiezioni su un piano di una visione del mondo reale. Si distinguono immagini fisse e in movimento. In ambedue i casi si supponga, per generalità, che le immagini siano a colori.

In questa ipotesi, l'informazione di una immagine fissa è associata alla luminosità e al colore in ogni areola nel piano S_2 dell'immagine. Se le immagini sono in movimento, le informazioni di luminosità e di colore debbono essere aggiornate nel tempo per riprodurre gli spostamenti dei soggetti rappresentati; sono quindi relative a ogni elemento di una successione continua di immagini.

In ambedue i casi l'informazione di luminosità è associata fisicamente all'intensità di luce, e cioè all'energia nel campo del visibile per unità di superficie sul piano S_2 . Invece, l'informazione di colore riguarda la distribuzione della luce nell'intervallo di lunghezze d'onda delimitato all'incirca dal rosso (780 nm) e dal violetto (380 nm).

D'altra parte, per le proprietà percettive dell'occhio umano, le informazioni di luminosità e di colore possono essere descritte da tre sole componenti di luce. Queste possono essere le intensità di luce monocromatica relative ai tre colori primari (rosso, verde e blu). In alternativa, come avviene nel caso di informazioni televisive, possono anche essere la luminanza da un lato e le crominanze del rosso e del blu dall'altro. Questo secondo insieme di componenti di luce è ottenibile dal primo tramite una trasformazione lineare.

Conseguentemente, se l'immagine è fissa, la relativa rappresentazione è costituita da tre funzioni di punto $f_i(P)$, $P \in S_2$ ($i=1,2,3$), corrispondenti ai valori delle tre componenti di luce che sono assunti al variare del punto $P(x_1, x_2)$ in S_2 : si può allora parlare di una funzione di punto $\mathbf{f}(P)$, $P \in S_2$, di natura vettoriale a tre componenti. I valori delle componenti di $\mathbf{f}(P)$ variano in un continuo e sono in corrispondenza, ad esempio, con la coordinata orizzontale x_1 e con quelle verticali x_2 di S_2 , definite nell'insieme dei numeri reali ($x_1, x_2 \in \mathbb{R}$).

Nel caso invece di immagini in movimento, la relativa rappresentazione è ancora una funzione di punto $\mathbf{f}(P, t)$, anch'essa di natura vettoriale a tre componenti, ma definita nello spazio tridimensionale S_3 : cioè, alle due coordinate x_1 e x_2 del piano S_2 occorre aggiungere una coordinata temporale t , anch'essa definita nell'insieme dei numeri reali ($t \in \mathbb{R}^+$).

Per scopi di trasferimento o di memorizzazione, dette funzioni vettoriali di punto negli spazi S_2 o S_3 debbono essere trasformate in funzioni ancora vettoriali, ma di una sola variabile definita nel dominio del tempo T . La trasformazione ha come obiettivo la generazione del segnale di immagine $\mathbf{g}(t)$, $t \in \mathbb{R}^+$ che è il sostegno fisico dell'informazione video: per quanto detto in precedenza, si tratta di un segnale vettoriale a tre componenti, che è analogico e a tempo-continuo. Ognuna di tali componenti è il risultato della trasformazione

operata su una singola componente di luce; ad esempio, nel caso dell'informazione televisiva, si hanno un segnale di luminanza e due segnali di cromaticità. Esaminiamo le modalità di queste trasformazioni dagli spazi S_2 o S_3 al dominio del tempo T .

Se l'immagine è fissa (come è il caso di una immagine fotografica), si opera un campionamento della dimensione verticale di S_2 (campionamento verticale), e cioè una discretizzazione della coordinata verticale x_2 . Ciò equivale a scandire S_2 lungo linee orizzontali (righe) e a prelevare l'andamento delle componenti di luce lungo ogni riga; in questa operazione si parte dal bordo dell'immagine in alto a sinistra e si termina in corrispondenza del bordo in basso a destra. I parametri di questo campionamento sono il numero di righe di scansione e il numero di righe esplorato nell'unità di tempo: il primo di questi parametri qualifica (a parità di dimensioni dell'immagine) la fedeltà della riproduzione rispetto all'originale; il secondo definisce unitamente al primo la durata dell'operazione di scansione.

Se le immagini sono in movimento (come si verifica, ad esempio, per le immagini televisive), la trasformazione dallo spazio S_3 al dominio del tempo T richiede di discretizzare, oltre alla dimensione verticale, anche quella temporale di S_3 e cioè di utilizzare, per la rappresentazione, un numero finito di immagini nell'unità di tempo. Ciò equivale ad effettuare, su ognuna delle componenti di luce, un campionamento temporale attuato a frequenza di quadro F_q .

La scelta del valore di F_q (uguale a 25 Hz nel sistema televisivo europeo) determina la risoluzione temporale della rappresentazione; è effettuata con l'obiettivo di una resa del movimento senza alterazioni (rispetto all'originale) percepibili da parte di un osservatore e tiene conto delle proprietà di persistenza di una immagine nel sistema visivo umano.

Dato poi che l'intervallo di tempo tra una immagine e la successiva (periodo di campionamento temporale) è utilizzato per effettuare la scansione orizzontale di una immagine, la frequenza di riga F_r (numero di righe esplorato nell'unità di tempo) è uguale al prodotto del numero N_{rt} di righe totali per ogni immagine ($N_{rt} = 625$ nel sistema televisivo europeo) e della frequenza di quadro F_q . Ne deriva un periodo di riga $1/F_r$ (periodo di campionamento verticale) che è dato da $1/N_{rt}F_q$ e che, per il sistema televisivo europeo, è uguale a 64 μs .

Riguardo al numero N_{rt} di righe totali, questo deve tenere conto degli intervalli di ritorno di riga sia in orizzontale che in verticale ed è quindi maggiore del numero N_{ra} di righe attive. Il numero N_{ra} determina la risoluzione verticale, che si consegue nella rappresentazione di ogni immagine. Nel sistema televisivo europeo, $N_{ra} = 576$.

Come risultato delle due operazioni di campionamento su immagini in movimento, la base dei tempi per le tre componenti del segnale di immagine $g(t)$ (e cioè ad esempio per il segnale di luminanza e i due segnali di cromaticità) è scandita dagli istanti di campionamento verticale con frequenza di riga e di campionamento temporale con frequenza di quadro. È però da osservare che i due segnali di cromaticità vengono normalmente rappresentati con una larghezza di banda ridotta rispetto a quella del segnale di luminanza, dato che esse hanno minore influenza sulla percezione soggettiva delle immagini.

Quanto detto chiarisce le modalità di generazione di un segnale di immagine (fissa o in movimento) che è analogico. Per effettuarne la conversione in un segnale numerico occorre effettuare un ulteriore campionamento, che riguarda la dimensione orizzontale del piano S_2 (campionamento orizzontale) e che consente di individuare un numero finito di punti per ogni riga. In corrispondenza di ognuno di questi punti vengono estratti i campioni delle tre componenti di luce. Questi tre campioni costituiscono l'elemento minimo di informazione

della sorgente video e cioè' il cosiddetto elemento di immagine o pixel (picture element).

Il numero di pixel attivi per ogni riga di scansione determina la risoluzione orizzontale dell'immagine fissa o in movimento, mentre il numero totale di pixel per riga definisce (tramite la frequenza di riga) la frequenza di pixel.

Dopo la discretizzazione delle ampiezze dei campioni delle componenti di luce, la forma numerica di immagini fisse o in movimento e' rappresentata da terne di matrici bidimensionali o da successioni temporali di terne di matrici bidimensionali, rispettivamente. Gli elementi di ognuna di queste matrici sono i pixel rappresentabili con parole di codice espresse in cifre binarie. La risoluzione spaziale (verticale e orizzontale) di immagini fisse o in movimento e' allora definita dai numeri di righe e di colonne che compongono ogni matrice bidimensionale, mentre il numero di matrici bidimensionali che si susseguono nell'unita' di tempo qualificano la risoluzione temporale delle immagini in movimento.

Infine, la serializzazione degli elementi di ognuna delle matrici bidimensionali associate a ciascuna componente di luce consente di ottenere una sequenza, ancora di natura vettoriale a tre componenti, avente elementi in forma quantizzata.

Se l'immagine e' fissa, questa sequenza e' autosufficiente per rappresentare completamente la relativa informazione video: si puo' cioe' prescindere da ogni riferimento temporale in modo analogo a quanto si e' visto per una sequenza di dati in serie o in parallelo.

Nel caso invece di immagini in movimento, a questa sequenza deve essere associata una base dei tempi in modo analogo a quanto si e' visto nella conversione A/N di un segnale vocale. Su questa base debbono essere precisati gli istanti relativi al campionamento temporale con frequenza di quadro, al campionamento verticale con frequenza di riga e al campionamento orizzontale con frequenza di pixel.

Dal punto di vista della temporizzazione di un segnale numerico, si distinguono:

- *segnali numerici isocroni*, in cui gli intervalli di tempo tra istanti significativi consecutivi hanno, almeno in media, la stessa durata ovvero durate che sono multipli interi della durata piu' breve; gli scostamenti dalla media debbono essere di valore massimo contenuto entro limiti specificati;
- *segnali numerici anisocroni*, in cui non sono verificate le condizioni che caratterizzano un segnale isocrono.

Al fine di rappresentare fisicamente la successione degli istanti significativi di un segnale numerico isocrono, a questo deve essere associato, in emissione e in ricezione, un *crono-segnale*, e cioe' un segnale periodico, in cui gli istanti significativi sono individuati da particolari valori dell'ascissa temporale: ad esempio, se il crono-segnale e' *sinusoidale*, gli istanti significativi sono rappresentati dagli attraversamenti dello zero con pendenza positiva (o negativa), mentre, se il crono-segnale e' *ad onda quadra periodica*, gli istanti significativi sono individuati dalla posizione temporale del fronte di salita (o di discesa).

Il processo di associazione tra una sequenza di parole binarie di codice e un segnale numerico consente di rappresentare fisicamente l'informazione emessa

da una sorgente. Al riguardo, e' esplicitativo considerare il caso di una sorgente di dati e dell'associato segnale numerico.

Il segnale di dati

L'informazione associata a una sequenza di dati in serie o in parallelo e' emessa sotto forma di un segnale numerico (segnale di dati), che sostiene la sequenza aggiungendo agli elementi di questa il riferimento a una base dei tempi.

Il processo di associazione tra una sequenza di dati seriali e un segnale di dati isocrono e' mostrato graficamente in figura I.2.1, nel caso in cui la sequenza di dati sia in serie. In particolare, la figura I.2.1a illustra la sequenza di dati seriali con la successione dei suoi stati logici "0" e "1", mentre la figura I.2.1c riproduce l'andamento del crono-segnaled, che definisce la base dei tempi.

Il segnale di dati e' poi descritto in figura I.2.1b ed e' ottenuto ponendo in corrispondenza gli stati "0" e "1" della sequenza di figura I.2.1a, ad esempio, con due valori di tensione uguali a $+V$ e a $-V$, rispettivamente. Tali valori sono assunti in corrispondenza degli istanti significativi del segnale e sono mantenuti costanti per tutto l'intervallo tra un istante significativo e il successivo. Gli istanti significativi sono equidistanziati (almeno in media) con intervallo che, al minimo, puo' assumere il valore del periodo T_s del crono-segnaled e che, in generale, puo' essere un multiplo intero di questo. Analogo processo di associazione puo' essere attuato operando su una sequenza di dati in parallelo.

La formazione del segnale di dati a partire da una sequenza di dati in serie o in parallelo puo' pero' essere svolta in due modi alternativi, a seconda che i caratteri vengano generati con cadenza regolare o meno. Con riferimento al caso di sequenza di dati in serie e nell'ipotesi che la cadenza di generazione dei caratteri sia regolare, l'emissione puo' avvenire con un segnale di dati isocrono, come si e' ipotizzato in figura I.2.1.

Si attua invece una emissione con un segnale di dati anisocrono quando vengono generati caratteri ad istanti aleatori e risulta allora conveniente emettere le corrispondenti parole di codice nel momento in cui queste escono dal processo di codifica, senza tenere conto dei caratteri che precedono o che seguono. In questo caso, il segnale di dati sostiene una successione di gruppi di cifre binarie, tali che l'intervallo tra due cifre binarie successive nello stesso gruppo e' sempre uguale a T_s (o a un suo multiplo intero), mentre l'intervallo tra due gruppi successivi puo' essere di durata qualunque. Ogni gruppo e' costituito da una parola di codice preceduta e seguita da elementi di segnale, che consentono di riconoscere l'inizio e la fine del carattere e che sono chiamati START e STOP, rispettivamente (fig. I.2.2). La durata dello START e' uguale a T_s , mentre quella dello STOP e' definita solo per la sua durata minima, che, in alternativa, e' fissata uguale a 1 o 1,5 o 2 volte l'intervallo T_s .

Anche dalle sequenze di dati in parallelo possono essere ottenuti segnali di dati isocroni o anisocroni, con modalita' di formazione ottenibili per generalizzazione di quanto detto in precedenza. Nella pratica, la emissione di dati in parallelo e' soprattutto utilizzata in modo anisocrono e su brevi distanze, come ad esempio all'interno di una apparecchiatura, ove l'insieme fisico degli n circuiti necessari per trasferire un carattere e' denominato bus.

E' da aggiungere che gli attributi "isocrono" e "anisocrono" per un segnale di dati sono comunemente sostituiti, nel linguaggio corrente, con quelli "sincrono" e "asincrono", rispettivamente.

Una sorgente numerica ovvero il suo equivalente a valle di una conversione analogico-numerica ha un comportamento che e' descrivibile mediante

l'andamento nel tempo del *ritmo binario di emissione*, e cioè' del numero di cifre binarie che sono emesse nell'unità' di tempo durante intervalli in cui questo numero rimanga invariato.

Nel caso in cui l'emissione sia sostenuta da un segnale numerico isocrono, in cui ogni elemento di segnale sia portatore di una singola cifra binaria e in cui il cronosegnale associato abbia periodo T_s , il ritmo binario di emissione è costante e uguale a $1/T_s$.

Se invece l'emissione è con segnale numerico anisocrono, in cui ogni elemento di segnale sia ancora portatore di una singola cifra binaria, si possono normalmente individuare tratti di segnale in cui l'intervallo tra due istanti significativi successivi è costante e che si susseguono con altri tratti in cui detto intervallo assume un valore diverso. Si può allora definire un ritmo binario di emissione che varia da tratto a tratto e, tra i valori assunti da questi ritmi, si può individuare un valore massimo, che è chiamato *ritmo binario di picco*. Questo è quindi il numero massimo di cifre binarie che la sorgente emette nell'unità' di tempo.

Esempio I.2.1 Per caratterizzare una emissione di dati in modo anisocrono è d'uso riferirsi alla cosiddetta *velocità di modulazione* F_v , espressa in *baud*. Questa esprime il ritmo binario nell'ambito di un carattere e, tenendo conto che il ritmo binario nell'intervallo tra due caratteri è nullo, è anche il ritmo binario di picco che la sorgente di dati emette. La velocità F_v è infatti definita da $1/T_v$, e cioè' come l'inverso della durata T_v , espressa in secondi, dell'intervallo più breve tra due istanti significativi successivi del segnale di dati. Si osserva che, nel caso di emissione con segnale numerico isocrono, il valore di F_v coincide con quello del ritmo binario espresso in bit/s.

I.2.2 Tecniche di compressione di banda

Su sorgenti che emettono informazione in forma numerica si opera talvolta una *codifica di sorgente*. Come è noto, questa ha lo scopo di minimizzare il ritmo binario che è richiesto per rappresentare l'informazione con una fedeltà specificata all'uscita della operazione di codifica ed è giustificata dalla possibile esigenza di ridurre, a parità di quantità di informazione emessa, il numero di cifre binarie che devono essere trasferite o memorizzate. Ciò vale sia per sorgenti di dati, sia nel caso di sorgenti originariamente analogiche, come quelle che emettono informazione audio o video in forma numerica. Per queste ultime il problema non è quindi solo quello di rappresentare con un segnale numerico l'informazione emessa, ma risiede anche nel ridurre la ridondanza del segnale di informazione che risulta da tale rappresentazione.

Nel caso dei segnali di informazione audio e video, una riduzione di ridondanza (o *compressione di banda*) e' possibile in quanto, nella maggior parte dei casi, questi segnali presentano una certa correlazione interna, tale che i campioni successivi estratti da questi segnali non sono statisticamente indipendenti. La codifica di sorgente consiste nel ridurre tale correlazione (*ridondanza statistica*) e se cio' non e' sufficiente, nello sfruttare le caratteristiche psicofisiche dei sistemi umani che presiedono alle percezioni auditive o visive. Queste caratteristiche consentono una qualita' di suono o di immagine che e' soddisfacente per chi ascolta o osserva anche eliminando, in modo irreversibile e quindi con perdita, parte dell'informazione originale (*ridondanza percettiva*).

La utilizzazione delle tecniche di elaborazione numerica dei segnali e le possibilita' oggi offerte dalle tecnologie circuitali VLSI rendono possibile realizzare:

- algoritmi di *compressione di banda* anche molto complessi, che permettono un trasferimento su un canale con larghezza di banda ridotta (rispetto a quella richiesta nel caso di assenza di compressione);
- algoritmi complementari di *espansione* che, a partire dall'informazione ricevuta, siano in grado di restituire in uscita una versione del segnale originario caratterizzata da qualita' sufficiente per l'ascoltatore o per l'osservatore.

L'efficacia di questi algoritmi e' normalmente qualificata mediante il *rapporto di riduzione di banda*, e cioe' mediante il rapporto tra le larghezze di banda di canale che sono richieste per il trasferimento dell'informazione emessa nei due casi di assenza e di presenza di compressione. Come e' intuitivo, la complessita' di un algoritmo di compressione cresce all'aumentare del rapporto di riduzione di banda che si desidera conseguire con una fissata fedelta'.

Dato pero' che l'attuazione di un algoritmo di compressione-espansione di banda introduce una certa degradazione del segnale riprodotto rispetto a quello originario (degradazione che e' tanto piu' elevata quanto maggiore e' il rapporto di riduzione di banda), la scelta della tecnica di codifica di sorgente e' anche legata al tipo di informazione e di servizio considerati, dato che a questo tipo corrisponde uno specifico requisito di qualita'. Le implicazioni di tale affermazione sono chiarite dai due casi di codifica su informazioni audio e video, rispettivamente.

Codifica di informazioni audio

Per la codifica delle informazioni audio occorre distinguere i due casi dell'informazione

vocale, che è scambiata nell'utilizzazione di servizi conversazionali, e dell'informazione musicale, che è distribuita nei servizi radiofonici a diffusione circolare. Tali casi si differenziano non solo per il tipo di segnale trattato (segnale vocale nel primo caso e segnale musicale nel secondo), ma anche per i diversi requisiti di qualità di servizio.

Nel caso delle applicazioni foniche si richiede in primo luogo l'intelligibilità del parlato e la riconoscibilità del parlatore con i suoi stati emotivi (qualità telefonica); ciò implica che la larghezza di banda del segnale vocale sia di 3,1 kHz, secondo quanto previsto per una conversazione telefonica. Per riprodurre la voce tramite altoparlante, come è richiesto nei servizi di tele-conferenza, è normalmente richiesta una qualità migliorata, per la quale detta larghezza di banda è di 7 kHz.

Per l'informazione vocale a qualità telefonica si impiega ugualmente la codifica PCM, secondo una delle due leggi di compressione (legge A e legge μ) previste nella Racc. G.711 senza alcun accorgimento di compressione di banda se non quello di una quantizzazione non uniforme. Per il conseguimento della qualità migliorata si adotta invece lo schema di codifica descritto nella Racc. G.722. Nel primo caso l'informazione emessa è con ritmo binario costante che è uguale a 64 kbit/s; nel secondo caso, l'emissione può essere, in alternativa, a 48, 56 o 64 kbit/s.

La voce a qualità telefonica può essere codificata anche con ritmi binari inferiori a 64 kbit/s. Ad esempio, il codificatore, che è descritto nella Racc. G.721 e che opera con codifica PCM differenziale adattativa (ADPCM - Adaptive Differencial PCM), rappresenta il segnale vocale con un ritmo binario di 32 kbit/s.

Tutti gli esempi sopra citati corrispondono a schemi di codifica di forma d'onda. Per conseguire ritmi binari di emissione inferiori a 32 kbit/s, si utilizzano codificatori di analisi per sintesi. Un primo esempio al riguardo è offerto dalla Racc. G.728, che prevede una emissione a 16 kbit/s. Per applicazioni nei sistemi radio-mobile, l'ITU-T ha poi iniziato recentemente la definizione di un nuovo codificatore a 8 kbit/s e si prevede che il passo successivo sia verso un ulteriore codificatore a 4 kbit/s.

Nel caso in cui l'informazione da trattare sia musicale, i requisiti di qualità sono estremamente più elevati rispetto a quanto necessario nelle applicazioni foniche. Il termine di paragone è la qualità offerta dal CD (Compact Disk), che produce, in condizioni stereofoniche, un flusso numerico di circa 1,41 Mbit/s (derivante da una frequenza di campionamento uguale a 44,1 kHz e da una quantizzazione con 16 bit/campione).

Le tecniche di riduzione di ridondanza proposte per questa applicazione hanno come obiettivo un numero di bit/campione uguale a circa 2, con un conseguente flusso numerico di circa 100 kbit/s per canale monofonico. Impiegando una codifica a sotto-bande (32 sotto-bande per canale monofonico), il flusso numerico in condizioni stereofoniche si riduce a 384 kbit/s. Se si adotta anche un modello psico-acustico della percezione uditiva rispetto al quale viene controllato il processo di quantizzazione, si ottiene un ritmo binario di codifica che è uguale a 256 kbit/s.

Codifica di informazioni video

Ci riferiamo alla codifica di sorgente su informazioni video e partiamo dal caso di immagini in movimento.

Secondo quanto previsto nella Racc. 601 del CCIR, il formato SDTV pieno (Standard Definition TV) nel sistema televisivo europeo è costituito da una matrice di pixel attivi di luminanza che ha 576 righe e 720 colonne, mentre l'analogica matrice per il formato HDTV (High Definition TV) sempre per il sistema europeo comprende 1152 righe e 1920 colonne. In ogni caso la frequenza di quadro è di 25 Hz. Inoltre, le matrici delle componenti di cromaticità hanno lo stesso numero di righe della matrice di luminanza, ma un numero di

colonne che e' ridotto a meta'. Con questi dati, se si effettua una discretizzazione delle ampiezze a 256 intervalli quantici (8 bit/campione), i segnali di immagine SDTV e HDTV in forma non compressa hanno ritmi binari uguali a 166 e a 885 Mbit/s, rispettivamente.

Rispetto a questi valori, occorre fare due osservazioni. In primo luogo, le attuali tecniche di codifica di sorgente, basate sulla riduzione di ridondanze sia statistiche che percettive, consentono di ottenere codificatori capaci di operare rapporti di compressione di banda intorno a 20-30:1. Inoltre, in molte applicazioni le immagini in movimento non devono essere rappresentate con le risoluzioni spaziale e temporale precedentemente precisate. Le implicazioni di queste osservazioni possono essere chiarite con vari esempi di algoritmi di codifica.

Un primo esempio e' offerto dall'algoritmo descritto nella Racc. H.261 del CCITT, nella quale si considera un codificatore video per applicazioni audiovisive (video-telefonia o video-conferenza). Riducendo la risoluzione spaziale rispetto al formato SDTV pieno, si ottiene un ritmo binario di emissione che puo' assumere valori uguali a $p \times 64$ kbit/s ($p=1,...,31$) e quindi compresi tra 64 e 1984 kbit/s.

Come secondo esempio, si puo' citare l'algoritmo di codifica MPEG-1, sviluppato in ambito ISO da un apposito gruppo di lavoro (Moving Picture Experts Group) per la memorizzazione e il recupero di immagini in movimento con forma numerica (ad esempio, con formato SDTV pieno) su supporti magnetici (ad esempio, su CD). Il ritmo binario emesso da un codificatore MPEG-1 e' nell'intervallo 1-1,5 Mbit/s. Cio' consente di assicurare una qualita' paragonabile o migliore di quella ottenibile dalla video-registrazione analogica.

E' anche importante menzionare l'attivita' attualmente in corso per definire un ulteriore codificatore video (MPEG-2) con una gamma di applicazioni decisamente piu' larga di quella prevista per il codificatore MPEG-1. L'obiettivo e' la codifica del formato SDTV pieno con ritmo binario di emissione uguale a circa 5 Mbit/s per scopi di diffusione e a circa 10 Mbit/s per applicazioni di contribuzione (cioe' per trasferimento tra studi di immagini destinate a successive elaborazioni). E' anche previsto il trattamento del formato HDTV con ritmi binari compresi tra 15 e 25 Mbit/s.

Per cio' che riguarda infine le immagini fisse di tipo fotografico, si puo' ricordare l'algoritmo di codifica JPEG, anch'esso sviluppato in ambito ISO dall'omonimo gruppo di lavoro (Joint Photographic Experts Group). Il formato scelto per il campionamento dell'immagine e' quello SDTV pieno, che e' considerato sufficiente per il trasferimento e per la memorizzazione di una immagine fissa di buona qualita'.

Nel cosiddetto "Sistema di base", l'algoritmo di compressione JPEG conserva la fedelta' dell'immagine con una codifica fino a circa 1 bit/pixel, ma con possibilita' di perdita rispetto all'immagine originaria: cio' consente, disponendo di un canale avente la capacita' di 64 kbit/s, di trasferire un'immagine con detto formato in poco piu' di 6 s. Nel "Sistema esteso", sono previste alcune potenzialita' addizionali destinate a soddisfare requisiti specifici, quali la formazione progressiva, che offre la possibilita' di ottenere una visione totale e riconoscibile dell'immagine dopo meno di 2 s, e una codifica senza perdite, che consente di ottenere una immagine identica a quella originale.

1.2.3 Caratteristiche di emissione

Dal punto di vista delle loro caratteristiche di emissione, le sorgenti di informazione possono essere a *ritmo binario costante* (CBR - Constant Bit Rate) e a *ritmo binario variabile* (VBR - Variable Bit Rate). In questo secondo caso, il ritmo binario emesso varia nel tempo tra un valore massimo (il ritmo binario di

picco) e un valore minimo, che può essere anche nullo. Tale variabilità è legata alla possibile variazione nel tempo del contenuto informativo all'ingresso del codificatore e alla convenienza di realizzare una codifica, che tenga conto di questa variazione, con l'obiettivo di utilizzare in modo più efficiente la capacità di canale richiesta per trasferire a distanza l'informazione emessa. Il caso delle sorgenti CBR può invece essere caratterizzato dal solo ritmo binario di picco.

In base al ritmo binario di picco emesso dalla sorgente associata, i servizi di telecomunicazione si possono suddividere in tre classi e cioè in quelle a *bassa*, *media* ed *alta* velocità (fig. I.2.3), con la precisazione che questi attributi sono in stretta relazione con le possibilità tecniche oggi offerte per il relativo trattamento.

La classe a bassa velocità comprende i servizi con ritmo binario di picco fino a 100 kbit/s, e cioè:

- i *servizi di telemetria* (per telelettura di contatori relativi ai consumi di energia, per impianti di sicurezza, per sondaggi di opinione, per telecontrolli con le finalità più varie) caratterizzati da ritmi binari orientativamente non superiori a 100 bit/s;
- i *servizi di comunicazione di testi* (Teletex, Telefax, Videotex) e alcuni *servizi di comunicazione di dati* con ritmo binario di emissione non superiore a 10 kbit/s (*dati in banda fonica*);
- i *servizi fonici* con ritmi binari compresi tra 10 e 100 kbit/s, in relazione alla qualità desiderata e, a parità di questa, alle modalità di codifica adottate per l'informazione vocale;
- il trasferimento di dati e di immagini fisse con ritmi binari compresi tra 10 e 100 kbit/s;
- i servizi che trattano immagini in movimento, aventi limitate esigenze di definizione e di riproduzione dei movimenti, come si verifica nella video-telefonia e nella video-conferenza.

Nella classe a media velocità, che comprende ritmi binari di picco tra 100 kbit/s e 2 Mbit/s, si incontrano servizi che hanno come obiettivo il trasferimento di:

- musica ad alta fedeltà;
- immagini fisse ad alta definizione, utilizzate in applicazioni con esigenze di tempo reale, quali la telemedicina;
- immagini in movimento con la qualità della televisione convenzionale (SDTV) e quindi con piena riproduzione dei movimenti, anche se con definizione limitata;

- dati per applicazioni particolari, quali il trasferimento di archivi di grossa mole in contesti di calcolo veloce.

Infine nella classe ad alta velocità, in cui le sorgenti hanno una capacità di emissione superiore a 2 Mbit/s, si trovano i servizi che trattano immagini in movimento con la qualità ancora della televisione convenzionale (SDTV) e di quella ad alta definizione (HDTV). Sono anche incluse comunicazioni di dati accoppiati ad immagini fisse, per applicazioni specifiche quali la progettazione assistita da calcolatore. I ritmi binari di questa classe, come anche quelli della precedente, potranno subire riduzioni anche sensibili in relazione ai progressi attualmente in atto nelle tecniche di elaborazione numerica dei segnali e nelle tecnologie circuitali VLSI.

CLASSE	SERVIZI	RITMO BINARIO DI PICCO (MBIT/S)
Bassa velocità	Telemetria	0,0001
	Dati/Testi	0,01
	Voce, Dati, Immagini fisse	0,1
Media velocità	Musica ad alta fedeltà	1
	Videotelefono/Videoconferenza	da 1 a 10
	Dati/Immagini fisse	da 1 a 10
Alta velocità	TV convenzionale	da 50 a 100
	TV ad alta definizione	da 150 a 1000
	Dati/Immagini fisse	da 10 a 1000

Tab. I-1 - Suddivisione dei servizi di telecomunicazione in classi di velocità.

Circa la variabilità o meno del ritmo binario emesso, un primo esempio di sorgente CBR è offerto dall'uscita di un codificatore che opera su un segnale analogico e che effettua su questo una conversione analogico-numerica senza ulteriori trattamenti. Sono questi i casi di sorgenti vocali a *qualità telefonica* o a *qualità migliorata*. Ulteriori esempi di sorgenti CBR si incontrano nei casi di codifica di immagini senza variazioni del grado di definizione in differenti regioni del quadro.

Sorgenti VBR si incontrano invece, ad esempio, nella codifica di immagini in cui si utilizza, per scopi di efficienza, la possibilità di variare il grado di definizione con cui vengono riprodotte differenti regioni di quadro. Il risultato è un'emissione in cui il ritmo binario varia nel tempo tra il suo valore di picco e valori inferiori scelti normalmente in un insieme discreto.

Per ciò che concerne poi la voce a qualità telefonica, è da osservare che le sue usuali caratteristiche CBR sono quelle risultanti da una operazione di codifica, in cui non si tiene conto degli intervalli di tempo in cui ognuno dei due parlatori è in fase d'ascolto ovvero introduce pause nelle sue interlocuzioni.

Conseguentemente, nella durata di questi intervalli, il codificatore emette sequenze di cifre binarie che non sono portatrici di informazione fonica.

Un risultato più efficiente si può conseguire se la codifica è accompagnata da una operazione di rivelazione dei *tratti vocali*, e cioè se si tiene conto sia dell'alternanza dei periodi di emissione e di ascolto, sia della struttura di ogni periodo di emissione in tratti vocali e in pause. In questo caso il risultato è una particolare sorgente VBR, in cui il ritmo binario emesso è a due valori, e cioè variabile tra un valore di picco e uno nullo. Sorgenti VBR di questo tipo sono chiamate "*tutto o niente*".

Il caso di sorgenti "tutto o niente" s'incontra anche in comunicazioni di dati di tipo interattivo, ove si hanno intervalli temporali in cui si verifica una emissione ininterrotta di cifre binarie e altri intervalli in cui l'emissione è sospesa. Ciò significa che la sequenza di dati emessi è strutturabile in intervalli di attività e di silenzio, che si alternano sull'asse dei tempi come descritto in Fig. I-9. Gli intervalli di attività sono chiamati *tratti informativi*, mentre quelli di silenzio sono detti *pause*.

Nelle sorgenti "tutto o niente" le durate dei tratti informativi e delle pause sono, in generale, quantità variabili in modo aleatorio e quindi descrivibili solo in modo probabilistico. Le distribuzioni di queste quantità sono strettamente legate al tipo di sorgente considerata e quindi al servizio nell'ambito del quale la sorgente opera.

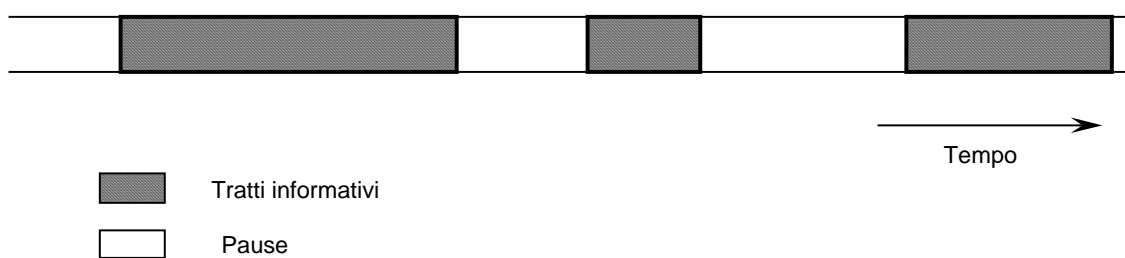


Fig. I-9 - Sequenza di emissione di una sorgente VBR "tutto o niente".

Nel caso di sorgenti VBR più generali di quelle "tutto o niente", l'emissione si presenta come una successione di tratti informativi in ognuno dei quali il ritmo binario emesso è costante e si modifica di valore passando da un tratto ad un altro. Circa le durate di questi tratti con i relativi ritmi binari emessi, valgono considerazioni analoghe a quelle fatte con riferimento alle sorgenti "tutto o niente": si è in presenza cioè di quantità variabili in modo aleatorio.

Conseguentemente una descrizione completa di una sorgente VBR o, in particolare, di una "tutto o niente", richiederebbe modelli probabilistici che possono essere di identificazione e di utilizzazione anche molto complesse. Se però la descrizione può essere ristretta al solo comportamento in media, si può fare riferimento a un *ritmo binario medio*, ottenuto come media dei valori dei ritmi binari che la sorgente emette nel tempo. Inoltre, per ogni sorgente la variabilità del ritmo binario può essere descritta dal *grado di intermittenza*, e cioè dal rapporto tra il ritmo binario di picco e quello medio.

Come è ovvio, il grado di intermittenza assume valore unitario nel caso di sorgenti CBR, mentre in quelle VBR il suo valore è maggiore dell'unità. Ad esempio, in sorgenti di dati operanti in comunicazioni interattive, il grado di intermittenza è dell'ordine di una decina, mentre assume valori mediamente di poco superiori a due per sorgenti di voce a qualità telefonica con rivelazione dei tratti vocali.

I.3 I servizi e le potenzialità di fornitura

Per fornire un'idea più precisa delle svariate possibilità di attuazione di una comunicazione, questo paragrafo è dedicato in primo luogo (§ I.3.1) alla descrizione di un ambiente di comunicazione secondo gli attuali orientamenti della tecnica.

Seguono

- una distinzione dei servizi in relazione alle potenzialità che occorre coinvolgere per la loro fornitura e che sono rese disponibili dalla infrastruttura di supporto in cooperazione o meno con gli apparecchi terminali (§ I.3.2);
- una illustrazione della tecnica degli attributi, che è utilizzata in contesti normativi per categorizzare i servizi e per descriverne le caratteristiche (§ I.3.3);
- una classificazione dei servizi basata sulle modalità di comunicazione, interattive o distributive, da essi offerte (§ I.3.4).

I.3.1 L'ambiente di comunicazione

Nell'evoluzione di un processo di comunicazione è richiesto in generale lo scambio di tre tipi di informazione, e cioè delle informazioni di utente, di segnalazione e di gestione.

L'*informazione di utente* include quanto viene emesso da una sorgente ed e' destinato a uno o piu' collettori di informazione per le finalita' di una particolare applicazione, ma comprende anche quanto viene in generale aggiunto al flusso informativo di sorgente (*extra-informazione*) per scopi di procedura o di protezione. Lo scambio dell'informazione di utente e' quindi l'obiettivo primario di un servizio di telecomunicazione.

Costituiscono l'informazione di utente, *in alternativa o in unione* parziale o totale, le forme codificate di voce, di suoni musicali, di dati, di testi, di immagini fisse o in movimento. A parita' di natura dell'informazione scambiata, e' la sua forma codificata che pone requisiti al relativo trattamento infrastrutturale. Al riguardo e' oggi d'uso parlare di *mezzo di rappresentazione* con riferimento a uno specifico tipo di informazione, come descritto dalla sua forma a valle di un operazione di codifica con o senza riduzione di ridondanza. Ad esempio la voce puo' essere individuata dal mezzo di rappresentazione descritto dalla Racc. G.711 dell'ITU-T.

In relazione poi alla capacita' di gestire un solo mezzo di rappresentazione o una pluralita' di questi, una comunicazione si dice *monomediale* nel primo caso o *multimediale* nel secondo. Gli stessi attributi sono utilizzati con riferimento a un servizio o a una applicazione.

L'informazione di utente puo' essere scambiata tra due o piu' utenti o tra utenti e centri di servizio e, nello scambio, puo' essere trattata dall'infrastruttura in modo trasparente oppure puo' essere elaborata, come accade nei casi di una archiviazione, di una conversione di mezzo di rappresentazione (ad esempio, da testo a voce sintetizzata), di una codifica crittografica svolta all'interno della rete.

L'*informazione di segnalazione* (o di controllo) e' di supporto affinche' possa avvenire lo scambio dell'informazione di utente. Essa ha lo scopo di consentire le interazioni tra cliente/utente, gestore di rete e fornitore di servizi nell'ambito di quanto previsto

- per inizializzare la comunicazione, per negoziarne le caratteristiche qualitative e quantitative iniziali e per modificare tali caratteristiche nel corso della comunicazione;
- per ottenere un arricchimento dei servizi di base con il coinvolgimento di risorse di elaborazione accessibili nell'ambiente di comunicazione.

Infine, l'*informazione di gestione* ha lo scopo di consentire il complesso di operazioni necessarie per gestire la fornitura dei servizi e i mezzi necessari allo scopo; in questi ultimi sono ovviamente incluse le risorse preposte al trasferimento delle informazioni di utente e di segnalazione. Tra le operazioni

sopra citate vanno menzionate quelle connesse all'erogazione del servizio (*operazioni di esercizio*), al suo mantenimento (*operazioni di manutenzione*) e al suo addebito (*operazioni di amministrazione*). Per lo svolgimento di tali operazioni deve essere previsto uno scambio di informazioni tra le apparecchiature di rete e quelle terminali. Oggetto dello scambio e' l'informazione di gestione, che e' quindi anch'essa di supporto al trattamento infrastrutturale dell'informazione di utente.

Il trasferimento delle informazioni di utente, di segnalazione e di gestione puo' essere attuato nell'ambito di un'unica infrastruttura. Questa e' la soluzione adottata in passato per le reti dedicate a un servizio. Gli attuali orientamenti verso l'integrazione dei servizi e verso "l'intelligenza" nelle reti sono a favore dell'impiego di infrastrutture separate. Per l'informazione di utente si attua allora una *rete di trasporto*, separata da quella di segnalazione (*rete di segnalazione a canale comune*, cfr. par. IV.6) e da quella di gestione (*rete di gestione delle telecomunicazioni*).

A queste tre infrastrutture corrispondono gruppi di funzioni tra loro distinti. Si fa riferimento alle:

- *funzionalita' di trasporto*, preposte al trattamento infrastrutturale dell'informazione di utente;
- *funzionalita' di controllo*, riguardanti gli obiettivi della informazione della segnalazione;
- *funzionalita' di gestione*, nell'ambito di quanto e' consentito dal trasferimento dell'informazione omonima.

Tali funzionalita' possono essere attuate con modalita' varie in relazione all'ambiente di comunicazione che si intende realizzare.

Se ci riferiamo all'insieme di utenti che possono fruire dei servizi offerti, una infrastruttura di telecomunicazione puo' essere:

- *pubblica*, se l'accesso e' consentito a chiunque provveda a stabilire un accordo contrattuale con il fornitore di servizi;
- *privata*, quando gli utenti abilitati all'accesso costituiscono un insieme chiuso con specifiche esigenze di comunicazione, che richiedono accordi tra cliente e fornitore non assimilabili a quelli in ambito pubblico.

In relazione poi alle possibilita' di accesso, da parte di utenti con o senza movimento, si e' gia' vista (par. I.1) la distinzione tra *reti mobili* e *reti fisse*, cosi' come, in base alla varieta' dei servizi offerti, si parla di *reti dedicate a un servizio* e di *reti integrate nei servizi*, a cui si e' accennato all'inizio di questo capitolo. Infine, se l'attenzione e' posta sull'estensione territoriale dell'infrastruttura, una distinzione importante, soprattutto per cio' che riguarda la

diversita' tra gli ambienti di comunicazione che vi corrispondono, e' tra *reti in area geografica* e *reti in area locale*.

Tramite queste infrastrutture un utente puo' usufruire di servizi di telecomunicazione. Per questi sono stati proposti, in seno agli organismi di normalizzazione, vari tipi di classificazioni, che e' utile prendere qui in considerazione per due motivi.

In primo luogo, a questo punto della trattazione e' opportuno fornire un quadro il piu' possibile ampio delle esigenze di servizio da parte dei clienti e delle possibilita' di offerta da parte dei fornitori. In secondo luogo l'insieme di queste classificazioni consentira' di iniziare a mettere in luce come un ambiente di comunicazione venga coinvolto nella fruizione e nella fornitura di un servizio.

Una prima distinzione, a cui si e' implicitamente accennato nel presentare gli scopi dell'informazione di segnalazione, e' tra *servizi di base* e *servizi supplementari*.

I primi possono essere offerti in modo indipendente dai secondi. Non e' pero' vero il contrario, in quanto i servizi supplementari modificano o complementano uno o piu' servizi di base. Ne segue che un servizio supplementare non e' offribile in modo autonomo da uno di base.

Esempio I.3-1 - Nella normativa internazionale, i servizi supplementari vengono distinti in *categorie*. Per alcune di queste verra' dato qui qualche esempio, limitando la citazione ai casi meno noti: infatti molti dei servizi supplementari sono gia' da tempo disponibili in reti private, ove sono normalmente di sostegno al servizio telefonico di base.

Nella categoria dei servizi supplementari *per l'identificazione del numero* rientrano, oltre alla *selezione passante*, le *presentazioni* (con o senza restrizioni) *dell'identificazione della linea chiamante* (a favore del chiamato) o *della linea connessa* (a favore del chiamante).

Tra i servizi *per l'offerta di chiamata*, vi sono quelli, ben noti in una rete privata, di *trasferimento di chiamata* e di *inoltro di chiamata*, quest'ultimo in condizione di utente occupato o in assenza di risposta del chiamato.

Alla categoria *per il completamento della chiamata* appartengono la *chiamata in attesa* e la *chiamata tenuta*, mentre nella categoria dei servizi *per parti multiple* e in quella *per comunita' di interessi* sono previsti, da un lato, la *chiamata in conferenza* e, dall'altro, il *gruppo chiuso di utenti*.

Il servizio "gruppo chiuso di utenti" (CUG-Closed User Group) consente ai suoi utenti di costituirsi in gruppi da e verso i quali l'accesso e' ristretto. Uno specifico utente puo' essere membro di uno o piu' CUG. I membri di uno specifico CUG possono comunicare tra di loro, ma non, in generale, con gli utenti esterni al gruppo. Membri specifici di un CUG possono avere la facolta' di originare chiamate verso l'esterno del gruppo e/o ricevere chiamate dall'esterno. Infine membri specifici di un CUG possono avere limitazioni addizionali, che impediscono loro di originare chiamate verso altri utenti del gruppo o ricevere chiamate da questi ultimi.

Un'ulteriore categoria e' quella preposta *all'addebito*. Di questa fanno parte l'*avviso di addebito* e, in particolare, la *chiamata con carta di credito*.

I.3.2 *Potenzialita' coinvolte*

Un servizio di comunicazione puo' essere classificato in base alle potenzialita' dell'ambiente di comunicazione, che sono coinvolte nella sua fornitura. Queste sono sostanzialmente di due tipi: *potenzialita' di rete* e, quando necessario, anche *potenzialita' di apparecchio terminale*.

Ciascuna potenzialita' e' definita come un insieme di funzioni che possono essere suddivise in due livelli per sottolineare la dipendenza gerarchica di un livello dall'altro: cioe' le *funzioni di basso livello* costituiscono la base per l'esecuzione delle *funzioni di alto livello* e queste ultime presuppongono lo svolgimento preventivo delle prime.

Le funzionalita' di basso livello sono preposte al trasferimento dell'informazione attraverso la rete. Le funzionalita' di alto livello riguardano invece gli aspetti connessi all'utilizzazione dell'informazione e possono anche includere funzionalita' di controllo e di gestione. Inoltre le funzionalita' di alto o basso livello sono *fondamentali* o *addizionali* a seconda che esse concorrano alla fornitura di servizi di base o di servizi supplementari, rispettivamente.

Sulla base di queste definizioni i servizi di telecomunicazioni possono essere distinti in *servizi portanti* e in *teleservizi*.

I servizi portanti forniscono, nella loro versione di base, la possibilita' di trasferire informazioni d'utente tra due o piu' terminazioni di rete. Essi coinvolgono quindi solo quelle potenzialita' di rete, che sono descritte da insiemi di funzioni fondamentali di basso livello. Nel caso in cui siano richiesti servizi supplementari, sono utilizzate funzioni addizionali di rete, ma sempre di basso livello.

I teleservizi forniscono, nella loro versione di base, una possibilita' di comunicare in senso lato, e cioe' comprendente oltre agli aspetti di puro trasferimento dell'informazione, anche quelli legati alla relativa utilizzazione. Cio' e' ottenuto per mezzo di una stretta cooperazione tra apparecchio terminale ed elementi di rete. Sono allora coinvolte, in aggiunta a potenzialita' di rete che corrispondono a funzioni fondamentali di basso livello, anche potenzialita' di apparecchio terminale che sono caratterizzate da insiemi di funzioni fondamentali di livello sia basso che alto. Per l'eventuale aggiunta di servizi supplementari, devono essere utilizzate funzioni addizionali di rete e di apparecchio terminale, che possono essere di livello sia basso che alto.

Nel caso di fornitura di un servizio portante, dato che la rete provvede con sole funzioni di livello basso, l'utente puo' scegliere un proprio insieme di

procedure relative alle interazioni tra funzioni di livello alto. D'altra parte, poichè la rete non assicura, in questo caso, alcun controllo di compatibilità, dovrà essere l'utente a verificare che l'insieme delle procedure scelte sia compatibile con le possibilità del proprio corrispondente.

Cio' non avviene nel caso di fornitura di un teleservizio, in quanto le procedure relative a interazioni tra funzioni di livello alto sono definite come parte del servizio.

Esempi di servizi portanti sono quelli *a circuito* e *a pacchetto*, su cui ci soffermeremo nel capitolo IV. Tra i teleservizi di base finora normalizzati si possono citare la *telefonia*, il *telex*, il *teletex*, il *telefax*, il *videotex*, avendo utilizzato per ognuno di questi la relativa denominazione internazionale.

Esempio I.3-2 - La *telefonia* e il *telex* sono servizi ben noti, ormai consolidati nelle loro modalità di impiego e nelle possibilità operative offerte ai loro utenti. In particolare la telefonia fornisce la possibilità di svolgere una conversazione vocale bidirezionale in tempo reale, mentre il telex consente una comunicazione interattiva di testi (insieme di caratteri).

Il *teletex* è un servizio per il trasferimento di testi scritti da terminale a terminale. Esso si presenta, rispetto al telex, come un servizio arricchito di alcune importanti caratteristiche quali il ritmo binario utilizzabile (non inferiore a 2400 bit/s), la riproduzione del documento trasmesso e le prestazioni locali del terminale.

Il *telefax* è un servizio per la riproduzione a distanza, normalmente in bianco e nero, di immagini fisse, che possono contenere dattiloscritti, manoscritti, grafici e fotografie. Questi documenti vengono acquisiti, nell'apparato trasmittente, mediante un'operazione di *scansione* ottica, e restituiti, nell'apparato ricevente, attraverso un'operazione di *registrazione*. La rappresentazione dell'informazione connessa all'immagine fissa è quindi attuata con modalità *a reticolo* (raster).

Il telefax ha subito vari aggiornamenti normativi. Tra questi quello relativamente più recente è il *telefax G4* (gruppo 4), nel quale, oltre al funzionamento tradizionale, è previsto anche quello a *modo misto*: cioè accanto a informazioni del tipo a reticolo sono trasferibili, in alternativa o in unione, anche informazioni di testi.

Il *videotex* è un servizio che consente ai suoi utenti, dotati di apparecchi terminali appositi, di accedere a una banca di informazioni. Gli elementi informativi interessati nel processo di scambio possono essere caratteri, elementi grafici e geometrici e, recentemente, anche elementi fotografici. Il servizio ha avuto realizzazioni specifiche nei vari Paesi; in Italia il videotex è commercializzato come *Videotel*.

I.3.3 La tecnica degli attributi

La caratterizzazione dei servizi portanti e dei teleservizi di base è fondata su un elenco di *attributi*, che riguardano il trasferimento dell'informazione di utente, le modalità di accesso e altre proprietà generali.

Tra i sette attributi relativi al trasferimento dell'informazione, se ne distinguono quattro, che sono detti *attributi dominanti* e che sono utilizzati per identificare una particolare categoria di servizi. I rimanenti tre sono gli *attributi*

secondari e sono impiegati per identificare un particolare servizio entro una categoria. Infine per caratterizzare le modalita' di accesso e altre proprieta' generali sono utilizzati gli *attributi qualificanti*, che meglio specificano un singolo servizio. Per ragioni di brevit  ci limitiamo qui a illustrare l'impiego degli attributi (dominanti e secondari) relativi al trasferimento dell'informazione.

Ognuno degli attributi puo' assumere un valore scelto in una gamma di alternative possibili. A questo riguardo conviene premettere alcune definizioni riguardanti l'informazione d'utente, che viene scambiata durante la fruizione di un servizio, e la sua relazione con l'interfaccia utente-rete (*punto di accesso*).

Nell'ambito di una comunicazione, la rete che ne consente lo svolgimento e' attraversata in primo luogo da *flussi informativi di utente*, che interessano almeno due interfacce utente-rete: l'una adiacente alla sorgente e l'altra prossima al collettore. Per ognuno di questi flussi, l'interessamento di una interfaccia utente-rete e' sommariamente caratterizzabile con

- il *verso di scorrimento*, che e' quello dell'informazione trasportata dal flusso: tale verso puo' essere *uscente* o *entrante* a seconda che il trasferimento sia dall'apparecchio terminale alla terminazione di rete o viceversa;
- il *ritmo di flusso*, che ha significativita' quando il flusso e' composto da cifre binarie con cadenza regolare e che e' misurato dal numero di cifre binarie componenti il flusso che attraversano l'interfaccia nell'unita' di tempo;
- la *portata media di flusso*, che e' un parametro significativo ogniqualvolta il flusso considerato e' composto da cifre binarie che non hanno cadenza regolare e che, in queste condizioni, misura (ad esempio in bit/s) la quantita' di informazione che viene mediamente trasportata dal flusso nell'unita' di tempo attraverso l'interfaccia; come e' evidente ritmo e portata media di flusso coincidono quando il flusso e' composto da cifre binarie con cadenza regolare.

Gli attributi dominanti sono:

- il *modo di trasferimento*;
- il *ritmo di trasferimento*;
- la *potenzialita' di trasferimento*;
- la *struttura dell'informazione trasferita*.

Piu' in particolare

- il *modo di trasferimento* descrive le modalita' operative che sono seguite nella fornitura di un servizio per trasferire l'informazione d'utente da un punto di accesso ad un altro;
- il *ritmo di trasferimento* riguarda il flusso informativo che il servizio attiva attraverso l'interfaccia utente-rete con verso uscente e e' caratterizzabile con un parametro dipendente dal tipo di modo di trasferimento utilizzato;
- la *potenzialita' di trasferimento* si riferisce alla idoneita' del servizio a trasferire differenti tipi di informazioni d'utente: cio' in relazione, ad esempio, all'impiego o meno di accorgimenti per salvaguardare l'integrita' di sequenza e di cifra binaria e alla esigenza di una specifica informazione d'utente per elevati gradi di integrita' informativa (cfr. § I.6.1);
- la *struttura dell'informazione trasferita* riguarda la invariabilita' o meno di specifici aspetti strutturali dell'informazione d'utente quando questa viene trasferita tra un punto di accesso di origine e uno di destinazione: cio' con riferimento ad aspetti strutturali che, se modificati per effetto delle modalita' di trasferimento, possono pregiudicare la qualita' del servizio.

Possibili valori dell'attributo "modo di trasferimento" emergeranno dalla trattazione nel capitolo IV. I valori del "ritmo di trasferimento" sono esprimibili con una misura del ritmo di flusso nel caso di un modo di trasferimento sincrono e della portata media di flusso se il modo di trasferimento e' asincrono.

Esempi di valori della "potenzialita' di trasferimento" sono l'*informazione numerica senza restrizioni*, la *voce* e l'*informazione audio nella banda 3,1 kHz*.

Si ha trasferimento di *informazione numerica senza restrizioni*, quando viene trasferita una sequenza di cifre binarie senza alterazioni di integrita': cio' richiede che il trasferimento

- sia indipendente dalla sequenza di cifre binarie;
- assicuri integrita' sia della sequenza che di cifra binaria (cfr. § I.6.1).

Ad esempio, nel caso in cui il ritmo di trasferimento sia di 64 kbit/s, ci si riferisce, indifferentemente, a informazione audio nella banda 3,1 kHz, alla voce, all'accesso a una rete per dati a pacchetto, ecc.

Il trasferimento dalla *voce* riguarda informazione fonica che e' stata codificata in accordo a una specificata regola, ad esempio secondo la Racc. G711. In questo caso, nel trasferimento possono essere utilizzati trattamenti del segnale vocale (come ad esempio quelli richiesti per il controllo di eco) e non e' assicurata l'integrita' di cifra binaria, in quanto non e' previsto un controllo di errore.

Infine l'*informazione audio nella banda 3,1 kHz* comprende voce come nel caso precedente o dati in banda fonica. In questo caso i trattamenti del segnale vocale sopra citati possono essere opportunamente modificati o funzionalmente rimossi quando sia necessario trasferire informazioni non foniche; anche in questo caso non e' assicurata l'integrita' di cifra binaria.

Per cio' che riguarda l'attributo "struttura", esempi di valori significativi sono l'*invarianza della struttura a 8 kHz* e l'*integrita' dell'unita' informativa*.

E' assicurata l'*invarianza della struttura a 8 kHz* quando si verificano due condizioni:

- a ogni interfaccia utente-rete, gli intervalli di 125 μ s sono delimitati esplicitamente o implicitamente;
- tutte le cifre binarie emesse entro un singolo intervallo delimitato di 125 μ s sono consegnate a destinazione entro un corrispondente singolo intervallo delimitato della stessa durata.

E' questo il tipo di struttura che deve essere assicurato per il trasferimento di voce in forma numerica; in tal modo, infatti, e' possibile estrarre in ricezione un segnale di campionamento che abbia uguale frequenza di quello utilizzato in emissione.

L'*integrita' dell'unita' informativa* si verifica quando:

- ad una interfaccia utente-rete, le funzionalita' di accesso forniscono un meccanismo per delimitare le *unita' informative* (UI) che debbono essere trasferite e quelle che vengono ricevute;
- tutte le cifre binarie che sono emesse entro una singola UI sono consegnate a destinazione entro una corrispondente UI.

Questo valore ha rilevanza ogniqualvolta l'informazione viene trasferita in "contenitori" strutturati di cifre binarie (UI) e questi contenitori hanno una identita' nei confronti delle funzionalita' di trasporto.

Passando poi agli *attributi secondari* di un servizio, questi sono

- la *simmetria della comunicazione*
- la *configurazione della comunicazione*
- la *inizializzazione della comunicazione*.

La *simmetria di una comunicazione* riguarda la relazione tra verso di scorrimento e portata media del relativo flusso attivati dal servizio tra due o più punti di accesso. In particolare, la comunicazione è:

- *unidirezionale*, quando si ha scorrimento solo in un verso;

- *bidirezionale simmetrica*, quando sono attivati entrambi i versi di scorrimento e le portate medie dei relativi flussi hanno valori paragonabili;

bidirezionale asimmetrica, quando, pur in presenza di entrambi i versi di scorrimento, la portata media di un flusso e' di valore decisamente prevalente rispetto a quello della portata media dell'altro flusso.

La *configurazione di una comunicazione* fa invece riferimento alla dislocazione spaziale e al numero dei punti di accesso che sono coinvolti nella fornitura di un servizio. Si parla allora di comunicazione:

- *punto-punto*, quando vengono interessati solo due punti di accesso;
- *multipunto*, quando i punti di accesso coinvolti sono in numero maggiore di due;
- *diffusiva*, quando, come nel caso precedente, i punti di accesso sono in numero maggiore di due, ma con la differenza che, in questo caso, l'informazione fluisce da un unico punto verso gli altri in modo unidirezionale.

Le modalità da seguire per dare inizio e conclusione a un trasferimento di informazione nell'ambito dell'utilizzazione di un particolare servizio di telecomunicazione (*inizializzazione di una comunicazione*) possono essere su basi *chiamata, prenotazione e permanente*.

In una *comunicazione su base chiamata* (nel seguito denominata, per brevità, *chiamata*) si distinguono tre fasi: una fase iniziale di *richiesta* del servizio, una fase intermedia di *utilizzo* e una fase finale di *chiusura*. In una chiamata è sempre possibile individuare, come attori, almeno due utenti: da un lato l'*utente chiamante*, che presenta la richiesta di fornitura di un servizio e, dall'altro, l'*utente chiamato*, con cui il primo desidera stabilire uno scambio di informazione.

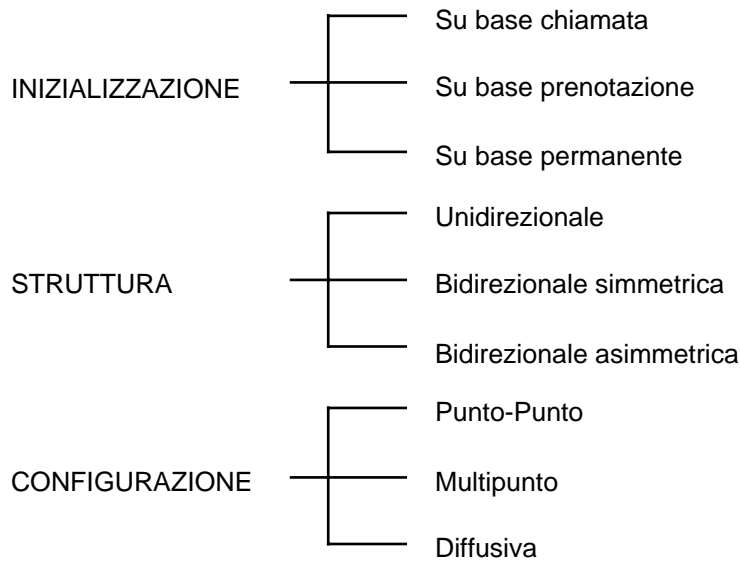


Fig. I-10 - Caratteristiche generali di una comunicazione e loro possibili attributi.

Esempio I.3-3 - In questa definizione è immediato, per qualunque utente del servizio telefonico, riconoscere le tre fasi tipiche di questo servizio. La fase di richiesta inizia quando l'utente sgancia il microtelefono e termina quando viene resa disponibile la connessione con l'utente desiderato. La fase di chiusura si attua invece a partire dall'istante in cui, da una delle due parti (ad esempio da quella chiamante), viene riagganciato il microtelefono. La fase di utilizzazione è ovviamente quella di conversazione.

Il termine "chiamata" viene normalmente utilizzato con riferimento a una interazione utente-rete in cui la rete risponde alla richiesta dell'utente *non appena possibile*, e cioè con un ritardo contenuto entro quanto consentito dalla tecnologia realizzativa e dalle condizioni di carico della rete. In queste condizioni la comunicazione può essere iniziata, non appena possibile, dopo la presentazione della richiesta da parte dell'utente e termina, non appena possibile, su richiesta di una delle parti in comunicazione.

Esiste però un'altra possibilità, che, seppure su una diversa scala temporale, ha una organizzazione in fasi del tipo di quella ora definita. E' questo il caso di una *comunicazione su base prenotazione*, in cui la comunicazione può iniziare a un istante che è stato definito in precedenza al momento di una prenotazione dell'utente e termina dopo un tempo che è stato prefissato al momento della prenotazione ovvero richiesto durante la svolgimento della comunicazione.

Nei due casi di comunicazioni su basi chiamata e prenotazione, e' anche necessario uno scambio di informazione di segnalazione tra le apparecchiature coinvolte nell'espletamento del servizio (apparecchi terminali, terminazioni di rete, apparati della sezione interna di rete). In particolare, nel caso di comunicazioni su base chiamata, l'informazione di segnalazione e' elemento

essenziale per svolgere la funzione di *trattamento di chiamata*. **Detta funzione ha lo scopo di:**

- mettere a disposizione degli utenti, quando ne fanno richiesta all'inizio della chiamata, quanto loro occorre per comunicare con altri utenti;
- supervisionare lo svolgimento della comunicazione;
- prendere atto della conclusione della chiamata.

Infine, in una *comunicazione su base permanente* non esiste una organizzazione in fasi. Esiste invece un contratto tra cliente e fornitore per la erogazione di un servizio senza vincoli sulla sua durata di fruizione. Nell'ambito di tale rapporto la comunicazione può iniziare a un istante qualunque successivo alla stipula del contratto e può continuare, a discrezione dell'utente, fino al termine stabilito contrattualmente.

I.3.4 Modalità di comunicazione

Come è sintetizzato in Fig. I-11, si possono distinguere due principali categorie di servizi di telecomunicazione: una include i *servizi interattivi* e l'altra i *servizi distributivi*.

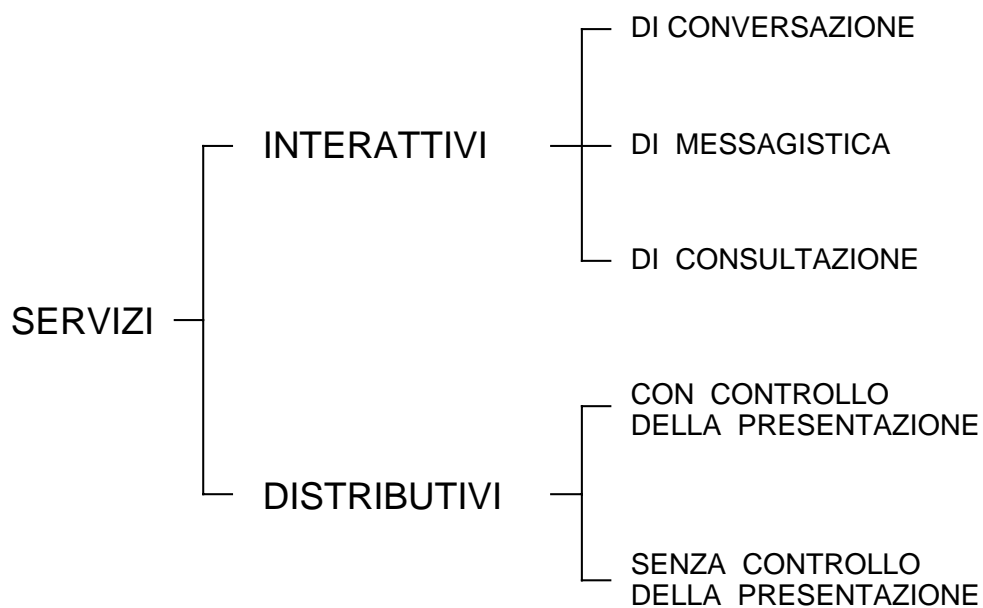


Fig. I-11 - Classificazione dei servizi di telecomunicazione in base alle modalità di comunicazione.

Nei servizi interattivi la comunicazione coinvolge due o più utenti (operatori umani, macchine di elaborazione, banche di informazione), che interagiscono tra loro per il conseguimento di uno scopo definito, quale un

dialogo in tempo reale, un trasferimento di messaggi in tempo differito o una consultazione di informazione archiviata in appositi centri di servizio.

Nei servizi distributivi esiste invece una sorgente centralizzata che distribuisce l'informazione a un gran numero di utenti senza richieste individuali.

Sono interattivi i servizi di conversazione, di messaggistica e di consultazione. Tra i servizi distributivi si distinguono poi quelli con controllo di presentazione e senza controllo di presentazione.

I *servizi di conversazione* forniscono il mezzo per un dialogo a distanza tra due o più utenti e richiedono, tra queste parti, un trasferimento di informazione *in tempo reale*, e cioè con un ritardo di transito che non pregiudichi la possibilità e l'efficacia del dialogo. L'entità tollerabile di questo ritardo dipende dal tipo di informazione scambiata e dall'applicazione ad essa connessa; può variare, in relazione a questi elementi, da valori dell'ordine di qualche decina di millisecondi a valori dell'ordine del secondo. La comunicazione può essere bidirezionale simmetrica o bidirezionale asimmetrica o, in alcuni casi specifici (quali, ad esempio, la video-sorveglianza), unidirezionale.

Nella categoria dei servizi di conversazione rientrano quelli che assicurano il *colloquio uomo-uomo* e altri che garantiscono quanto è richiesto in molte e diverse applicazioni di tele-elaborazione: per esempio, nel colloquio a distanza tra apparecchiature che siano in grado di elaborare e/o di memorizzare l'informazione (*colloquio macchina-macchina*) e nel colloquio tra queste ultime e operatori umani (*colloquio uomo-macchina*).

I *servizi di messaggistica* offrono una comunicazione da utente a utente per mezzo di uno scambio, in tempo differito, di messaggi aventi come contenuto, in alternativa o in unione, testi, voce o immagini. La comunicazione avviene per il tramite di dispositivi di memorizzazione, che possono svolgere funzioni di immagazzinamento e rilancio, di casella postale e di trattamento di messaggio (ad esempio, per scopi redazionali).

Nei *servizi di consultazione* l'utente ha la possibilità di reperire l'informazione memorizzata in appositi centri di servizio. L'informazione è inviata all'utente solo a sua domanda e può essere consultata su base individuale. Inoltre l'istante in cui la sequenza delle informazioni richieste deve iniziare è sotto il controllo dell'utente.

Per ciò che riguarda poi i *servizi distributivi con controllo di presentazione*, l'informazione, diffusa circolarmente, è strutturata in una sequenza di unità con ripetizione ciclica. In tal modo l'utente ha la possibilità di accedere

individualmente a tale informazione e può controllarne sia l'inizio che l'ordine di presentazione.

Nel caso infine dei *servizi distributivi senza controllo di presentazione*, la sorgente centralizzata emette, senza soluzione di continuità, un flusso informativo che l'utente riceve senza però avere la possibilità di controllarlo. L'utente, cioè, non può determinare l'inizio e l'ordine di presentazione dell'informazione diffusa, perchè tali funzioni sono di esclusiva competenza del fornitore del servizio.

Esempio I.3-4 - Esempi di *servizi di conversazione* sono la telefonia, la video-telefonia, la video-conferenza, alcuni tipi di comunicazioni di dati e lo scambio di documenti semplici o misti.

La *video-telefonia* e la *video-conferenza* consentono il trasferimento, oltre che della voce, anche di

- immagini in movimento che permettono di arricchire lo scambio di informazioni foniche con la visione dei parlatori e dell'ambiente in cui questi operano;
- immagini fisse e documenti che i parlatori desiderano utilizzare come corredo della loro conversazione.

In particolare, la video-telefonia riguarda la comunicazione tra due utenti, mentre la video-conferenza è rivolta alle esigenze di un insieme di utenti che, situati nelle loro rispettive sedi, desiderano riprodurre le condizioni di un incontro di lavoro in un prestabilito luogo di riunione. Applicazioni della video-telefonia e della video-conferenza si hanno in attività educative (*teledidattica*) o pubblicitarie (*tele-pubblicità*). La video-telefonia può anche essere utilizzata in operazioni di acquisto effettuate da posizione remota, mentre la riunione di lavoro è l'applicazione tipica della video-conferenza.

Tra i servizi di conversazione nell'ambito delle comunicazioni di dati si possono citare quelli connessi all'interazione conversazionale con banche di dati (ad esempio per la prenotazione di posti su vettori di trasporto) o tra stazioni di lavoro (ad esempio, per scopi di lavoro cooperativo, di progetto o di produzione assistiti da calcolatore).

I documenti che sono scambiati sono *in forma semplice*, e cioè costituiti da un solo mezzo di rappresentazione, o *in forma mista*, e cioè costituiti da testo, grafici, dati, immagini fisse o in movimento, unitamente ad annotazioni vocali. Le applicazioni dello scambio interattivo di documenti possono essere svariate; esempi al riguardo si hanno in attività di diagnostica medica (con trasferimento di immagini che debbono essere normalmente ad alta risoluzione) e nei giochi a distanza.

Per lo scambio di documenti in forma semplice si possono infine citare il Telex, il Teletex e il Telefax (cfr. Esempio I.3-2).

Esempio I.3-5 - Tra i *servizi di messaggistica* rientra la *Posta Elettronica*. Questo servizio, che è stato normalizzato nei cosiddetti *sistemi di trattamento dei messaggi* (Message Handling Systems o MHS), consente, tra gli utenti che ne possono usufruire, lo scambio di messaggi di varia natura (corrispondenza, relazioni, ecc.), la redazione di un documento e la possibilità di interconnessione tra terminali tra loro anche non compatibili. Anche in questo caso i documenti scambiati possono essere in forma semplice o mista.

Esempio I.3-6 - La categoria dei *servizi di consultazione* include, ad esempio, il Videotex (cfr. Esempio I.3-2) e la consultazione di filmati, di immagini ad alta definizione e

di informazioni foniche, oltre che l'interrogazione di banche di dati per applicazioni varie.

Le informazioni scambiate possono essere, anche in questo caso, testi, dati, grafici, suoni, immagini fisse o in movimento sotto forma di documenti semplici o misti. Campi applicativi di questi scambi si hanno in attività educative e addestrative, nell'intrattenimento, nella consultazione di notiziari, nel cosiddetto "tele-software" e in settori specialistici quali la medicina.

Esempio I.3-7 - Un esempio tipico dei *servizi distributivi con controllo di presentazione* è il *Teletext*, che è commercializzato in Italia con il termine *Televideo*.

Infine, alla categoria dei *servizi distributivi senza controllo di presentazione* appartengono i *servizi a diffusione circolare*, come, ad esempio, quelli per programmi audio e televisivi.

Per ambedue queste categorie di servizi, i tipi di informazioni trasferite possono essere suoni, dati, testi, grafici, immagini fisse o in movimento. I campi applicativi sono sostanzialmente quelli già citati con riferimento ai servizi di consultazione.

I.4 Le esigenze di comunicazione

L'informazione è ormai entrata in tutte le nostre attività quotidiane, sia nell'ambiente domestico, sia, soprattutto, in quello di lavoro.

Risulta quindi sempre più arduo, per ciascuno di noi, gestire e controllare la mole di informazioni che ci viene offerta attraverso gli usuali mezzi di diffusione (testi scritti, supporti di registrazione per musica e filmati, fotografie, notizie di stampa, programmi radio e televisivi, ecc.) o mediante le comunicazioni inter-personali. È poi praticamente impossibile elaborare, con sole nostre risorse, l'informazione offerta, in modo da poterne fruire nella misura richiesta dalle nostre attività lavorative e ricreative.

Le tecniche di telecomunicazioni possono già oggi (e potranno ulteriormente nel prossimo futuro) rendere disponibile quanto necessario per superare queste difficoltà. Ciò, da un lato, è dovuto ai progressi conseguiti, negli ultimi 30 anni, dalla tecnologia dei componenti e dei sistemi e, dall'altra, è una risposta alle nostre esigenze di servizio, che, in alcuni casi, si sono ormai manifestate in modo completo e, in altri, sono attualmente in una fase emergente.

Tra queste esigenze, esiste, già da tempo, una diffusa necessità di migliorare la produttività dell'informazione, e cioè di utilizzare quest'ultima per una razionalizzazione delle nostre attività. Una seconda esigenza deriva dalla convenienza, da molti sentita, di effettuare a distanza operazioni che oggi richiedono uno spostamento fisico, e quindi l'impegno di tempo che potrebbe essere dedicato ad attività di maggiore interesse. Come terza esigenza significativa, si può citare la richiesta di forme di intrattenimento di tipo

personalizzato, in aggiunta a quanto offerto, in forme tradizionali, da spettacoli in concerti pubblici e da altre attività ricreative, quali giochi ed esercizi sportivi.

Gli esempi ora forniti non sono certo comprensivi di tutto ciò che già da tempo si prospetta nella presentazione di scenari sulle modalità di vita umana nei prossimi decenni. Sono tuttavia sufficienti a consentire di individuare altre esigenze oggi emergenti. Si allude alla *multimedialità*, alla *ubiquità* e alla *personalizzazione* della comunicazione, e cioè a quanto sta guidando l'attuale fase di sviluppo dei servizi di telecomunicazioni.

Per chiarire come queste esigenze possono essere soddisfatte, nel seguito del paragrafo si considererà, dapprima (§ I.4.1), la *multimedialità* e, successivamente (§ I.4.2), l'insieme di *ubiquità* e *personalizzazione*. Si passerà poi (§ I.4.3) ai requisiti di sicurezza e di protezione delle informazioni scambiate o manipolate: tali requisiti stanno infatti acquistando un ruolo sempre più importante in uno scenario di comunicazioni personalizzate. Infine (§ I.4.4) verrà affrontato il problema della gestione di reti e di servizi, mettendo in evidenza la complessità di azioni richieste per gestire i molteplici aspetti che riguardano la fornitura di servizi di telecomunicazioni.

I.4.1 La multimedialità

Lo sviluppo delle telecomunicazioni è stato caratterizzato per oltre mezzo secolo dalla richiesta e dall'offerta di *servizi monomediali*. Ognuno di questi consente il trattamento di un singolo mezzo di rappresentazione, che lo caratterizza rispetto a altri servizi e che è ovviamente diversificabile passando da un servizio all'altro.

Oggi, in aggiunta a servizi monomediali, si manifesta l'esigenza di nuove forme di comunicazione, in cui siano trattati due o più mezzi di rappresentazione e in cui questi ultimi conservino, nell'operazione di trasporto, le caratteristiche di concatenazioni fisica e logica presentate all'origine. Si parla di *servizi multimediali*.

Questa esigenza deriva in primo luogo dalla *multimedialità* della comunicazione umana e dall'obiettivo di conservare questa caratteristica anche in un colloquio tra utenti in posizioni remote. Si faccia infatti riferimento a una comunicazione interattiva tra due utenti umani. Come è ragionevole attendersi, questi desidererebbero che nella interazione a distanza siano riprodotte il più fedelmente possibili le condizioni ambientali di un colloquio faccia a faccia. D'altra parte l'informazione scambiata in un colloquio siffatto include in generale, oltre alla voce e ai suoni, anche:

- immagini fisse e/o in movimento risultanti dall'interazione con lo spazio tridimensionale in cui avviene il colloquio;
- dati emessi da elaboratori sotto forma di informazioni alfa-numeriche e video.

Riprodurre questo insieme completo di informazioni equivale a trasferire a distanza una *realta' virtuale*, e cioe' la simulazione dell'ambiente tridimensionale in cui opera e con cui interagisce ogni interlocutore utilizzando tutti i mezzi connessi ai suoi organi sensori (udito, vista, tatto, ecc.) e agli strumenti di elaborazione a sua disposizione. Ma, affinche' questa simulazione sia una riproduzione sufficientemente fedele della realta' lontana, occorre che il trasporto dei mezzi di rappresentazione sia effettuato in modo da assicurare loro la conservazione delle caratteristiche di concatenazioni fisica e logica presentate all'origine.

Se la nostra esperienza quotidiana rivela quanto multimediale sia la comunicazione umana, con analoghe caratteristiche si presenta l'interazione uomo-macchina. Da questo punto di vista si possono ravvisare due scopi principali della multimedialita'.

Da un lato questa deve consentire un rapporto di naturale colloquio tra uomo e macchina facilitando l'accesso del primo alla seconda senza richiedere competenze non facilmente conseguibili dalla generalita' delle persone: si pensi ad esempio al diffondersi delle tecniche iconiche e di quelle basate su metafore della realta' esterna. Da un altro lato la multimedialita' deve supportare lo sviluppo di nuovi metodi di comunicazione uomo-macchina, che consentano di potenziare e di amplificare le naturali capacita' dell'uomo: cio' si potra' conseguire con il miglioramento e con l'arricchimento dell'interfaccia tra uomo e macchina.

Esempio I.4-1 - Un esempio di comunicazione multimediale tra soggetti umani e' l'interazione di due progettisti lontani che colloquiano tra di loro. Base del colloquio sono rappresentazioni grafiche o fotografiche di un progetto in corso, relazioni scritte su un elaboratore di testi, altri documenti reperibili su una banca di dati, immagini in movimento per illustrare il comportamento dell'oggetto a cui il progetto si riferisce o per mostrare l'andamento di rilievi in corso su un banco sperimentale. Finalita' del colloquio e' lo svolgimento di un *lavoro cooperativo*, nell'ambito del quale puo' essere necessario, da parte di ambodue le parti, apportare varianti sulla documentazione del progetto (schemi, disegni, relazioni, ecc.), accedere alla banca di dati per il reperimento di ulteriore documentazione, valutare con simulazioni le prestazioni ottenibili con le scelte progettuali effettuate, e cosi' via.

Un secondo esempio di comunicazione multimediale interattiva, che ha scopi di intrattenimento e che riguarda un colloquio uomo-macchina, e' il servizio di *video su domanda* (VOD - Video On Demand), per il quale sussistono interessanti prospettive sia per la sua attuazione tecnica che per i suoi riflessi commerciali. Il servizio VOD prevede che un

utente possa richiedere, per una visione in ambito domestico, un programma televisivo specifico. La richiesta è effettuata a un fornitore di servizi, che rende disponibile l'accesso a una teleteca in cui l'utente possa scegliere un programma corrispondente ai suoi gusti, alla sua cultura e ai suoi desideri. Oltre alla visione del programma, l'utente deve avere possibilità di controllo analoghe a quelle offerte da un video-registratore casalingo, e cioè, ad esempio, l'arresto momentaneo della proiezione, il ritorno su particolari sequenze, l'effetto "moviola", e così via.

La differenza sostanziale del servizio VOD rispetto a una visione basata sull'uso di nastri o di dischi letti con apparati domestici di video-registrazione (che in prospettiva potranno offrire alta definizione, presentazione su grande schermo ed audio stereo ad alta fedeltà) risiede nella diversità dell'operazione con cui si trasporta l'informazione video: nel caso del servizio VOD il trasporto è attuato per mezzo di una rete di telecomunicazione, mentre in quello della videoregistrazione viene invece trasportato materiale di supporto in memoria.

I.4.2 L'ubiquità e la personalizzazione dei servizi

Una seconda esigenza di comunicazione è l'*ubiquità del servizio*, e cioè la possibilità per l'utente di comunicare in qualunque luogo si trovi e con le stesse caratteristiche di reperibilità e di sicurezza a lui offerte nei suoi ambienti domestico o lavorativo. A tale scopo occorre che un utente possa essere chiamante o chiamato in qualsiasi momento della giornata o in qualsiasi posto si trovi nell'ambito dei suoi spostamenti a piedi o su mezzi di trasporto.

L'*ubiquità* del servizio comporta il soddisfacimento del requisito di *mobilità*, nel quale possono distinguersi due aspetti principali, connessi alla persona che comunica e all'apparecchio terminale che viene utilizzato.

La *mobilità della persona* consente a un utente di usufruire di tutti i servizi che sono di suo interesse, indipendentemente dalla terminazione di rete a cui è fisicamente connesso.

La *mobilità del terminale* consente all'utente di comunicare in condizioni di effettivo movimento in un'area geografica che dovrebbe essere il più possibile vasta e, in particolare, in un'area più o meno estesa nell'intorno di singole terminazioni di rete.

Se la *mobilità della persona* è conseguibile prevalentemente con funzionalità infrastrutturali, la *mobilità del terminale* richiede, oltre a queste ultime, anche l'uso di mezzi trasmissivi radio nella sezione di accesso.

È evidente che il requisito di *mobilità della persona* è più cogente di quello di *mobilità del terminale*; anzi è possibile immaginare il secondo contenuto nel primo, ma non il viceversa. Infatti, mentre un utente dotato di *mobilità della persona* può godere anche di quella del terminale, un altro utente che possiede un terminale radiomobile non gode necessariamente della *mobilità personale*.

Una terza esigenza di comunicazione e' la *personalizzazione dei servizi*, e cioe' la possibilita' per l'utente di poter comunicare secondo modalita' che sono ritagliate sulle sue esigenze specifiche e che possono essere direttamente sotto il suo controllo. A tale scopo e' necessario soddisfare la larga varieta' di interessi di ogni persona, che da un lato manifesta necessita' di razionalizzare le sue funzioni di lavoro e che dall'altro cerca di coltivare i suoi interessi culturali e ricreativi.

Per attuare la personalizzazione di una comunicazione i tre attori principali che intervengono nel conseguimento di questo obiettivo, e cioe' il cliente, il gestore di rete e il fornitore dei servizi, assumono ruoli specifici.

In particolare, al cliente deve essere assicurata una maggiore quota di controllo sulle modalita' di creazione e di gestione dei servizi. Esempi al riguardo, e in ordine di crescente complessita' di controllo, possono essere la possibilita' per l'utente di:

- modificare i parametri che descrivono il suo profilo entro i limiti stabiliti dal contratto con il fornitore dei servizi;
- variare, in funzione delle sue esigenze di qualita' di servizio nella fruizione di informazioni audio-visive, i parametri degli algoritmi di codifica;
- costruire uno specifico servizio mediante la composizione dinamica di "elementi" di servizio, assemblati anche in tempo reale durante la svolgimento di una comunicazione.

Circa poi il punto di vista del gestore di rete, un primo obiettivo da raggiungere e' poter manipolare, in modo flessibile e sotto il suo controllo, le componenti funzionali presenti nell'infrastruttura di rete. Inoltre un altro obiettivo oggi giudicato di particolare interesse e' riuscire a integrare i servizi all'utenza con le applicazioni gestionali.

Riguardo infine al fornitore di servizi, e' da sottolineare la sua convenienza a operare in un contesto di modularita' e di inter-operabilita' degli elementi di servizio atti a comporre servizi personalizzati. Le applicazioni di interesse per l'utenza, anche in una prospettiva a lungo termine, potranno diventare entita' da comporre e da integrare tra loro per dare origine ad altri servizi piu' complessi, a loro volta ulteriormente componibili e integrabili con altri.

E' da sottolineare che le comunicazioni personalizzate verranno offerte a breve-medio termine in regime di concorrenza, nel quale sara' stimolata la competizione tra vari fornitori di servizio. Cio' impone gia' da ora che le modalita' di composizione a cui si e' sopra accennato vengano attuate nel

rispetto di una normativa tale da garantire uno sviluppo ordinato del mercato e una effettiva utilizzabilità da parte dell'utenza.

Esempio I.4-2 - Per dare un'idea delle possibilità di comunicazione offerte dalla composizione di elementi di servizio, si faccia riferimento al rapporto tra una futuribile agenzia di viaggio e un cliente interessato a usufruire dei servizi da questa offerti. Queste parti sono in posizioni tra loro remote: il loro rapporto si attua quindi utilizzando mezzi di telecomunicazione e può articolarsi in vari passi.

Si ha dapprima, a cura di una compagnia di gestione di viaggi, la presentazione di uno o più programmi di itinerario turistico. Dopo la scelta da parte del cliente, si effettua la stipula del contratto tra questi e un apposito ufficio legale: la firma di entrambe le parti è effettuata con mezzi elettronici.

Successivamente il pagamento coinvolge una banca a cui il cliente dà l'ordine di prelevare dal proprio conto l'ammontare dell'operazione. Può poi essere richiesta la stipula di una polizza assicurativa e ciò richiede l'intervento di un ulteriore fornitore rappresentato da una compagnia di assicurazione.

Per la eventuale emissione di documenti di viaggio (carte di identità, rilascio di visti, passaporti) può poi essere richiesto l'intervento di uffici pubblici, di consolati, ecc.. Infine il cliente può aver bisogno di altre informazioni, quali quelle connesse al clima, alla moneta e a quant'altro può essere di suo interesse per lo svolgimento del viaggio prescelto. Queste informazioni, se non direttamente disponibili, possono essere fornite con la consultazione di enti vari.

Come è chiaro da questa esemplificazione, il servizio offerto dalla agenzia di viaggio può essere arricchito senza limiti, se non quelli imposti dalla disponibilità dei servizi componenti.

Multimedialità e ubiquità di servizio, pur essendo esigenze autonome, possono essere viste come particolari aspetti di uno scenario di *comunicazioni mobili e personalizzate*.

Al riguardo si osserva in primo luogo che i servizi basati sul trasporto combinato di più mezzi di rappresentazione (come, ad esempio, si verifica nella messaggistica di documenti multimediali, nella consultazione di banche di informazioni audio, video o di dati, nei servizi conversazionali di video-telefonia e di video-conferenza con lo scambio di dati) possono essere visti come costituiti dalla composizione di elementi di servizio nei quali la comunicazione evolve con lo scambio di un singolo mezzo di rappresentazione (ad esempio, audio o video o di dati). Questa impostazione degli obiettivi di multimedialità guiderà anche gli sviluppi futuri.

Inoltre la multimedialità dei servizi personalizza la comunicazione in quanto offre all'utenza una possibilità di scelta nell'ambito di una larga varietà di servizi/applicazioni. D'altra parte l'ubiquità è un aspetto della personalizzazione in quanto consente la comunicazione anche ad utenti che si spostano all'interno o all'esterno di edifici, nella propria sede domestica o in

quella lavorativa, in ambienti pubblici o privati, con le proprie gambe o a bordo di veicoli (auto, aerei o navi).

Un'ulteriore e rilevante caratteristica della personalizzazione si manifesta nell'attribuzione di un *numero personale* o PIN (Personal Identification Number) al singolo utente. Il PIN identifica l'utente in tutti i rapporti di questi con l'infrastruttura di comunicazione e sostituisce quindi i vari numeri che sono indirizzo delle terminazioni di rete a cui l'utente fa capo per comunicare nel corso delle proprie attività domestiche e lavorative. Viene rotto in tal modo il tradizionale rapporto utente-terminazione di rete, che è tuttora una caratteristica costante nell'accesso alle reti di telecomunicazioni.

Ogniqualevolta l'utente dotato di PIN desidera interagire con la rete a cui fa capo tramite un apparecchio terminale fisso o mobile, egli dichiara il proprio PIN alla rete. Questa provvede allora a individuare la terminazione di rete a cui fa capo l'apparecchio terminale utilizzato e quindi la posizione dell'utente. Ciò consente a quest'ultimo di essere chiamato o di chiamare e, in questo secondo caso, di ricevere l'addebito della chiamata indipendentemente dalla circostanza che questa sia stata effettuata da un apparecchio terminale fisso o mobile, nel proprio ambiente domestico o lavorativo.

Esempio I.4-3 - Due utenti, che per comodità chiamiamo X ed Y, sono in grado di accedere a servizi personalizzati e come tali dispongono dei loro numeri personali *PINX* e *PINY* rispettivamente.

In un certo giorno Y attende da X un messaggio sotto forma di testo scritto, ma desidera che questo servizio venga opportunamente personalizzato in relazione al programma della propria giornata. Ad esempio desidera che, se arriva da X un testo dopo le 11 antimeridiane e fino alle 2 pomeridiane, questo testo debba essere instradato verso il suo apparecchio terminale veicolare; ma che in questo caso non debba apparire in forma alfa-numerica su uno schermo, ma piuttosto debba essere tradotto in forma vocale attraverso voce sintetizzata.

Per conseguire questo scopo, Y, dopo essersi fatto riconoscere attraverso il suo *PINY*, comunica alla rete questa personalizzazione del servizio e la rete provvede a tenerne memoria in una banca di dati e, più precisamente, in una sezione di questa che è associata al numero *PINY*.

Sempre ad esempio supponiamo che X, alle ore 1 pomeridiane e da un apparecchio terminale connesso alla rete fissa, dopo essersi fatto riconoscere dalla rete tramite il suo *PINX*, invii un messaggio all'indirizzo *PINY*. A questo punto la rete:

- riconosce che il messaggio è indirizzato a un utente avente accesso a servizi personalizzati;
- interroga allora la banca di dati in cui sono potenzialmente reperibili le istruzioni personalizzate di Y;
- dopo aver preso conoscenza del trattamento desiderato da Y sul messaggio inviato da X, manda all'autocommutatore a cui fa capo X un comando per convertire il testo in voce e un numero convenzionale per poter instradare la comunicazione verso l'apparecchio veicolare di Y.

Come conseguenza di queste azioni, il testo viene tradotto in voce sintetica e viene

inoltrato verso la destinazione prescelta da Y.

In questo esempio viene considerata la comunicazione tra un primo utente che chiama operando sulla rete fissa e un secondo utente che e' chiamato mentre si trova in movimento. Il trasferimento del messaggio dal primo al secondo utente avviene per il tramite dell'interrogazione di una banca di dati effettuata in tempo reale e, ulteriormente, per mezzo di una conversione da testo a voce eseguita da una apparecchiatura periferica.

Sono evidenti le funzionalita' coinvolte. Innanzitutto quelle di trasporto in cui si ipotizza siano disponibili criteri di instradamento di tipo selettivo e in cui e' possibile una modifica in relazione alla personalizzazione desiderata dall'utente chiamato. E' poi necessaria una interconnessione tra due sezioni di rete, quella fissa e quella mobile. Infine e' richiesto l'impiego di una risorsa addizionale che consenta la conversione del mezzo di rappresentazione.

Circa poi le funzionalita' di controllo, queste sono l'elemento portante dello scenario ipotizzato. Attraverso di esse e' possibile:

- il riconoscimento degli utenti X ed Y da parte della rete (*procedura di autenticazione*);
- la localizzazione dell'utente Y;
- la attuazione di quanto previsto nella personalizzazione del servizio.

I.4.3 La sicurezza e la protezione delle informazioni

Le informazioni scambiate o manipolate in una comunicazione debbono essere adeguatamente protette in modo da conseguire sicurezza relativamente alla autenticazione, alla integrita', alla confidenzialita' e alla imputabilita'.

L'*autenticazione* dei dati o della loro origine e' necessaria per garantire che le parti impegnate in una comunicazione siano effettivamente quelle che dichiarano di essere. Questo requisito di sicurezza e' ottenibile con una procedura apposita.

L'*integrita'* dei dati rappresenta, invece, la garanzia contro accidentali o indebite modifiche del flusso informativo di utente durante il suo trasferimento. Su questo requisito si tornera' nel seguito (cfr. § I.6.1).

La *confidenzialita'* e' poi la garanzia che il contenuto informativo di una comunicazione o la conoscenza delle parti in questa coinvolte non siano divulgate a terze parti, se non debitamente autorizzate. Questa garanzia rappresenta quindi il fine tradizionale delle tecniche di *cifratura dell'informazione* quali sono trattate dalla *crittografia*.

La *imputabilita'* rappresenta infine la garanzia, per l'utente, di vedersi attribuito un addebito effettivamente commisurato ai servizi fruiti e, per il fornitore, di poter effettuare una tariffazione che sia incontestabile.

Esempio I.4-4 - I principali meccanismi a disposizione per assicurare la confidenzialita' di una comunicazione sono basati sull'impiego di algoritmi crittografici di vario tipo.

Nel passato sono state adottate soluzioni cosiddette *a chiave segreta*, che sono

caratterizzate dall'esistenza di una sola chiave la cui conoscenza e' ristretta solo alle parti in comunicazione; un esempio al riguardo e' offerto dall'*algoritmo DES* (Data Encryption Standard).

Piu' di recente si sono affermate soluzioni piu' flessibili, cosiddette *a chiave pubblica*. Queste offrono maggior opportunita' di impiego e impongono requisiti di gestione meno stringenti. Sono basate sull'impiego di algoritmi del tipo RSA (dal nome dei suoi ideatori: River, Shamir e Adleman), in cui sono previste due differenti chiavi: una segreta e' assegnata individualmente, l'altra invece e' pubblica.

I.4.4 La gestione della rete e dei servizi

La gestione di un ambiente di comunicazione comprende l'insieme delle attivita' volte al perseguimento della migliore qualita' dei servizi offerti e al contenimento dei costi di fornitura. Questo secondo obiettivo puo' essere conseguito con una piu' efficiente utilizzazione delle risorse impiantistiche e operative.

Le principali funzioni di gestione riguardano:

- l'esercizio e la manutenzione degli elementi di rete;
- la gestione dei servizi, compreso il controllo, ove richiesto, da parte dell'utente; questi agira' sui parametri di gestione dei servizi a lui forniti;
- la supervisione, la misura e la gestione del traffico di telecomunicazione (cfr. § II.2.6 e II.3.2);
- la gestione della tassazione, e cioe' degli addebiti al cliente per i servizi forniti;
- la pianificazione e la progettazione di reti e di servizi.

Per lo svolgimento di queste funzioni, la tendenza ormai consolidata e' quella di sovrapporre alla infrastruttura gestita una *rete di gestione* (TMN-Telecommunication Management Network). Questa comprende *entita' funzionali* situate negli elementi di rete e *sistemi di mediazione*, che cooperano mediante l'impiego di appropriate funzioni di trasporto dell'informazione di gestione. A questa rete sovrapposta si agganciano i *sistemi di gestione* (OS-Operation System), che hanno quindi tutte le informazioni provenienti dalla infrastruttura gestita o contenute nella rete di gestione: possono pertanto inviare opportuni comandi a tutti gli apparati che sono sotto le loro cure gestionali.

I.5 Evoluzione delle infrastrutture

I contenuti di questo paragrafo sono dedicati a presentare, in modo sintetico e del tutto preliminare, gli attuali sviluppi delle infrastrutture di telecomunicazione e le previsioni circa gli ulteriori sviluppi nel futuro.

In particolare, con riferimento alle esigenze di multimedialita', si considerera' dapprima (§ I.5.1) la tendenza attualmente in atto verso infrastrutture che consentano *integrazione dei servizi*. Successivamente (§ I.5.2) si mettera' in evidenza il ruolo che ha assunto lo sviluppo, attualmente in corso, delle *reti intelligenti*: cio' allo scopo di consentire la fornitura di servizi particolarmente evoluti con specifico riferimento a quelli di tipo personalizzato.

Si passera' poi (§ I.5.3) a descrivere come le *reti mobili* si siano affiancate a quelle fisse (dedicate o integrate) per risolvere i problemi connessi alla mobilita' del terminale e siano attualmente in una fase evolutiva verso obiettivi di integrazione dei servizi e di ulteriore integrazione con le reti fisse. L'ultimo argomento trattato (§ I.5.4) riguardera' infine l'obiettivo di *comunicazione personale universale* e le sue prospettive di attuazione.

I.5.1 Verso l'integrazione di servizi

La risposta all'esigenza di comunicazioni multimediali si e' manifestata tramite almeno due azioni.

In primo luogo, si e' gia' da tempo riconosciuta la convenienza di una nuova forma di *integrazione delle tecniche*. Cioe' da funzioni di trasporto nelle reti fisica e logica realizzate con tecnologia completamente analogica si sta passando a un impiego generalizzato di tecnologie numeriche, basate oggi su componentistica di tipo elettronico e, nel prossimo futuro, anche di tipo ottico. La convenienza di questa trasformazione di natura tecnica risiede in una migliore qualita' dei servizi offerti e in ridotti costi di fornitura.

In secondo luogo, con accordo ormai totale nel mondo della tecnica, e' giunta ad una prima fase attuativa l'idea di soddisfare le nuove esigenze dell'utenza con una infrastruttura di trasporto che preveda:

- a) un accesso comune in forma numerica e di tipo normalizzato per tutti i servizi mono e multimediali (*accesso integrato*);
- b) un trasporto di informazione attuato con modalita' esclusivamente numeriche da utente a utente (*connettivita' numerica da estremo a estremo*) e quindi senza i limiti di qualita' insiti nell'impiego di tecnologie analogiche nella sezione interna e/o in quella di accesso dell'infrastruttura.

Queste caratteristiche qualificano un obiettivo di sviluppo delle funzionalita' di trasporto, che viene chiamato *rete numerica integrata nei servizi* (ISDN-Integrated Services Digital Network) e che prevede l'integrazione dei

servizi offerti in un ambiente di comunicazione con impiego generalizzato di tecnologie realizzative di tipo numerico.

Il primo passo verso questo obiettivo e' stato gia' in larga parte attuato nei Paesi industrialmente piu' evoluti ed e' rappresentato dalla *ISDN a banda stretta* (N-ISDN, Narrowband-ISDN). Questa sta integrando gradualmente le esistenti reti dedicate a un servizio per consentire il trasferimento di voce, di dati e di immagini fisse o con limitate esigenze di movimento: cio' nell'ambito di servizi che, orientativamente, appartengono alle classi a bassa e a media velocita' (cfr. § I.2.3) e che quindi richiedono capacita' di trasferimento entro 2 Mbit/s.

Il punto di partenza del percorso verso la N-ISDN e' stato, come gia' detto, la realizzazione di una connettivita' numerica sia nella sezione interna che in quella di accesso. La rete da cui si e' sviluppata la N-ISDN e' stata quindi di tipo *integrato nelle tecniche numeriche*. Nella sua sezione interna deve cioe' sussistere comunanza delle tecniche numeriche impiegate per svolgere le funzioni di trasmissione e di commutazione. E anche la sezione di accesso utilizza trasferimenti in tecnica solo numerica, seppure su mezzi trasmissivi in rame.

Per cio' che riguarda poi gli sviluppi successivi alla N-ISDN, e' oggi riconosciuto che, in una prospettiva ormai a breve termine anche se con avanzamenti gradualmente, le potenzialita' della N-ISDN debbono essere ampliate per soddisfare ulteriori esigenze di comunicazione negli ambienti domestico e lavorativo. L'obiettivo finale di questo ulteriore processo evolutivo e' la *ISDN a larga banda* (B-ISDN, Broadband-ISDN).

Tali nuove esigenze comprendono, in aggiunta a quelle consentite da una N-ISDN, comunicazioni nell'ambito di servizi ad alta velocita' e cioe' con esigenza di capacita' di trasferimento superiore a 2 Mbit/s.

Esempio I.5-1 - Per dare un'idea quantitativa dell'esigenza di larga banda in un ambiente di comunicazioni multimediali evolute, si consideri l'interazione dei due progettisti lontani ipotizzata nel precedente Esempio I.4-1. Immaginiamo che ognuno dei due disponga di una stazione di lavoro.

Se supponiamo che la visualizzazione sia su uno schermo a definizione elevata, diciamo con 2000×2000 punti a colori, il numero di cifre binarie che rappresentano l'informazione associata al contenuto dello schermo e' prossimo a 100 Mbit. Questo numero si riduce a circa 30 Mbit se ci riferiamo alla rappresentazione di una immagine fotografica e se si adottano tecniche di riduzione di ridondanza che consentano il recupero completo dell'informazione originaria; si puo' pero' contenere entro circa 2 Mbit con una riduzione di ridondanza che lasci all'osservatore una qualita' percepita sostanzialmente inalterata rispetto a quella originale.

In tali ultime condizioni, se si vuole contenere il tempo di trasferimento della immagine fotografica entro un secondo cosi' come e' richiesto dalla interattivita' della comunicazione, la capacita' di trasferimento necessaria non puo' essere inferiore a 2 Mbit/s; si situa cioe' al confine tra le capacita' richieste da servizi a media e alta velocita'.

Ulteriori dettagli sui servizi offerti da una B-ISDN e sulle differenti esigenze prestazionali che ne conseguono saranno oggetto di trattazione in § I.6.4.

I.5.2 Le reti intelligenti

Alcuni servizi di telecomunicazioni, come ad esempio quello supplementare di numerazione abbreviata, sono forniti impegnando le risorse del solo nodo che e' di accesso per l'utente richiedente. Altri servizi, piu' evoluti dei precedenti e esemplificabili tramite il servizio supplementare di identificazione della linea chiamante/chiamata ovvero quello di gruppo chiuso di utenti, coinvolgono invece due o piu' nodi ma sempre senza distinzione tra la logica del servizio e le modalita' di fornitura.

Per lo sviluppo di servizi supplementari ancora piu' evoluti, come ad esempio quelli rispondenti alle esigenze di personalizzazione e di mobilita' della persona, si e' riconosciuta l'importanza di separare la logica del servizio dalle modalita' di fornitura. Su questo principio e' basata una infrastruttura, chiamata *rete intelligente* (IN-Intelligent Network), che costituisce la piena valorizzazione delle funzionalita' di controllo e che e' attualmente in corso di sviluppo utilizzando le funzionalita' di trasporto della rete telefonica o di quella ISDN.

I servizi supplementari fornibili tramite una IN offrono quanto e' assicurato dalle funzionalita' di controllo con il coinvolgimento di risorse di elaborazione condivise; sono chiamati *servizi di rete intelligente*.

La IN comprende due tipi di nodi e alcuni sistemi periferici specializzati. I nodi sono quelli "intelligenti" e quelli di accesso; mentre i primi sono apparecchiature centralizzate e in poche unita' nell'ambito di una IN, i secondi sono disseminati nella rete in relazione alla consistenza numerica dell'utenza.

I nodi "*intelligenti*" o SCP (Service Control Point) costituiscono la *parte controllante* della IN e, come tali,

- contengono la logica dei servizi offerti sotto forma di appositi programmi (SLP-Service Logic Program) e i dati relativi al profilo degli utenti che possono accedere alla IN;
- eseguono l'SLP rispondendo a una richiesta esplicita da parte dei nodi di accesso;
- istruiscono questi ultimi circa il trattamento di chiamata.

I *nodi di accesso* o SSP (Service Switching Point) sono invece la *parte controllata* della IN e costituiscono il tramite con gli utenti per la fornitura dei servizi; quindi:

- riconoscono la richiesta di un servizio di rete intelligente;
- interrogano conseguentemente i nodi "intelligenti";
- svolgono il trattamento di chiamata sulla base delle istruzioni ricevute da questi ultimi.

Esempio I.5-2 - Per l'espletamento di un servizio di rete intelligente, il nodo SSP riconosce la chiamata come relativa a uno specifico servizio: cioè sulla base delle cifre digitate dall'utente (*codice di accesso al servizio*) o in base alla caratterizzazione della linea a cui l'utente fa capo.

A seguito di questo riconoscimento, l'SSP chiede istruzione a un SCP. Quest'ultimo accede ai dati relativi al servizio richiesto (l'SLP) e al profilo dell'utente richiedente. L'SCP esegue poi l'SLP e istruisce opportunamente l'SSP per il trattamento di chiamata.

E' da osservare che la "intelligenza" di una IN permette il trattamento di chiamata in modo dinamico e sotto il controllo dell'utente. Cio' e' dovuto all'interazione stretta di controllo da parte degli utenti sui loro profili che vengono memorizzati nei nodi SCP e che possono essere modificati dall'utente in tempo quasi reale.

La separazione già sottolineata tra logica del servizio e modalità di fornitura si riflette, da un punto di vista funzionale, in una caratteristica fondamentale delle IN, e cioè nella *distinzione netta delle funzionalità di controllo* (logica del servizio, profili di utente, ecc.) *da quelle di trasporto*: le prime sono collocate in apparecchiature centralizzate a livello di rete quali sono i nodi SCP, le seconde nei nodi SSP. Come vantaggio principale di questa distinzione, l'introduzione di nuovi servizi e la modifica di quelli già forniti richiedono, in linea di principio, interventi solo sui nodi "intelligenti", senza dover modificare la logica di comando dei nodi di accesso.

Infine, il dialogo tra SSP e SCP, oltre che con i sistemi periferici della IN, e' assicurato dallo scambio di informazioni di segnalazione. Per questo scopo assume un ruolo di fondamentale importanza l'impiego di sistemi di segnalazione a canale comune (cfr. par. IV.6).

Esempio I.5-3 - Per dare un'idea sui servizi offribili tramite una IN, se ne descrivono qui di seguito alcuni, che verranno presentati come arricchimento di un servizio di base telefonico.

Il *Numero Verde* e' il servizio che consente di trasferire in modo automatico il costo della chiamata telefonica dal chiamante (a cui viene di norma addebitata la comunicazione) al chiamato che e' il cliente del servizio. Oltre ad essere in grado di offrire la chiamata gratuita ai propri corrispondenti, il cliente di questo servizio puo' specificare destinazioni di chiamata multiple per un solo "numero verde". Le chiamate possono essere inoltrate a queste varie destinazioni in base a varie condizioni, quali la ubicazione geografica del chiamante, l'ora e il

giorno in cui la chiamata è effettuata. Ciò aumenta per il chiamante la possibilità di effettuare una chiamata che vada a buon fine.

La *Rete Privata Virtuale* consente di caratterizzare un gruppo di utenti della IN come facenti capo a una rete privata, pur utilizzando risorse della rete pubblica. Il vantaggio in termini di servizio risiede nella possibilità di offrire agli utenti benefici di personalizzazione del tutto paragonabili a quelli offribili tramite una rete privata. In termini di costo si ha poi il vantaggio di condividere le risorse di una rete pubblica: ciò può consentire risparmi globali, che sono dovuti all'uso più efficiente dei servizi e alle economie di scala.

La *Tariffa a Premio* (che è un servizio commercializzato in Italia come Audiotel) permette di girare parte della tariffa della comunicazione al chiamato, che assume il ruolo di fornitore del servizio. Viene utilizzato da quest'ultimo per offrire particolari servizi di fornitura di informazione ai propri clienti (ad esempio, per consulenza fiscale, per assistenza medica, ecc.). La tariffa pagata dal chiamante copre il prezzo della chiamata e quello del servizio di informazione; il suo importo è poi indipendente dalla destinazione alla quale la chiamata viene inoltrata. Infine, come nel caso del servizio numero verde, la destinazione può dipendere dall'ubicazione geografica del chiamante.

Il servizio *Chiamate di Massa* è destinato al cliente che desidera ricevere solo un numero controllato di chiamate completate con successo: ciò in presenza di eventi che fanno prevedere una molteplicità di tentativi di chiamata aventi come destinazione quella del cliente del servizio. Il servizio in questione consente, durante questi eventi, di ottenere la protezione dell'infrastruttura da fenomeni di congestione e di sovraccarico.

L'*Addebito Ripartito* permette la ripartizione del prezzo della chiamata tra chiamante e chiamato. Il cliente del servizio è il chiamato, che desidera offrire ai propri corrispondenti la possibilità di comunicare con lui a costi suddivisi. Il prezzo del servizio è indipendente dalla destinazione alla quale la chiamata deve essere inoltrata.

Il *Televoto* è il servizio che consente di effettuare sondaggi d'opinione usando la rete telefonica. Il cliente del servizio può proporre un voto a distanza su argomenti di suo interesse. I partecipanti al voto sono richiesti di esprimere la loro opinione. Il servizio include anche, a beneficio del suo cliente, il conteggio dei voti espressi.

Il *Numero di Accesso Universale* consente al suo cliente, che fa capo a una molteplicità di terminazioni di rete in varie località geografiche, di poter essere raggiunto da chiamate utilizzanti un unico numero di elenco su una destinazione che dipende dalla localizzazione geografica del chiamante. La tassazione della chiamata è completamente a carico di quest'ultimo. Dato che il chiamante non conosce la destinazione della sua chiamata, deve essere previsto un annuncio fonico per informare il chiamante sul tipo di tariffa che gli verrà addebitata: ciò ovviamente prima che la sua chiamata venga inoltrata e con la possibilità per il chiamante di autorizzare o meno il proseguimento del trattamento di chiamata.

Infine il servizio di *Numero Personale* fornisce un meccanismo di inoltro della chiamata verso un suo cliente, che ha necessità di muoversi frequentemente fra ubicazioni diverse e di essere raggiunto dalle chiamate a lui dirette. Il cliente può indirizzare queste ultime verso un qualunque numero della rete richiedendo la prestazione da qualsiasi postazione telefonica. Il chiamante può raggiungere il cliente del servizio utilizzando il PIN attribuito a questo ultimo.

Ognuno di tali servizi può essere corredato da un ampio numero di prestazioni opzionali, che hanno lo scopo di personalizzare un servizio in base alle specifiche necessità degli utenti. Esempi di prestazioni opzionali sono costituiti da

- *Instradamento Variabile*, che permette di instradare le chiamate in base a una serie di parametri (ora, giorno della settimana, giorno feriale o festivo, giorno dell'anno, area di origine della chiamata) a scelta dell'utente;
- *Reinstradamento su Occupato*, che abilita l'utente alla definizione di una serie di destinazioni alternative nel caso in cui la chiamata verso la prima destinazione non

trovi libero il chiamato.

I.5.3 *Le reti mobili*

Come si e' visto, le reti intelligenti consentono di soddisfare, almeno entro certi limiti, le esigenze di personalizzazione e di mobilita' della persona.

Diverso provvedimento e' invece richiesto per soddisfare le esigenze di mobilita' del terminale. Per questo scopo bisogna ricorrere a una *rete mobile* (MN-Mobile Network).

Una MN, qualunque sia la sua tecnica realizzativa, comprende sempre una sezione di accesso in cui i legamenti di utente utilizzano il mezzo radio come supporto trasmissivo e in cui una opportuna *funzione di accesso* rende disponibile un canale radio all'utente in movimento quando questi ne ha necessita' per chiamare o per essere chiamato.

Oltre alla funzione di accesso, una MN deve essere in grado di svolgere altre tre funzioni fondamentali, e cioe':

- *l'identificazione dell'utente*, ossia la possibilita' di individuare univocamente l'apparecchio terminale a cui fa capo l'utente mobile, quando questi richiede un accesso;
- *l'aggiornamento della posizione dell'utente*, ossia la possibilita' di avere l'aggiornamento continuo ed automatico, in un'appropriata banca di dati, della posizione dell'apparecchio terminale a cui fa capo l'utente mobile: cio' nonostante gli spostamenti di quest'ultimo in un'area anche molto estesa, purché nei limiti della regione di copertura della MN;
- *il passa-mano dei canali radio* (handover), ossia la possibilita' per l'utente mobile di mantenere il legame con la MN pur nella necessita' di cambiare il canale radio che lo connette alla rete; questa necessita' puo' manifestarsi nel passaggio da un'area ad un'altra adiacente che e' caratterizzata da una differente copertura radio ovvero nella degradazione della qualita' del canale radio che e' stato precedentemente assegnato.

Le reti mobili pubbliche sono nate come infrastrutture dedicate alla fornitura del servizio telefonico in ambito geografico. Si sono anche sviluppate reti mobili private, che offrono servizi di fonia e di dati in area locale. Le evoluzioni piu' recenti sono verso *reti mobili integrate*, in cui i servizi offerti sono sostanzialmente dei tipi voce e dati con esigenze di limitata larghezza di banda.

In prospettiva si sta lavorando per la realizzazione di MN capaci di trasferire informazioni con esigenze di larga banda. L'obiettivo finale e' la realizzazione di una infrastruttura per applicazioni mobili a cui si possa accedere

con apparecchi terminali mobili di basso costo e di ingombro limitato e che consenta la fornitura di un vasto spettro di servizi ivi compresi quelli di tipo multimediale.

Circa il legame tra reti mobili e reti fisse, finora queste due reti mobili e quelle fisse si sono sviluppate come strutture separate tra le quali sussistono solo i rapporti di inter-lavoro necessari per consentire le comunicazioni tra due utenti facenti capo alle due reti. In prospettiva però è prevedibile l'integrazione tra rete fissa (rete telefonica o ISDN) e rete mobile. Requisito di tale integrazione è il reciproco beneficio in termini di efficienza e di flessibilità nella fornitura di nuovi servizi con elevate prestazioni.

I.5.4 Comunicazione personale universale

L'obiettivo di *comunicazione personale universale* (UPT-Universal Personal Telecommunication) può essere considerato uno dei più ambiziosi tra quelli oggi perseguiti nello sviluppo delle telecomunicazioni. Si tratta di un servizio che, secondo la definizione datane negli organismi di standardizzazione internazionale, "consente l'accesso a servizi di telecomunicazione di base permettendo la mobilità della persona; esso consente di effettuare o di ricevere chiamate

- mediante un unico numero personale (*numero UPT*), che è indipendente dalla rete;
- attraverso molteplici reti fisse o mobili;
- da un qualunque punto di accesso,

con una disponibilità dei servizi di telecomunicazione che è limitata solo dalla capacità dell'apparecchio terminale utilizzato e dal tipo di accesso alla rete."

In questa definizione sono ravvisabili i seguenti principi di base:

- l'utente del servizio è identificato mediante un unico numero UPT, che è trasparente rispetto alla rete e all'apparecchio terminale;
- l'utente può godere della mobilità della persona, e cioè può spostarsi da un apparecchio terminale ad un'altro e può effettuare o ricevere chiamate trovandosi in una qualunque località geografica; ciò implica che l'infrastruttura per il servizio UPT sia in grado di localizzare l'utente in base al suo numero personale, per consentire il corretto instradamento delle chiamate e per rispondere alle esigenze di tariffazione;
- è reso flessibile l'accesso alla rete da parte dell'utente, che può quindi essere raggiunto ovunque utilizzando sempre lo stesso numero;

- sono definite procedure unificate in ambito mondiale per l'accesso al servizio UPT; quest'ultimo consente di utilizzare tutti i servizi di base;
- e' possibile usufruire del servizio UPT attraverso piu' tipi di reti fisse o mobili, anche affidate a gestori diversi o nell'ambito di infrastrutture private, purché queste abbiano concordato le opportune procedure di inter-lavoro con le reti pubbliche;
- e' possibile un controllo personale e flessibile dei servizi di base scelti da parte dell'utente attraverso l'uso di un *profilo personalizzato* (UPT Service Profile);
- la tassazione e' effettuata sulla base della identita' dell'utente piuttosto che di quella del terminale;
- la sicurezza della tassazione e' garantita dall'impiego di procedure di autenticazione e di tecniche per la riservatezza.

Per la realizzazione degli obiettivi di comunicazione personale universale, attualmente si ritiene che la soluzione piu' conveniente sia attraverso una rete intelligente evoluta.

I.6 Caratteristiche prestazionali

Tra i vari tipi di requisiti prestazionali che caratterizzano la qualita' di un servizio di telecomunicazione ne esistono due di particolare rilevanza, e cioe' quelli riguardanti l'integrita' informativa e la trasparenza temporale. Alla descrizione di queste prestazioni sono dedicati le due prime parti (§ I.6.1 e I.6.2) di questo paragrafo.

Successivamente (§ I.6.3) si parlera' della intesa o meno che si puo' stabilire fra le parti in comunicazione prima che avvenga lo scambio informativo previsto dal servizio. Infine (§ I.6.4) si presenteranno quattro classi di servizio corrispondenti a diverse modalita' di trattamento nell'ambito di una ISDN a larga banda.

I.6.1 Integrita' informativa

Ci riferiamo alle diversita' che si possono manifestare, in relazione a una particolare modalita' di trasferimento, tra

- la sequenza di cifre binarie emessa dalla sorgente nell'ambito del servizio considerato (*sequenza di emissione*);

- la corrispondente sequenza di cifre binarie ricevute, e cioè a valle dell'operazione di trasferimento dell'informazione dalla sua origine alla sua destinazione (*sequenza di ricezione*).

L'integrità informativa fa riferimento alle diversità, che si possono manifestare, in modo aleatorio,

- tra le singole cifre binarie emesse e quelle corrispondenti ricevute (*integrità di cifra binaria*);
- tra le intere sequenze di emissione e di ricezione o tra segmenti corrispondenti di queste (*integrità di sequenza di cifre binarie*).

Nel primo caso le diversità sono dovute a *errori*, normalmente di tipo trasmissivo; nel secondo sono addebitabili a errori procedurali o a specifiche modalità di trasferimento. Il *grado di integrità informativa* è quindi tanto più elevato quanto minore è la frequenza media di errore, misurata, ad esempio nel primo caso, dal *tasso di errore binario*.

Un elevato grado di integrità informativa è un requisito particolarmente importante quando l'informazione scambiata è costituita da dati. Conseguentemente, nei servizi di comunicazione di dati, occorre normalmente mettere in atto procedure protettive, che siano in grado di recuperare gli errori quando questi vengano rivelati ovvero che riescano a correggere gli errori quando questi si manifestano. L'obiettivo di queste procedure protettive è contenere il tasso di errore binario *residuo* a valori orientativamente inferiori a 10^{-10} .

Nel caso invece in cui si tratti di informazione audio, il grado di integrità informativa che è richiesto dipende dal mezzo di rappresentazione che si è scelto. Se a tale mezzo è associata una forma di codifica senza riduzione di ridondanza (ad esempio quella secondo la Racc. G.711 dell'ITU-T), detto grado può essere di valore decisamente inferiore al caso dell'informazione di dati. Deve invece essere via via aumentato man mano che si riduce la ridondanza dell'informazione sottoposta a codifica.

I.6.2 Trasparenza temporale

La trasparenza temporale riguarda invece i *ritardi di transito* che differenti segmenti della sequenza di ricezione possono presentare rispetto ai corrispondenti segmenti della sequenza di emissione. Questi ritardi sono dovuti a cause di natura varia agenti, in generale, in modo aleatorio.

Il *grado di trasparenza temporale* può essere quindi valutato quantitativamente con un parametro che qualifichi la variabilità dei ritardi di

transito e che sia di valore tanto piu' elevato quanto minore e' tale variabilita'. Questo parametro ha allora valore massimo quando la variabilita' dei ritardi di transito e' nulla o di entita' trascurabile, cioe' quando i ritardi di transito sono di valore praticamente costante. Non lo ha quando tale condizione non e' soddisfatta. La distinzione tra queste due caratteristiche prestazionali e' chiarita graficamente in Fig. I-12.

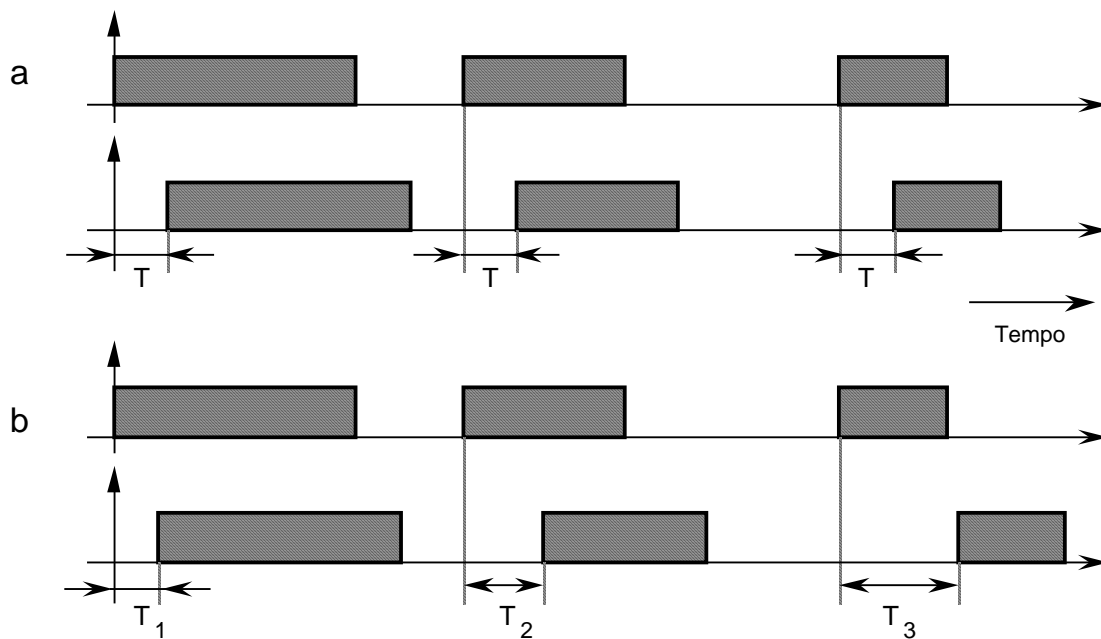


Fig. I-12 - Sequenze di emissione e di ricezione nei due casi di modo di trasferimento (a) temporalmente trasparente e (b) non temporalmente trasparente.

Esistono servizi, per i quali una corretta interpretazione dell'informazione in ricezione richiede che il trasferimento avvenga con un massimo grado di trasparenza temporale. Ne esistono invece altri che non hanno esigenze analoghe. Nel primo caso si dice che e' richiesta una *relazione tra le temporizzazioni all'origine e alla destinazione*; nel secondo caso tale relazione non e' invece richiesta.

Esempi significativi di servizi che richiedono un elevato grado di trasparenza temporale si incontrano in tutti quei casi in cui l'informazione da trasferire sia il risultato di una operazione di conversione analogico-numerica, e cioe' in cui dal segnale analogico originario si estrae una sequenza di campioni, che occorre restituire in ricezione con la stessa periodicit  dell'operazione di campionamento. Pertanto esempi tipici in questo ambito sono offerti da servizi audio e video. Non rientrano invece in queste esigenze i servizi di dati.

Un servizio per il quale e' richiesta una relazione tra le temporizzazioni all'origine e alla destinazione puo' utilizzare un trasferimento in cui sussista variabilita' dei ritardi di transito. Cio' puo' avvenire solo se la distribuzione dei ritardi di transito e' tale da presentare un valore medio sufficientemente basso e deviazioni intorno a questo sufficientemente contenute. In tal caso, in ricezione, si puo' operare una *equalizzazione dei ritardi*, che eleva il grado di trasparenza temporale senza pregiudicare quello di integrita' informativa.

Ai fini della qualita' del servizio e' di interesse anche il valore di picco dei ritardi di transito. Questo deve essere inferiore a un limite che dipende dal servizio considerato e che e' tanto piu' basso quanto maggiori sono le esigenze di interattivita' della comunicazione, come nel caso dei servizi di conversazione (cfr. § I.3.4). Nel caso della telefonia questo limite deve essere inferiore a circa 300 ms.

I.6.3 Modi di connessione

Un servizio puo' essere fruito dalle parti interessate con o senza una loro intesa preliminare. Nel caso in cui l'intesa sussista si parla di *servizio con connessione*, facendo riferimento con questo termine a un legame, almeno *logico* e in alcuni casi anche *fisico*, che viene stabilito tra l'origine e la destinazione della comunicazione. Nel caso contrario si tratta di un *servizio senza connessione*.

Un servizio con connessione ha alcune caratteristiche distintive. In primo luogo ha una *organizzazione temporale ben definita*. Si espleta infatti attraverso tre distinte fasi in sequenza: l'*instaurazione della connessione*, il *trasferimento dell'informazione* e l'*abbattimento della connessione*. La vita di una connessione puo' essere di lunga durata e manifestarsi in molti scambi separati tra le parti connesse; ovvero puo' essere compressa in una interazione di durata molto breve, in cui l'informazione necessaria per instaurare la connessione, quella da trasferire e quella necessaria per abbattere la connessione sono trasportate in un piccolo numero di scambi.

L'instaurazione di una connessione richiede il preventivo accordo tra almeno tre parti: le due che desiderano comunicare e il fornitore che mette a disposizione i mezzi per soddisfare questa esigenza. Queste parti debbono preliminarmente manifestare la loro volonta' di partecipare alla comunicazione. Percio', per tutta la durata della connessione, tali parti sono vincolate all'accordo iniziale.

Nell'ambito di questo accordo, debbono essere negoziati i parametri e le opzioni che governeranno il trasferimento di informazione. Percio' una richiesta di instaurazione puo' essere rifiutata da una parte se i parametri e le opzioni scelti dalla controparte sono per lei inaccettabili.

La negoziazione puo' consentire alle parti di scegliere specifiche procedure, quali quelle riguardanti la sicurezza e la protezione dell'informazione (cfr. § I.4.3). Inoltre l'accordo che risulta dalla negoziazione puo', in alcuni casi, essere modificato (e cioe' rinegoziato) dopo che la connessione sia stata instaurata e sia iniziata la fase di trasferimento dell'informazione.

Una volta che una connessione sia stata instaurata, essa puo' essere utilizzata per trasferire segmenti di una sequenza informativa fintantoche' la connessione non venga rilasciata da una delle parti in comunicazione. Questi segmenti sono mutuamente legati in virtu' del fatto che sono trasferiti su una particolare connessione. Piu' specificamente, per effetto del trasferimento ordinato sulla stessa connessione

- possono essere facilmente rivelate e recuperate condizioni di fuori-sequenza, di perdita e di duplicazione riguardanti differenti segmenti della sequenza informativa;
- possono essere impiegate tecniche di controllo di flusso per assicurare che il ritmo di trasferimento tra le parti in comunicazione non superi quello che queste parti sono in grado di trattare.

Infine deve essere sottolineata la distinzione tra l'inizializzazione di una comunicazione su base chiamata e la connessione che nell'ambito della chiamata puo' essere instaurata. Infatti nell'ambito di una chiamata possono essere instaurate anche due o piu' connessioni in relazione alle esigenze di differenti flussi informativi che debbono essere scambiati. Inoltre ogni connessione instaurata inizialmente puo' essere modificata nel corso della chiamata.

Passando ora ai servizi senza connessione e alle loro caratteristiche distintive, osserviamo che in questo caso il trasferimento avviene a seguito di accordi *tra due sole parti*. Tra le parti in comunicazione esiste solo un conoscenza mutua a priori. Esistono poi accordi individuali tra ogni parte e il fornitore del servizio. Non viene invece scambiata preliminarmente tra le parti in comunicazione alcuna informazione che riguardi la volonta' mutua di queste parti a impegnarsi in una comunicazione.

Poiche' gli scambi informativi avvengono senza la preventiva instaurazione di una connessione, l'infrastruttura deve essere sempre disponibile per il trasferimento richiesto, mentre le parti potenzialmente coinvolte nella

comunicazione possono non essere contemporaneamente attive. E' sufficiente infatti che la parte di origine sia attiva solo per il tempo necessario all'emissione delle informazioni, mentre la parte di destinazione lo deve essere solo al momento della ricezione.

In un servizio senza connessione debbono essere sottolineate la *indipendenza* e l'*autoconsistenza* dei segmenti informativi trasferiti. Circa l'indipendenza, questa implica che una sequenza di segmenti informativi puo' essere recapitata a destinazione in un ordine diverso da quello di consegna all'origine.

Per cio' che riguarda poi l'autoconsistenza, ogni segmento informativo deve contenere tutta l'informazione necessaria per essere consegnata a destinazione. Questa caratteristica, se da un lato migliora la robustezza del servizio, dall'altro comporta una quota di extra-informazione che e' maggiore di quella richiesta nel caso di un servizio con connessione.

1.6.4 I servizi nella B-ISDN

Come si e' gia' visto (§ I.5.1), una B-ISDN rappresenta il risultato attualmente piu' evoluto nello sviluppo delle funzionalita' di trasporto presenti nelle reti fisse: cio' allo scopo di soddisfare le esigenze di comunicazione, in particolare multimediale, che sono in corso di maturazione negli ambienti domestico e lavorativo.

Tali nuove esigenze comprendono, in aggiunta a quelle gia' soddisfatte dalla N-ISDN:

- servizi su base chiamata con possibilita' di trasferire anche informazione di dati e di immagine con movimento completo, nell'ambito di emissioni ad alta velocita';
- servizi senza connessione, per trasferire dati tra reti in area locale e stazioni di lavoro ad alta velocita' elaborativa;
- servizi multimediali, con trasferimento, in forma correlata, di flussi composti da due o piu' mezzi di rappresentazione;
- servizi con configurazione multipunto per comunicazioni sia interattive che distributive.

Il soddisfacimento di queste esigenze deve tuttavia tenere conto delle caratteristiche dei flussi informativi che, conseguentemente, dovrebbero essere trattati. In particolare, questi flussi

- hanno origine da sorgenti numeriche che emettono con ritmi binari di picco compresi in una dinamica di almeno 6 decadi, e cioè tra un limite inferiore minore di 1 kbit/s e un limite superiore di circa 1 Gbit/s;
- possono essere CBR o VBR e, in questo secondo caso, il grado di intermittenza può assumere valori che vanno da qualche unità ad alcune decine;
- presentano requisiti prestazionali che possono essere non omogenei; ad esempio, come già discusso, i requisiti di trasparenza temporale e di integrità informativa possono essere decisamente diversi se si passa da servizi di dati a comunicazioni in cui l'informazione è di tipo audio o video; inoltre è chiaramente intuibile la diversità di trattamento richiesta da servizi con e senza connessione.

Circa la dinamica dei ritmi binari di picco emessi, la B-ISDN deve rendere disponibile una potenzialità di trasferimento ad alta capacità in corrispondenza dell'interfaccia utente-rete e delle apparecchiature della sezione interna. Tale capacità deve infatti essere superiore di almeno due ordini di grandezza a quella disponibile in una N-ISDN. Il conseguimento di questa capacità dal punto di vista fisico è oggi possibile con un impiego generalizzato delle fibre ottiche, sia nella sezione interna che in quella di accesso. Riguardo poi ai provvedimenti di natura logica, questi saranno oggetto di trattazione nel capitolo IV.

Si consideri poi come la B-ISDN può fronteggiare le differenti esigenze prestazionali di servizi diversi. Da un punto di vista di economia di sistema è con evidenza non conveniente

- adottare, per ogni attributo prestazionale, il valore che soddisfa le esigenze di qualunque servizio (esigenze di caso peggiore);
- trattare ogni servizio secondo le sue specifiche esigenze prestazionali.

È preferibile invece (ed è quanto si è deciso di fare) suddividere i servizi offerti dalla B-ISDN in un insieme di classi, ove i requisiti prestazionali sono sufficientemente omogenei. In tal modo per ogni classe può essere applicato con successo un trattamento che soddisfa le esigenze di quella classe.

La classificazione in parola è stata effettuata rispetto ai seguenti attributi

- la relazione tra le temporizzazioni all'origine e alla destinazione;
- le caratteristiche di emissione della sorgente;
- il modo di connessione.

Il primo attributo può assumere i due valori "relazione richiesta" o "non richiesta"; il secondo considera i due casi di sorgenti CBR o VBR (cfr. § I.2.3); il terzo riguarda la distinzione fra servizi con o senza connessione.

Preso atto che non tutte le possibili combinazioni tra queste alternative hanno significato, l'orientamento è stato verso quattro classi possibili (fig. I.6.2):

- la *classe A* comprende servizi in cui è richiesta relazione tra le temporizzazioni all'origine e alla destinazione, il ritmo binario emesso è costante e il servizio è con connessione;
- la *classe B* comprende servizi in cui è ancora richiesta relazione tra le temporizzazioni all'origine e alla destinazione, il servizio è ancora con connessione, ma il ritmo binario emesso è variabile;
- la *classe C* non richiede relazione tra le temporizzazioni all'origine e alla destinazione, il servizio è con connessione e il ritmo binario è variabile;
- la *classe D* è come la classe C, ma con la differenza che i servizi inclusi sono senza connessione.

Della classe A fanno parte i servizi che richiedono una emulazione del modo di trasferimento a circuito (cfr. § IV.5.1). Alla classe B appartengono i servizi audio e video con sorgenti che emettono a ritmo binario variabile. Infine le classi C e D includono le comunicazioni di dati, che hanno svolgimento con e senza connessione, rispettivamente.

II LE RISORSE DI RETE

Le risorse di una rete sono di natura varia; se tuttavia ci limitiamo all'insieme delle funzioni di tipo logico, esse possono identificarsi in *risorse di trasferimento* e in *risorse di elaborazione*.

Le prime risiedono nei rami e nei nodi della sezione interna, oltre che nelle potenzialità della sezione di accesso. Le seconde interessano principalmente i nodi della sezione interna, ma si incontrano anche nella sezione di accesso al livello di interfaccia utente-rete; per esse possono individuarsi almeno tre finalità principali, e cioè il trattamento di chiamata nel caso di comunicazioni su base chiamata, il controllo dello scambio dell'informazione di utente nel rispetto di opportune regole di procedura e di protezione e, infine, la gestione della rete.

Più in generale si può affermare, che, per l'espletamento delle loro funzioni, una rete e le sue parti componenti devono svolgere *attività* e devono utilizzare *risorse*.

Una attività è un insieme coerente di azioni elementari che perseguono uno scopo definito e che utilizzano le risorse all'uopo necessarie. Una attività è quindi l'associazione evolutiva di risorse che concorrono al conseguimento di uno scopo comune.

Con riferimento all'interazione tra attività e risorse, questo capitolo chiarisce dapprima (par.II.1) la distinzione tra risorse indivise e risorse condivise: a conferma del ruolo fondamentale che queste ultime hanno nello sviluppo di una infrastruttura per telecomunicazioni, vengono esaminati i problemi posti dalla condivisione e alcuni dei criteri gestionali che consentono di risolverli.

Seguono tre paragrafi dedicati alla introduzione di modelli che rappresentano l'interazione tra attività e risorse condivise: ciò allo scopo di fornire primi elementi sulla metodologia per

- valutare le prestazioni conseguibili con date risorse (*problema di analisi*);
- individuare, in termini qualitativi e quantitativi, le risorse necessarie al fine di ottenere le prestazioni desiderate (*problema di sintesi*).

In particolare, il par. II.2 riguarda il caso semplice in cui l'interazione da schematizzare possa considerarsi indipendente da altre interazioni che coinvolgono altre attività e altre risorse. Il modello che può identificarsi per fronteggiare questo caso è un sistema di natura stocastica e può presentare, come un sistema deterministico, una evoluzione temporale in cui possono distinguersi un regime transitorio e, seppure con particolari limitazioni, un

regime permanente: il par. II.3 e' dedicato a chiarire come questo secondo regime (detto di *equilibrio statistico*) sia di riferimento per la descrizione dei fenomeni di traffico e per la caratterizzazione prestazionale della realta' che il modello rappresenta.

Il terzo tema modellistico (par.II.4) riguarda il caso di risorse multiple tra loro cooperanti in azioni concatenate e quindi interdipendenti: anche in questo caso, piu' complesso di quello trattato nei paragrafi II.2 e II.3, viene proposto un modello e ne vengono fornite alcune proprieta' con riferimento a condizioni di equilibrio statistico.

Sempre con riferimento alla interazione tra attivita' e risorse, il capitolo si conclude (par II.5) con la introduzione di alcuni concetti che sono di base per la comprensione della varieta' di scelte sistemistiche da considerarsi per la realizzazione di un ambiente di comunicazione. Si tratta di concetti che, isolatamente, hanno gia' trovato applicazione in contesti consolidati quali quelli delle reti dedicate a un servizio e che vengono qui riproposti in un quadro organico per sottolineare i vantaggi e gli svantaggi dei differenti criteri di gestione di una risorsa.

II.1 Risorse indivise e risorse condivise

Come gia' sottolineato nell'introduzione del capitolo I, le risorse di una rete di telecomunicazione debbono essere commisurate in termini di quantita' e di potenzialita' (*vincolo di costo*) alle esigenze di un servizio qualitativamente accettabile (*vincolo di qualita' di servizio*).

A questo riguardo, come anticipato nel par. I.1, un criterio fondamentale, che trova la sua giustificazione nel rispetto del vincolo di costo, suggerisce di limitare allo stretto indispensabile l'impiego di *risorse indivise*, e cioe' di risorse assegnate, in modo permanente o semi-permanente, allo svolgimento di una specifica attivita'.

E' invece preferibile, ove possibile, prevedere l'impiego di *risorse condivise*, che costituiscono un insieme, i cui elementi sono utilizzabili da piu' attivita', anche se queste perseguono scopi distinti. Dato pero' che ogni risorsa puo' essere al servizio di una sola attivita' alla volta, una sua utlizzazione da parte di un insieme di attivita' deve avvenire, per ciascuna di queste, in intervalli di tempo distinti.

Per le risorse dei tipi indiviso e condiviso e' possibile fornire svariati esempi significativi con riferimento alle funzionalita' di una rete di

telecomunicazione. Qui, tenendo conto della fase ancora introduttiva della trattazione, ci si limita a citare un esempio per ognuno dei due tipi di risorse.

Esempio II.1-1 - Nelle attuali reti telefoniche fisse, la linea d'utente è individuale (cfr. § I.1.3). Essa utilizza, come mezzo trasmissivo, una coppia bifilare in rame (*doppino*) e connette uno o più apparecchi telefonici disponibili in una postazione di utente con l'autocommutatore locale di competenza. L'interfaccia tra la linea d'utente e detto autocommutatore è chiamata *attacco d'utente*; le sue funzioni verranno presentate nel seguito. La linea individuale d'utente e l'attacco d'utente costituiscono un esempio di risorsa di trasferimento indivisa.

Esempio II.1-2 - Un processore per il trattamento di chiamate telefoniche è una risorsa di elaborazione che è condivisa nell'espletamento delle funzioni decisionali svolte da un autocommutatore.

Il lettore, come utente abituale del servizio telefonico, conosce bene le modalità operative da seguire per effettuare una chiamata telefonica. Deve infatti sganciare il microtelefono, attendere il cosiddetto *invito a selezionare* e, dopo aver ricevuto il *tono* corrispondente, deve agire su un disco o su una tastiera componendo il numero del chiamato. Le azioni successive dipendono dai casi che si possono presentare: se riceve il *tono di libero*, deve attendere la risposta del chiamato; se invece riceve un *tono di occupato*, deve desistere o effettuare una nuova chiamata.

Condizioni altrettanto ben note si verificano quando un utente è chiamato. L'avviso di questa situazione è dato, tramite la suoneria dell'apparecchio telefonico, da un *segnale di chiamata*; l'attivazione della comunicazione si verifica all'atto dello sgancio del microtelefono.

Questo insieme di modalità operative coinvolge l'azionamento di varie apparecchiature della rete telefonica, e precisamente l'apparecchio telefonico del chiamante e quello del chiamato, l'autocommutatore di origine e quello di destinazione (che può o meno essere distinto dal precedente), nonché eventuali autocommutatori intermedi.

Detti azionamenti avvengono sulla base di informazioni di segnalazione che vengono scambiate tra le apparecchiature sopra citate allo scopo di rendere possibile una comunicazione telefonica o di metterne in evidenza la temporanea impossibilità. Di questo scambio di informazioni l'utente chiamante percepisce solo i risultati intermedi o finali, che lo guidano nei suoi comportamenti.

I compiti svolti dalle risorse di elaborazione negli autocommutatori coinvolti rientrano nella funzione di trattamento di chiamata (cfr. § I.3.3).

L'interazione tra attività e risorse si manifesta con una *domanda* e con una *risposta*. La domanda è quella connessa alle esigenze delle attività per l'assolvimento dei compiti a cui sono preposte: si presenta con *richieste di servizio*, e cioè con richieste di utilizzazione di una o più risorse a cui le attività possono accedere.

La risposta è legata, oltre che alla domanda, anche alla potenzialità di ogni risorsa a fronteggiare e a soddisfare le richieste di servizio che le vengono presentate; nel caso di risorse condivise, intervengono poi altri fattori che influenzano la risposta: su questi si tornerà nel seguito.

Circa la domanda, l'evoluzione temporale di una attività richiede, in generale, l'impegno di una risorsa solo durante *intervalli di vitalità*, che si alternano ad altri *di latenza* nel corso dell'intera *durata dell'attività* e cioè nell'intervallo tra l'inizio e la conclusione di detta evoluzione. Inoltre occorre considerare la aleatorietà delle esigenze di utilizzazione da parte di ogni attività interessata all'impiego di una risorsa. Questa aleatorietà si manifesta sotto vari aspetti; in particolare:

- la presentazione degli intervalli di vitalità, così come quella delle durate di ogni attività, è in generale di tipo aleatorio;
- altrettanto aleatoria è, sempre in generale, la quantità di lavoro che le risorse sono chiamate a svolgere a seguito delle richieste di servizio.

Il lavoro richiesto e l'unità assunta per misurarne la quantità variano in relazione alla natura della risorsa e del suo utilizzatore. Ad esempio, nel caso in cui si consideri una risorsa di trasferimento e l'utilizzatore sia un messaggio di dati, l'unità di lavoro corrisponde al trasferimento di una singola cifra binaria o a quello di una sequenza strutturata (e quindi identificabile in modo univoco) di cifre binarie che compongono il messaggio. Quando invece si tratta di una risorsa di elaborazione e l'utilizzatore è un programma, come unità di lavoro può essere assunto il trattamento di una singola istruzione tra quelle che compongono il programma.

Per ciò che riguarda la risposta e, più in particolare, le sue limitazioni connesse alla potenzialità della risorsa, questa fronteggia la domanda entro quanto consentito dalla sua *capacità*, e cioè dal numero massimo di unità di lavoro che la risorsa è in grado di svolgere nell'unità di tempo.

Quanto ora detto non è però sufficiente a chiarire i problemi connessi a una condivisione di risorse. A questo chiarimento è dedicato il resto del paragrafo.

II.1.1 Il ruolo delle attività di gestione

L'accesso a un insieme di una o più risorse condivise deve essere opportunamente controllato. Per questo scopo, accanto alle *attività di utilizzazione* a cui ci si è riferiti finora, devono essere presenti anche *attività di gestione*.

Le attività di utilizzazione perseguono gli scopi che sono a loro propri senza curarsi dei problemi posti dalla condivisione delle risorse con altre attività di utilizzazione. Per ciò che riguarda la loro interazione con le attività di

gestione, esse si limitano a formulare a queste ultime le richieste del servizio che e' loro necessario.

Le attivita' di gestione debbono invece provvedere a controllare gli accessi alle risorse a cui sono preposte. Il loro obiettivo e' quello di soddisfare la domanda in modo da rispettare i vincoli di costo e quelli di qualita' di servizio.

Con questi obiettivi, compiti fondamentali delle attivita' di gestione sono:

- 1) risolvere le *condizioni di contesa*, dovute alla concorrenza delle richieste di servizio da parte delle attivita' di utilizzazione;
- 2) minimizzare il rischio di *situazioni di stallo*, che si possono manifestare nell'evoluzione di due o piu' attivita' di utilizzazione aventi possibilita' di accesso allo stesso insieme di risorse condivise;
- 3) attuare le *strategie di assegnazione* delle risorse alle attivita' di utilizzazione che ne fanno richiesta;
- 4) evitare, nei limiti del possibile, l'insorgere di *fenomeni di sovraccarico* delle risorse, in quanto questi determinano un drastico peggioramento delle prestazioni di qualita' di servizio.

Per i compiti di cui in 1) e 2) verranno forniti qui di seguito alcuni dettagli, mentre sui compiti di cui in 3) e 4) si ritornera' nel paragrafo II.5.

II.1.2 Gestione delle condizioni di contesa

La concorrenza delle richieste di servizio presentate dalle attivita' di utilizzazione puo' determinare *condizioni di contesa*. Queste si verificano quando tutte le risorse disponibili sono occupate a favore di altrettante attivita' di utilizzazione in corso di evoluzione e vengono presentate nuove richieste di servizio. Le condizioni di contesa debbono essere risolte assicurando equita' di trattamento nei confronti delle attivita' di utilizzazione. Come gia' si e' detto, questo e' uno dei compiti delle attivita' di gestione.

Questo compito puo' essere assolto con varie modalita'. Qui se ne presentano alcune senza entrare nel merito delle soluzioni tecniche che possono essere adottate per la loro realizzazione. Si preferisce invece effettuarne una descrizione su un piano astratto, che meglio si presta all'identificazione dei modelli a cui si e' fatto cenno nell'introduzione del capitolo.

Con quest'ultima precisazione, le principali modalita' per risolvere condizioni di contesa sono quella *orientata al ritardo* e quella *orientata alla perdita*. Se ne chiarisce il significato facendo riferimento, per semplicita', alla disponibilita' di una sola risorsa condivisa.

Nella *modalita' orientata al ritardo*, l'accesso alla risorsa e' attuato tramite una fila d'attesa, che si assume di capienza illimitata. Si pongono allora in fila d'attesa tutte le attivita' di utilizzazione che presentano una richiesta di servizio quando la risorsa e' gia' occupata a favore di un'altra attivita' di utilizzazione (*stato di blocco*). Non appena la risorsa si libera da questo impegno, una delle attivita' di utilizzazione in fila d'attesa, scelta con un opportuno ordine di accesso, viene ammessa alla risorsa.

L'evento che caratterizza questa modalita' e' allora quello di *ritardo*, che si presenta quando la risorsa e' in uno stato di blocco con il condizionamento che arrivi una nuova richiesta di servizio. In tali situazioni, detta richiesta puo' essere accolta solo dopo lo stazionamento della attivita' di utilizzazione richiedente in fila d'attesa; cio' comporta un *ritardo di attesa*, corrispondente all'intervallo di tempo tra l'istante di arrivo della nuova richiesta e quello di ammissione alla risorsa.

Nella *modalita' orientata alla perdita* l'accesso alla risorsa e' attuato senza fila d'attesa. Conseguentemente viene rifiutata ogni richiesta di servizio che sia presentata quando la risorsa e' in uno stato di blocco. L'attivita' di utilizzazione che ha subito questo trattamento dovra' ripresentare la sua richiesta di servizio in tempi successivi senza alcuna garanzia a priori di vederla accolta. Pertanto, in questa modalita' l'evento caratterizzante e' quello di *rifiuto*. Si tratta ancora di un evento condizionato, le cui condizioni di presentazione sono analoghe a quelle dell'evento di ritardo, ma con la differenza sostanziale, gia' sottolineata, che riguarda il trattamento delle attivita' di utilizzazione che ne subiscono gli effetti.

E' ovviamente anche possibile adottare una modalita' intermedia tra le due ora descritte; si puo' cioe' procedere con una *modalita' orientata al ritardo con perdita*. A tale scopo e' sufficiente che l'accesso alla risorsa avvenga tramite una fila d'attesa avente capienza limitata. In tal modo si possono verificare ambedue gli eventi di rifiuto e di ritardo. In particolare, una attivita' di utilizzazione, che presenta la sua richiesta di servizio quando, oltre allo stato di blocco della risorsa desiderata, si verifica anche la completa occupazione della fila d'attesa, subira' lo stesso trattamento che si ha nella modalita' orientata alla perdita. Se, invece, la fila d'attesa non e' completamente occupata, il trattamento e' quello della modalita' orientata al ritardo.

II.1.3 Gestione degli stati di stallo

Gli stalli sono un fenomeno tipico in un ambiente di risorse condivise e possono compromettere l'evoluzione di due o più attività di utilizzazione.

Esaminiamo un esempio dell'insorgere di stati di questo tipo con riferimento a due sole attività di utilizzazione. Se l'attività di utilizzazione A conserva il possesso di una risorsa α in attesa della risorsa β che è in possesso dell'attività di utilizzazione B e se quest'ultima ha necessità di accedere alla risorsa α per proseguire la sua evoluzione, le due attività si bloccano mutuamente.

Per minimizzare il rischio di stati di stallo tra le attività di utilizzazione in una rete di telecomunicazione, possono essere adottati vari accorgimenti, che ben si differenziano da quelli largamente studiati in sistemi con controllo centralizzato: in una rete, infatti, l'interazione tra attività di utilizzazione e risorse condivise avviene solitamente nell'ambito di controlli distribuiti, che sono compito delle singole attività di gestione.

Tra tali accorgimenti uno dei più frequentemente utilizzati nel caso di risoluzione delle condizioni di contesa orientata al ritardo, consiste nell'adozione sistematica di *temporizzatori* per il ritardo di attesa nell'accesso ad ogni risorsa. In queste condizioni, quando tale ritardo relativo ad una risorsa specifica ha superato un limite opportunamente fissato (*fuori-tempo massimo*), si suppone che si sia verificato uno stato di stallo. Si libera allora quella risorsa, interrompendo l'attività di utilizzazione che ne era in possesso, e si effettua un recupero dell'errore procedurale che conseguentemente si determina.

II.2 Modelli dell'interazione tra attività e risorse.

L'accesso ad una risorsa condivisa può essere descritto da particolari modelli matematici, chiamati *sistemi di servizio* (o *sistemi a coda*). Tali modelli schematizzano l'interazione tra richieste di servizio e soddisfacimenti di tali richieste; sono utilizzati per la risoluzione dei problemi di analisi e di sintesi a cui si è accennato nell'introduzione del capitolo.

In generale un sistema di servizio, come illustrato in Fig. II-1, ha come parti costituenti una *fila d'attesa* e un *servizio*, mentre le parti agenti sono gli *utenti potenziali* del sistema, che, nel nostro caso, si identificano con le attività d'utilizzazione. Il *controllo del sistema* rientra invece nei compiti delle attività di gestione. Mentre il servizio è una parte sempre presente, la fila d'attesa può mancare in alcuni casi particolari. Per semplicità espositiva ci riferiremo dapprima al caso più generale che è quello in cui la fila d'attesa sia presente con un numero limitato di posti disponibili e possa quindi ospitare un numero finito di utenti.

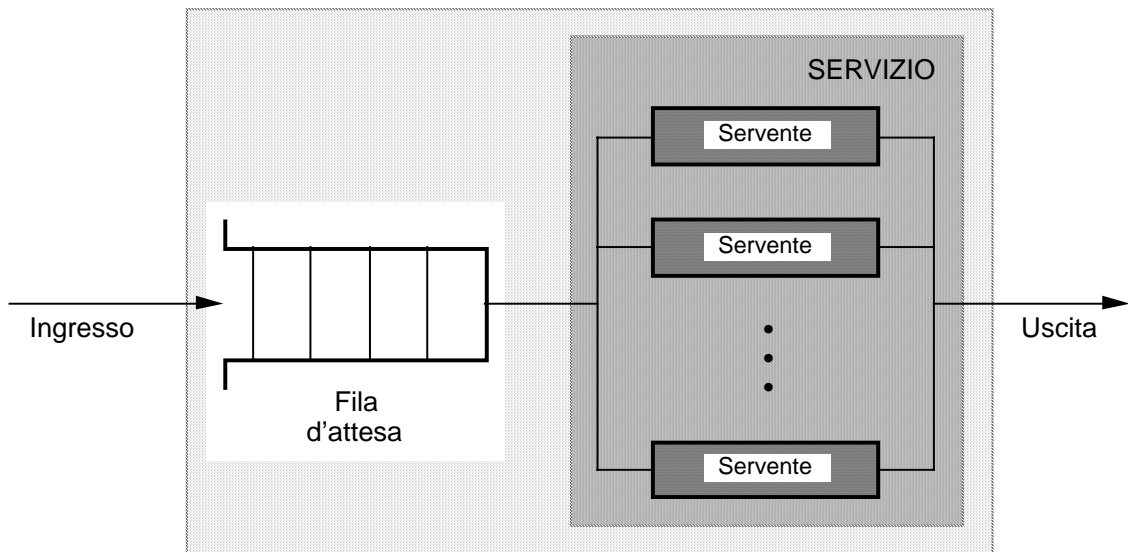


Fig. II-1 - Struttura generale di un sistema di servizio.

Il servizio comprende i *serventi* del sistema, e cioè, nel caso specifico, le risorse condivise, che sono finalizzate ad uno scopo definito. Un servente è libero quando la risorsa corrispondente non è impegnata; è invece occupato in caso contrario.

Fintantochè esiste almeno un servente libero, un utente che richiede servizio viene ammesso direttamente all'utilizzazione della risorsa disponibile. Quando invece tutti i serventi sono occupati, gli utenti che sopraggiungono (e cioè che richiedono servizio mentre si verifica questo stato) si dispongono in fila d'attesa. Essi potranno quindi essere ammessi al servizio con un *ritardo* corrispondente al tempo trascorso in fila di attesa.

Se, oltre ai serventi, anche la fila d'attesa è completamente occupata, gli utenti che sopraggiungono sono *rifiutati* dal sistema e dovranno quindi ripresentare la loro richiesta di servizio in tempi successivi.

Non appena si libera uno dei serventi occupati, un utente in fila d'attesa, scelto con un'opportuna regola gestita dal controllo del sistema, comincia ad essere servito. Al compimento del lavoro richiesto, l'utente esce dal sistema.

Circa l'utilizzazione dei sistemi di servizio come modelli dell'interazione tra attività di utilizzazione e risorse, si possono distinguere due forme principali:

- la prima corrisponde al caso in cui l'interazione da modellare sia schematizzabile come indipendente da altre interazioni che possono coinvolgere le stesse attività e altre risorse;

- la seconda si presenta quanto l'interazione coinvolge risorse multiple tra loro cooperanti in azioni concatenate e quindi interdipendenti.

Nel primo caso i problemi di contesa e di assegnazione che ne sorgono possono essere affrontati con un *singolo sistema di servizio*. Nel secondo caso occorre ricorrere a modelli piu' complessi, in cui due o piu' sistemi di servizio sono tra loro interconnessi nell'ambito di una *rete di code*. Questo paragrafo e il successivo sono dedicati al primo di questi casi, mentre la trattazione del secondo caso verra' effettuata nel paragrafo II.4.

Si precisa che i contenuti di questi tre paragrafi hanno scopo solo introduttivo per le esigenze della trattazione che segue: non possono quindi essere considerati sostitutivi di uno sviluppo organico della *teoria dei sistemi di servizio* (o teoria delle code) con le sue basi matematiche (calcolo delle probabilita' e teoria dei processi aleatori). Per la presentazione di questa teoria, si rinvia ai testi citati in bibliografia, ove l'argomento e' sviluppato con riferimento specifico al contesto applicativo che e' qui considerato. Da questi testi possono essere acquisite le metodologie matematiche per risolvere i problemi di analisi e di sintesi a cui si e' accennato in precedenza.

Un singolo sistema di servizio e' definito in modo completo qualora siano precisati i seguenti elementi:

- la composizione della *popolazione degli utenti potenziali*;
- la *struttura del sistema*;
- la *domanda di servizio* da parte degli utenti potenziali;
- le *funzionalita' di controllo* del sistema.

Alla descrizione di questi elementi e' dedicato cio' che segue.

II.2.1 *La popolazione degli utenti potenziali*

In generale, la popolazione degli utenti potenziali di un sistema di servizio e' costituita da una o piu' classi, che presentano differenti comportamenti, in particolare per cio' che riguarda le caratteristiche della domanda di servizio (cfr. § II.2.3). Nel seguito si indichera' con C ($C=1, 2, \dots$) il numero di queste classi e con $U^{(c)}$ la cardinalita' dell'insieme di utenti appartenenti alla *c-classe* ($c=1, 2, \dots C$). Questa *c-cardinalita'* puo' essere finita o infinita. Inoltre un utente appartenente alla *c-classe* verra' identificato per brevitaa' come *c-utente*.

Nei confronti del sistema di servizio, ogni *c-classe* e' origine di una *corrente di richieste di servizio*, mentre il singolo *c-utente* e' una *sorgente di traffico*. Riguardo al trattamento richiesto, i *c-utenti* possono differenziarsi da

altri utenti appartenenti a differenti classi per due elementi: in ogni caso, per la particolare domanda di servizio che e' espressa dalla *c*-classe; talvolta, anche per una discriminazione effettuata dalle funzionalita' di controllo del sistema.

Questa discriminazione puo' essere richiesta per tenere conto delle possibili diversita' di qualita' di servizio che possono richiedere certe attivita' di utilizzazione rispetto ad altre. Ad esempio, possono esistere attivita' per le quali il "costo" dell'attesa risulta maggiore di quello relativo ad altre. Il rispetto del vincolo di costo richiedera' allora, in questi casi, di attribuire una piu' elevata priorita' di servizio alle attivita' del primo tipo rispetto a quelle del secondo.

Il *c-grado di priorita'*, che e' attribuito ai *c*-utenti, puo' o meno essere funzione dell'istante di osservazione del sistema. Nel primo caso si parla di *priorita' dipendente dal tempo* e nel secondo di *priorita' fissa*. Comunque il *c-grado di priorita'* e' definito da una numero (dipendente o meno dal tempo), che aumenta all'aumentare della priorita' di trattamento che e' richiesta dalla *c*-classe. In § II.2.4 si descriveranno le possibili modalita' di utilizzazione del grado di priorita' da parte delle funzionalita' di controllo del sistema di servizio.

II.2.2 Caratteristiche strutturali di un sistema di servizio

Una delle possibili classificazioni dei sistemi di servizio riguarda le possibilita' che sono previste per l'accesso ai serventi del sistema. Al riguardo si distinguono due tipi di sistemi:

- a) *a piena accessibilita'* (full availability), in cui tutti i serventi, quando non sono occupati, sono disponibili a soddisfare la richiesta di servizio di un utente qualunque indipendentemente dalla classe di appartenenza;
- b) *ad accessibilita' limitata* (limited availability), in cui l'utente appartenente a una generica classe ha accesso solo ad un sottoinsieme dei serventi presenti nel sistema e i sottoinsiemi relativi alle varie classi possono avere intersezione non nulla.

Ovviamente, nel caso particolare in cui i sottoinsiemi citati in b) coincidono con l'intero insieme dei serventi presenti nel sistema, si rientra nel caso di piena accessibilita'. Questo e' il caso che verra' considerato nel seguito.

Esempio II.2-1 - I sistemi di servizio sono modelli utilizzabili in svariati campi applicativi in cui si configurano situazioni di richieste di servizio e di risorse condivise messe a disposizione per soddisfare tali richieste. Esistono svariati esempi vicini alle nostre abitudini quotidiane.

Si puo' fare riferimento alla organizzazione di un ente di servizi (ad esempio, di una banca) che interagisce con i propri utenti attraverso sportelli a cui gli utenti accedono per la richiesta di particolari prestazioni (ad esempio, di operazioni bancarie).

Se gli addetti ai vari sportelli sono abilitati a fornire qualunque prestazione che sia di interesse per l'utenza, il sistema di servizio corrispondente è a piena accessibilità. In caso contrario esisteranno sportelli specializzati a cui si rivolgeranno solo gli utenti interessati alle prestazioni di specializzazione. In questo caso si è in presenza di un sistema di servizio ad accessibilità limitata.

I parametri strutturali di un sistema di servizio a piena accessibilità sono il numero S di serventi che sono disponibili nel servizio e il numero Q di posti occupabili in fila d'attesa (*lunghezza della fila d'attesa*). Dato che ogni servente e ogni posto in fila d'attesa è sede potenziale di stazionamento per un utente, la somma $S+Q$ rappresenta la *capienza* del sistema, e cioè il numero massimo di utenti che possono ivi trovare posto contemporaneamente.

I valori che vengono fissati per questi parametri e per le cardinalità $U^{(c)}$ ($c=1, 2, \dots, U$) di ognuna delle classi di utenza consentono di definire vari tipi di modelli, in relazione ai possibili modi di interazione tra attività e risorse che si intendono modellare. A questo scopo e per semplicità espositiva, ci riferiamo al caso di una sola classe di utenti potenziali, per la quale indichiamo con U la relativa cardinalità.

Tutti i tre numeri U , S e Q sono interi non negativi, che possono assumere, a seconda dei casi e in modo indipendente, valori finiti o infiniti. Tuttavia i numeri U e S debbono essere positivi, dato che nel caso di valore nullo (anche per uno solo di questi numeri) il modello perderebbe di significato. Invece il numero Q può essere scelto anche di valore nullo, quando si desidera modellare una situazione in cui l'accesso ai serventi avvenga in modo diretto, e cioè senza possibilità di stazionamento in fila d'attesa.

Come esempio dei criteri per scegliere i numeri U , S e Q consideriamo i modelli con i quali è possibile rappresentare le modalità per risolvere condizioni di contesa (cfr. § II.1.2).

Con questo scopo, il numero S di serventi è posto uguale al numero finito di risorse condivise disponibili, mentre il numero U di utenti potenziali deriva dalla scelta fatta per modellare la domanda di servizio presentata dalle attività di utilizzazione. Rimane allora da scegliere il numero Q .

Al riguardo, sono riconoscibili tre casi significativi, che corrispondono al rispetto di una tra le seguenti relazioni:

$$Q + S \geq U \quad (\text{II.2.1})$$

$$Q + S < U \quad (\text{II.2.2})$$

$$Q = 0; \quad U > S. \quad (\text{II.2.3})$$

Il caso descritto dalla (II.2.1) conduce a modelli *ad attesa senza perdita*, in quanto la capienza del sistema non è inferiore al numero U e quindi non possono verificarsi eventi di rifiuto. Questa è la situazione che si presenta se si sceglie Q di valore infinito anche quando è tale il valore di U . Questi modelli sono quindi atti a rappresentare una modalità di risoluzione delle contese orientata al ritardo.

Nel caso descritto dalla (II.2.2) rientrano invece i modelli *ad attesa con perdita*, che rappresentano una modalità orientata al ritardo con perdita. In particolare, se il valore di U è infinito, è sufficiente assumere Q di valore finito.

Se, infine, l'interesse è rivolto alla rappresentazione di una modalità orientata alla perdita, si farà riferimento al caso descritto dalla (II.2.3) per ottenere modelli *a perdita in senso stretto*. Si nota che la scelta di una fila d'attesa con lunghezza nulla vale indipendentemente dal valore di U , che può essere finito o infinito purché maggiore di S .

II.2.3 La domanda di servizio

Limitiamoci al caso di una sola classe di utenti potenziali ed esaminiamo il modo con cui è possibile modellare la domanda posta da questi utenti in relazione alle loro esigenze di servizio. Al riguardo si osserva che le componenti di tale domanda sono descrivibili, in generale, solo in termini probabilistici: ciò è la naturale conseguenza della corrispondenza modellistica tra utenti del sistema di servizio e attività di utilizzazione e va posto in relazione alla aleatorietà della domanda da parte di queste ultime.

Sia $\{t_i, i=1, 2, \dots\}$ una sequenza degli istanti di richiesta di servizio (Fig. II-2a), quali sono osservabili come possibile manifestazione del comportamento degli utenti potenziali appartenenti a una popolazione che indichiamo con U . Gli elementi di questa sequenza, nel seguito indicati come *istanti di arrivo*, sono distribuiti aleatoriamente sull'asse dei tempi e costituiscono un insieme numerabile: cioè ad ogni istante di arrivo può essere attribuito un numero d'ordine i , che appartiene all'insieme I dei numeri interi naturali e che si suppone crescente nel verso dell'asse dei tempi.

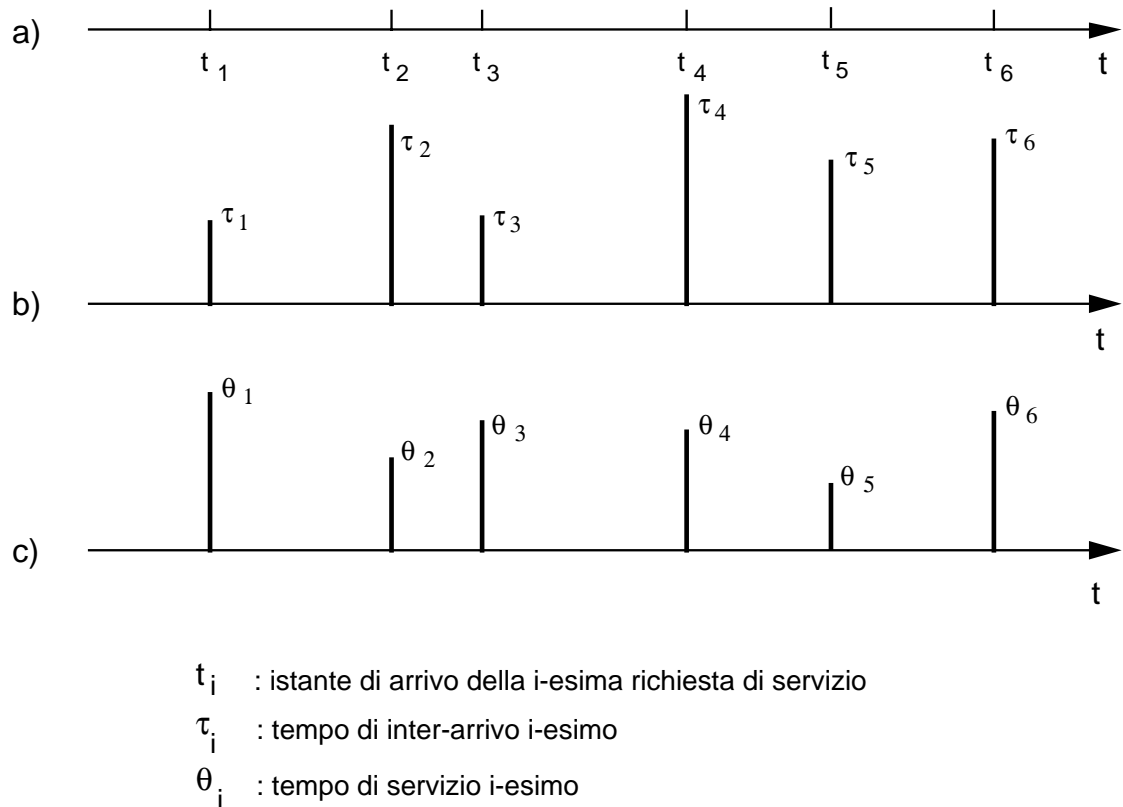


Fig. II-2 - Sequenza di istanti di richiesta di servizio (a) e rappresentazione di una realizzazione del processo di ingresso (b) e di quello di servizio (c).

Se ci riferiamo a un generico istante di arrivo t_i ($i \in I$), si possono osservare le realizzazioni di due variabili aleatorie (nel seguito indicate come V.A.) non negative $T(i)$ e $\Theta(i)$, usualmente continue, che sono associate alla richiesta di servizio presentata nell'istante t_i ; tali V.A. sono:

- 1) l' i -esimo *tempo di inter-arrivo* $T(i)$, e cioè l'intervallo di tempo che intercorre tra l'istante t_{i-1} dell'arrivo immediatamente precedente e quello t_i dell'arrivo considerato; cioè

$$T(i) = t_i - t_{i-1} \quad (i \in I), \quad (\text{II.2.4})$$

ove si pone $t_0=0$;

- 2) l' i -esimo *tempo di servizio* $\Theta(i)$, e cioè l'intervallo di tempo che un servente deve dedicare per soddisfare la richiesta di servizio pervenuta nell'istante t_i ;

nella corrispondenza modellistica, la V.A. $\Theta(i)$ è il rapporto tra la quantità di lavoro $Z(i)$ richiesta nell'istante d'arrivo t_i e la capacità C della risorsa, purché queste due grandezze siano misurate con la stessa unità di lavoro; cioè

$$\Theta(i) = \frac{Z(i)}{C} \quad (i \in I) . \quad (\text{II.2.5})$$

Le V.A. $T(i)$ e $\Theta(i)$ sono i membri i -esimi di due sequenze, in ognuna delle quali gli elementi componenti sono definiti in un dato spazio di probabilità. Tali sequenze, che descrivono l'evoluzione temporale della presentazione degli eventi d'arrivo e della quantità di lavoro richiesta in occasione di ognuno di questi, identificano due processi aleatori (nel seguito indicati come P.A.):

- il primo, definito da $\{T(i), i \in I\}$, e' il *processo di ingresso*;
- il secondo, definito da $\{\Theta(i), i \in I\}$, e' invece il *processo di servizio*.

Nelle figure Fig. II-2b e Fig. II-2c e' mostrata una realizzazione per ognuno di questi due P.A., avendo indicato con τ_i e θ_i le realizzazioni delle V.A. $T(i)$ e $\Theta(i)$, rispettivamente. I processi di ingresso e di servizio, così definiti, sono i modelli matematici delle due componenti della domanda di servizio presentata dalla popolazione U .

Per i P.A. in esame, lo spazio di stato e' continuo se, come qui si suppone, le V.A. $T(i)$ e $\Theta(i)$ hanno realizzazioni che appartengono all'insieme R^+ dei numeri reali positivi. Lo spazio di indice, che coincide con l'insieme I , e' invece discreto. Perciò entrambi i P.A. sono a *stato-continuo* e a *tempo-discreto*.

Circa la caratterizzazione probabilistica di questi P.A., la teoria di base per i sistemi di servizio utilizza una drastica ipotesi semplificativa: si assume infatti che i processi di ingresso e di servizio siano entrambi *di rinnovo*, e cioè che le V.A. non negative che compongono ognuno di essi siano *equidistribuite* e *statisticamente indipendenti*.

Questa ipotesi e' ben rispondente alle situazioni reali, quando i modelli da trattare si riferiscono a particolari contesti applicativi (ad esempio, al caso delle reti telefoniche); lo e' in misura minore o addirittura e' del tutto inadeguata in altri casi.

Il vantaggio dell'ipotesi e' evidente: per una descrizione completa di ciascuno dei due P.A., che richiederebbe in generale la conoscenza delle funzioni di distribuzione di qualunque ordine, e' sufficiente fornire, per ogni P.A., la *funzione di distribuzione del primo ordine* per una singola V.A., assumendo quest'ultima come paradigma dell'intera sequenza di V.A. che definiscono quel P.A. Tale singola V.A. e' il generico tempo di inter-arrivo T nel caso del processo di ingresso e il generico tempo di servizio Θ in quello del processo di servizio.

Alle ipotesi ora dette con riguardo al processo di ingresso e a quello di servizio considerati separatamente, le trattazioni elementari di teoria delle code aggiungono una ulteriore ipotesi riguardante la relazione tra i processi di ingresso e di servizio: si assume che *tali processi siano mutuamente*

indipendenti. Ciò consente di affrontare, in modo matematicamente più semplice, lo studio dell'evoluzione temporale del sistema di servizio, ma introduce anche un serio vincolo alla utilizzabilità dei modelli basati su questa ipotesi in specifici contesti applicativi.

Accanto alle definizioni basate sui tempi di inter-arrivo e di servizio, per i processi di ingresso e di servizio possono essere fornite altre definizioni, che è utile considerare.

Al riguardo si consideri la V.A. N_t : con questa indichiamo il numero di arrivi che si verificano nell'intervallo temporale $(0, t)$, $t \in R^+$. Il P.A. di conteggio $\{N_t, t \in R^+\}$, che è a stato-discreto e a tempo-continuo, può essere assunto come modello alternativo del processo di ingresso. Infatti gli eventi $\{N_t = n\}$ e $\{t_n < t < t_{n+1}\}$ con $n \in I$ sono equivalenti. Tenuto allora conto che, in base alla (II.2.4), si ha

$$t_n = \sum_{i=1}^n T(i) \quad (n \in I) \quad (\text{II.2.6})$$

e che, per definizione, $N_0 = 0$, è immediato ricavare il legame tra le descrizioni probabilistiche dei due P.A. $\{N_t, t \in R^+\}$ e $\{T(i), i \in I\}$; risulta infatti:

$$P\{N_t = 0\} = P\{T(I) > t\} \quad (t \in R^+) \quad (\text{II.2.7})$$

e per $n \geq 1$

$$P\{N_t = n\} = P\left\{\sum_{i=1}^n T(i) < t\right\} - P\left\{\sum_{i=1}^{n+1} T(i) < t\right\} \quad (t \in R^+; n \in I). \quad (\text{II.2.8})$$

Un processo di conteggio può essere assunto anche come definizione alternativa del processo di servizio: basta fare riferimento a una V.A. che rappresenti il numero di terminazioni di servizio occorrenti in un intervallo temporale di durata $t \in R^+$.

Riguardo alle possibili scelte della funzione di distribuzione del primo ordine per le V.A. T e Θ , si forniranno alcuni esempi; in questi, per ragioni di brevità, si farà riferimento, in luogo di T e di Θ , a una V.A. continua non negativa X , che potrà, a seconda dei casi, assumere il significato di generico tempo di inter-arrivo e di generico tempo di servizio.

Per tale V.A. si utilizzeranno le seguenti notazioni:

- $f_X(t)$ è la funzione di densità di probabilità (f.d.p.) di X ;
- $F_X(t)$ è la funzione di distribuzione cumulativa (F.D.P.) di X ;
- $E[X]$ è il valore atteso (o medio) di X ;
- $Var[X]$ è la varianza di X ;
- C_X è il coefficiente di variazione di X , e cioè il rapporto tra deviazione standard e valore medio di questa V.A..

Alla V.A. X e' associata tra l'altro la frequenza media α , che e' di arrivo o di servizio a seconda del significato di X . Tale frequenza media puo' essere definita come l'inverso del valore atteso $E[X]$. Nel seguito le notazioni utilizzate per le frequenze medie associate ai processi di ingresso e di servizio saranno le seguenti

- λ = frequenza media di arrivo = $1/E[T]$;
- μ = frequenza media di servizio = $1/E[\Theta]$. (II.2.9)

Un primo insieme di modelli merita citazione dato che comprende distribuzioni largamente studiate in letteratura per le loro caratteristiche di applicabilita' in svariati contesti e per la possibilita' che esse consentono di approssimare la distribuzione del processo di ingresso e di quello di servizio con una legge semplice. Per questo scopo si dispone usualmente di pochi elementi; ad esempio, da misure eseguite sui fenomeni reali, si conoscono il valor medio e la varianza della V.A. considerata e, per motivi di semplicita', si assume sufficiente una approssimazione che tenga conto dei soli momenti del primo e del secondo ordine per la distribuzione di quella V.A.. Per questo scopo il valore medio $E[X_{ap}]$ e il coefficiente di variazione $C_{X_{ap}}$ della *distribuzione approssimante* debbono essere opportunamente adattati agli analoghi valori $E[X_{as}]$ e $C_{X_{as}}$ della *distribuzione assegnata*.

L'insieme delle distribuzioni a cui si e' fatto ora riferimento comprende:

- la distribuzione *esponenziale negativa*;
- la distribuzione *deterministica*;
- la distribuzione *erlangiana di ordine r* ($r \in I, 1 < r < \infty$);
- la distribuzione *iper-esponenziale di ordine r* ($r \in I, 1 < r < \infty$).

Qui di seguito sono sottolineate alcune delle proprieta' piu' significative di queste distribuzioni.

La distribuzione *esponenziale negativa* di parametro α per la V.A. continua X e' definita da:

$$f_X(t) = \alpha \exp[-\alpha t] \quad (t \geq 0), \quad (\text{II.2.10})$$

ove α e' una quantita' positiva, che ha le dimensioni dell'inverso di un tempo e che, quando X si particolarizza nelle V.A. T e Θ , dipende in generale dallo stato del processo di coda. Dalla (II.2.10) deriva che

$$E[X] = \frac{1}{\alpha}; \quad \text{Var}[X] = \frac{1}{\alpha^2}; \quad C_X = 1. \quad (\text{II.2.11})$$

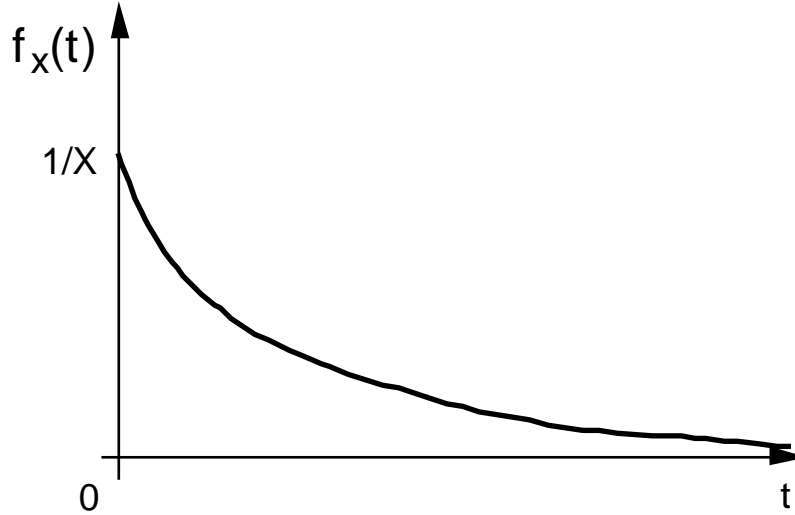


Fig. II-3 - Andamento della funzione di densita' di probabilita' di una distribuzione esponenziale negativa

Se la distribuzione della V.A. X e' *deterministica* con valor medio uguale a $1/\alpha$, allora

$$f_X(t) = \delta\left(t - \frac{1}{\alpha}\right) \quad (t \geq 0), \quad (\text{II.2.12})$$

ove $\delta(\cdot)$ denota l'impulso di Dirac centrato sul valore dell'ascissa che annulla l'argomento dell'impulso. Come e' ovvio, in questo caso si ha:

$$E[X] = \frac{1}{\alpha}; \quad \text{Var}[X] = 0; \quad C_X = 0. \quad (\text{II.2.13})$$

Si osserva che ambedue le distribuzioni definite dalle (II.2.10) e (II.2.12) sono caratterizzate da un unico parametro. La distribuzione *erlangiana di ordine* r ($r \in I$, $1 < r < \infty$) e di parametro α e' invece identificata dalla coppia di valori (r, α) , in cui α e' una quantita' positiva che, in generale, si assume costante. Tale distribuzione riguarda una V.A. definita come somma di un numero fisso r di V.A. equidistribuite e statisticamente indipendenti, con distribuzione esponenziale negativa di parametro $r\alpha$. Si ha allora:

$$f_X(t) = r\alpha \frac{(r\alpha t)^{r-1}}{(r-1)!} \exp[-r\alpha t] \quad (t \geq 0) \quad (\text{II.2.14})$$

e

$$E[X] = \frac{1}{\alpha}; \quad \text{Var}[X] = \frac{1}{r\alpha^2}; \quad C_X = \frac{1}{\sqrt{r}} < 1. \quad (\text{II.2.15})$$

La distribuzione *iper-esponenziale di ordine* r ($r \in I$, $1 < r < \infty$) e' ottenuta come combinazione lineare di r distribuzioni esponenziali negative con differenti parametri α_i ($i = 1, 2, \dots, r$); i coefficienti della combinazione lineare, indicati con π_i ($i = 1, 2, \dots, r$), sono quantita' positive che debbono soddisfare la relazione

$$\sum_{i=1}^r \pi_i = 1.$$

Si ha quindi

$$f_X(t) = \sum_{i=1}^r \pi_i \alpha_i \exp[-\alpha_i t] \quad (t \geq 0) \quad (\text{II.2.16})$$

e

$$E[X] = \sum_{i=1}^r \frac{\pi_i}{\alpha_i} = \frac{1}{\alpha}; \quad \text{Var}[X] = \frac{1}{\alpha^2} + \sum_{i=1}^r \sum_{j=1}^r \pi_i \pi_j \left(\frac{1}{\alpha_i} - \frac{1}{\alpha_j} \right)^2; \quad C_X \geq 1, \quad (\text{II.2.17})$$

ove il segno di uguaglianza si presenta se e solo se $\alpha_i = \alpha$ per ogni i .

Circa il problema di approssimazione, si possono presentare quattro casi. Se C_{Xas} e' prossimo all'unita' ovvero all'incirca uguale a zero si puo' assumere, come approssimante, la distribuzione esponenziale negativa [cfr. la (II.2.11)] ovvero quella deterministica [cfr. la (II.2.13)], rispettivamente: il relativo parametro α e' ottenibile dal valor medio della distribuzione assegnata.

Se invece $C_{Xas} < 1$, l'approssimazione e' attuabile con una distribuzione erlangiana di ordine r , in cui l'ordine r viene scelto uguale all'intero piu' prossimo a $(1/C_{Xas})^2$ e in cui il parametro α e' deducibile, come in precedenza, dal valor medio della distribuzione assegnata [cfr. la (II.2.15)].

Se, infine, $C_{Xas} > 1$, come distribuzione approssimante puo' essere assunta quella iper-esponenziale di ordine r . La determinazione dei relativi parametri (in numero uguale a $2r - 1$) richiede pero' la risoluzione di un problema di "adattamento" della distribuzione approssimante nei confronti di quella approssimata.

Le distribuzioni esponenziale negativa, erlangiana di ordine r e iper-esponenziale di ordine r sono casi particolari di una classe di distribuzioni per le quali la funzione di densita' di probabilita' ha trasformata di Laplace-Stieltjes

$$\hat{f}_X(s) = \int_0^\infty f_X(t) \exp(-st) dt \quad (\text{II.2.18})$$

che e' una funzione razionale fratta della variabile complessa s . Si tratta della classe di *distribuzioni di Cox*. Queste si riferiscono a una V.A. che risulta dalla somma di un numero aleatorio di V.A. con distribuzioni esponenziali negative di parametri costanti $\alpha_1, \alpha_2 \dots \alpha_r$ ($r \in I$; $1 < r < \infty$) tra loro in generale diversi. Si ha

$$\hat{f}_X(s) = \gamma_1 + \sum_{i=1}^r \gamma_{i+1} A_i \prod_{j=1}^i \frac{\alpha_j}{s + \alpha_j}, \quad (\text{II.2.19})$$

ove

$$A_i = \beta_1 \beta_2 \dots \beta_i; \quad \gamma_i + \beta_i = 1 \quad (i \cdot i \cdot r) \quad e \quad \gamma_{r+1} = 1.$$

Inoltre

$$E[X] = \sum_{i=1}^r A_i \alpha_i^{-1}; \quad \text{Var}[X] = \sum_{i=1}^r A_i (\alpha_i^{-1})^2. \quad (\text{II.2.20})$$

La (II.2.19) consente di schematizzare un processo d'ingresso (o di servizio) con la struttura di figura II.2.3, in cui un utente che inizia il suo arrivo (o il suo servizio) entra nel punto A e, dopo uno stazionamento in essa che ha durata T (o Θ), esce dal punto B. In ogni istante di osservazione, la struttura può contenere un solo utente.

La struttura è composta da r stazioni, nella i -esima ($1 \leq i \leq r$) delle quali il tempo di stazionamento è distribuito con legge esponenziale negativa di parametro α_i . Inoltre:

- un utente che entra nella struttura si dirige con probabilità β_1 verso la prima stazione e con probabilità $\gamma_1 = 1 - \beta_1$ verso l'uscita B;
- un utente che invece esce dalla stazione i -esima si dirige con probabilità β_{i+1} verso la stazione $(i+1)$ -esima e con probabilità $\gamma_{i+1} = 1 - \beta_{i+1}$ verso l'uscita.

Questa interpretazione della (II.2.19) è la base per il cosiddetto *metodo degli stadi*.

La distribuzione esponenziale negativa gode della importante proprietà di essere *senza memoria*: cioè la conoscenza che la V.A. X è maggiore di t_0 ($t_0 \geq 0$) non modifica in alcun modo la probabilità che X sia maggiore di $t_0 + t$, ($t > 0$). Questa proprietà si esprime dicendo che, se vale la (II.2.10), risulta necessariamente

$$P\{X > t_0 + t / X \geq t_0\} = g(t), \quad (t \geq 0). \quad (\text{II.2.21})$$

ove $g(t)$ è una funzione che dipende solo dal tempo t . Si può anche dimostrare che l'unica distribuzione senza memoria è esponenziale negativa di parametro $\alpha = f_X(0)$. Conseguentemente il P.A. di rinnovo, che è descritto dalla (II.2.10), è un P.A. *markoviano*.

Inoltre, se il processo d'ingresso è caratterizzato da una distribuzione esponenziale negativa di parametro λ , e cioè se

$$f_T(t) = \lambda \exp[-\lambda t] \quad (t \geq 0), \quad (\text{II.2.22})$$

ove λ è una quantità indipendente dallo stato del processo di coda, allora la V.A. discreta N_t , che esprime il numero di arrivi che si verificano in un intervallo temporale di durata t , ha una distribuzione di Poisson con parametro λt ; cioè

$$P\{N_t = k\} = \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \quad (t \geq 0). \quad (\text{II.2.23})$$

Dato che vale anche la proposizione inversa [e cioè che la (II.2.23) implica la (II.2.22)], il caso ora considerato viene normalmente riferito come processo di ingresso *poissoniano*.

Un processo di ingresso poissoniano gode di due ulteriori proprietà relative alla *sovrapposizione* e alla *estrazione*:

- dati due o più processi di ingresso (*processi componenti*), il *processo di sovrapposizione* si ottiene da quelli dati considerando l'insieme di tutti gli arrivi dei processi componenti; si può dimostrare che, se questi ultimi sono poissoniani, il processo di sovrapposizione è ancora di Poisson e la sua frequenza media di arrivo è uguale alla somma delle analoghe frequenze dei processi componenti;
- dato un processo di ingresso, il *processo di estrazione* si ottiene da quello dato selezionando gli arrivi mediante una *scelta casuale e indipendente da arrivo ad arrivo*; si può dimostrare che, se il processo dato è poissoniano, tale risulta anche il processo di estrazione; inoltre se γ è la probabilità di successo nell'operazione di scelta e se λ è la frequenza media di arrivo del processo dato, allora la analoga frequenza media del processo di estrazione è uguale a $\gamma\lambda$.

II.2.4 Funzionalità di controllo

Le funzionalità di controllo di un sistema di servizio includono la *gestione dell'accesso al sistema* e quella della *disciplina di servizio*.

Circa l'accesso al sistema, si procede, come già detto, con il meccanismo del rifiuto, che può o meno discriminare tra le classi di utenti potenziali. In assenza di discriminazione, il criterio applicato per questo meccanismo è quello di rifiutare l'ammissione di un nuovo utente all'interno del sistema, se la relativa richiesta di servizio viene presentata quando lo stato di occupazione ha raggiunto la capienza del sistema. In presenza, invece, di discriminazione si applica lo stesso criterio, ma con riferimento a una capienza del sistema che può variare da classe a classe di utenti.

La disciplina di servizio è il criterio con cui viene assegnato un nuovo lavoro ad un servente non appena questi si libera da un impegno precedente.

Alcune discipline di servizio sono *conservative*, se non è creato né distrutto lavoro entro il sistema: cioè il servente è sempre occupato quando il sistema non è vuoto e gli utenti lasciano il sistema solo dopo aver ricevuto il lavoro richiesto.

Un primo gruppo di discipline di servizio si basa sugli istanti di arrivo di ogni utente nel sistema. Esempi al riguardo, senz'altro conservativi, sono offerti dalle discipline:

- del *primo arrivato-primo servito* (PAPS), secondo cui si servono gli utenti in fila d'attesa secondo il loro ordine di arrivo;
- dell'*ultimo arrivato-primo servito* (UAPS), secondo cui si servono gli utenti in fila d'attesa in ordine inverso a quello di arrivo.

Come caso limite di criteri di questo tipo, va anche citata la disciplina dell'*ordine casuale*, secondo cui ogni utente in fila d'attesa viene scelto a caso per essere ammesso nel servizio.

Un secondo gruppo di discipline di servizio si riferisce a una certa misura (valore esatto o stima) del tempo di servizio richiesto o della quantità di lavoro già ricevuta da un utente. In questo gruppo rientrano ad esempio le discipline

- del *lavoro più breve-primo servito* (SJF - Shortest Job-First), secondo cui viene scelto l'utente che richiede il più breve tempo d'impegno per il servente;
- del *lavoro più lungo-primo servito* (LJF - Longest Job-First), secondo cui la scelta cade sull'utente che richiede il più lungo tempo di impegno per il servente;
- *a condivisione temporale* (Time Sharing), secondo cui un servente dedica ad un utente in fila d'attesa una porzione di lavoro, fissa o variabile, in funzione o meno del lavoro già svolto a favore di quell'utente; terminata l'esecuzione di questa porzione di lavoro, il servente si dedica a un altro utente in fila d'attesa, ponendo momentaneamente in attesa l'utente precedente se il lavoro richiesto da questi non è stato ancora completato; una possibile modalità di esecuzione di questa disciplina si riferisce ad un sistema di servizio con un solo servente: se a un dato istante la fila d'attesa contiene k utenti, ciascuno di questi riceve k^{-1} unità di tempo di servizio per unità di tempo.

Infine un terzo gruppo di discipline di servizio è basato sull'uso di priorità. Un diritto di priorità si può manifestare sotto due modalità principali:

- nel caso di *priorità senza interruzione* (non preemptive priority), un utente che è entrato nel servizio continua a rimanervi fintanto che il servizio da lui richiesto non è stato completato; l'utente che viene successivamente servito è quello a più elevato grado di priorità presente in fila d'attesa ovvero, a parità di grado di priorità, un utente scelto secondo una delle discipline di servizio precedentemente descritte (ad esempio, la PAPS);

- nel caso di *priorita' con interruzione* (preemptive priority), l'utente che chiede servizio ha immediata precedenza su tutti gli utenti di priorita' minore presenti nel sistema compreso quelli che stanno ricevendo servizio; se uno di questi e' di grado di priorita' minore, il suo servizio viene interrotto e inizia subito il servizio dell'utente con grado di priorita' piu' elevato; anche in questo caso, eventuali condizioni di uguaglianza del grado di priorita' possono essere risolte con una delle discipline di servizio precedentemente descritte.

Sempre con riferimento al caso di priorita' con interruzione e per cio' che riguarda l'utente che ha subito l'interruzione del servizio, questi attende entro il servizio fintantoche' il servente non si rende nuovamente a lui disponibile e potra' vedere accolta la propria richiesta con due alternative:

- nel caso di *priorita' con interruzione e senza ripetizione*, l'utente riceve l'esecuzione della sola parte mancante per il completamento del servizio gia' iniziato in precedenza;
- nel caso di *priorita' con interruzione e con ripetizione*, l'utente riceve la ripetizione completa dell'intero servizio.

Per sistemi multi-serventi ($S > 1$) ad accessibilita' piena o limitata, puo' essere talvolta richiesto di precisare il criterio con cui un utente viene assegnato a uno dei serventi liberi che siano per lui accessibili. Si tratta cioe' di definire la modalita' di ricerca di un servente libero. Al riguardo esistono due modelli significativi:

- la *ricerca sequenziale* (sequential hunting), che presuppone un ordinamento dei serventi, coincidente anche con l'ordine di ricerca; allora lo stato libero o occupato dei serventi viene rilevato partendo dal primo servente e passando poi via via a quelli che lo seguono in detto ordinamento; inoltre un servente non puo' essere impegnato se ne esistono altri liberi che lo precedono nell'ordine della ricerca e la ricerca viene arrestata non appena si incontra un servente libero;
- la *ricerca casuale* (random hunting), in cui la scelta del servente libero e' effettuata a caso.

Come e' ovvio, la ricerca sequenziale comporta una probabilita' di impegno per ognuno dei serventi che risulta decrescente nel senso del loro ordinamento; nella ricerca casuale, detta probabilita' di impegno e' uguale per tutti i serventi del sistema.

II.2.5 Il processo di coda

Il processo aleatorio che risulta dall'interazione dei processi di ingresso e di servizio

- per assegnate caratteristiche strutturali del sistema;
- per una definita composizione della popolazione di utenti potenziali;
- per una fissata disciplina di servizio,

è il *processo di coda*.

Le condizioni di sistema entro le quali si ha l'evoluzione di questo processo sono usualmente descritte tramite una notazione sintetica, che fa riferimento al caso in cui il sistema tratta una sola classe di utenti. Tale notazione utilizza cinque simboli in sequenza. Il *primo* e il *secondo* simbolo si riferiscono, rispettivamente, al processo di ingresso e a quello di servizio e denotano il tipo di distribuzione dei tempi T e Θ , secondo le corrispondenze riportate nella tabella seguente.

Il *terzo*, il *quarto* e il *quinto* simbolo individuano, rispettivamente, il numero S dei serventi, la lunghezza Q della fila d'attesa e la cardinalità U della popolazione di utenti potenziali. Si precisa che gli ultimi due simboli della notazione sono talvolta omessi quando i parametri Q e U sono scelti di valore infinito.

Simbolo	Distribuzione
M	<i>Esponenziale negativa</i>
E_r	<i>Erlangiana di ordine r</i>
H_r	<i>Iper-esponenziale di ordine r</i>
D	<i>Deterministica</i>
G	<i>Generale</i>

Tab. II-1 – Simboli usati nella notazione sintetica.

Per la definizione del processo di coda possono essere considerate, in alternativa, varie famiglie di variabili aleatorie; per ogni famiglia, le variabili componenti sono definite in uno spazio di probabilità. Le famiglie di maggiore interesse sono raggruppabili a coppie, ognuna delle quali è associata ad un ambiente A di posti disponibili all'interno del sistema di servizio.

Le componenti della prima famiglia associata ad un generico ambiente A sono *variabili aleatorie discrete*, che sono contrassegnate dagli istanti di osservazione t del sistema. Pertanto lo spazio di stato del processo di coda è discreto, mentre lo spazio di indice è quello R^+ dei numeri reali positivi: la famiglia in esame costituisce quindi un insieme non numerabile. In particolare,

per uno specifico ambiente A , la variabile aleatoria $N_A(t)$, che è contrassegnata da un fissato valore t ($t \in R^+$) del parametro temporale, esprime il numero d'utenti che sono contemporaneamente presenti all'interno di A in quell'istante di osservazione. Il processo aleatorio $\{N_A(t), t \in R^+\}$ è allora *a stato-discreto e a tempo-continuo*.

La seconda famiglia associata all'ambiente A comprende variabili aleatorie continue, che sono contrassegnate dagli istanti di arrivo t_i ovvero, in termini equivalenti, dai corrispondenti numeri d'ordine i che identificano anche gli utenti che arrivano all'ingresso del sistema in quegli istanti. Pertanto lo spazio di stato del processo di coda è continuo, mentre lo spazio di indice è quello I dei numeri interi naturali: la famiglia in questione è allora un insieme numerabile. In particolare, con riferimento a uno specifico ambiente A , la variabile aleatoria $\Omega_A(i)$, che è contrassegnata da un fissato valore i ($i \in I$) del parametro temporale, esprime il tempo di soggiorno dell'utente i -esimo all'interno di A , e cioè l'intervallo di tempo che intercorre tra l'entrata di questo utente in tale ambiente e la sua uscita. Il processo aleatorio $\{\Omega_A(i), i \in I\}$ è allora *a stato-continuo e a tempo-discreto*.

L'ambiente A può coincidere con l'intero sistema o con la fila d'attesa o con il servizio. In relazione a queste tre possibilità, le variabili aleatorie $N_A(t)$ e $\Omega_A(i)$ verranno qui di seguito particolarizzate con le notazioni precisate nella seguente tabella di corrispondenza:

A	$N_A(t)$	$\Omega_A(i)$
intero sistema	$K(t)$	$\Delta(i)$
fila d'attesa	$L(t)$	$H(i)$
servizio	$M(t)$	$\Theta(i)$

Tab. II-2

Il significato di queste notazioni è immediatamente deducibile dalle definizioni di $N_A(t)$ e di $\Omega_A(i)$ per un ambiente A generico. È opportuno tuttavia aggiungere che, con riferimento all' i -esimo utente ($i \in I$), $\Delta(i)$ è una V.A. denominata *ritardo* (o tempo) *di coda*, mentre $H(i)$ è il *ritardo* (o tempo) *di attesa*, di cui si è parlato genericamente in § II.1.2. Si osserva anche che $\Theta(i)$ è, per l' i -esimo utente ($i \in I$), il relativo tempo di servizio che assume qui il significato di tempo di stazionamento di questo utente all'interno del servizio.

Circa il legame tra le variabili ora introdotte, è ovvio che, per un fissato istante t , deve essere

$$K(t) = L(t) + M(t) \quad (t \in \mathbb{R}^+), \quad (\text{II.2.24})$$

mentre, se ci si riferisce ad un fissato utente i -esimo, deve risultare:

$$\Delta(i) = H(i) + \Theta(i) \quad (i \in I). \quad (\text{II.2.25})$$

Riguardo poi alle famiglie di V.A. $\{K(t), t \in \mathbb{R}^+\}$ e $\{\Delta(i), i \in I\}$ che sono associate all'intero sistema, le distribuzioni di probabilit  delle V.A. appartenenti alla seconda famiglia possono essere dedotte, almeno in linea di principio, da quelle della prima. Cio  consente di risolvere il problema dell'analisi di un sistema di servizio determinando in primo luogo le distribuzioni delle V.A. appartenenti alla famiglia $\{K(t), t \in \mathbb{R}^+\}$. Da queste ultime possono essere poi ricavate anche quelle delle V.A. appartenenti alle altre famiglie precedentemente definite con riferimento a un diverso ambiente A .

Pertanto, nel seguito, rivolgeremo la nostra attenzione solo alla famiglia $\{K(t), t \in \mathbb{R}^+\}$, la cui componente generica, lo ricordiamo, descrive il numero di utenti contemporaneamente presenti all'interno del sistema di servizio in un fissato istante di osservazione t . Ci possiamo domandare sotto quali condizioni questa famiglia puo' definire un processo di Markov.

Al riguardo, si osserva che, per un fissato t , la V.A. $K(t)$ riassume la storia passata del processo di coda solo se si aggiungono informazioni sull'istante di arrivo immediatamente precedente rispetto all'istante t e sugli istanti in cui sono iniziati i servizi in corso di espletamento nello stesso istante t .

Si puo' quindi affermare che il processo di coda puo' essere trattato come markoviano se il suo stato generico all'istante t e' descritto dai seguenti dati:

- il numero $K(t)$ di utenti contemporaneamente presenti all'interno del sistema in quell'istante t ;
- l'eta' dell'ultimo arrivo, e cioe' l'intervallo di tempo tra l'istante dell'arrivo immediatamente precedente e l'istante t ;
- l'eta' di un servizio in corso, e cioe' l'intervallo di tempo tra l'inizio di un servizio in corso e l'istante t ; se $S > 1$, questo dato deve essere fornito per ognuno dei servizi in corso nell'istante di osservazione.

Si osserva che, mentre il primo dato ha valori variabili in un insieme discreto, gli altri due assumono valori in un insieme continuo. Conseguentemente il processo di coda cosi' descritto e' markoviano *a stato-continuo/discreto* e *a tempo-continuo*.

II.2.6 Il concetto di traffico

Nel linguaggio corrente la parola "traffico" e' largamente utilizzata anche in ambienti diversi da quelli qui considerati, ma e' in ogni caso associata a una

descrizione qualitativa o quantitativa di fenomeni di impegno di risorse condivise. A questo tipo di descrizione, ma con scopi di rappresentazione quantitativa, sono dedicate le note seguenti.

Il concetto di traffico e' qui introdotto con riferimento a un contesto modellistico, ma le quantita' rappresentative del traffico, seppure con le limitazioni che verranno esplicitamente segnalate, possono essere utilizzate anche in contesti reali, purché si tenga conto della corrispondenza tra realta' e sua schematizzazione modellistica.

Si consideri la sequenza degli istanti di arrivo $\{t_i, i \in I\}_o$, che e' gia' stata definita in § II.2.3 e che viene qui riproposta con il pedice "o" per sottolineare che il generico elemento componente e' l'istante in cui viene offerta una domanda di servizio da parte della popolazione U verso il sistema S . Questa sequenza di V.A. e' una manifestazione delle esigenze di servizio che sorgono aleatoriamente in U e che non sono normalmente condizionate da S .

Esistono pero' casi in cui questo condizionamento sussiste: cio' si verifica se S e' con perdita e se il processo di ingresso dipende dallo stato del processo di coda (cfr. § II.3.4). Se queste ipotesi sono ambodue verificate, gli eventi di rifiuto che si presentano nell'evoluzione di S provocano una reazione nel comportamento della popolazione U : piu' precisamente, un condizionamento da parte di S sugli istanti di arrivo t_i . Il risultato e' una modifica di $\{t_i, i \in I\}_o$; per sottolineare il condizionamento da parte di S , la sequenza modificata verra' indicata con $\{t_i, i \in I\}_{oS}$.

Alle due sequenze degli istanti di arrivo

$$\{t_i, i \in I\}_o, \quad , \quad \{t_i, i \in I\}_{oS}.$$

- possono essere associati, tramite la (II.2.4), altrettanti processi di ingresso

$$\{T(i), i \in I\}_o, \quad , \quad \{T(i), i \in I\}_{oS}$$

il cui significato e' evidente: il pedice "o" riguarda le offerte da U ad S , mentre l'aggiunta del pedice "S" indica offerte che sono condizionate;

- puo' essere associato, tramite la (II.2.5), lo stesso processo di servizio $\{\Theta(i), i \in I\}_o$; quest'ultimo infatti non subisce modifiche a seguito del condizionamento da parte di S sulla popolazione U .

Nel seguito, per generalita', assumiamo che S sia con perdita e che il processo di ingresso sia $\{T(i), i \in I\}_{oS}$. Con questa ipotesi la coppia di P.A.

$$\{T(i), i \in I\}_{oS} \quad , \quad \{\Theta(i), i \in I\}_o \quad (II.2.26)$$

descrive in modo completo l'evoluzione delle domande di servizio offerte al sistema S da parte della popolazione U .

A questa evoluzione esponiamo un sistema di servizio fittizio S^* con un numero *illimitato* di serventi e consideriamo, all'istante di osservazione t , il numero di serventi che sono contemporaneamente occupati in S^* per soddisfare tutte le domande di servizio presentate dagli utenti in U . Al variare di t , tale numero $Y_o(t)$ è elemento componente di una famiglia di V.A. discrete $\{Y_o(t), t \in R^+\}$, aventi come parametro l'istante di osservazione t . Questa famiglia è un P.A. a stato discreto e a tempo continuo che è chiamato *processo di traffico offerto*, precisando, quando ci sia possibilità di equivoco, che l'offerta è da parte della popolazione U e rivolta al sistema S .

Relativamente a questa definizione, osserviamo che

- il riferimento al sistema fittizio S^* è meno artificioso di quanto non appaia a prima vista: è infatti solo un espediente modellistico per considerare il numero di serventi che sarebbero impegnati nell'ipotesi che ogni domanda di servizio offerta da U ad S venisse accolta senza ritardo;
- il processo di traffico offerto è la rappresentazione sintetica dell'evoluzione delle domande di servizio da U a S , così come descritta dalla coppia di P.A. nella (II.2.26);
- la rappresentazione di cui sopra è effettuata tramite il numero di serventi che sarebbero contemporaneamente impegnati in ogni istante di osservazione, ove si tenga conto di tutte le domande di servizio da U ad S e ove si prescinda dall'esito che tali domande possono incontrare quando sono presentate a S ;
- la famiglia di V.A. $\{Y_o(t), t \in R^+\}$ può essere vista come il processo di coda, che
 - è relativo al sistema S in cui si faccia tendere all'infinito il numero S dei serventi;
 - è indotto dai processi di ingresso e di servizio nella (II.2.26);
 - è rappresentato dalle famiglie di V.A. discrete $\{K(t), t \in R^+\}$ e $\{M(t), t \in R^+\}$, che in questo caso ($S = \infty$) coincidono in base alla definizione datane in § II.2.5;
- il traffico offerto a un insieme di risorse condivise in un contesto reale è la risultante di fenomeni aleatori che *non possono essere osservati*; infatti, mentre le realizzazioni del processo di ingresso sono osservabili senza difficoltà, non altrettanto accade per quelle del processo di servizio; cioè in quanto l'attribuzione di un tempo di servizio a una domanda che non ha modo di esprimersi (in quanto viene rifiutata dal sistema) è possibile solo in un modello matematico.

Ancora nell'ipotesi di sistema S con perdita, si consideri nuovamente la sequenza degli istanti di arrivo $\{t_i, i \in I\}_{oS}$ e si effettui su questa una separazione in due insiemi disgiunti: il primo riguardante gli istanti in corrispondenza a eventi di domanda accolta; il secondo comprendente gli istanti di arrivo rimanenti in corrispondenza dei quali si verifica un rifiuto. Si ottengono in tal modo, dopo opportuna modifica della numerazione degli elementi componenti, due nuove sequenze

$$\{t_i, i \in I\}_s, \quad \{t_i, i \in I\}_p \quad (\text{II.2.27})$$

che sono associate alle domande di servizio *accolte* (pedice " s ") e *rifiutate* (pedice " p "), rispettivamente. A ognuna di queste sequenze puo' essere associata una coppia di processi di ingresso e di servizio. L'associazione e' immediata: se, per fissare le idee, ci riferiamo alla sequenza relativa alle domande di servizio accolte, i tempi di inter-arrivo possono essere ottenuti dalla sequenza $\{t_i, i \in I\}_s$ applicando la (II.2.4), mentre i tempi di servizio sono ricavabili dalla (II.2.5), in cui le quantita' di lavoro da considerare sono solo quelle richieste da utenti che vengono accolti nel sistema.

Si ottengono in tal modo due coppie di P.A.: la prima, che indichiamo con

$$\{T(i), i \in I\}_s, \quad \{\Theta(i), i \in I\}_s \quad (\text{II.2.28})$$

riguarda le domande di servizio accolte; la seconda, denotata con

$$\{T(i), i \in I\}_p, \quad \{\Theta(i), i \in I\}_p \quad (\text{II.2.29})$$

corrisponde alle domande di servizio rifiutate. Si nota che, mentre i due processi di servizio nelle (II.2.28) e (II.2.29) sono ottenuti da $\{\Theta(i), i \in I\}_o$ tramite la stessa separazione che e' stata effettuata su $\{t_i, i \in I\}_{oS}$ per pervenire alle due sequenze in (II.2.26), i due processi di ingresso nelle (II.2.27) e (II.2.28) sono ambodue sostanzialmente diversi da $\{T(i), i \in I\}_{oS}$. dato che sono ottenuti da differenti sequenze dei tempi di arrivo.

A questo punto e' possibile definire il *processo di traffico smaltito*, precisando, quando sussista possibilita' di equivoco, che lo smaltimento e' da parte del sistema S e a favore della popolazione U . Tale processo e' costituito da una famiglia di V.A. $\{Y_s(t), t \in R^+\}$, tra cui quella $Y_s(t)$ rappresenta il numero di serventi nel sistema S contemporaneamente impegnati nell'istante t a seguito delle domande di servizio presentate dagli utenti in U e accolte da S . Come e' evidente, tale P.A. e' la sintesi della coppia di P.A. nella (II.2.28).

La sintesi della coppia di P.A. nella (II.2.29) e' invece il *processo di traffico perduto*, ove, quando sussista possibilita' di equivoco, conviene precisare che la perdita avviene da parte del sistema S e riguarda domande di servizio aventi origine dalla popolazione U . In questo caso la famiglia di V.A. da considerare e'

data da $\{Y_p(t), t \in R^+\}$, ove $Y_p(t)$ e' il numero di serventi nel sistema fittizio S^* (gia' utilizzato per la definizione del processo di traffico offerto) contemporaneamente impegnati nell'istante d'osservazione t a seguito delle domande di servizio presentate dagli utenti in U e rifiutate da S .

A commento della definizione di processo di traffico smaltito, si rileva che

- tale processo e' la rappresentazione sintetica della evoluzione delle domande di servizio che vengono accolte da S , cosi' come descritta dalla coppia di P.A. nella (II.2.28);
- detta rappresentazione e' attuata tramite il numero *effettivo* di serventi che, in ogni istante di osservazione, sono contemporaneamente impegnati in S a seguito delle domande di servizio presentate da U ;
- la famiglia di V.A. $\{Y_s(t), t \in R^+\}$ puo' essere vista come il processo di coda che
 - e' relativo al sistema S ;
 - e' indotto dal processo di ingresso e di servizio nella (II.2.28);
 - e' rappresentato dalla famiglia di V.A. $\{M(t), t \in R^+\}$, definita in § II.2.5;
- il traffico smaltito da un insieme di risorse condivise in un contesto reale e' la risultante di fenomeni aleatori che, a differenza di quelli connessi con il traffico offerto, *possono essere osservati*.

II.3 I fenomeni di traffico in equilibrio statistico

Facciamo riferimento al processo di coda definito in § II.2.5. Dall'analisi di questo P.A., utilizzando la teoria dei processi stocastici markoviani, si possono ricavare, almeno in linea di principio, le distribuzioni di probabilita' delle V.A. componenti la famiglia $\{K(t), t \in R^+\}$. Tali distribuzioni dipendono in generale dal parametro temporale di ogni famiglia e dalle *condizioni iniziali* dell'evoluzione temporale del processo di coda.

Esistono tuttavia situazioni per le quali si possono distinguere due fasi di questa evoluzione temporale:

- una fase iniziale, che e' il *regime transitorio* del sistema di servizio e in cui il processo di coda ha il comportamento generale sopra descritto;
- una fase successiva, che e' il *regime permanente* o la *condizione di equilibrio statistico* del sistema e in cui le distribuzioni del processo di coda diventano indipendenti sia dal parametro temporale, che dalle condizioni iniziali.

Questa seconda fase e' raggiungibile teoricamente solo dopo un tempo infinitamente grande rispetto all'istante iniziale dell'evoluzione, ma in pratica puo' attuarsi in intervalli di tempo finiti.

Inoltre la possibilita' di raggiungere condizioni di equilibrio statistico pone opportuni condizionamenti sui parametri del processo di ingresso e di servizio nell'ambito di assegnate caratteristiche strutturali del sistema. Possibilita' di questo tipo, che corrispondono alla ergodicita' del processo di coda, sono quelle di prevalente interesse nelle applicazioni qui considerate. Pertanto nel seguito ci riferiremo esclusivamente a *processi di coda ergodici*. Cio' ci consentira' di limitare la nostra attenzione all'evoluzione del processo di coda in condizioni di equilibrio statistico.

Alla giustificazione e alle implicazioni di tale ultima affermazione e' dedicato il contenuto di questo paragrafo.

II.3.1 L'ergodicita' del processo di coda

Si consideri il processo di coda $\{K(t), t \in R^+\}$, che per semplicita' supporremo essere di tipo markoviano e di cui indichiamo con $I_K \in I$ il relativo spazio di stato. La caratterizzazione completa di questo P.A. richiederebbe la conoscenza dell'intera gerarchia di probabilita'. Concentriamo pero' l'attenzione sulla gerarchia del primo ordine, e cioe' sulla distribuzione di probabilita'

$$p_K(t, k) = P\{K(t) = k\}, \quad (t \in R^+; k \in I_K). \quad (II.3.1)$$

Il processo di coda e' *ergodico*, se il limite

$$\lim_{t \rightarrow \infty} P\{K(t) = k \mid K(t_0) = j\}, \quad (t, t_0 \in R^+; k, j \in I_K) \quad (II.3.2)$$

esiste, e' finito ed e' indipendente da j per ogni t_0 . Se cio' si verifica, esiste anche il limite

$$\lim_{t \rightarrow \infty} p_K(t, k), \quad (t \in R^+; k \in I_K) \quad (II.3.3)$$

e ha lo stesso valore del limite (II.3.2).

Questo comune valore dei limiti (II.3.2) e (II.3.3)

$$p_K(k) = \lim_{t \rightarrow \infty} p_K(t, k), \quad (k \in I_K) \quad (II.3.4)$$

e' denominato *distribuzione-limite* e i suoi elementi componenti sono chiamati *probabilita'-liite di stato*.

Circa l'indipendenza della distribuzione-limite dall'istante di osservazione, si sottolinea (a scanso di equivoci) che tale indipendenza non riguarda l'evoluzione del P.A. in regime permanente: infatti ciascuna realizzazione del

P.A. varia comunque nel tempo e conserva in ogni istante il suo carattere di V.A.

Esempio II.3-1 - Ci riferiamo a un processo di coda markoviano ergodico, di tipo M/M/1/0/∞ con ingresso poissoniano. Indichiamo inoltre con λ e μ le frequenze medie di arrivo e di servizio, rispettivamente. In queste condizioni, consideriamo la distribuzione di probabilit  $p_K(t, k)$, $k=0,1$, in cui lo stato $k=0$ si riferisce all'evento di servente libero e lo stato $k=1$ a quello di servente occupato.

Esaminiamo l'andamento di questa distribuzione in funzione dell'istante di osservazione t ; come origine dell'evoluzione temporale del processo di coda viene scelto l'istante $t=0$.

Gli andamenti di tale distribuzione, che sono mostrati in Fig. II-4, dipendono dai parametri λ e μ e dalle condizioni iniziali dell'evoluzione. In particolare la Fig. II-4a si riferisce al caso in cui, nell'istante $t=0$, il servente   libero con probabilit  unitaria [$p_K(0,0)=1$; $p_K(0,1)=0$], mentre la Fig. II-4b rappresenta una condizione che   complementare della precedente [$p_K(0,0)=0$; $p_K(0,1)=1$]. Gli andamenti in questione sono descritti analiticamente da

$$p_K(t, 0) = p_{0\infty} + p_{1\infty}e^{-\alpha t}; \quad p_K(t, 1) = p_{1\infty}(1 - e^{-\alpha t})$$

nel caso considerato in Fig. II-4a e da

$$p_K(t, 0) = p_{0\infty}(1 - e^{-\alpha t}); \quad p_K(t, 1) = p_{1\infty} + p_{0\infty}e^{-\alpha t}$$

nel caso di Fig. II-4b. In tali espressioni $p_{0\infty}$, $p_{1\infty}$ e α sono parametri che dipendono unicamente da quelli dei processi di ingresso e di servizio e che sono espressi da

$$p_{0\infty} = \frac{\mu}{\lambda + \mu}, \quad p_{1\infty} = \frac{\lambda}{\lambda + \mu}, \quad \alpha = \lambda + \mu.$$

Dall'esame delle figure II.3.1a e II.3.1b possiamo osservare che, al crescere dell'istante di osservazione t e indipendentemente dalle condizioni iniziali, le due probabilit  $p_K(t, 0)$ e $p_K(t, 1)$ tendono asintoticamente ai valori-limite $p_{0\infty}$ e $p_{1\infty}$, rispettivamente. Questi ultimi rappresentano quindi le probabilit -limite di stato del processo di coda; cio 

$$p_K(0) = p_{0\infty} \quad p_K(1) = p_{1\infty}.$$

La legge di convergenza verso i due asintoti orizzontali   di tipo esponenziale con rapidit  dipendente dal parametro α .

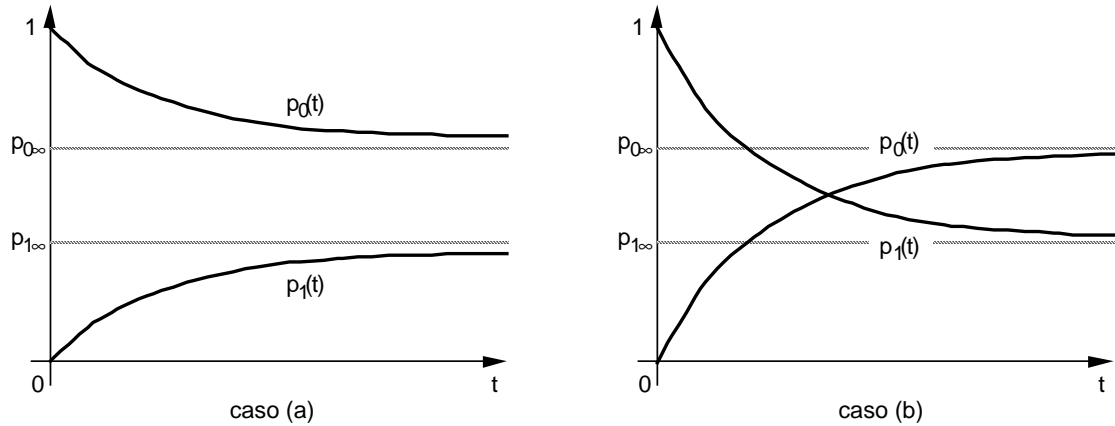


Fig. II-4 - Per l'illustrazione dei regimi transitorio e permanente con differenti condizioni iniziali:

caso a): $p_0(0) = 1$; $p_1(0) = 0$

caso b): $p_0(0) = 0$; $p_1(0) = 1$.

Proprietà fondamentali connesse alla ergodicità sono le seguenti:

- 1) come immediata conseguenza della (II.3.2) e come sottolineato nell'esempio II.3.1, la condizione di equilibrio statistico e la associata distribuzione-limite sono raggiunti, nell'evoluzione di un P.A. ergodico, *indipendentemente dalle condizioni iniziali*, e cioè dal valore assegnato alle probabilità $p_K(0,k)$, $k \in I_K$, avendo fissato l'istante di osservazione iniziale nell'origine dei tempi;
- 2) la distribuzione-limite è anche *stazionaria*: infatti, se le condizioni iniziali $p_K(0,k)$, $k \in I_K$ sono assunte coincidenti con le corrispondenti probabilità-limite, e cioè, se

$$p_K(0,k) = p_K(k) \quad (k \in I_K), \quad (\text{II.3.5})$$

allora

$$p_K(t,k) = p_K(k) \quad (t \in \mathbb{R}^+; k \in I_K); \quad (\text{II.3.6})$$

conseguentemente, con le condizioni iniziali espresse dalla (II.3.5), il P.A. raggiunge senza transitorio la condizione di equilibrio statistico;

- 3) si verifica coincidenza numerica (con probabilità 1) tra le *medie di insieme* e quelle *temporali*: ciò equivale a dire che l'operazione di media effettuata tramite l'osservazione dell'intera evoluzione temporale di una singola realizzazione di un P.A. ergodico porta (con probabilità 1) allo stesso risultato che si ottiene.
 - fissando un istante di osservazione sull'asse dei tempi per l'evoluzione del P.A.;
 - considerando tutte le possibili realizzazioni del P.A. in quell'istante;

- 4) la probabilit -limite di un dato stato k -esimo ($k \in I_K$) del P.A. puo' essere interpretata come la frazione media di tempo che una realizzazione del P.A. trascorre in quello stato; cio' consente di effettuare una stima della generica probabilit -limite $p_K(k)$, ($k \in I_K$)
- osservando una realizzazione del processo che abbia raggiunto in pratica l'equilibrio statistico;
 - misurando e totalizzando, rispetto a una fissata base dei tempi, gli intervalli temporali trascorsi nello stato k -esimo ($k \in I_K$).

Le propriet  ora richiamate possono aiutare a chiarire la significativita' dei processi di coda ergodici nei problemi di analisi e di sintesi riguardanti un insieme infrastrutturale che sia schematizzabile con un sistema di servizio. Al riguardo, le motivazioni per assumere l'equilibrio statistico come modello di condizione operativa ai fini di una valutazione delle prestazioni o di un dimensionamento di progetto sono cosi' riassumibili:

- a) si tratta di un riferimento ben definito, alla cui realizzazione concorrono solo la domanda di servizio e i parametri intrinseci del sistema sotto esame (quali quelli connessi a caratteristiche strutturali e a funzionalita' di controllo); non intervengono invece le modalita' di innesco che determinano l'evoluzione temporale del sistema, almeno nella sua fase transitoria;
- b) il modello e', almeno in alcuni casi, sufficientemente aderente a condizioni operative reali;
- c) la descrizione quantitativa della distribuzione di probabilit  di un P.A. nella fase di transitorio e' normalmente ardua da determinare anche se si introducono ipotesi di modello fortemente semplificatrici; essa offre in ogni caso una tale mole di dati di osservazione da rendere particolarmente complessa una loro utilizzazione;
- d) invece la descrizione probabilistica di condizioni di equilibrio statistico e' relativamente piu' agevole e puo', in alcuni casi significativi, essere ottenuta in forma analitica anche sotto ipotesi non eccessivamente cogenti; inoltre e' possibile sintetizzare le prestazioni del sistema in parametri che sono indipendenti dall'istante di osservazione e quindi piu' agevolmente utilizzabili da parte dell'osservatore.

Vale la pena commentare quanto affermato nei punti a) e b). Identificare l'equilibrio statistico del sistema modellante o modellato come modello di condizione operativa (punto a) assicura riproducibilit  delle condizioni di osservazione e quindi confrontabilit  dei dati di prestazione e di

dimensionamento che si possono ottenere. Confrontabilità dei dati e riproducibilità delle condizioni di osservazione sono d'altra parte obiettivi essenziali per svolgere attività di progettazione e di esercizio.

Circa la riproducibilità delle condizioni di osservazione, ci riferiamo all'elaborazione di un sistema modellante in forma analitica o simulativa e alla verifica di un sistema modellato operante in esercizio. Sia in questo secondo caso, sia in quello della metodologia simulativa è sufficiente esaminare una singola realizzazione del P.A. che è d'interesse. Se la durata dell'osservazione è sufficiente per raggiungere in pratica il regime permanente, si possono ottenere i dati che sono desiderati, e cioè ad esempio:

- probabilità-limite stimate come frequenze di presentazione;
- momenti di insieme con qualunque ordine valutati tramite i corrispondenti momenti temporali.

Le operazioni da effettuare sono di misura e vanno attuate secondo le metodologie tipiche della statistica. La significatività dei dati ottenuti è assicurata dalle considerazioni svolte in precedenza.

Relativamente al punto b), si osserva poi, che l'aderenza tra il modello dell'equilibrio statistico e le condizioni operative reali è assicurata in modo accettabile solo se la domanda di servizio è a sua volta caratterizzabile come stazionaria o quasi-stazionaria. Se ciò non si verifica, come talvolta accade anche in casi molto significativi da un punto di vista applicativo, lo scostamento tra modello e realtà diventa via via maggiore quanto più la domanda di servizio si allontana dalla stazionarietà. Anche in questi casi però si considerano prevalenti le motivazioni di cui in a), c) e d) e si continua normalmente ad assumere l'equilibrio statistico come condizione operativa di riferimento.

II.3.2 *Correnti di traffico in equilibrio statistico*

Ci poniamo nel caso più generale di un *sistema ad attesa con perdita*; sarà poi immediato estendere le conclusioni raggiunte ai casi particolari di sistemi a perdita in senso stretto e a quelli ad attesa senza perdita. Assumiamo che il processo di coda sia in condizioni di equilibrio statistico e analizziamo come si ripartiscono le domande di servizio in base agli eventi di rifiuto e di ritardo.

Si è già sottolineato in § II.2.6 che le domande di servizio offerte possono essere *accolte* ovvero *rifiutate*. Inoltre una richiesta di servizio accolta può a sua volta essere trattata con due diverse modalità: può essere *accolta senza ritardo* quando l'utente viene ammesso direttamente nel servizio oppure può essere

accolta con ritardo, in quanto l'utente viene posto in fila d'attesa prima di essere ammesso nel servizio.

Le domande di servizio offerte e i risultati della loro ripartizione in domande accolte o rifiutate, così come i risultati della ripartizione delle domande accolte danno luogo ad altrettante *correnti di traffico*, che sono descritte graficamente in Fig. II-5.

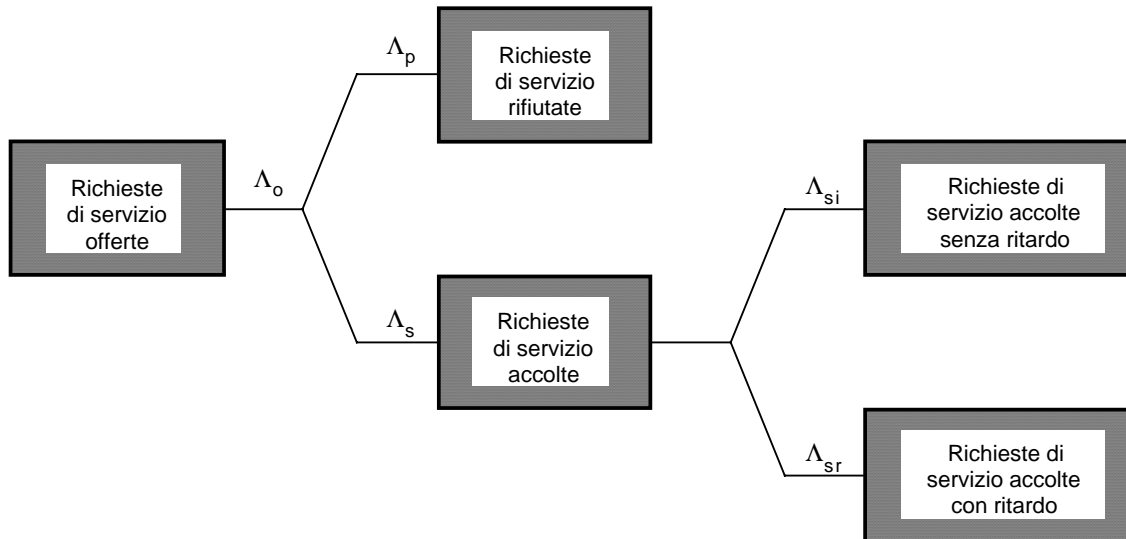


Fig. II-5 - Decomposizione dei flussi di richiesta di servizio in un sistema ad attesa con perdita, operante in condizioni di equilibrio statistico.

Dal punto di vista dei soli momenti del primo ordine, ognuna di tali correnti può essere caratterizzata mediante una grandezza frequenziale media, e cioè mediante il numero medio di domande di servizio che appartengono alla corrente considerata e che vengono trattate nell'unità di tempo. In base all'ipotesi di condizione di equilibrio statistico, queste grandezze medie sono indipendenti dall'istante di osservazione.

Si può allora parlare di *frequenza media delle domande di servizio offerte* o, in termini equivalenti, di *carico medio*. Tale frequenza, indicata con Λ_o , è uguale alla somma di due componenti indicate con Λ_s e Λ_p . La prima di queste rappresenta la *frequenza media delle domande di servizio accolte*, mentre la seconda rappresenta la *frequenza media delle domande di servizio rifiutate*. Si ha quindi

$$\Lambda_o = \Lambda_s + \Lambda_p. \quad (\text{II.3.7})$$

D'altra parte la frequenza Λ_s è uguale alla somma di due ulteriori componenti Λ_{si} e Λ_{sr} , che rappresentano le frequenze medie delle domande di servizio accolte *senza* e *con* ritardo, rispettivamente. Cioè

$$\Lambda_s = \Lambda_{si} + \Lambda_{sr}. \quad (\text{II.3.8})$$

Osserviamo che un elemento distintivo delle condizioni di equilibrio statistico è rappresentato dalla *conservazione delle correnti di traffico* attraverso il sistema. Cioè, ad esempio, la frequenza media degli utenti che entrano nel sistema è uguale a quella degli utenti che escono dal sistema dopo aver ricevuto il servizio richiesto. In base a tale proprietà, la frequenza Λ_s è anche chiamata *portata media* (throughput) del sistema di servizio.

La frequenza Λ_s stabilisce inoltre un importante legame tra i valori medi delle due famiglie di V.A. (una continua e l'altra discreta), che definiscono il processo di coda associato a un ambiente A del sistema di servizio, secondo quanto precisato in § II.2.5. Se si opera in condizioni di equilibrio statistico, tali valori medi sono indipendenti dagli indici temporali delle variabili aleatorie corrispondenti.

Questo legame, che ha validità molto generale e che è noto come *risultato di Little*, stabilisce che, in condizioni di equilibrio statistico, il numero medio $E[N_A]$ di utenti che sono contemporaneamente presenti all'interno dell'ambiente A è uguale al prodotto della frequenza media Λ_s delle domande di servizio accolte e del valore medio $E[\Omega_A]$ del tempo di soggiorno di un utente all'interno dello stesso ambiente A ; si ha cioè

$$E[N_A] = \Lambda_s E[\Omega_A]. \quad (\text{II.3.9})$$

Ad ogni corrente costituita da specifiche domande di servizio può essere associato un processo di traffico che sintetizza l'effetto di impegno di risorse (effettive o fittizie) risultante dalla evoluzione delle domande pertinenti. Ad esempio alla corrente costituita dalle domande di servizio offerte è associabile il processo di traffico offerto.

Nell'ipotesi che i processi di traffico siano ergodici e che si operi in condizione di equilibrio statistico, ogni corrente di traffico può essere descritta in modo completo assegnando l'intera gerarchia di probabilità del processo di traffico ad essa associato. Usualmente però l'assegnazione è limitata alla gerarchia del primo ordine e talvolta anche a quella del secondo ordine. Inoltre spesso ci si limita ai due primi momenti e cioè al valore atteso $E[Y]$ e alla varianza $\text{Var}[Y]$, ambedue indipendenti dall'istante d'osservazione, data la condizione di equilibrio statistico.

Il valore atteso $E[Y]$ è chiamato *intensità media di traffico* ed è indicato con la lettera A , a cui si aggiunge un pedice per indicare la corrente di traffico a cui ci si riferisce. Il risultato di Little ci consente di dare per i parametri A una definizione operativa, che ha validità del tutto generale.

Consideriamo dapprima l'*intensità media di traffico smaltito*, che indichiamo con A_s . Per definizione (cfr. § II.2.6) A_s deve essere uguale al numero medio $E[M]$ di serventi che sono contemporaneamente occupati in condizioni di equilibrio statistico. In base al risultato di Little, A_s deve essere quindi uguale al prodotto della frequenza media delle domande di servizio accolte e del valor medio del tempo di servizio; cioè

$$A_s = \Lambda_s E[\Theta]. \quad (\text{II.3.10})$$

In maniera del tutto analoga può essere ricavata la *intensità media di traffico offerto* A_o . Infatti A_o può essere ottenuto come numero medio di serventi che sarebbero contemporaneamente occupati in condizioni di equilibrio statistico se il numero S dei serventi fosse aumentato senza limiti. Invece l'*intensità media di traffico perduto* A_p può essere espressa come differenza tra A_o e A_s . In definitiva si ha

$$A_o = \Lambda_o E[\Theta]; \quad A_p = \Lambda_p E[\Theta]. \quad (\text{II.3.11})$$

Si osserva che i parametri A_s , A_o e A_p sono anche il risultato di una normalizzazione effettuata sulle frequenze Λ_s , Λ_o e Λ_p , rispettivamente: la grandezza di normalizzazione è la frequenza media di servizio $\mu = 1/E[\Theta]$.

Sulla base di questa interpretazione il parametro, A_s è anche chiamato *portata media normalizzata* del sistema di servizio a cui il parametro si riferisce, mentre l'*intensità* A_o è denominata anche *carico medio normalizzato*. Questo carico è offerto al sistema di servizio e coincide con A_s solo se il sistema è senza perdita.

Tutti i tre parametri di intensità di traffico ora introdotti sono numeri puri: essi sono infatti definiti, in base alle (II.3.10) e (II.3.11), come prodotto di due grandezze, aventi dimensione di un tempo l'una e dell'inverso di un tempo l'altra. Se per la misura di ambedue queste grandezze si sceglie un'identica unità di tempo (secondi, minuti, ore, ecc.), si dice, convenzionalmente, che il valore di A_o , A_s e A_p è espresso in *erlang* (Erl).

Se la caratterizzazione di una corrente di traffico è limitata ai due primi momenti, è d'uso suddividere i processi ad essa associabili in tre classi, basate sul valore che assume il rapporto varianza/valore medio

$$RVR = \frac{Var[Y]}{E[Y]}. \quad (II.3.12)$$

Se $RVR < 1$, si parla di traffico *livellato*; se invece $RVR > 1$ il traffico è detto *a picchi*; se infine $RVR = 1$, il traffico appartiene a una classe che è intermedia tra le due precedenti e che viene qualificato *casuale puro*.

Una corrente di traffico è allora livellata quando l'impegno di risorse ad essa corrispondente ha variazioni nel tempo che sono di valore piccolo rispetto al suo valore medio; al limite se $RVR = 0$, la corrente diventa deterministica.

Nel caso di traffico a picchi, l'impegno di risorse presenta invece sporadiche e subitanee variazioni che sono di valore decisamente maggiore rispetto al suo valore medio.

Infine si può considerare la composizione additiva di due o più correnti di traffico (*correnti componenti*). La *corrente risultante* è data dalla somma delle domande di servizio appartenenti alle correnti componenti. È immediato verificare che l'intensità media del traffico associato alla corrente risultante è sempre uguale alla somma delle intensità media di traffico associate alle correnti componenti. Se invece si fa riferimento alla varianza, l'addittività vale solo se si ipotizza che le correnti componenti siano statisticamente indipendenti.

II.3.3 Caratteristiche prestazionali di un sistema di servizio

Un primo parametro prestazionale che qualifica un sistema di servizio è il *rendimento di utilizzazione* del servente. Tale parametro, indicato nel seguito con ρ , è definito, in condizioni di equilibrio statistico, dal rapporto tra l'intensità media del traffico smaltito e il numero dei serventi, e cioè da:

$$\rho = \frac{A_s}{S}. \quad (II.3.13)$$

Circa questa definizione, vale la pena osservare che, in base a quanto detto in precedenza sul significato di A_s , il rendimento di utilizzazione ρ è uguale al rapporto tra il numero medio $E[M]$ di serventi che sono contemporaneamente occupati in condizioni di equilibrio statistico e il numero S di serventi disponibili nel sistema. Ne consegue che ρ esprime la quota-parte di tempo in cui mediamente ogni servente del sistema risulta impegnato per espletare i servizi richiesti e, come tale, deve essere non superiore all'unità; cioè'

$$\rho \leq 1. \quad (II.3.14)$$

La (II.3.14) senza il segno di uguaglianza è condizione necessaria affinché un sistema ad attesa senza perdita possa operare in condizioni di equilibrio

statistico e cioè affinché il processo di coda sia ergodico. Se si aumenta il rendimento di utilizzazione, si può tendere a un valore unitario, ma questo costituisce un limite non raggiungibile, in quanto al tendere di ρ ad 1, le prestazioni del sistema di servizio degradano in modo intollerabile; ad esempio, il valor medio $E[H]$ del ritardo di attesa tende all'infinito.

Altri due parametri prestazionali tipici di un sistema di servizio ad attesa con perdita, operante in condizioni di equilibrio statistico, possono essere definiti in base alle considerazioni svolte in § II.1.2 e alla definizione delle correnti di traffico effettuata in § II.3.2.

Il primo di questi parametri, indicato con Π_p , è la *probabilità di rifiuto*: è definita come la probabilità dell'evento di rifiuto (cfr. § II.1.2) e può essere interpretata come la frazione delle richieste di servizio offerte che subiscono rifiuto. Il secondo, denotato con Π_r , è invece la *probabilità di ritardo*: è definita come la probabilità dell'evento di ritardo (cfr. § II.1.2) ed è interpretabile come la frazione delle richieste di servizio accolte che subiscono ritardo. Risulta allora:

$$\Pi_p = \frac{\Lambda_p}{\Lambda_p}; \quad \Pi_r = \frac{\Lambda_{sr}}{\Lambda_s}. \quad (\text{II.3.15})$$

Ovviamente la probabilità di rifiuto è nulla nel caso di sistema ad attesa senza perdita, mentre la probabilità di ritardo è identicamente uguale a zero nel caso di sistema a perdita senza attesa.

In generale, i requisiti di prestazione di un sistema di servizio conducono, in relazione alle caratteristiche della domanda, a dimensionare la struttura in modo che tali due probabilità abbiano valori il più possibile contenuti. A parità di Λ_o , la probabilità di rifiuto può diminuire se si aumenta il numero massimo di utenti che possono essere contenuti contemporaneamente all'interno del sistema; d'altra parte, per un fissato valore di Π_p , la probabilità di ritardo può essere ridotta aumentando il numero dei serventi disponibili nel sistema.

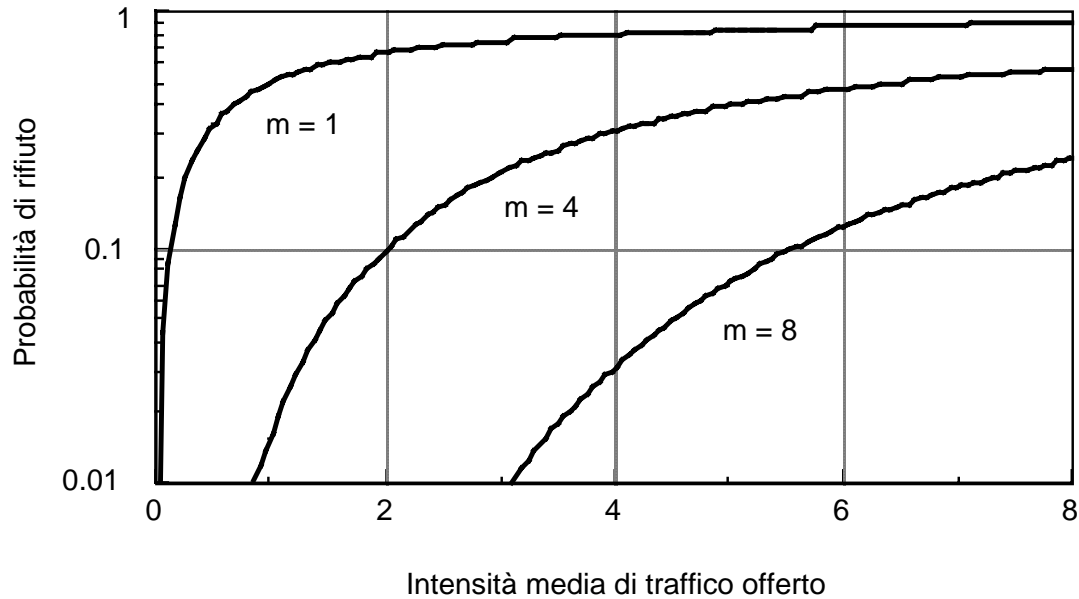


Fig. II-6 - Andamento della probabilità di rifiuto al variare dell'intensità media di traffico offerto, per alcuni valori del numero m di serventi.

Esempio II.3-2 - Analizziamo la risposta di un dato sistema di servizio al variare delle condizioni medie di carico e a parità di tutti gli altri parametri in gioco. Consideriamo pertanto l'andamento della portata media normalizzata A_s al variare del carico medio normalizzato A_o , che è offerto al sistema. Supponiamo, per generalità, che il sistema in esame sia *ad attesa con perdita* e assumiamo condizioni di equilibrio statistico.

Per studiare questo andamento, partiamo da condizioni di sistema scarico ($A_o=0$). All'aumentare di A_o , si può osservare un corrispondente incremento lineare della portata media normalizzata del sistema; risulta infatti $A_s \cong A_o$. Ciò si verifica per tutto il campo di variazione di A_o , in corrispondenza del quale la probabilità di rifiuto ha valore trascurabile.

Se tuttavia A_o aumenta al di sopra di questo campo, la portata media normalizzata del sistema si discosta da A_o (dato che cresce la intensità media A_p) e tende asintoticamente a un valore-limite A_{sM} , che è dato da:

$$A_{sM} = S. \quad (\text{II.3.16})$$

Questo andamento è descritto graficamente in Fig. II-7.

Per giustificare intuitivamente la (II.3.16), basta osservare che il rapporto A_s/S è, in base alla (II.3.13), il rendimento di utilizzazione del servente e non può quindi essere maggiore dell'unità, come espresso dalla (II.3.14). Conseguentemente deve risultare $A_s < A_{sM}$ e pertanto A_{sM} è il limite superiore della portata media normalizzata del sistema.

Si sottolinea che l'andamento di A_s in funzione di A_o descritto in Fig. II-7 è strettamente condizionato dall'aver considerato un sistema ad attesa con perdita. Questa ipotesi introduce implicitamente, all'ingresso del sistema, un controllo di accesso, che consente di commisurare il numero medio degli utenti entranti nell'unità di tempo alla capacità media che il sistema ha di smaltire il carico offertogli. Ciò assicura che il relativo processo di coda sia di tipo ergodico e che l'evoluzione temporale del sistema possa raggiungere condizioni di equilibrio statistico.

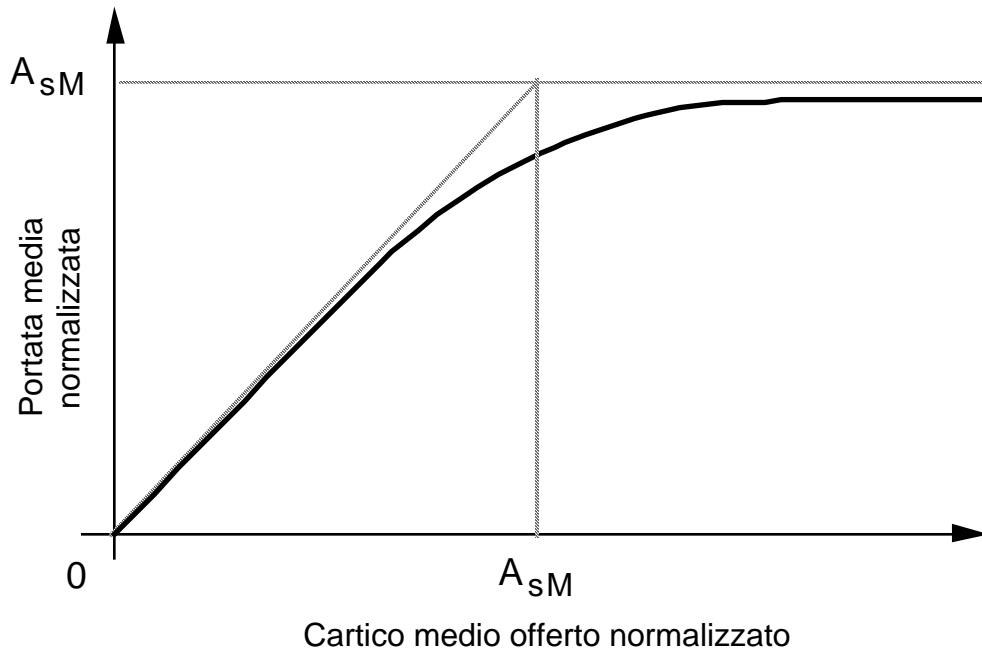


Fig. II-7 - Andamento della portata media normalizzata in un sistema di servizio ad attesa con perdita, al variare del carico medio offerto normalizzato.

Se invece l'accesso non è controllato e cioè se si considera un sistema di servizio ad attesa senza perdita, con una popolazione di utenti potenziali infinitamente grande, lo studio del comportamento della portata media normalizzata A_s all'aumentare di A_o richiede qualche cautela.

In questo caso, infatti, all'aumentare di A_o , la portata media normalizzata presenta sempre il valore-limite dato dalla (II.3.16), ma tale valore non è più un limite asintotico. Esso rappresenta invece il limite superiore del campo di variazione di A_o entro cui il processo di coda è ergodico.

Per convincersene in termini intuitivi, basta osservare che, non verificandosi eventi di rifiuto, i valori di A_o e di A_s coincidono fintantochè $A_o < A_{sM}$. Se invece $A_o \geq A_{sM}$, non sussiste più una conservazione dei flussi entranti ed uscenti. Si verifica allora un incremento indefinito degli utenti contemporaneamente presenti all'interno del sistema. Ciò equivale a dire che il processo di coda ha perso le sue caratteristiche di ergodicità.

In queste condizioni si evidenzia che, in assenza di controlli di accesso, il sovraccarico di un sistema di servizio, al di sopra di opportune condizioni-limite, determina un collasso delle sue potenzialità di smaltimento. Questa, infatti, diminuisce bruscamente rispetto a A_o e tende rapidamente ad annullarsi.

Esempio II.3-3 - Un ulteriore aspetto della risposta di un sistema di servizio al variare delle condizioni medie di carico riguarda l'andamento del valor medio $E[H]$ del ritardo d'attesa in funzione della portata media normalizzata A_s o del rendimento di utilizzazione ρ .]

A tale proposito si può dimostrare, con riferimento ad un sistema *ad attesa senza perdita* e sotto ipotesi molto generali, che $E[H]$ è una funzione monotonamente crescente di A_s o di ρ e tendente asintoticamente all'infinito al tendere di A_s al suo valore-limite dato dalla (II.3.16) ovvero al tendere di ρ al valore unitario.

Questo andamento, che è qualitativamente descritto in Fig. II-8, chiarisce che il

contenimento del valor medio del ritardo di attesa deve essere ottenuto con una limitazione del valore del rendimento di utilizzazione della risorsa.

Circa infine la dipendenza di $E[H]$ dal carico medio normalizzato A_0 , questa è immediatamente deducibile dalla combinazione dei legami funzionali rappresentati nelle Fig. II-7 e Fig. II-8.

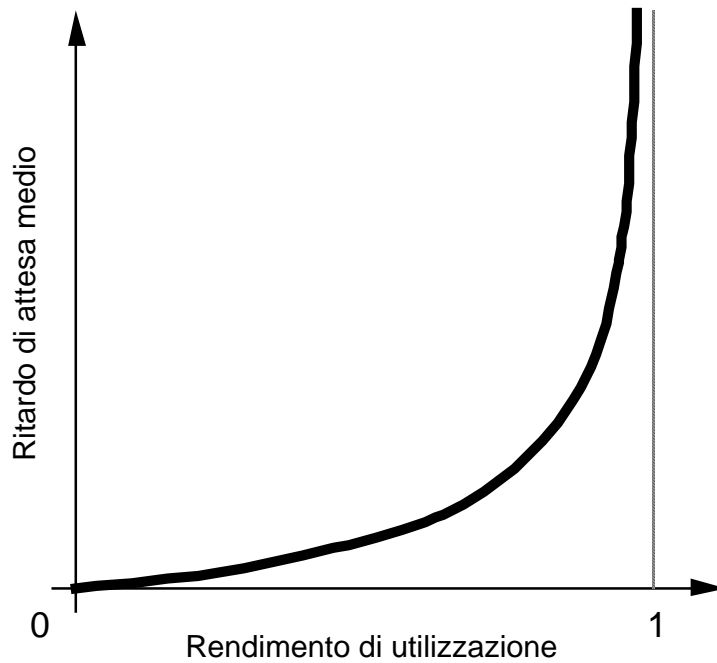


Fig. II-8 - Andamento del valor medio del ritardo di attesa in funzione del rendimento di utilizzazione

II.3.4 Code con ingresso e servizio markoviani

Si sono già discusse le proprietà del processo di coda con riferimento alle modalità di descrizione di un suo stato generico e a una sua tipizzazione markoviana (cfr. § II.2.5). Esistono però casi particolari (ma molto significativi da un punto di vista applicativo), in cui lo stato del processo di coda è descritto in modo completo dal solo numero $K(t)$ di utenti contemporaneamente presenti all'interno del sistema nell'istante di osservazione. Ciò si verifica quando entrambi i processi di ingresso e di servizio sono markoviani, e cioè caratterizzati da distribuzioni esponenziali negative. In questi casi infatti, tenuto conto dell'assenza di memoria che è tipica per questi P.A., sono superflui i dati relativi all'età dell'ultimo arrivo e dei servizi in corso.

Più in particolare, supponiamo che il numero intero naturale k rappresenti la generica realizzazione del numero $K(t)$ di utenti contenuto nel sistema all'istante di osservazione $t \in \mathbb{R}^+$: k può quindi assumere il valore 0 (sistema

vuoto) e variare inoltre in un insieme $I_K \in I$ che, in relazione alla capienza del sistema di servizio, ha cardinalità $\omega = |I_K|$ finita o infinita. Si può allora dimostrare che

- 1) se i tempi di inter-arrivo sono equidistribuiti e statisticamente indipendenti con comune distribuzione esponenziale negativa avente parametro $\lambda(k)$ eventualmente dipendente da k ; cioè, se

$$f_T(t) = \lambda(k) \exp[-\lambda(k)t] \quad (t \in \mathbb{R}^+; k \in 0 \cup I_K); \quad (\text{II.3.17})$$

- 2) se i tempi di servizio sono equidistribuiti e statisticamente indipendenti con comune distribuzione esponenziale negativa avente parametro $\mu(k)$ eventualmente dipendente da k ; cioè se

$$f_\Theta(t) = \mu(k) \exp[-\mu(k)t] \quad (t \in \mathbb{R}^+; k \in I_K); \quad (\text{II.3.18})$$

- 3) se i processi di ingresso e di servizio, definiti come in 1) e in 2), sono mutuamente indipendenti,

allora il processo di coda $\{K(t), t \in \mathbb{R}^+\}$ che (secondo le notazioni in § II.2.5 e in base alle ipotesi in 1. e in 2.) è di tipo M/M, è una particolare catena di Markov a tempo continuo e a tempo omogeneo, per la quale $0 \cup I_K$ è lo spazio di stato.

La particolarità risiede nel fatto che, se si ordinano gli stati del processo di coda con il numero k , le transizioni da uno stato all'altro possono avvenire solo tra stati adiacenti. Le relative frequenze di transizione di stato, che sono indipendenti dall'istante di osservazione t (proprietà di tempo-omogeneità), sono immediatamente deducibili dalle (II.3.17) e (II.3.18).

Si tratta quindi di un *processo di nascita e di morte*, nel quale le frequenze di transizione dallo stato di numero d'ordine k verso i due stati adiacenti con numeri d'ordine $k+1$ e $k-1$ sono date da $\lambda(k)$ (*frequenza di nascita*) e da $\mu(k)$ (*frequenza di morte*) a seconda che la transizione corrisponda a un evento di arrivo e a uno di terminazione di servizio, rispettivamente.

Si può ulteriormente dimostrare che il processo di coda così definito è ergodico senza condizionamenti se lo spazio di stato I_K ha cardinalità finita. Se invece questa cardinalità è infinita, la ergodicità vale se si stabiliscono opportune condizioni sull'intero insieme delle frequenze di nascita $\{\lambda(k), k \in 0 \cup I_K\}$ e di morte $\{\mu(k), k \in I_K\}$.

D'altra parte, le probabilità-limite di stato di un processo di nascita e di morte a tempo-continuo e a tempo-omogeneo, se esistono come avviene nel caso di ergodicità, possono essere determinate in forma chiusa.

Si conclude allora che:

- a) l'evoluzione di un processo di coda di tipo M/M in condizioni di equilibrio statistico puo' essere analizzata senza condizionamenti quando la capienza $S+Q$ del sistema e' limitata;
- b) quando invece tale capienza $S+Q$ e' illimitata, l'analisi di un processo di coda di tipo M/M richiede preliminarmente di accertarne l'ergodicita' o meno; se il processo e' ergodico, la sua evoluzione in condizioni di equilibrio statistico puo' essere analizzata come nel caso di cui in a).

Consideriamo allora un sistema di servizio avente capacita' finita, quale risulta da un numero S di serventi e da una lunghezza Q della fila d'attesa entrambi finiti. In questo caso le probabilita'-limite di stato sono ottenibili dalla seguente equazione di ricorrenza

$$\mu(k)p_K(k) = \lambda(k-1)p_K(k-1). \quad (k \in I_K), \quad (\text{II.3.19})$$

che risolta fornisce

$$p_K(k) = \frac{A(k)}{M(k)} p_K(0) \quad (k \in I_K), \quad (\text{II.3.20})$$

ove

$$A(k) = \prod_{i=1}^k \lambda(i-1) \quad (k \in I_K)$$

$$M(k) = \prod_{i=1}^k \mu(i) \quad (k \in I_K)$$

e, ponendo $\omega = I_K = S+Q$,

$$p_K(0) = \left[1 + \sum_{i=1}^{\omega} \frac{A(i)}{M(i)} \right]^{-1}. \quad (\text{II.3.21})$$

Se invece la capienza del sistema e' infinita ($\omega=\infty$), la (II.3.20) ha ancora il significato di distribuzione-limite di stato del processo di coda $\{K(t), t \in R^+\}$ se la serie

$$1 + \sum_{i=1}^{\infty} \frac{A(i)}{M(i)} = 1 + \sum_{i=1}^{\infty} \prod_{k=1}^i \frac{\lambda(k-1)}{\mu(k)} \quad (\text{II.3.22})$$

converge. Questa e' quindi la condizione necessaria affinche' un processo di coda M/M ammetta condizioni di equilibrio statistico.

II.3.5 Il servizio singolo con distribuzione generale e con vacanze

.....

II.4 Modelli per l'accesso a risorse multiple: le reti di code

Quando, come accade in vari contesti applicativi, sono coinvolte risorse (non necessariamente omogenee) tra loro cooperanti in azioni concatenate e asincrone, i problemi di contesa e di assegnazione che ne sorgono possono essere affrontati con modelli piu' complessi (rispetto a quelli finora considerati), e cioe' con le *reti di code*.

Una rete di code e' un sistema stocastico, che puo' o meno avere relazioni con il suo esterno. Esso puo' essere descritto innanzitutto dalla sua *struttura topologica*, ove i nodi sono rappresentativi di sistemi di servizio, mentre i rami sono le vie percorribili dagli *utenti della rete*, che si trasferiscono sia tra un nodo e l'altro della rete, sia dall'esterno verso la rete e viceversa; questo secondo caso si verifica solo quando la rete ha relazioni con l'esterno.

Rispetto al modello di Fig. II-1, un sistema di servizio che e' nodo di una rete di code presenta due importanti diversita' per cio' che riguarda il suo ingresso e la sua uscita. Infatti, con riferimento a una rete di code che ha relazioni con l'esterno, l'ingresso del nodo e', in generale, la *combinazione* di due componenti di traffico:

- le *correnti di traffico esterne*, che corrispondono, come nel modello di Fig. II-1, a utenti che entrano nella rete dall'esterno di questa;
- le *correnti di traffico interne*, che sono in relazione a utenti che si spostano da un nodo all'altro della rete.

All'uscita di un nodo di rete si ha poi, sempre in generale, la *ripartizione* della relativa corrente di traffico secondo due o piu' direttrici verso l'interno e/o verso l'esterno della rete. Tale ripartizione e' effettuata in base a una opportuna *regola di instradamento*.

Circa l'esistenza o meno di relazioni con l'esterno si possono distinguere *reti chiuse* e *reti aperte*. Nel caso di reti chiuse, il numero degli utenti all'interno della rete, in un qualsiasi istante di osservazione, e' costante: cioe' nessun nuovo utente puo' essere accolto all'interno della rete ovvero uscirne. Nel caso di reti aperte, invece, il numero di utenti all'interno della rete e' variabile: cioe' gli utenti possono sia entrare nella rete che uscirne.

La descrizione di una rete di code richiede la specificazione dei seguenti dati:

- la struttura topologica, che e' descritta da un *grafo orientato*; questo comprende l'insieme $V = \{1, 2, \dots, N\}$ dei nodi della rete e quello $E = \{(i, j), i \in V, j \in V\}$ dei rami; per l'insieme E si intende che il ramo orientato (i, j) esiste solo nel caso in cui la geometria della rete preveda la possibilita' di

una via diretta tra l'uscita del nodo i -esimo e l'ingresso di quello j -esimo. Inoltre, da un punto di vista notazionale e limitatamente alle reti di code aperte, e' comodo includere nell'insieme V un ulteriore nodo fittizio $(N+1)$ -esimo, che indica l'esterno della rete;

- la *popolazione di utenti potenziali*, che puo' o meno essere suddivisa in classi;
- le *caratteristiche strutturali* di ogni sistema di servizio, e cioe' il numero dei serventi e la lunghezza della fila d'attesa per ognuno dei nodi della rete;
- le *caratteristiche della domanda di servizio*, e cioe', per ognuno dei nodi della rete, il processo di servizio e il processo di ingresso dall'esterno della rete; come e' evidente, il secondo di questi processi deve essere considerato nel solo caso di reti di code aperte;
- la *disciplina di servizio*, per ognuno dei nodi della rete; tale disciplina puo' rientrare in uno dei criteri descritti in § II.2.4;
- la *regola di instradamento*, e cioe' il percorso che un utente puo' seguire all'interno della rete.

A tale ultimo riguardo, questo percorso e' descritto da una sequenza di V.A. $\{G(k); k \in I\}$, in cui k e' un parametro temporale e il valore assunto dalla V.A. $G(k)$ rappresenta il nodo visitato al passo k -esimo da un utente nel suo cammino attraverso la rete; ad esempio, se $G(k) = i$, cio' significa che il nodo visitato al passo k -esimo e' quello di numero d'ordine i ; se invece $G(k) = N+1$, come puo' verificarsi nel caso di reti aperte, si e' in presenza, sempre al passo k -esimo, di una uscita verso l'esterno.

Nel seguito verranno dapprima descritte sinteticamente le proprieta' di due modelli semplici, che, sulla base di particolari scelte per i dati descrittivi di una rete di code, consentono di analizzare l'evoluzione temporale del cosiddetto *processo di rete*. Tali modelli sono quello di Jackson (§ II.4.1) e quello di Gordon & Newell (§ II.4.2) per le reti chiuse.

II.4.1 *Modello di Jackson*

Con riferimento a una rete aperta, le ipotesi del modello di Jackson sono:

- 1) la topologia della rete e' di qualsivoglia ordine di complessita', ma *completamente connessa*, in modo tale che sia possibile, partendo da qualunque nodo, raggiungere un qualsiasi altro nodo della rete ovvero uscire verso l'esterno;
- 2) esiste una sola classe di utenti potenziali;

- 3) le correnti di utenti provenienti dall'esterno e dirette verso i singoli nodi della rete sono descritte da processi di sola nascita e sono mutuamente indipendenti; se si indica con λ_i la frequenza media di arrivo dall'esterno nel nodo i -esimo e se conseguentemente

$$\lambda = \sum_{i=1}^N \lambda_i \quad (\text{II.4.1})$$

e' la frequenza media di arrivo dall'esterno nell'intera rete, non si esclude che λ sia un parametro dipendente dal numero M complessivo di utenti nel sistema; nel seguito ci si limitera' a considerare il caso in cui i processi di ingresso dall'esterno sono poissoniani e cioe' indipendenti da M ;

- 4) per il nodo i -esimo, che si assume senza perdita, la sequenza dei tempi di servizio $\{\theta_i(k), k \in I\}$, relativi agli utenti che passano attraverso quel nodo, e' costituita da V.A. statisticamente indipendenti, non negative e distribuite con una legge comune che e' esponenziale negativa; il parametro di questa distribuzione e' in generale dipendente dal numero n di utenti nel nodo in questione e come tale verra' denotato con $\mu_i(n)$; inoltre le V.A. $\theta_i(h)$ e $\theta_j(k)$ sono statisticamente indipendenti per ogni coppia $i, j \in V$ e per $h, k \in I$;
- 5) la disciplina di servizio puo' essere qualunque purché di tipo conservativo;
- 6) per la regola di instradamento, la sequenza $\{G(k), k \in I\}$ e' una catena di Markov tempo-omogenea a tempo-discreto, che si assume *irriducibile*; in particolare le relative probabilita' di transizione sono date da

$$P\{G(k+1) = j | G(k) = i\} = r_{ij}; \quad 1 \leq i, \quad j \leq N+1; \quad k \geq 1, \quad (\text{II.4.2})$$

mentre, per la distribuzione delle probabilita' iniziali ($k = 1$), si ha

$$P\{G(1) = i\} = q_i; \quad 1 \leq i \leq N. \quad (\text{II.4.3})$$

Inoltre si pone anche

$$\begin{aligned} r_{i,N+1} &= 1 - \sum_{j=1}^N r_{ij} = \hat{r}_i; & 1 \leq i \leq N, \\ r_{N+1,N+1} &= 1 \\ r_{N+1,i} &= 0; & 1 \leq i \leq N. \end{aligned} \quad (\text{II.4.4})$$

Secondo la (II.4.3), gli utenti esterni si dirigono con probabilita' costante q_i verso il nodo i -esimo; cio', in base alle notazioni di cui all'ipotesi 3), equivale a dire che la frequenza media di arrivo dall'esterno λ_i nel nodo i -esimo e' uguale a $q_i \lambda$.

Interpretando poi la (II.4.2), r_{ij} ($1 \leq i \leq N$; $1 \leq j \leq N+1$) e' la probabilita' costante che un utente, avendo terminato il suo servizio nel nodo i -esimo, si diriga verso il nodo j -esimo se $1 \leq j \leq N$ ovvero verso l'esterno se $j = N+1$.

Infine, dalle (II.4.4) si vede che il nodo fittizio $(N+1)$ -esimo e' *assorbente*. Cio' equivale a dire che, una volta uscito dalla rete, un utente non puo' farvi ritorno.

Con le ipotesi 1) ÷ 6), si puo' dimostrare che il processo di rete e' di natura vettoriale N -dimensionale ed e' di nascita e morte a tempo-omogeneo e a tempo-continuo. Il suo stato nell'istante di osservazione t e' completamente definito dal vettore N -dimensionale

$$\mathbf{K}(t) = [K_1(t), K_2(t), \dots, K_N(t)], \quad (\text{II.4.5})$$

in cui l'elemento i -esimo rappresenta il numero di utenti all'interno del nodo i -esimo nell'istante t .

Il P.A. $\{\mathbf{K}(t), t \in \mathbb{R}^+\}$ puo' ammettere condizioni di equilibrio statistico. Se indichiamo con

$$p(\mathbf{k}, t) = P\{\mathbf{K}(t) = \mathbf{k}\} \quad (t \in \mathbb{R}^+), \quad (\text{II.4.6})$$

la probabilita' che la rete sia nello stato $\mathbf{k} = [k_1, k_2, \dots, k_N]$ all'istante t , la distribuzione di regime permanente (probabilita'-limite di stato)

$$p(\mathbf{k}) = \lim_{t \rightarrow \infty} p(\mathbf{k}, t) \quad (\text{II.4.7})$$

se esiste, e' unica e puo' essere posta in forma chiusa. Per presentarla, introduciamo le seguenti notazioni:

\mathbf{R} e' una matrice quadrata N -dimensionale, chiamata *matrice di instradamento*; il suo elemento (i, j) e' la probabilita' r_{ij} definita dalla (II.4.2); inoltre la somma degli elementi della sua riga generica i -esima e' uguale a

$$\sum_{i=1}^N r_{ij} = 1 - \hat{r}_j \quad (j \in I) \quad (\text{II.4.8})$$

in accordo con la prima delle (II.4.4);

$\lambda = [\lambda_1, \lambda_2, \dots, \lambda_N]$ e' un vettore riga N -dimensionale, il cui elemento i -esimo esprime la frequenza media di arrivo dall'esterno nel nodo i -esimo;

$\Lambda = [\Lambda_1, \Lambda_2, \dots, \Lambda_N]$ e' un vettore riga N -dimensionale, il cui elemento i -esimo esprime la frequenza media di arrivo complessiva nel nodo i -esimo ovvero la portata media di questo; tale vettore e' ottenibile come soluzione della seguente equazione vettoriale algebrica *non omogenea*:

$$\Lambda = \lambda + \Lambda \mathbf{R}; \quad (\text{II.4.9})$$

$M_i(k_i)$ e' una quantita' dipendente dai parametri $\mu_i(n)$ per tutti i valori di n compresi tra 1 e la componente i -esima della realizzazione \mathbf{k} del vettore di stato; cioe'

$$M_i(k_i) = \prod_{n=1}^{k_i} \mu_i(n). \quad (\text{II.4.10})$$

Si ha allora

$$p(\mathbf{k}) = \prod_{i=1}^N p_i(k_i), \quad (\text{II.4.11})$$

ove

$$p_i(k_i) = \frac{\Lambda_i^{k_i}}{M_i(k_i)} p_i(0), \quad (\text{II.4.12})$$

in cui

$$[p_i(0)]^{-1} = \sum_{n=0}^{\infty} \frac{\Lambda_i^n}{M_i(n)}. \quad (\text{II.4.13})$$

Il processo di rete ammette condizioni di equilibrio statistico se risulta

$$\sum_{n=0}^{\infty} \frac{\Lambda_i^n}{M_i(n)} < \infty. \quad (\text{II.4.14})$$

Si osserva innanzitutto che la (II.4.11) e' in forma prodotto e puo' essere cosi' interpretata: la distribuzione congiunta $p(\mathbf{k})$ e' scomponibile nel prodotto di N distribuzioni marginali. La i -esima di queste e' la distribuzione in regime permanente di un processo di coda derivante da un processo di ingresso poissoniano con frequenza di arrivo Λ_i e da un processo di servizio con distribuzione esponenziale negativa avente parametro $\mu_i(n)$, che dipende dal numero di utenti contenuto nel nodo i -esimo.

Cio' consente di affermare che ogni nodo della rete si comporta come se fosse indipendente dagli altri nodi e come se, oltre a un servizio di tipo esponenziale negativo (secondo l'ipotesi 4)), il suo ingresso fosse poissoniano. Il carattere poissoniano degli ingressi e' pero' solo apparente, dato che, come si puo' dimostrare, il processo di ingresso nei nodi della rete aperta sopra considerata non e' in generale poissoniano.

Per concludere vale la pena osservare che

$$V_i = \frac{\Lambda_i}{\lambda}. \quad (\text{II.4.15})$$

e' il valor medio del numero di visite che un utente effettua al nodo i -esimo durante la sua permanenza all'interno della rete.

II.4.2 Modello di Gordon & Newell

.....
.....

II.5 Traffico e prestazioni nell'interazione tra attività e risorse

In questo paragrafo, sulla base dei contenuti dei paragrafi II.2 e II.3, viene completata la trattazione iniziata nel paragrafo II.1. In particolare si effettua dapprima (§ II.5.1) un confronto tra soluzioni con risorse indivise e con risorse condivise. Si passa poi (§ II.5.2) a chiarire il concetto di risorsa virtuale, che ha particolare importanza nella trattazione del capitolo IV. Con riferimento a un ambiente di risorse condivise, vengono infine presentati due ulteriori compiti delle attività di gestione citati in § II.1.1, e cioè l'attuazione delle strategie di assegnazione (§ II.5.3) e il controllo dei sovraccarichi (§ II.5.4).

Elemento comune di questi argomenti è il legame tra fenomeni di traffico nell'interazione tra attività e risorse da un lato e le conseguenti prestazioni di efficienza e di accessibilità dall'altro.

II.5.1 Campi di impiego delle risorse indivise e di quelle condivise

E' naturale domandarsi a questo punto quali siano i campi di conveniente impiego delle risorse indivise e di quelle condivise. Per rispondere a questa domanda si farà riferimento ai parametri che sono stati definiti nel paragrafo II.3 per qualificare il traffico e le prestazioni derivanti dall'interazione tra attività di utilizzazione e risorse condivise. Questi parametri, con opportune interpretazioni, sono infatti utilizzabili anche nel caso della interazione tra una risorsa indivisa e l'attività di utilizzazione che ne ha disponibilità esclusiva.

Circa la caratterizzazione della domanda in ambiente di risorse condivise, si è già visto che questa può essere effettuata, in termini medi e in condizioni di equilibrio statistico, tramite il carico medio. In modo analogo, la portata media può essere considerata, sempre in termini medi e in condizioni di equilibrio statistico, la risposta di una risorsa condivisa con i suoi controlli di accesso.

Analoghe definizioni possono anche essere date per una risorsa indivisa, con la precisazione che carico medio e portata media hanno valori uguali almeno in un campo di variazione che non superi il limite posto dalla capacità della risorsa.

Relativamente poi ai parametri prestazionali, il rispetto del vincolo di costo indica, come significativo parametro prestazionale di una risorsa indivisa o condivisa, il suo rendimento di utilizzazione, che misura, in termini medi e in condizioni di equilibrio statistico, l'efficienza di impiego della risorsa. L'obiettivo prestazionale consiste allora nell'assicurare, per ogni risorsa, un elevato valore del rendimento di utilizzazione nei limiti del campo di variazione di questo.

Circa poi il rispetto del vincolo di qualità di servizio, si può fare generico riferimento al *livello di accessibilità*, come misura delle limitazioni all'accesso poste dalla risoluzione delle condizioni di contesa. L'obiettivo è quello di garantire, per una risorsa, un livello di accessibilità, che sia sufficientemente elevato in relazione alle esigenze connesse all'evoluzione di ognuna delle attività di utilizzazione richiedenti quella risorsa. Questo obiettivo è ovviamente soddisfatto senza limitazioni nel caso di risorsa indivisa, dato che l'accesso a questa è senza contese.

Nel caso invece di risorsa condivisa, il livello di accessibilità ha limitazioni in relazione agli eventi connessi alla modalità di risoluzione delle condizioni di contesa: una sua misura a priori può, ad esempio, essere data (cfr. § II.3.3) da:

- la probabilità di ritardo Π_r nel caso di modalità orientata al ritardo;
- la probabilità di rifiuto Π_p se la modalità è orientata alla perdita;
- entrambe queste probabilità se ci si riferisce a contese risolte con modalità orientata al ritardo con perdita.

Conseguentemente, un aumento del livello di accessibilità per risorse condivise può essere conseguito mediante interventi tali da ridurre il valore di una o di entrambe le probabilità di ritardo e di rifiuto. Si tratta però, in generale, di un obiettivo contrastante con quello di un aumento del rendimento di utilizzazione. Infatti, a questo aumento corrisponde anche una crescita delle probabilità Π_r e Π_p .

Quanto ora premesso consente di pervenire a qualche prima conclusione circa il confronto tra una soluzione con risorse indivise e una con risorse condivise.

Se una risorsa è indivisa, il carico medio che la riguarda è determinato dalle sole richieste dell'attività di utilizzazione che ne ha disponibilità esclusiva. Invece, il carico medio su una risorsa condivisa è la risultante delle richieste presentate da tutte le attività di utilizzazione che hanno accesso alla risorsa.

Circa la portata media, questa, nel caso di risorsa indivisa, può risultare decisamente inferiore alla relativa capacità se, come talvolta si verifica, la domanda comporta un basso valore del carico medio. Inoltre non possono

presentarsi condizioni di sovraccarico della risorsa, se l'attività che ne ha disponibilità esclusiva contiene la sua domanda entro limiti compatibili con la capacità della risorsa. Una situazione contraria si presenta invece nel caso di risorse condivise: qui, ancora in relazione alla modalità di risoluzione delle condizioni di contesa, può intervenire l'esigenza di controllare i fenomeni di sovraccarico.

Nel caso di soluzione con accesso indiviso, il rendimento di utilizzazione di una risorsa può essere di piccolo valore in relazione alle caratteristiche della domanda dell'unica attività avente diritto di accesso. Quindi, dato che è invece senza limitazioni il livello di accessibilità, la soluzione con risorse indivise dovrà essere ristretta a quei casi in cui sia soddisfatta almeno una delle condizioni seguenti:

- sia assicurato un elevato carico medio da parte dell'attività avente accesso indiviso;
- risulti fortemente prevalente l'esigenza di conseguire un accesso senza contese e, quindi, senza ritardi o senza rifiuti.

Nel caso invece di soluzione con accesso condiviso, il rendimento di utilizzazione di una risorsa può essere di valore più soddisfacente rispetto alla soluzione precedente. Ma, corrispondentemente, il livello di accessibilità si riduce e in misura tanto maggiore quanto più elevato è il rendimento di utilizzazione della risorsa. Questa soluzione è quindi, in generale, preferibile alla precedente per il rispetto del vincolo di costo, ma richiede, per il rispetto del vincolo di qualità di servizio, una attenta soluzione dei relativi problemi di dimensionamento e di gestione delle risorse.

II.5.2 Risorse virtuali

In un ambiente di condivisione, una *risorsa fisica* R condivisa, sotto il controllo di una attività di gestione, può essere impegnata, in intervalli di tempo distinti, da parte di più attività di utilizzazione. Si supponga che la gestione di R sia tale che la domanda di ognuna di queste attività sia pienamente soddisfatta e che il livello di accessibilità sia rispondente alle esigenze connesse all'evoluzione di ogni attività di utilizzazione.

In queste condizioni, ciascuna delle attività di utilizzazione, ad esempio, la i -esima, dato che non percepisce limitazioni al soddisfacimento della sua domanda e alle sue esigenze di livello di accessibilità, utilizza la risorsa fisica R come se questa fosse a lei riservata, e cioè come se fosse una risorsa R_i a sua disposizione esclusiva.

Tale tipo di disponibilità è però solo apparente, in quanto la risorsa R_i è solo l'immagine di R vista dalla i -esima attività di utilizzazione con il condizionamento determinato dal suo traffico offerto e dalle sue esigenze di livello di accessibilità. Per sottolineare tale distinzione, l'immagine R_i di R è chiamata *risorsa virtuale associata alla i -esima attività di utilizzazione*. In Fig. II-9 è illustrata la corrispondenza tra la risorsa fisica R e le risorse virtuali associate alle n attività di utilizzazione che accedono a R .

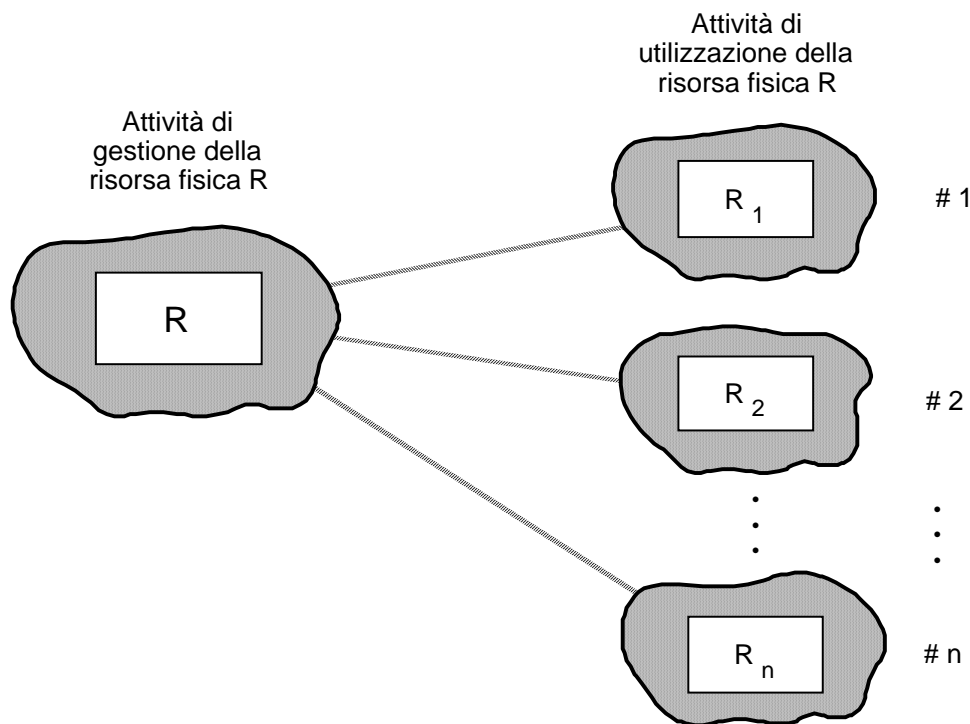


Fig. II-9 - Corrispondenza tra la risorsa fisica R condivisa e le relative risorse virtuali R_i ($i = 1, 2, \dots, n$)

Si rimarca che l'immagine di cui si parla è di tipo condizionato dalla domanda offerta e dalle prestazioni desiderate da parte di una particolare attività di utilizzazione. Pertanto, è possibile parlare di risorsa virtuale per detta attività solo con la precisazione a priori di questi dati di richiesta e solo dopo la verifica a posteriori della possibilità di soddisfare quanto richiesto nell'interazione tra la risorsa fisica R e le attività che accedono ad essa.

II.5.3 Strategie di assegnazione di risorse condivise

Le strategie di assegnazione definiscono le modalità di accesso a un insieme di risorse condivise da parte di un gruppo di attività di utilizzazione

sotto il controllo delle attività di gestione. Possono attuarsi secondo le alternative schematizzate in Fig. II-10.

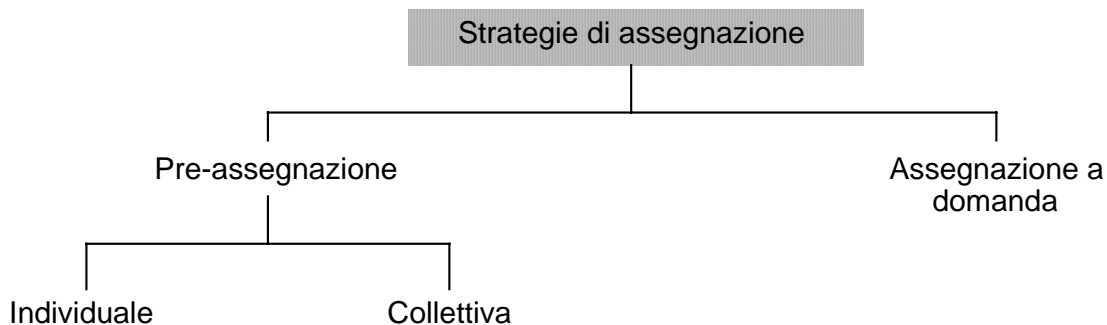


Fig. II-10 - Alternative di attuazione per le strategie di assegnazione delle risorse.

Nella strategia di *assegnazione a domanda*, una risorsa viene assegnata ad una attività di utilizzazione solo a seguito di una richiesta di servizio effettuata immediatamente a monte di un intervallo di vitalità e viene restituita non appena questo intervallo è terminato. Cioè, nel corso della durata di una attività di utilizzazione, questa può impegnare la risorsa più volte per intervalli di tempo minori di detta durata e, negli intervalli residui, la risorsa può essere impegnata da altre attività. Ciò può assicurare, per la risorsa in questione, un rendimento di utilizzazione anche elevato, a spese però di un livello di accessibilità che può risultare insufficiente.

Quando i requisiti prestazionali relativi al livello di accessibilità sono più stringenti è preferibile adottare differenti strategie di assegnazione, e cioè quelle di *pre-assegnazione*. In questo caso una attività di utilizzazione presenta una richiesta di assegnazione immediatamente a monte della sua evoluzione temporale. Se la risposta a tale richiesta è favorevole, l'assegnazione è mantenuta per tutta la durata dell'attività. In questo ambito si possono però distinguere almeno due alternative, che si distinguono in base all'oggetto dell'operazione di pre-assegnazione.

Nella prima di queste, chiamata *pre-assegnazione individuale*, la risorsa fisica viene assegnata in modo indiviso ad una sola attività per tutta la sua durata. Il vantaggio di tale alternativa è quello di prospettare una probabilità di ritardo o una di rifiuto non nulle solo all'inizio dell'evoluzione dell'attività di utilizzazione e di assicurare successivamente, per la durata di questa, una accessibilità senza ritardi o senza rifiuti. Lo svantaggio è quello di monopolizzare una risorsa a vantaggio di una sola attività di utilizzazione per

volta, seppure limitatamente alla durata di questa. Il risultato possibile è allora quello di un insufficiente rendimento di utilizzazione.

Una differente alternativa di pre-assegnazione, che cerca di combinare i vantaggi offerti dalle due strategie precedentemente descritte, è la *pre-assegnazione collettiva*. In questo caso la risorsa fisica può essere impegnata da più attività di utilizzazione, seppure in intervalli di tempo distinti, con un meccanismo di accesso del tutto analogo a quello della assegnazione a domanda. Ciò è però possibile solo per un *numero controllato* di attività di utilizzazione, che sono contemporaneamente in possesso di una *autorizzazione di accesso*.

Per ottenere quest'ultima, una attività di utilizzazione interessata presenta, immediatamente a monte della sua evoluzione, una esplicita richiesta in cui siano precisate le caratteristiche della sua domanda e le sue esigenze di livello di accessibilità. Detta richiesta, che è trattata dalle attività di gestione, viene accolta o meno in base all'applicazione di una opportuna *regola di decisione* (cfr. Esempio II.5-1). Il risultato favorevole è la pre-assegnazione di una risorsa virtuale.

Esempio II.5-1 - La regola di decisione per una pre-assegnazione collettiva, su cui torneremo nel capitolo IV, si basa su dati forniti dalle attività di utilizzazione richiedenti e su una loro elaborazione effettuata dall'attività di gestione preposta alla decisione.

I dati in questione sono:

- a) le caratteristiche della domanda (*dati di traffico*) e le esigenze di livello di accessibilità (*requisiti prestazionali*) per ognuna delle attività di utilizzazione le cui richieste di accesso sono già state accolte in precedenza;
- b) dati analoghi relativi all'attività di utilizzazione la cui richiesta è oggetto della decisione.

L'elaborazione riguarda in primo luogo i dati di traffico di cui in a), in modo da valutare lo stato di carico della risorsa, risultante dalle richieste già accolte (*stato di carico di riferimento*). Vengono poi utilizzati i dati di traffico di cui in b) per valutare come verrebbe modificato lo stato di carico di riferimento se la nuova richiesta fosse accolta (*stato di carico di riferimento modificato*).

Questo secondo stato costituisce successivamente l'ingresso per una valutazione delle prestazioni. Se i livelli di accessibilità per le attività di utilizzazione già accolte e per quella presentante la nuova richiesta, quali risultano da questa valutazione, rispettano i requisiti prestazionali di cui in a) e b), l'autorizzazione di accesso viene concessa; in caso contrario, viene negata.

In tutti i tre tipi di strategie di assegnazione considerate in Fig. II-10, le risorse fisiche o virtuali, che sono assegnabili, appartengono a un insieme di dimensioni limitate. Le attività di gestione debbono allora risolvere le condizioni di contesa, che si verificano quando tutte le risorse disponibili sono state assegnate e vengono presentate nuove richieste di assegnazione.

In particolare, condizioni di contesa si possono presentare in due casi:

- in una operazione di pre-assegnazione individuale o collettiva (*contesa di pre-assegnazione*) all'inizio dell'evoluzione di una attività di utilizzazione;
- nel corso dell'evoluzione di varie attività di utilizzazione, se agenti nell'ambito di strategie con assegnazione a domanda o con pre-assegnazione collettiva (*contesa di utilizzazione*).

Per valutare le prestazioni e per dimensionare le risorse con riferimento a ciascuna di queste condizioni di contesa, si possono utilizzare modelli del tipo illustrato nei due precedenti paragrafi. La struttura di questi modelli e' ovviamente legata al numero di risorse (fisiche o virtuali) in gioco e alle modalita' di gestione delle relative condizioni di contesa. Se il tipo di interazioni tra attivita' e risorse puo' essere modellato con un singolo sistema di servizio, e' interessante precisare quali siano le considerazioni per identificare i processi di ingresso e di servizio.

Nel caso di modello che rappresenti contese di pre-assegnazione, il processo di ingresso deve descrivere l'arrivo delle richieste di pre-assegnazione di una risorsa fisica (nel caso di pre-assegnazione individuale) o virtuale (nel caso di pre-assegnazione collettiva). Se invece il modello e' rivolto a studiare le contese di utilizzazione, la rappresentazione deve riguardare l'arrivo delle domande di accesso, presentate dall'insieme delle attivita' interessate, nel corso della loro durata e a monte di ogni intervallo di vitalita' (cfr. par. II.1).

Circa poi il processo di servizio, i relativi tempi di servizio si identificano con le durate delle attivita' di utilizzazione richiedenti, nel caso di modello che rappresenti contese di pre-assegnazione, e con gli intervalli di vitalita' di ogni risorsa, nel caso in cui il modello riguardi contese di utilizzazione.

Si sottolinea che lo studio di una strategia di pre-assegnazione collettiva richiede l'impiego di due modelli: uno per le contese di pre-assegnazione, l'altro per quelle di utilizzazione; cio' a differenza di quanto accade per gli altri due tipi di strategie di assegnazione, per i quali e' sufficiente un solo modello, che riguarda le contese di pre-assegnazione nel caso di pre-assegnazione individuale e quelle di utilizzazione nel caso di assegnazione a domanda.

II.5.4 Controllo dei sovraccarichi

Quando si fa crescere il carico medio su un insieme di risorse condivise, si può osservare che, al di sopra di un carico-limite (*soglia di sovraccarico*), le prestazioni dell'insieme peggiorano rapidamente. In particolare diminuisce bruscamente il livello di accessibilita' per le singole attività di utilizzazione e diminuisce, altrettanto bruscamente, la portata media dell'insieme (cfr. § II.3.3).

Per evitare condizioni di questo tipo (*condizioni di congestione*), che possono condurre rapidamente al completo collasso dell'insieme, occorre affidare alle attività di gestione compiti di controllo che consentano di "filtrare" le domande di accesso in modo che:

- l'insieme operi con un carico medio che sia sempre inferiore alla soglia di sovraccarico;
- le risorse componenti siano utilizzate in modo efficiente.

Si tratta, come e' evidente, di obiettivi tra loro contrastanti, dato che questa seconda condizione indurrebbe a lavorare in prossimità della soglia. La soluzione deve essere quindi di compromesso.

Il criterio di filtraggio delle domande d'accesso è normalmente basato sulla individuazione di risorse critiche e consiste nel limitare, per queste, il rendimento di utilizzazione che si conseguirebbe nel caso in cui una nuova domanda fosse accolta.

III ARCHITETTURE E PROTOCOLLI DI COMUNICAZIONE

La comunicazione tra due o più parti richiede *cooperazione*, e cioè collaborazione per il conseguimento di uno scopo comune. Da un lato occorre assicurare il rispetto di opportune regole procedurali nel trasferimento di informazione e negli adempimenti richiesti per l'utilizzazione di questa. Dall'altro bisogna mettere in atto, quando è necessario, provvedimenti che possano fronteggiare eventi di natura aleatoria (ad es. disturbi trasmissivi, errori procedurali, guasti di apparecchiature, ecc.), che potrebbero compromettere lo scambio di informazione.

Si tratta quindi di svolgere, in forma collaborativa tra le parti coinvolte, una sequenza di funzioni che rendano possibile a una parte, non solo di essere *fisicamente connessa* con un'altra parte, ma anche di *comunicare* con quest'ultima nonostante impedimenti di natura varia, quali errori di origine fisica o logica, diversità di linguaggi, ecc..

Una descrizione astratta delle modalità di comunicazione tra due o più parti, in posizioni tra loro remote, attraverso una rete di telecomunicazione può essere effettuata definendo opportuni *modelli di riferimento*. Questi sono oggi riconosciuti di importanza basilare per lo studio progettuale di specifici ambienti di comunicazione e, in particolare, per normalizzare le procedure da impiegare nelle interazioni tra le varie parti in gioco (apparecchi terminali, terminazioni di rete, nodi di accesso e di transito) allo scopo di consentirne la cooperazione.

Per queste ragioni gli organi di standardizzazione internazionale (CCITT, ISO, ETSI, ecc.), che hanno il compito di emettere normative per lo sviluppo di nuovi ambienti di comunicazione, provvedono a definire per questi i relativi modelli di riferimento. Il primo risultato di queste attività è stato il cosiddetto *modello di riferimento per l'interconnessione di sistemi aperti*, che riguarda in modo specifico le comunicazioni di dati attraverso una rete dedicata. Questo modello, che nel seguito verrà chiamato, per brevità, *modello OSI* (Open System Interconnection), non comprende alcune funzionalità specifiche di altri ambienti, quale, ad esempio, quello delle reti integrate nei servizi. Per questi casi si è provveduto alla definizione di altri modelli di riferimento, che, comunque, sono basati sugli stessi principi fondamentali del modello OSI.

L'identificazione di un modello di riferimento da associare a uno specifico ambiente di comunicazione può avvenire in vari passi logici. Il primo di questi

corrisponde al piu' elevato livello di astrazione nell'identificazione e riguarda la definizione degli oggetti che sono utilizzati per descrivere il processo di comunicazione sotto esame, delle relazioni generali tra questi oggetti e dei vincoli generali tra questi tipi di oggetti e di relazioni. In particolare, debbono essere definite le funzioni da svolgere per consentire la comunicazione e le relative modalita' organizzative per permetterne uno svolgimento coordinato.

Il risultato di questo primo passo di identificazione e' l'*architettura della comunicazione*. Il suo elemento distintivo (ad esempio, rispetto ad architetture di elaborazione) e' la presenza costante di interazioni tra due o piu' parti. E' opportuno sottolineare che il termine "architettura" viene qui utilizzato con un significato astratto, che non deve essere confuso con quello di struttura di una particolare realizzazione. Ad esempio, con analoga distinzione terminologica, l'architettura barocca e' un insieme di regole e di convenzioni stilistiche che caratterizzano una particolare forma architettonica; un edificio barocco e' invece una realizzazione che rispetta e applica queste regole e queste convenzioni. Come nel nostro caso, si fa quindi distinzione tra il tipo di oggetto (architettura) e un esempio realizzativo di esso (edificio).

L'ultimo passo per definire un modello di riferimento, e cioe' quello che corrisponde al piu' basso livello di astrazione nell'identificazione, riguarda la descrizione dettagliata delle modalita' di esecuzione delle funzioni identificate nel primo passo e consente di specificare le procedure operative che debbono essere seguite per ognuna delle interazioni tra le parti in gioco nell'architettura di comunicazione.

Tali procedure sono i *protocolli di comunicazione*. Elementi costituenti di questi sono: a) la *semantica*, e cioe' l'insieme delle richieste che una parte puo' emettere (*comandi*), delle azioni conseguenti da effettuare e delle *risposte* di ritorno dalla parte remota; b) la *sintassi*, e cioe' la struttura dei comandi e delle risposte; c) la *temporizzazione*, e cioe' la specificazione delle possibili sequenze temporali di emissione dei comandi e delle risposte.

I protocolli di comunicazione consentono interazioni tra le apparecchiature di rete e tra queste ultime e gli apparecchi terminali. Essi comportano che, all'interno della rete e tra quest'ultima e l'utenza, si attivi uno scambio di ulteriori informazioni, dette *di controllo*. Queste debbono essere distinte da quelle di utente (cfr. § I.2.1), nel senso che ne costituiscono il necessario sussidio con l'aggiunta di extra-informazione (overhead). Percio', per esigenze di efficienza, una volta fatto salvo il conseguimento degli obiettivi di cooperazione sopra descritti, e' opportuno contenere, nei limiti del possibile, la quota parte di extra-informazione rispetto a quella di utente (*quota di extra-informazione*).

Nel seguito di questo capitolo descriveremo, dapprima (par. II.1), un possibile insieme di funzioni da attivare in un processo di comunicazione. Successivamente (par. II.2) verrà introdotta una metodologia oggi largamente impiegata per definire architetture di comunicazione, e cioè quella basata sul principio della stratificazione. Infine, come esempio di applicazione di questa metodologia, si descriveranno (par. II.3) le caratteristiche generali del modello OSI.

III.1 Funzioni di un processo di comunicazione

Con riferimento prevalente (ma non esclusivo) a un ambiente di comunicazione di dati, ci proponiamo di descrivere le principali funzioni che debbono essere eseguite per permettere a due o più *sistemi terminali*, che desiderano comunicare, di scambiarsi informazioni attraverso una rete di telecomunicazioni a cui hanno accesso. In questo contesto per sistema terminale si intende una qualsiasi entità fisica o logica che sia in grado di emettere e/o di ricevere informazioni. Esempi di coppie di sistemi terminali sono offerti da un apparecchio terminale per dati (DTE, cfr. esempio I.1.2) e un programma applicativo residente in un elaboratore in posizione remota, ovvero da due programmi applicativi interagenti con evoluzione in macchine di elaborazione tra loro distanti, ovvero da un programma applicativo che interroga un archivio remoto.

Le funzioni in esame verranno nel seguito distinte in due classi: quelle preposte al trasferimento dell'informazione tra i sistemi terminali (*funzioni orientate al trasferimento*) e quelle che, in ognuno di questi, sono finalizzate a consentire l'utilizzazione dell'informazione trasferita (*funzioni orientate all'utilizzazione*). Le prime debbono assicurare che le seconde non siano condizionate dalle specifiche modalità del trasferimento. All'analisi degli elementi funzionali che compongono queste due classi è dedicato il seguito del paragrafo (§ II.1.1 e II.1.2).

III.1.1 Funzioni orientate al trasferimento

In questa classe possono distinguersi le funzionalità preposte al trasferimento fisico da quelle legate ad aspetti logici.

Tra le prime rientra la *funzione di connessione fisica*, e cioè la fornitura di un insieme di risorse, che connettano fisicamente il sistema terminale di origine a quello di destinazione. Tali risorse consistono in una o più tratte in cascata di

mezzo trasmissivo (ad es. in rame, in fibra ottica o su portante radio) e sono scelte in un insieme piu' ampio nell'ambito della rete fisica (cfr. § I.1.1).

Una volta garantita la disponibilita' di un mezzo fisico tra i punti terminali della comunicazione, e' necessario trasferire fisicamente l'informazione emessa dai sistemi terminali. Tale obiettivo e' conseguito con un insieme di apparati e di dispositivi, che realizzano la *funzione di trasmissione* e che, in unione al mezzo fisico, costituiscono il *canale trasmissivo* tra le parti in comunicazione. Il compito di questa funzione e' duplice: in emissione, deve formare un segnale che sia supporto dell'informazione da trasferire e che abbia un contenuto in frequenza armonizzato con la banda passante del canale trasmissivo; in ricezione, deve estrarre dal segnale ricevuto una informazione che sia la replica piu' fedele possibile di quella emessa all'origine.

Le due funzioni ora illustrate garantiscono il trasferimento fisico dell'informazione tra i sistemi terminali in comunicazione. Occorre pero' considerare che, in questo trasferimento, l'informazione puo' essere deteriorata per cause fisiche di natura varia, che sono dovute alle caratteristiche non ideali del canale trasmissivo. L'effetto e' la presentazione casuale di *errori di trasmissione*, che, anche se interessano solo brevi porzioni dell'informazione, possono comunque pregiudicare la sua corretta interpretazione. Per fronteggiare questi eventi, che sono particolarmente indesiderabili nel caso di comunicazioni di dati, e' necessario prevedere un meccanismo di accordo tra le parti in comunicazione, che assicuri la *rivelazione* o la *correzione* degli errori di trasmissione. A tale scopo sono impiegate tecniche di *codifica di canale* e, nel caso di sola rivelazione, puo' essere prevista anche una *funzione di recupero* degli errori che sono stati rivelati.

Circa poi la struttura dell'informazione da trasferire, occorre tenere conto della natura intermittente che normalmente caratterizza l'emissione da parte delle sorgenti della comunicazione (cfr. § I.2.3). La lunghezza (in cifre binarie) dei tratti informativi e l'intervallo temporale tra due tratti successivi sono variabili aleatorie e sono, in ogni caso, dati caratteristici di ciascuna sorgente, in quanto dipendono dal tipo di informazione emessa.

Per rendere piu' efficiente il trasferimento e' opportuno prevedere che, ad esempio in corrispondenza ai punti di accesso alla rete, i tratti informativi subiscano una ristrutturazione, e cioe' una *suddivisione* o una *aggregazione*, per costituire *unita' informative*. Il formato di queste (e in particolare la loro lunghezza in cifre binarie) e' definito in relazione alle caratteristiche di trasferimento attraverso la rete e non e', in generale, legato alla lunghezza dei

tratti informativi. Le unità informative costituiscono l'oggetto di alcune delle funzioni che verranno illustrate nel seguito del presente paragrafo.

Le risorse utilizzate per un trasferimento sono costituite dai mezzi forniti dalla rete logica. Tra questi si possono distinguere i nodi di commutazione, i rami di giunzione e quelli di utente (cfr. par. I.1). Utilizzando questi elementi è possibile formare un *percorso di rete* da un sistema terminale di origine ad uno di destinazione. Ciò si ottiene, di volta in volta, individuando un insieme di rami e di nodi che una unità informativa deve attraversare per essere trasferita da un punto di accesso di origine ad uno di destinazione.

Tale insieme non è in generale univoco e quindi la sua scelta dovrà essere fatta con un'opportuna *strategia di instradamento*, che deve, tra l'altro, tenere conto dello stato di impegno degli elementi (nodi e rami) che compongono il percorso di rete e delle prestazioni desiderate nel trasferimento. Individuato allora, per ogni nodo, quale sia il ramo uscente da mettere in corrispondenza con un particolare ramo entrante, si tratta poi di mettere in atto quanto è necessario perché ogni unità informativa possa attraversare il nodo verso la destinazione desiderata.

I rami, che per definizione hanno un ingresso e una uscita, possono essere caratterizzati dalla loro *capacità di trasferimento*, e cioè, ad esempio, dal numero massimo di cifre binarie che possono essere trasferite nell'unità di tempo sul canale trasmissivo ad essi associato.

I nodi, invece, presentano una pluralità di ingressi e di uscite. Essi costituiscono quindi una risorsa che è caratterizzata da due aspetti: il primo riguarda la capacità di trasferimento per ogni singola coppia ingresso-uscita; il secondo si riferisce invece alla *capacità di elaborazione* necessaria per gestire la pluralità di scelte connessa a questi possibili accoppiamenti.

Tali risorse sono di tipo condiviso e sono quindi soggette a richieste concorrenti. Supponiamo allora che venga richiesto il trasferimento di unità informative tra i punti di accesso *A* e *B* e che tale richiesta venga presentata quando la rete è già impegnata in altre operazioni dello stesso tipo, che, in generale, avranno differenti origine e destinazione. In queste condizioni una parte (al limite la totalità) della capacità di trasferimento dei rami e dei nodi disponibili può essere già utilizzata. Per decidere allora quale sia il percorso di rete più conveniente (dal punto di vista del conseguimento di fissati obiettivi prestazionali), e cioè quale sia l'insieme dei rami e dei nodi che conviene attraversare per effettuare il trasferimento da *A* a *B*, bisogna considerare le strategie adottate per l'utilizzazione condivisa dei rami e dei nodi.

L'utilizzazione della capacita' di trasferimento dei rami da parte di una molteplicita' di unita' informative aventi origine da sorgenti diverse e' legata alla funzione di *multiplazione*. Questa, come si vedra' nel par. III.2, puo' essere attuata in accordo a diverse modalita' alternative.

Invece le modalita' di attraversamento dei nodi e di utilizzazione della relativa capacita' di elaborazione rientrano nelle funzionalita' tipiche di un nodo di rete e caratterizzano il tipo di *commutazione* da esso svolta. Anche la funzione di commutazione puo' essere svolta con varie soluzioni tecniche, su cui ci soffermeremo nel par. III.3.

Un'ulteriore funzione tipica in un trasferimento di informazione e' quella riguardante il *controllo di flusso*. Tale funzione regola il trasferimento delle unita' informative su una connessione; provvede cioe' a variare il loro ritmo di trasferimento in modo da assicurare la migliore utilizzazione possibile delle risorse disponibili in ogni condizione di carico, prevenendo eventuali possibilita' di degradazione delle prestazioni. In altre parole, un efficiente metodo di controllo di flusso ha lo scopo di evitare l'insorgere di fenomeni di *congestione* (cfr. § I.3.7).

Le funzioni fino ad ora esaminate sono eseguite all'interno di una rete di telecomunicazione e si riferiscono al trasferimento delle unita' informative. E' evidente che, in relazione alle specifiche modalita' con cui questo insieme di funzioni e' eseguito, puo' essere definita una *qualita' di servizio* che la rete e' in grado di offrire. Reti diverse, basate su procedure operative diverse, offriranno agli utenti una qualita' di servizio diversa.

In relazione al tipo di applicazione che rappresenta l'oggetto dello scambio di informazione tra sistemi terminali, una particolare qualita' di servizio offerta da una rete puo' non essere sufficiente al soddisfacimento delle esigenze dell'applicazione in questione. In questi casi, e' necessario prevedere un complesso di funzioni che hanno l'obiettivo di *incrementare* la qualita' di servizio complessiva per portarla al livello richiesto per il corretto funzionamento dell'applicazione. Tali funzioni sono ovviamente eseguite, in aggiunta a quelle di rete, sfruttando il servizio da questa offerto.

Di particolare interesse e' il caso in cui il percorso di rete tra due sistemi terminali attraversi, in successione, due o piu' reti (*reti in cascata*). In tal caso e' essenziale che sia eseguita una funzione di *equalizzazione* delle qualita' di servizio delle varie reti in modo da ottenere una qualita' di servizio che risulti accettabile in relazione alle specifiche esigenze dell'applicazione. Si precisa che, a differenza delle funzioni illustrate in precedenza, la funzione di equalizzazione

delle qualità di servizio non è eseguita all'interno della rete, ma deve essere svolta ai bordi di questa o direttamente dai sistemi terminali.

III.1.2 Funzioni orientate all'utilizzazione

Saranno ora introdotte alcune funzioni che più direttamente riguardano l'utilizzazione dell'informazione. Queste dipendono in modo diretto dalla particolare applicazione che è lo scopo dello scambio informativo e sono di esclusiva competenza dei sistemi terminali. Con il termine "applicazione" intenderemo, d'ora in avanti, l'insieme dei compiti, che un sistema terminale deve svolgere in base alle informazioni ricevute, e delle procedure relative alla loro esecuzione.

La funzione, che logicamente segue quelle descritte in precedenza (cfr. § II.1.1), concerne la *gestione del dialogo* tra i sistemi terminali. Così come nel dialogo tra esseri umani sono implicitamente rispettate regole di colloquio che consentono il corretto svolgimento di una conversazione, allo stesso modo, nel caso di dialogo tra sistemi di elaborazione, è indispensabile eseguire le funzioni necessarie a regolare e a sincronizzare l'emissione di informazione da parte dei due interlocutori. In questo modo lo scambio informativo può svolgersi ordinatamente senza che si verifichino situazioni tali da pregiudicarne il risultato.

Affinché le informazioni scambiate possano essere opportunamente utilizzate dai sistemi terminali a colloquio, occorre che tali informazioni possano essere correttamente interpretate da un punto di vista sia sintattico che semantico. Le informazioni devono essere quindi trasferite in una forma che sia compatibile con il sistema terminale di destinazione.

Per la riuscita della comunicazione è quindi indispensabile la scelta e l'utilizzazione di una *sintassi comune* tra i sistemi. In altre parole devono essere fissate a priori le regole secondo cui l'informazione deve essere codificata, in modo che entrambi i sistemi a colloquio possano, sotto l'aspetto sintattico, interpretare univocamente le informazioni ricevute. In particolare, le informazioni scambiate tra sistemi terminali devono essere codificate in accordo ad uno stesso alfabeto.

L'ultima funzione che deve essere eseguita in un processo di comunicazione è l'interpretazione delle informazioni ricevute, e cioè la traduzione dell'informazione in una azione o in un compito da svolgere. Tale funzione è possibile solo se il sistema ricevente è in grado di riconoscere il significato delle informazioni ricevute e cioè la loro *semantica*.

L'insieme delle funzioni di un processo di comunicazione che e' stato ora illustrato non e' esaustivo, ma rappresenta solo un esempio della complessita' delle operazioni necessarie affinche' una comunicazione vada a buon fine. Da tale esemplificazione emerge pero' una osservazione fondamentale che e' opportuno sottolineare e che sara' largamente utilizzata nel seguito del capitolo.

Le funzioni elementari sopra descritte, pur essendo reciprocamente autonome, sono logicamente collegate tra loro. Cio' significa che l'esecuzione di una qualsiasi funzione elementare presuppone, da un lato, lo svolgimento preventivo di altre funzioni e, dall'altro, costituisce la base per l'esecuzione di altre. Ad esempio, la funzione di rivelazione o di correzione degli errori di trasmissione presuppone che sia stato garantito il trasferimento fisico dell'informazione tra le estremita' del canale trasmissivo e, contemporaneamente, consente di iniziare il processo di interpretazione dell'informazione trasferita.

III.2 Architettura a strati

Sulla base di quanto detto nella conclusione del precedente paragrafo, si puo' affermare che, per il corretto svolgimento di una comunicazione tra due o piu' sistemi terminali, e' necessario svolgere un insieme di funzioni tra le quali e' possibile identificare una relazione gerarchica. Per chiarire le implicazioni di questa relazione, indichiamo con A, B e C tre sottoinsiemi di funzioni in ordine gerarchico crescente. La dipendenza gerarchica di questi sottoinsiemi significa che lo svolgimento di B presuppone la preventiva esecuzione di A e che l'unione di A e di B costituisce il presupposto per l'esecuzione di C. Il sottoinsieme B offre quindi un servizio a C e, per questo scopo, opera in modo da aggiungere valore al "servizio" che gli e' offerto da A. Il risultato e' che il "servizio" offerto a C si presenta con caratteristiche arricchite rispetto a quello offerto a B.

In queste precisazioni sono utilizzati due concetti fondamentali, su cui si fondano tutte le architetture oggi utilizzate per descrivere, in termini modellistici, processi di comunicazione sia monomediali che multimediali: si allude al *raggruppamento* e alla *stratificazione*. Il raggruppamento implica che funzioni simili per logica e per tecnologia realizzativa siano raggruppate in insiemi funzionali omogenei. La stratificazione impone invece che tali insiemi siano gerarchicamente organizzati in modo che ognuno di essi arricchisca di un ulteriore insieme di funzioni il servizio fornito dall'insieme funzionale gerarchicamente inferiore. Il complesso dei due insiemi funzionali da' luogo a

un nuovo servizio che sarà fornito a sua volta all'insieme funzionale gerarchicamente superiore.

La stratificazione consente di suddividere il problema generale riguardante la definizione delle modalità operative del colloquio tra sistemi terminali in un insieme di problemi più semplici, ognuno dei quali si riferisce esclusivamente a un insieme limitato di funzioni.

La trattazione di questo paragrafo sviluppa, in modo dettagliato, i principali argomenti riguardanti una *architettura a strati*. La terminologia impiegata è quella introdotta inizialmente per il modello OSI e oggi usualmente impiegata anche per altri modelli di riferimento.

III.2.1 Elementi architetturali

In Fig. III-1 sono mostrati i tre elementi fondamentali di una architettura di comunicazione; si distinguono:

- i *sistemi*, che costituiscono un insieme autonomo capace di effettuare il trattamento e/o il trasferimento dell'informazione in vista di specifiche applicazioni;
- gli aspetti dei *processi applicativi*, che risiedono nei sistemi e che sono coinvolti da esigenze di interazione con altri processi nell'ambito dell'architettura;
- i *mezzi trasmissivi*, che rappresentano la struttura fisica di interconnessione tra i sistemi.

Per ognuno dei sistemi interconnessi, l'architettura considera solo gli aspetti che riguardano il funzionamento verso l'esterno, e cioè quelli volti alla cooperazione con altri sistemi.

Ogni sistema è visto come logicamente composto da una successione ordinata di *sottosistemi*. L'ordine definisce il *rango* di ognuno di questi. Un sottosistema è quella parte dell'intero sistema che, per lo svolgimento delle funzioni appartenenti a uno specifico raggruppamento funzionale, interagisce solo con i sottosistemi di ranghi immediatamente superiore e immediatamente inferiore.

Tutti i sottosistemi, che appartengono a qualunque sistema tra quelli interconnessi e che sono caratterizzati da uguale rango (*sottosistemi omologhi*), formano uno *strato*. Questo rappresenta quindi l'unione di tutti i sottosistemi omologhi appartenenti ai sistemi interconnessi.

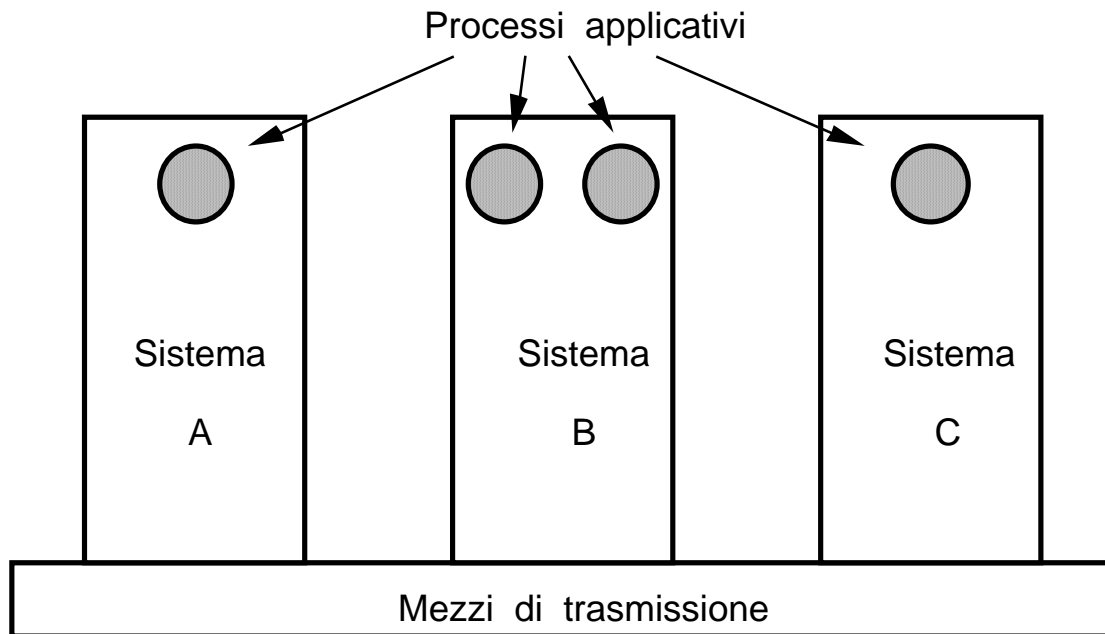


Fig. III-1 - Elementi fondamentali di una architettura di comunicazione.

Un aspetto fondamentale di questa architettura è l'indipendenza funzionale dei vari strati. Il servizio fornito da un generico strato è definito in modo del tutto indipendente dalle procedure con cui è effettivamente realizzato. Ciò consente, ad esempio, di assimilare, nella realizzazione di un singolo strato, eventuali evoluzioni tecnologiche o variazioni nella sua definizione senza che ciò influenzi il funzionamento degli altri strati.

All'interno di un sottosistema possono identificarsi una o più *entità* che rappresentano gli elementi attivi di ogni singolo strato. Un'entità corrisponde a quella parte di un sottosistema che provvede all'esecuzione di una o più tra le funzioni dello strato. Durante la comunicazione, entità appartenenti allo stesso strato e residenti nei sistemi interconnessi (*entità alla pari*) interagiscono tra loro per il corretto espletamento delle funzioni a cui sono preposte.

Nel seguito di questo capitolo, per indicare qualsiasi elemento della architettura, sarà usata una particolare notazione in cui il nome dell'elemento è preceduto dal numero d'ordine dello strato a cui si riferisce. Ad esempio, il termine *(N)-entità* identificherà una entità appartenente all'*(N)-strato*. In Fig. III-2 sono illustrati i principi della stratificazione funzionale dei sistemi.

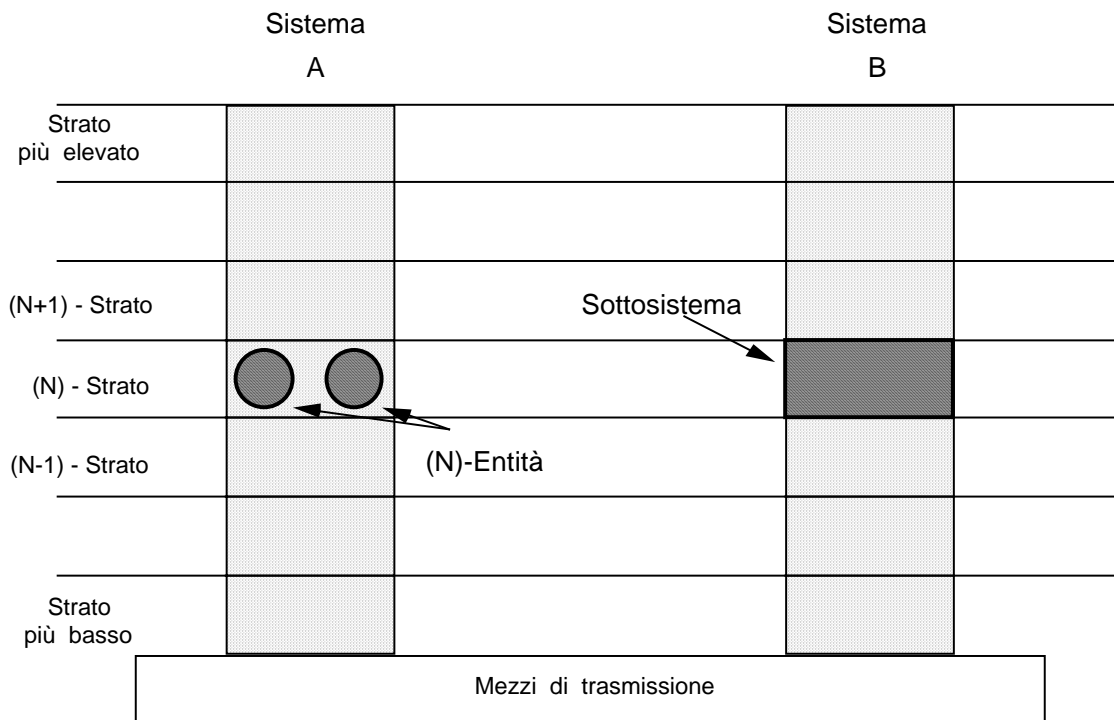


Fig. III-2 - Stratificazione funzionale di sistemi.

In taluni casi è necessario estendere la tecnica della stratificazione anche all'interno di un singolo strato. In tal modo si possono identificare *sottostrati*, corrispondenti a particolari raggruppamenti di funzioni che possono eventualmente non essere eseguite nel corso della comunicazione.

Ogni (N)-strato, sfruttando il servizio dello strato inferiore, fornisce un (N)-servizio alle (N+1)-entità. Un (N)-servizio rappresenta un particolare sottoinsieme delle funzioni eseguite dall'(N)-strato; in particolare esso contiene solo quelle funzioni che sono visibili all'interfaccia tra l'(N)-strato e l'(N+1)-strato. Tale visibilità è da intendersi nel senso che lo strato superiore riceve una esplicita notifica dell'esecuzione di tali funzioni da parte dello strato inferiore.

Dall'esame della Fig. III-2 emerge che allo strato funzionale più elevato non corrisponde nessuno strato superiore; esso quindi non offre servizio a un altro strato, ma, al contempo, riceve la somma dei servizi di tutti gli altri strati. La stessa figura mette in risalto che lo strato funzionale più basso si interfaccia direttamente con i mezzi trasmissivi e quindi non riceve servizio da alcuno strato.

Una (N)-entità può fornire servizi ad una (N+1)-entità ed usare i servizi offerti da una o più (N-1)-entità. Gli (N)-servizi sono offerti alle (N+1)-entità negli (N)-punti d'accesso al servizio [(N)-SAP, Service Access Point], che

rappresentano le interfacce logiche tra le (N)-entita' e le (N+1)-entita'. Una (N+1)-entita' comunica con una (N)-entita', residente nello stesso sistema, attraverso un (N)-SAP. Un (N)-SAP puo' essere servito da una sola (N)-entita' e essere utilizzato da una sola (N+1)-entita'. Tuttavia una (N)-entita' puo' servire vari (N)-SAP e una (N+1)-entita' puo' utilizzare vari (N)-SAP.

Sulla base di tali definizioni, gli utenti di un (N)-servizio [(N)-utenti] sono le (N+1)-entita' alla pari che ne usufruiscono per i loro scopi di cooperazione. Fornitore di un (N)-servizio [(N)-fornitore] sono invece le (N)-entita', che sono allacciate agli (N)-SAP, attraverso cui viene presentata dagli (N)-utenti la richiesta del servizio. Si sottolinea il significato che qui viene introdotto per il termine "utente" di un servizio di strato; esso ben si differenzia da quello di utente di un servizio di telecomunicazione, che e' stato definito nell'introduzione del cap.I.

Una (N)-funzione e' parte delle attivita' di una (N)-entita' ed e' svolta mediante la cooperazione di entita' alla pari dell'(N)-strato che, a questo scopo, comunicano tra di loro. Come gia' detto, una (N)-funzione puo' rientrare tra le funzionalita' che caratterizzano l'(N)-servizio, e quindi sono visibili in uno o piu' tra gli (N)-SAP, ovvero puo' appartenere all'insieme delle funzioni atte a conseguire gli obiettivi di prestazione specifici dell'(N)-strato e quindi svolte all'interno di questo senza richieste specifiche da parte delle (N+1)-entita'.

La cooperazione tra (N)-entita' residenti in sistemi diversi e' governata da un insieme di regole che prende il nome di (N)-protocollo. In particolare la comunicazione diretta tra (N)-entita' residenti nello stesso sistema non e' visibile all'esterno di questo e non e' quindi oggetto di interesse nella definizione dell'architettura.

Un sistema impegnato in una comunicazione puo' essere terminale o di rilegamento. Si parla di *sistema terminale*, quando questo e' origine o destinazione finale delle informazioni. Quando i mezzi fisici non connettono direttamente due sistemi terminali, la comunicazione tra questi deve avvenire attraverso un *sistema di rilegamento*, che deve eseguire le funzioni necessarie alla concatenazione delle varie tratte della connessione fisica.

I sistemi di rilegamento hanno una organizzazione architettuale di tipo analogo a quella dei sistemi terminali, ma con un numero di sottosistemi che e' solitamente inferiore; in particolare, partendo dal sistema di rango piu' basso e procedendo verso ordini gerarchici crescenti, essi includono tutti i sottosistemi fino a quello che e' in grado di svolgere una *funzione di rilegamento*. Questa, se

svolta nell'(N)-strato, consente a una (N)-entita' di inoltrare le informazioni ricevute da una (N)-entita' a monte verso un'altra (N)-entita' a valle.

Gli oggetti entro uno strato o alla frontiera tra due strati adiacenti debbono essere identificati in modo univoco. A tale scopo vengono definiti *identificatori* per le entita' e per i SAP.

Una (N)-entita' e' identificata con un *titolo globale*, che la individua in modo univoco in tutto l'insieme dei sistemi interconnessi. Entro un dominio piu' limitato, una (N)-entita' puo' essere identificata con un *titolo locale*, che la individua in modo univoco entro quel dominio. Per esempio, entro il dominio corrispondente all'(N)-strato, le (N)-entita' sono identificate con (N)-titoli locali, che sono univoci entro quello strato.

Ogni (N)-SAP e' identificato da un (N)-*indirizzo*, che localizza in modo univoco l'(N)-SAP a cui e' allacciata una specifica (N+1)-entita'. Data la corrispondenza uno ad uno tra un (N)-SAP e una (N+1)-entita', l'uso dell'(N)-indirizzo di questo (N)-SAP per identificare questa (N+1)-entita' e' il meccanismo di indirizzamento piu' efficiente, se puo' essere assicurato un allacciamento permanente tra questa (N+1)-entita' e questo (N)-SAP.

Gli allacciamenti tra (N)-entita' e gli (N-1)-SAP che esse utilizzano per comunicare tra loro sono definiti in una particolare (N)-funzione, detta di (N)-*guida* (directory). Questa traduce i titoli globali delle (N)-entita' negli (N-1)-indirizzi, attraverso i quali dette entita' possono essere raggiunte e quindi cooperare tra loro.

Infine l'interpretazione della corrispondenza tra gli (N)-indirizzi serviti da una (N)-entita' e gli (N-1)-indirizzi utilizzati per accedere agli (N-1)-servizi che l'(N)-entita' utilizza (e quindi per identificare quest'ultima) e' effettuata da un'altra particolare (N)-funzione, che e' detta di *corrispondenza di indirizzo* (address mapping).

III.2.2 Comunicazione tra entita'

Affinche' possa aver luogo uno scambio informativo tra due o piu' (N+1)-entita', l'(N)-strato deve offrire a queste un (N)-servizio, che puo' essere di due tipi: con *connessione* ovvero *senza connessione*.

Nel caso di (N)-servizio con connessione, tra le (N+1)-entita' interessate a comunicare viene preventivamente stabilita una associazione logica. Questa e' realizzata per mezzo dei servizi offerti dall'(N)-strato ed e' detta (N)-*connessione*. Una (N)-connessione e' quindi l'associazione dinamica stabilita tra

due o piu' (N+1)-entita' per controllare il trasferimento dell'informazione tra queste. Piu' in particolare, una (N)-connessione congiunge i due o piu' (N)-SAP a cui sono allacciate le (N+1)-entita'. La Fig. III-3 mette in evidenza le relazioni tra entita', protocolli di strato e relative connessioni.

Nel caso di (N)-servizio senza connessione, si ha il trasferimento di unita' informative da un (N)-SAP di origine ad uno o piu' (N)-SAP di destinazione in assenza di una connessione.

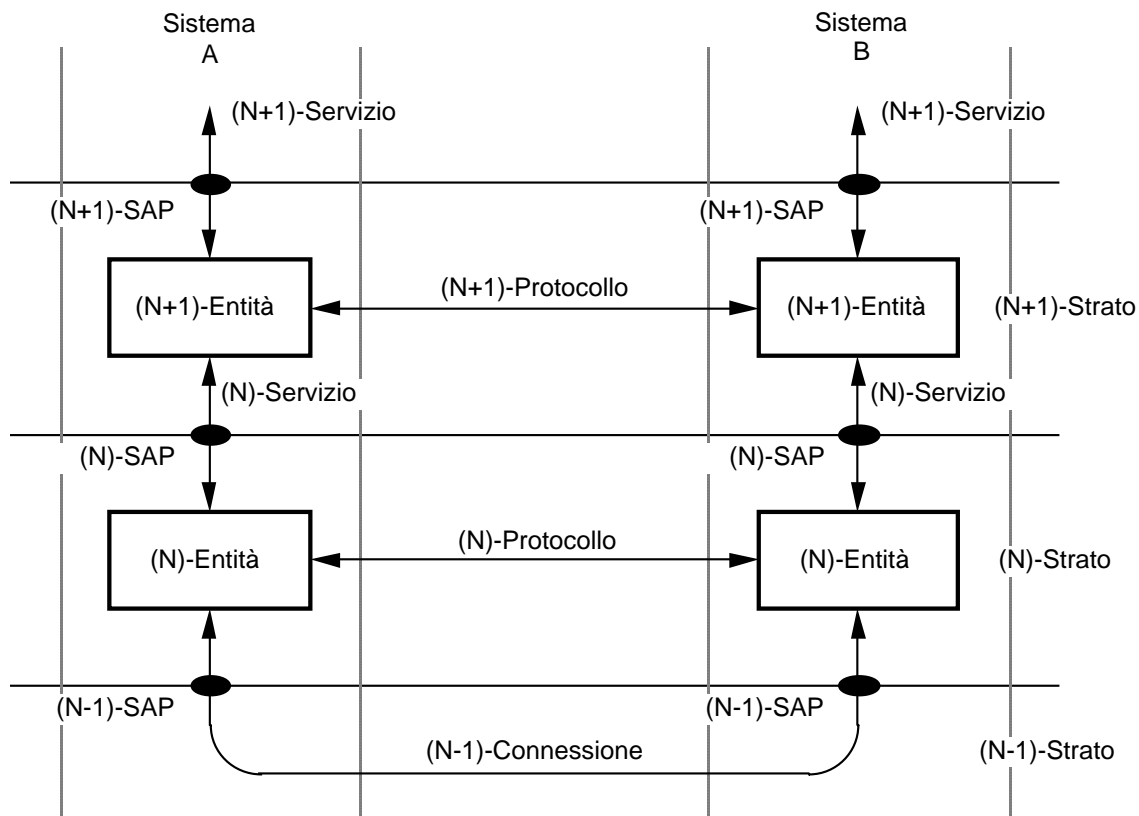


Fig. III-3 - Relazione tra gli elementi di una architettura di comunicazione con trasferimento nel modo con connessione.

Un trasferimento di informazione nel servizio con connessione ha cinque caratteristiche fondamentali, e cioè:

- 1) una durata chiaramente valutabile;
- 2) un accordo che coinvolge tre parti;
- 3) una negoziazione dei parametri e delle opzioni che governeranno il trasferimento di informazione;
- 4) una utilizzazione di identificatori di connessione;
- 5) una relazione tra le unita' informative scambiate.

Circa la prima di queste caratteristiche, si osserva che, in base alla definizione precedente, interazioni tramite il servizio con connessione

avvengono attraverso tre distinte fasi in sequenza: l'*instaurazione della connessione*, il *trasferimento dell'informazione* e l'*abbattimento della connessione*. La vita di una connessione puo' essere di lunga durata e manifestarsi in molti scambi separati tra le (N+1)-entita' connesse; ovvero puo' essere compressa in una interazione molto breve, in cui l'informazione necessaria per instaurare la connessione, quella da trasferire e quella necessaria per abbattere la connessione sono trasportate in un piccolo numero di scambi, anzi, comunemente, in un solo scambio per ogni direzione di trasferimento.

L'instaurazione di una (N)-connessione richiede il preventivo accordo tra almeno tre parti: le due (N+1)-entita' che desiderano comunicare e l'(N)-servizio che fornisce loro i mezzi per soddisfare questa esigenza. Queste parti debbono preliminarmente manifestare la loro volonta' di partecipare al trasferimento di informazione. Percio', per tutta la durata della connessione, tali parti sono vincolate all'accordo iniziale e debbono quindi accettare il trasferimento di informazione su tale connessione.

Nell'ambito di questo accordo, debbono essere negoziati i parametri e le opzioni che governeranno il trasferimento di informazione. Percio' una richiesta di instaurazione puo' essere rifiutata dalla parte ricevente se i parametri e le opzioni scelti dalla parte emittente sono inaccettabili per la parte ricevente; in tal caso quest'ultima puo', nell'ambito del suo rifiuto, suggerire scelte alternative. I motivi di un eventuale rifiuto di instaurazione di una connessione sono vari; ad esempio, per mancanza delle risorse necessarie.

La negoziazione puo' consentire alle parti di scegliere specifiche procedure, quali quelle riguardanti la sicurezza, l'addebito e la verifica di identita'. Inoltre l'accordo che risulta dalla negoziazione puo', in alcuni casi, essere modificato (e cioe' rinegoziato) dopo che la connessione sia stata instaurata e sia iniziata la fase di trasferimento dell'informazione.

All'atto in cui una (N)-connessione deve essere instaurata, ogni (N+1)-entita' e' identificata, per cio' che riguarda l'(N)-servizio, dall'indirizzo dell'(N)-SAP attraverso il quale l'(N+1)-entita' interagisce con l'(N)-servizio. Quest'ultimo utilizza quindi gli (N)-indirizzi per instaurare la connessione richiesta.

Invece, nella fase di trasferimento dell'informazione e in quella di abbattimento della connessione, non vengono utilizzati gli (N)-indirizzi degli (N)-SAP connessi. Si impiega, in loro luogo, un *identificatore di connessione*, che individua in modo univoco una connessione durante tutta la sua durata. Cio' consente di ridurre la quota di extra-informazione che sarebbe necessaria per

trasferire la totalita' degli (N)-indirizzi. Sulle modalita' per identificare una connessione si tornera' nel seguito di questa sezione.

Una volta che una connessione sia stata instaurata, essa puo' essere utilizzata per trasferire unita' informative in sequenza temporale, fintantoche' la connessione non venga rilasciata da una delle parti interagenti. Queste unita' informative sono mutuamente legate in virtu' del fatto che sono trasferite su una particolare connessione. Piu' specificamente, per effetto del trasferimento ordinato sulla stessa connessione:

- possono essere facilmente rivelate e recuperate condizioni di fuori sequenza, di perdita e di duplicazione riguardanti le unita' informative;
- possono essere impiegate tecniche di controllo di flusso per assicurare che il ritmo di trasferimento tra entita' alla pari non superi quello che le entita' in comunicazione sono in grado di trattare (cfr. § II.2.5).

Passando ora al trasferimento di informazione nel modo senza connessione e alle sue caratteristiche distintive, osserviamo che in questo caso il trasferimento avviene a seguito di accordi tra due sole parti. Esiste cioe' un accordo a priori tra le (N+1)-entita' in corrispondenza, che non e' noto all'(N)-servizio e che consiste, almeno, di una conoscenza mutua a priori. Esistono poi accordi individuali tra ogni (N+1)-entita' e il fornitore dell'(N)-servizio. Non viene invece scambiata, tra le (N)-entita', informazione dell'(N)-protocollo che riguardi la volonta' mutua delle (N+1)-entita' di impegnarsi in una comunicazione.

Poiche' le entita' di uno strato possono scambiarsi informazioni senza la preventiva instaurazione di una connessione nello strato inferiore, questo, a meno di avviso esplicito contrario, risulta sempre disponibile per il trasferimento richiesto. Inoltre le due entita' coinvolte nella comunicazione possono non essere contemporaneamente attive; e' sufficiente infatti che l'entita' di origine sia attiva solo per il tempo necessario all'emissione delle informazioni, mentre l'entita' di destinazione lo deve essere solo al momento della ricezione.

La caratteristica piu' appariscente di un trasferimento con modo senza connessione e' la singolarita' di accesso di servizio richiesta per iniziare la emissione di unita' informative. Piu' precisamente, tutta l'informazione di controllo che e' richiesta per recapitare a destinazione una unita' informativa (e cioe' l'indirizzo della destinazione, la qualita' di servizio scelta, eventuali opzioni) e' presentata, al fornitore dell'(N)-servizio, in un'unica istanza insieme all'informazione da trasferire.

Questa istanza non e' in alcun modo legata a quelle che la precedono e la seguono; inoltre non implica ulteriore comunicazione tra l'(N)-fornitore e l'(N)-utente, circa la sorte dell'unita' informativa emessa. Tuttavia gli accordi bilaterali che, istanza per istanza, si stabiliscono tra l'(N)-fornitore e l'(N)-utente consentono a quest'ultimo di specificare i valori dei parametri richiesti per il trasferimento (ad esempio del ritmo di trasferimento, del tasso di errore accettabile, ecc.). In relazione al modo in cui il servizio senza connessione e' realizzato, l'(N)-fornitore puo' o meno essere in grado di decidere se la richiesta dell'(N)-utente puo' essere accolta con le condizioni specificate.

Ulteriore caratteristica di un (N)-servizio senza connessione e' rappresentata dall'assenza di negoziazioni tra le parti in gioco. L'associazione a priori tra le (N+1)-entita' definisce i protocolli per il controllo delle informazioni di loro interesse, mentre l'(N)-servizio agisce solo come portatore passivo di tali informazioni. Questa caratteristica contribuisce alla relativa semplicita' dei protocolli utilizzati, in quanto e' limitata l'influenza delle interazioni tra gli (N)-utenti sul funzionamento dell'(N)-protocollo.

Infine, in un (N)-servizio senza connessione debbono essere sottolineate la indipendenza e l'auto-consistenza delle unita' informative trasferite. Circa l'indipendenza, questa implica che una sequenza di unita' informative consegnate a un servizio senza connessione puo' essere recapitata a destinazione in un ordine diverso da quello di consegna all'origine; cio' in quanto l'(N)-fornitore non svolge alcuna funzione per legare logicamente le unita' informative che gli vengono consegnate.

Per cio' che riguarda poi l'auto-consistenza, ogni unita' informativa deve contenere, come gia' detto, tutta l'informazione necessaria per essere consegnata a destinazione. Questa caratteristica introduce vantaggi e svantaggi. Da un lato migliora la robustezza del servizio e riduce la quantita' di informazione, in aggiunta alle unita' informative da scambiare, che deve essere memorizzata entro l'(N)-fornitore. Dall'altro comporta una quota di extra-informazione che e' maggiore di quella richiesta nel caso delle unita' informative scambiate nel modo con connessione durante la fase di trasferimento.

Nel seguito, per semplicita' di trattazione, ci riferiremo esclusivamente al modo con connessione.

Una (N)-connessione puo' essere instaurata tra due o piu' (N)-SAP. Se sono interessati solo due (N)-SAP, la connessione e' detta punto-punto; in caso contrario e' detta punto-multipunto.

Una (N+1)-entità può gestire simultaneamente una molteplicità di connessioni con altre (N+1)-entità. In questo caso su di uno stesso (N)-SAP si attesterà più di una (N)-connessione; il punto terminale di ognuna di queste è detto Punto terminale di una (N)-connessione [(N)-CEP, Connection End Point]. All'interno di uno stesso (N)-SAP i punti terminali delle (N)-connessioni sono individuati dai cosiddetti Identificatori degli (N)-CEP. Questi ultimi identificano quindi le singole connessioni che si attestano ad uno specifico (N)-SAP. Nella Fig. III-4 sono illustrate le definizioni ora fornite.

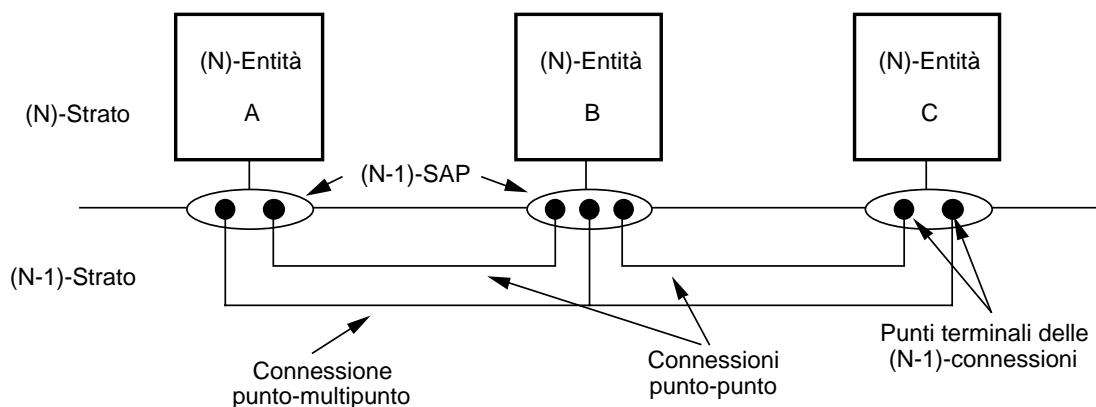


Fig. III-4 - Esempi di connessioni di strato.

Da quanto detto risulta che una (N)-entità è impegnata nella gestione di due tipi di flussi informativi:

- a) con entità appartenenti agli strati adiacenti;
- b) con entità alla pari.

Tuttavia, solo nel primo caso si ha un trasferimento diretto di informazioni; nel secondo queste sono invece trasferite in modo indiretto usando la connessione messa a disposizione dallo strato inferiore. Un trasferimento diretto di informazioni tra entità alla pari avviene solo nel caso delle entità appartenenti allo strato gerarchicamente più basso; queste si scambiano infatti informazione utilizzando il mezzo trasmissivo che le connette.

In entrambi i casi a) e b), lo scambio informativo è regolato da specifici protocolli. L'attività di normalizzazione riguarda però esclusivamente quelli relativi al colloquio tra entità alla pari, mentre le modalità di trasferimento tra strati adiacenti sono lasciate alla completa discrezione dei singoli costruttori.

III.2.3 Unita' informative

Nell'ambito dell'interazione tra entita', che e' richiesta per il conseguimento degli obiettivi di un processo di comunicazione, debbono avvenire scambi di informazione sia tra entita' alla pari residenti in sistemi diversi, sia tra entita' appartenenti a strati adiacenti e residenti nello stesso sistema.

Questa informazione puo' essere di due tipi:

- e' l'oggetto primario dello scambio per lo svolgimento delle funzioni dell'architettura di comunicazione e verra' denominata, nel seguito di questo capitolo, *informazione di dati*;
- ha scopi di controllo per consentire alle entita', che sono origine e destinazione dello scambio o che sono preposte a svolgere funzioni di rilegamento, di coordinare il loro funzionamento secondo gli obiettivi architetturali; in questo caso si parla di *informazione di controllo*.

A seconda del tipo di scambio, le informazioni di dati si distinguono a loro volta in due categorie: i *dati di utente* e i *dati di interfaccia*. Nel caso in cui lo scambio avvenga tra (N)-entita' alla pari residenti in sistemi diversi, a favore di (N+1)-entita' per le quali le prime stanno fornendo un servizio, si fa riferimento ai cosiddetti *dati di (N)-utente*, ove il termine "utente" deriva dal ruolo che le (N+1)-entita' assumono nei confronti delle (N)-entita' fornitrici di servizio.

Quando invece lo scambio si verifica tra entita' appartenenti a strati adiacenti e residenti nello stesso sistema, le informazioni di dati possono essere trasferite, in alternativa, secondo due casi:

- da una (N+1)-entita' a una (N)-entita' perche' questa provveda al loro inoltro verso un'altra (N+1)-entita' in corrispondenza con la prima su una (N)-connessione;
- da una (N)-entita' a una (N+1)-entita' dopo la loro ricezione da parte della (N)-entita', avendo come origine un'altra (N+1)-entita' con cui la prima e' in corrispondenza su una (N)-connessione.

In ambedue questi casi ci si riferisce a *dati di (N)-interfaccia*.

Analoga distinzione e' effettuabile per le informazioni di controllo. Nel caso di scambio tra (N)-entita' alla pari con l'utilizzazione di una (N-1)-connessione, si parla di *informazione di controllo dell'(N)-protocollo* [(N)-PCI, Protocol Control Information], alludendo al fatto che questo tipo di informazione e' quella corrispondente alle regole di interazione previste nel protocollo dell'(N)-strato. Quando invece lo scambio e' effettuato tra una (N+1)-entita' e una (N)-entita' residenti nello stesso sistema e cioe' utilizza un (N)-SAP, si fa riferimento a *informazione di controllo dell'(N)-interfaccia*.

Indipendentemente dalla loro natura, le informazioni scambiate nell'ambito di una architettura di comunicazione sono strutturate in *unita'*. In relazione al tipo di scambio, si possono avere (Tab. III-1) *unita' di dati dell'(N)-protocollo* [(N)-PDU, Protocol Data Unit] e *unita' di dati dell'(N)-interfaccia* [(N)-IDU, Interface Data Unit].

	Informazione di controllo	Informazione di dati	Unita' di dati
(N)-(N) Entita' alla pari	Informazione di controllo dell'(N)-protocollo	Dati dell'(N)-utente	Unita' di dati dell'(N)-protocollo
(N+1)-(N) Entita' di strati adiacenti	Informazione di controllo dell'(N)-interfaccia	Dati dell'(N)-interfaccia	Unita' di dati dell'(N)-interfaccia

Tab. III-1 - Relazioni tra tipi di informazione e unita' di dati.

Una (N)-PDU e' una unita' informativa che e' specificata in un (N)-protocollo. Essa contiene, in ogni caso, una (N)-PCI e, possibilmente, dati di (N)-utente.

Una (N)-IDU e' invece una unita' informativa che e' scambiata attraverso un (N)-SAP in una singola interazione tra una (N+1)-entita' e una (N)-entita'. E' costituita dalla combinazione di informazioni di controllo e di dati dell'(N)-interfaccia. La lunghezza di una (N)-IDU non e' necessariamente la stessa ad ogni estremita' della (N)-connessione. Il significato di una (N)-IDU e' limitato all'interfaccia tra l'(N+1)-strato e l'(N)-strato.

Quando sia stata instaurata una (N)-connessione, le (N+1)-entita' alla pari che cosi' sono poste in corrispondenza possono scambiarsi informazioni di dati e/o di controllo mediante le cosiddette *unita' di dati dell'(N)-servizio* [(N)-SDU, Service Data Unit]. Una (N)-SDU e' la porzione di dati dell'(N)-interfaccia, la cui identita' e' preservata da un estremo all'altro di una (N)-connessione. L'(N)-servizio di connessione garantisce cioe' l'integrita' di ogni (N)-SDU trasferita, nel senso che ne assicura la delimitazione e garantisce l'ordine dei dati al suo interno.

L'(N)-servizio di connessione trasferisce le (N)-SDU come dati di (N)-utente nelle (N)-PDU. La lunghezza di una (N)-SDU non e' vincolata da quella delle (N)-PDU scambiate tra le (N)-entita' che realizzano il servizio. A questo riguardo, le (N)-SDU possono essere in *corrispondenza uno a uno* con le (N)-PDU; ovvero piu' (N)-SDU possono essere raggruppate per il trasferimento in una singola (N)-PDU; ovvero una singola (N)-SDU puo' essere suddivisa in segmenti e trasferita in una sequenza di (N)-PDU. In Fig. III-5 e' considerato il

caso di corrispondenza uno a uno, in cui, a conferma di quanto detto in precedenza (cfr. Tab. III-1), l'unione di una (N)-PCI e di una (N)-SDU forma una (N)-PDU. Altre corrispondenze saranno considerate in § II.2.5.

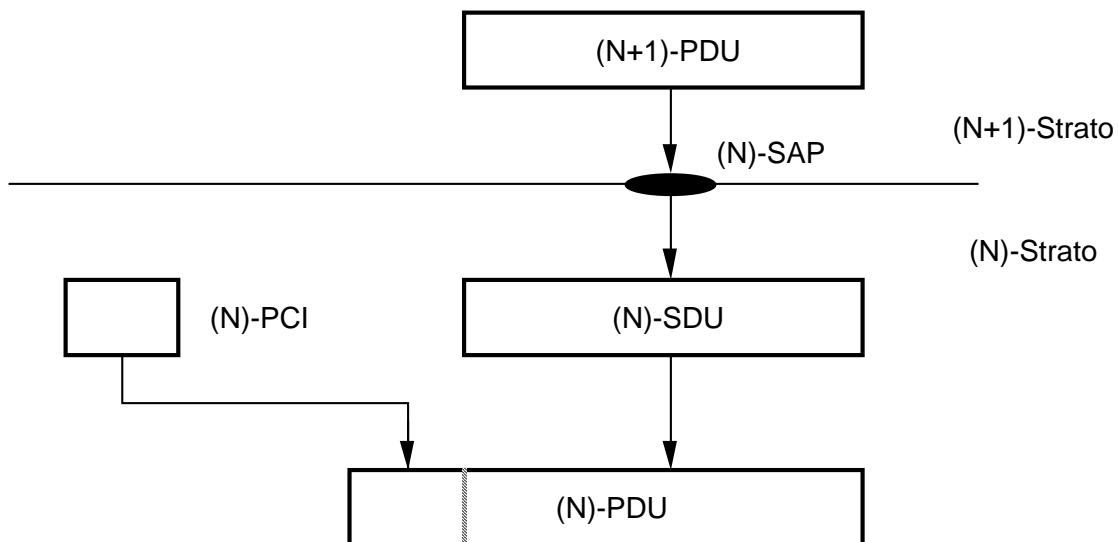


Fig. III-5 - Esempio di relazione tra unità di dati in strati adiacenti.

III.2.4 Gestione delle connessioni

Ci riferiamo dapprima alle funzioni di instaurazione e di abbattimento di una (N)-connessione.

Affinche' due (N+1)-entita' possano comunicare, una di esse deve richiedere all'(N)-strato l'*instaurazione* di una connessione attraverso gli (N)-SAP che le identificano. Le (N)-entita' raggiunte attraverso questi (N)-SAP provvedono a instaurare la connessione richiesta attraverso lo scambio (controllato da un opportuno protocollo) di (N)-PDU su una (N-1)-connessione che, a loro volta, richiedono all'(N-1)-strato. Conseguentemente la richiesta di connessione da parte delle (N+1)-entita' da' luogo verso il basso alla formazione di *N* connessioni attraverso la gerarchia degli strati fino a raggiungere quello piu' basso, in cui le entita' possono comunicare direttamente attraverso il mezzo fisico.

Una (N)-connessione e' instaurata precisando, esplicitamente o implicitamente, un (N)-indirizzo per la (N+1)-entita' che ne fa richiesta e un (N)-indirizzo per ognuna delle (N+1)-entita' che ne sono la destinazione. Una volta instaurata, una (N)-connessione ha due o piu' (N)-CEP. Ognuno di questi e' indiviso per ogni (N+1)-entita' e per ogni (N)-connessione. A esso viene quindi

attribuito, come già detto, un identificatore di riferimento [identificatore di (N)-CEP] a uso esclusivo dell'(N+1)-entità e dell'(N)-servizio.

L'instaurazione di una (N)-connessione presuppone che sia disponibile una (N-1)-connessione tra le (N)-entità che agiscono di supporto e che queste ultime siano in uno stato nel quale possono svolgere lo scambio informativo protocollare per l'instaurazione della connessione. Se la prima condizione non è verificata, deve essere preventivamente instaurata una (N-2)-connessione tra le entità alla pari dell'(N-1)-strato. Ciò richiede, per l'(N-1)-strato, il soddisfacimento delle stesse condizioni richieste per l'(N)-strato. Analoghe considerazioni si applicano per gli strati dell'architettura verso il basso fintanto che si trova una connessione già disponibile ovvero il mezzo fisico.

L'*abbattimento* di una (N)-connessione è normalmente richiesto al termine dello scambio informativo da una delle (N+1)-entità che l'hanno utilizzata. In presenza di condizioni anomale di funzionamento dell'(N)-strato (ad es. a causa di guasti o di errori procedurali), l'abbattimento di una (N)-connessione può essere iniziato anche da parte di una (N)-entità. In questo caso saranno comunicati, alle (N+1)-entità impegnate nella comunicazione, l'avvenuto abbattimento della (N)-connessione e i motivi che hanno portato a tale provvedimento. Alle (N+1)-entità spetta allora il compito di attivare le procedure necessarie per l'eventuale proseguimento della comunicazione.

Per quanto riguarda la corrispondenza tra connessioni appartenenti a strati adiacenti, possono essere individuati tre tipi di relazioni:

- 1) una singola (N)-connessione corrisponde ad una singola (N-1)-connessione;
- 2) una molteplicità di (N)-connessioni corrisponde ad una sola (N-1)-connessione;
- 3) una singola (N)-connessione corrisponde ad una molteplicità di (N-1)-connessioni.

La corrispondenza descritta nel secondo caso è detta funzione di *multiplazione*, mentre nel terzo caso si parla di funzione di *suddivisione*; nella corrispondenza ipotizzata, ambedue queste funzioni sono svolte nell'(N)-strato.

L'esecuzione della funzione di multiplazione è necessaria per rendere più economica ed efficiente l'utilizzazione di un (N-1)-servizio. La funzione di suddivisione è invece impiegata per aumentare l'affidabilità e le prestazioni di una (N)-connessione.

Le funzioni di multiplazione e di suddivisione comportano, all'interno dello strato in cui sono svolte, l'esecuzione di altre funzioni, che possono non essere

necessarie per una corrispondenza uno ad uno tra le connessioni di strato. In particolare, per quanto riguarda la multiplazione svolta nell'(N)-strato, occorre:

- identificare l'(N)-connessione per ogni (N)-PDU che viene trasferita sulla (N-1)-connessione; cio' al fine di assicurare che i dati dell'(N)-utente, supportati da (N)-connessioni tra loro multiplate, non perdano la loro identita' e non possano quindi essere consegnati a una destinazione errata; l'identificazione della connessione e' effettuata mediante l'identificatore di (N)-CEP;
- controllare il flusso su ogni (N)-connessione (cfr. § II.2.5); cio' allo scopo di non compromettere le prestazioni della (N-1)-connessione con l'invio di unita' informative con ritmo superiore alla sua capacita' di trasferimento;
- programmare la prossima (N)-connessione da servire sulla (N-1)-connessione, quando piu' di una (N)-connessione e' pronta a inviare dati.

Nel caso invece della funzione di suddivisione, occorre provvedere a:

- programmare l'ordine di utilizzazione delle (N-1)-connessioni multiple, che sono utilizzate nella suddivisione di una singola (N)-connessione;
- effettuare, in ricezione, una eventuale riseenquenziamento delle (N)-PDU associate a una stessa (N)-connessione, dato che queste sono trasferite su (N-1)-connessioni diverse e quindi possono arrivare alla (N)-entita' di destinazione in una sequenza diversa da quella di emissione, anche se ogni (N-1)-connessione garantisce la sequenzialita' di consegna.

III.2.5 Trasferimento delle informazioni

Si distinguono trasferimenti di *dati normali* e di *dati veloci*. Nel caso di dati normali, si e' gia' visto che le informazioni di controllo e i dati d'utente sono scambiati tra (N)-entita' sottoforma di (N)-PDU. Una (N)-SDU e' trasferita tra una (N+1)-entita' e una (N)-entita', attraverso un (N)-SAP, nella forma di una o piu' (N)-IDU; il suo trasferimento tra (N+1)-entita' alla pari avviene invece sottoforma di dati dell'(N)-utente in una o piu' (N)-PDU.

Una *unita' di dati veloce* (Expedited Data Unit) e' una SDU, che, richiedendo una particolare urgenza di consegna, e' trasferita e/o elaborata con priorita' rispetto alle SDU normali; essa puo' essere utilizzata per scopi di segnalazione o di interruzione. Concettualmente una connessione che supporta anche flussi di dati veloci puo' essere vista come composta di due vie, una per i dati normali, l'altra per i dati veloci; i dati inoltrati sulla via per i dati veloci hanno priorita' sui dati normali. Conseguentemente l'(N)-strato assicura che una

(N)-SDU veloce non sia consegnata a destinazione dopo una (N)-SDU veloce o normale, che sia inoltrata susseguentemente sulla stessa (N)-connessione.

Nel trasferimento di informazione possono essere richieste le funzioni di controllo di flusso, di segmentazione, di aggregazione, di concatenazione e, infine, di sequenzializzazione.

Il *controllo di flusso*, se effettuato, esercita una regolazione del ritmo di trasferimento dell'informazione e può operare, in alternativa, su PDU e su IDU. Nel primo caso, il controllo di flusso operato nell'(N)-strato regola il ritmo con il quale le (N)-PDU sono scambiate tra (N)-entità alla pari su una (N)-connessione. Nel secondo caso, tale funzione regola il ritmo di trasferimento con il quale le (N)-IDU sono scambiate tra una (N+1)-entità e una (N)-entità attraverso un (N)-SAP. Si è già visto che la moltiplicazione operata in uno strato può richiedere un controllo di flusso del primo tipo per ognuno dei flussi moltiplicati.

In generale le unità di dati relative ai vari strati possono avere lunghezze non necessariamente compatibili. In relazione a queste possibili incompatibilità di formato tra unità informative relative a strati adiacenti, uno strato è chiamato a eseguire le funzioni di:

- a) *segmentazione*, se, in emissione, un'unica (N)-SDU deve essere suddivisa in diverse (N)-PDU per il trasferimento su una (N)-connessione; in ricezione, deve allora essere eseguita la funzione complementare, e cioè la *riunificazione*;
- b) *aggregazione*, se il formato delle (N)-PDU è maggiore di quello delle (N)-SDU; in questo caso è necessario riunire varie (N)-SDU in un'unica (N)-PDU; la funzione complementare, svolta all'estremità ricevente è quella di *disaggregazione*.

Nel caso di segmentazione, ogni (N)-PDU, in cui è suddivisa una (N)-SDU, contiene la stessa (N)-PCI, mentre, nel caso di aggregazione, ogni (N)-PDU è formata da due o più (N)-SDU, ognuna delle quali mantiene la propria (N)-PCI. Questi accorgimenti assicurano il mantenimento dell'identità di ogni (N)-SDU segmentata o aggregata.

Un'ulteriore funzione, detta di *concatenazione*, permette di unire in un'unica (N-1)-SDU due o più (N)-PDU; in questo caso la funzione complementare da svolgere in ricezione è quella di *separazione*.

La Fig. III-6 illustra le possibili relazioni tra (N)-SDU, (N)-PDU e (N-1)-SDU entro uno strato e le funzioni ora introdotte. Si osserva che tali relazioni

sono possibili in alternativa alla corrispondenza uno a uno considerata in Fig. III-5.

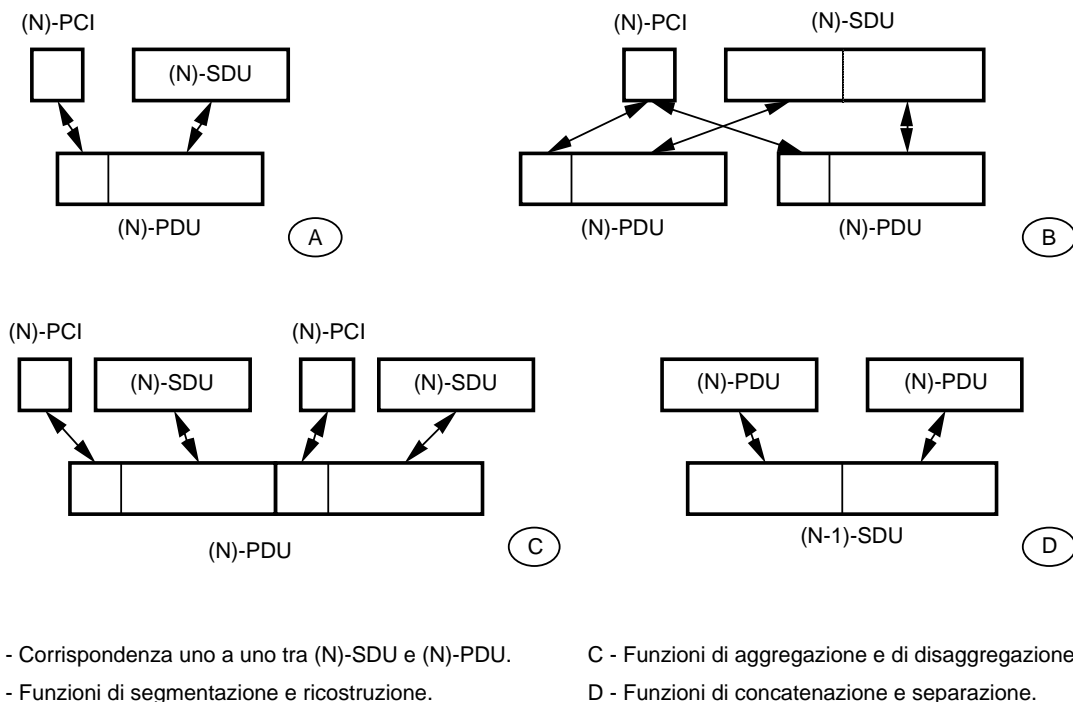


Fig. III-6 - Possibili relazioni tra unità di dati in strati adiacenti.

Consideriamo infine il caso in cui l'(N-1)-servizio non possa garantire la consegna delle (N-1)-SDU nello stesso ordine secondo cui queste sono state consegnate dall'(N)-strato. Se quest'ultimo richiede che venga preservato l'ordine delle (N-1)-SDU, l'(N)-strato deve prevedere una funzione di *sequenzializzazione*, e cioè di riordino delle (N-1)-SDU che sono state trasferite su una (N-1)-connessione.

III.2.6 Condizioni di errore

Durante l'evoluzione della comunicazione all'interno di uno strato, possono verificarsi condizioni di errore. Queste possono essere causate da malfunzionamenti degli apparati, da errori procedurali o da altre cause accidentali, quali errori trasmissivi.

Per fronteggiare condizioni di questo tipo, un (N)-strato deve poter prevedere funzioni di riscontro, di rivelazione e notifica di errore e di reinizializzazione.

La funzione di *riscontro* può essere utilizzata da (N)-entità alla pari, con l'impiego di un (N)-protocollo, per ottenere una più elevata probabilità di

rivelare la perdita di (N)-PDU di quanto non consenta l'(N-1)-servizio. A tale scopo ogni (N)-PDU trasferita tra (N)-entita' e' resa identificabile in modo univoco, in modo che il ricevitore possa informare l'emettitore circa la ricezione di ogni (N)-PDU. Una funzione di riscontro e' anche in grado di inferire la mancata ricezione di (N)-PDU e di attuare appropriate azioni di recupero.

Una (N)-entita' puo' essere in grado di rivelare eventuali condizioni di malfunzionamento dell'(N)-strato. Se cio' si verifica, la (N)-entita' deve notificare l'avvenuto errore alle entita' dello strato superiore. Questo insieme di funzioni, detto di *rivelazione e notifica di errore*, incrementa la qualita' del servizio offerto dallo strato inferiore, poiche' aumenta la probabilita' che un errore sia rivelato e quindi recuperato. Una volta che l'errore sia stato rivelato e notificato, spetta alle (N+1)-entita' adottare le opportune contromisure per recuperare le normali condizioni di funzionamento dello strato.

In alcuni casi, se il tipo di errore e' particolarmente grave, puo' essere necessario operare una *reinizializzazione* (reset) della (N)-connessione. Cio' si verifica, ad esempio, a seguito di una perdita di sincronizzazione tra le (N)-entita' in comunicazione. La funzione di reinizializzazione ha il compito di riportare la (N)-connessione a uno stato precedente, fino a cui si puo' assumere che lo scambio di informazioni sia avvenuto in modo esente da errori.

III.2.7 *Modello astratto del servizio di uno strato*

L'(N)-servizio è visto come logicamente composto da un insieme di *elementi di servizio* che corrispondono a particolari funzioni svolte dall'(N)-strato. Le singole interazioni elementari tra utente e fornitore del servizio attraverso un (N)-SAP, necessarie per attuare un elemento di servizio, sono dette *primitive di servizio*. Queste possono essere classificate secondo quattro tipi:

- a) *primitiva di richiesta*, che e' emessa da un (N)-utente per richiedere l'attivazione di un particolare elemento di servizio;
- b) *primitiva di indicazione*, che e' emessa dall'(N)-fornitore per indicare che l'(N)-utente remoto ha presentato una richiesta di attivazione di un elemento di servizio; lo stesso tipo di primitiva può essere utilizzata dall'(N)-fornitore per indicare all'(N)-utente l'attivazione di una particolare procedura interna al servizio stesso (ad esempio nel caso di notifica d'errore);
- c) *primitiva di risposta*, che e' emessa dall'(N)-utente per completare la procedura relativa ad un elemento di servizio precedentemente attivato da una primitiva di indicazione;

- d) *primitiva di conferma*, che è emessa dall'(N)-fornitore per completare la procedura relativa ad un elemento di servizio precedentemente attivato da una primitiva di richiesta.

È necessario osservare che la definizione delle primitive di servizio non si riferisce necessariamente ad un tipo particolare di realizzazione, ma rappresenta esclusivamente un modello astratto secondo il quale l'utente e il fornitore del servizio interagiscono tra loro. Inoltre può non esistere una corrispondenza tra primitive di servizio e procedure del protocollo interno ad uno strato.

Un elemento di servizio, in relazione alle primitive necessarie alla sua attuazione, può essere classificato nel modo seguente:

- *servizio confermato*, se è richiesto dall'(N)-utente e se necessita di un'esplicita conferma da parte dell'(N)-fornitore; la sua attuazione può richiedere sia l'impiego di tutti i tipi di primitive, sia essere limitata all'impiego delle primitive di richiesta e di conferma;
- *servizio non-confermato*, se è richiesto dall'(N)-utente, ma non necessita di conferma da parte dell'(N)-fornitore; richiede l'utilizzazione delle sole primitive di richiesta e di indicazione;
- *servizio iniziato dal fornitore*, se è attivato autonomamente dall'(N)-fornitore; richiede l'utilizzazione delle sole primitive di indicazione.

La Fig. III-7 illustra il modello astratto delle interazioni tra utente e fornitore di un generico (N)-servizio, mentre la Fig. III-8 illustra gli esempi di servizio confermato, non confermato e iniziato dal fornitore.

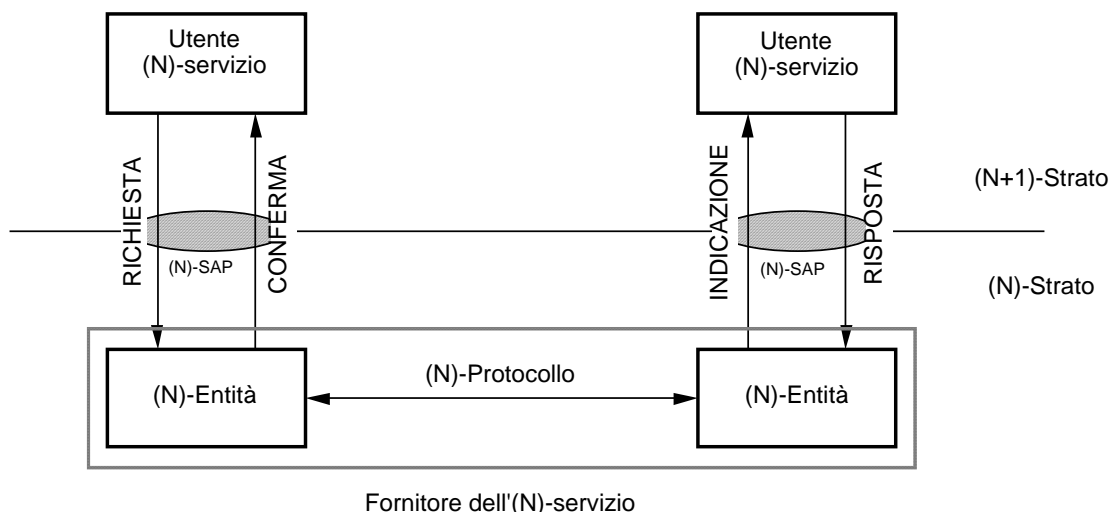


Fig. III-7 - Modello e primitive di servizio di un generico (N)-strato.

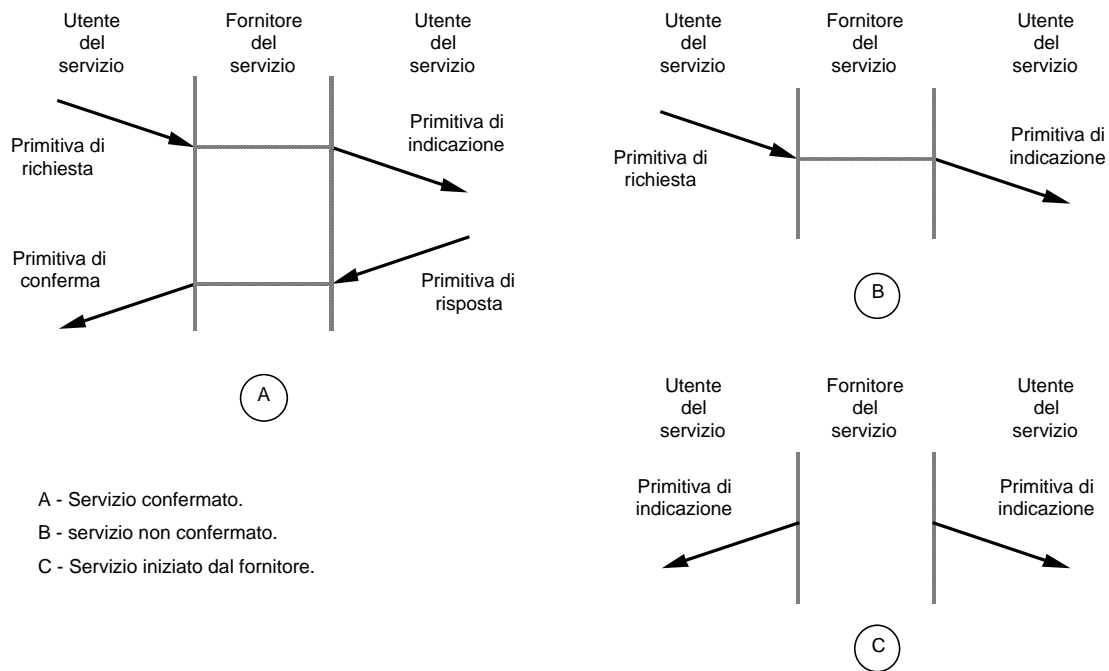


Fig. III-8 - Classificazione dei servizi di strato.

III.3 Modello OSI

Saranno ora presentate le caratteristiche fondamentali del modello OSI, il cui acronimo fa riferimento alla "interconnessione di sistemi aperti". Questo è un termine ormai correntemente utilizzato per definire la comunicazione tra sistemi di elaborazione che adottano un insieme di norme (le *norme OSI*), tali da consentire la cooperazione indipendentemente dalla natura dei sistemi e quindi con lo scopo di rendere possibile l'interconnessione e il colloquio tra sistemi eterogenei (cioè tra sistemi che hanno origine da costruttori diversi), oltre che tra sistemi omogenei.

Infatti il colloquio e la cooperazione tra sistemi di elaborazione può avvenire se, e solo se, le informazioni di dati scambiate sono univocamente interpretabili dai sistemi a colloquio e le procedure seguite nelle varie fasi della comunicazione sono le stesse. Deve cioè sussistere una compatibilità funzionale reciproca tra i sistemi. Questo requisito è soddisfatto a priori nel caso di sistemi di elaborazione omogenei, mentre deve essere ottenuto attraverso particolari accorgimenti nel caso di sistemi eterogenei.

Per assicurare la piena compatibilità funzionale tra sistemi eterogenei è quindi indispensabile stabilire le modalità secondo cui deve avvenire la comunicazione. I costruttori devono conformarsi a queste modalità se

desiderano assicurare alle proprie macchine la possibilità di interconnessione con ambienti diversi.

Questo compito di normalizzazione è stato dapprima affrontato dall'ISO (International Organization for Standardization) e successivamente anche dal CCITT. Il punto di partenza della relativa attività di normalizzazione è consistito nell'analisi dell'intero processo di comunicazione tra due o più sistemi di elaborazione remoti. Tale analisi è stata effettuata con due scopi: evidenziare, da un lato, le molteplici funzioni che debbono essere espletate per rendere possibile il colloquio tra i sistemi di elaborazione; formulare, dall'altro, un modello astratto della comunicazione, che possa essere usato per specificare le particolari modalità di esecuzione delle funzioni stesse.

L'attività dell'ISO, iniziata alla fine degli anni '70, ha condotto alla definizione del modello OSI. Questo è completamente descritto nella norma IS 7498 dell'ISO e nella Racc. X.200 del CCITT.

Il modello OSI adotta i principi dell'architettura a strati descritti in generale nel par.II.2. Anzi si può affermare che tali principi hanno la formulazione attuale proprio come risultato dell'attività di normalizzazione che ha condotto alla definizione di questo modello.

In questo paragrafo vengono dapprima (§ II.3.1) introdotte le caratteristiche generali del modello. Successivamente (§ II.3.2) vengono chiariti i principi seguiti per l'identificazione dei sette strati che compongono il modello OSI. Infine (§ II.3.3-II.3.9) vengono brevemente discussi i compiti di ognuno di questi strati. In questa trattazione si assumerà che i servizi di strato siano con modo a connessione.

III.3.1 Caratteristiche generali del modello

Il modello OSI ha lo scopo di fornire una descrizione astratta delle modalità di comunicazione tra processi applicativi residenti in sistemi di elaborazione diversi, tra loro interconnessi. L'architettura di riferimento è utilizzata per definire e per normalizzare le procedure di comunicazione tra tali sistemi, allo scopo di consentirne la cooperazione.

Nella terminologia OSI, un sistema è definito come un insieme di uno o più elaboratori, comprensivi di periferiche ed, eventualmente, degli operatori umani, in grado di eseguire funzioni di utilizzazione e di trasferimento dell'informazione. Un sistema è detto *aperto* se utilizza le procedure di comunicazione definite dal modello OSI. Un sistema aperto è quindi in grado di

cooperare con qualsiasi altro sistema, la cui architettura e' basata sui principi del modello.

Un processo applicativo e' definito come quell'elemento di un sistema, che esegue elaborazioni delle informazioni per una particolare applicazione.

III.3.2 Identificazione degli strati componenti

L'identificazione degli strati che compongono l'architettura OSI è stata effettuata in base al soddisfacimento delle seguenti regole generali:

- a) funzioni simili sono state raggruppate in uno stesso strato;
- b) funzioni diverse, sia per logica che per tecnologia impiegata nella loro realizzazione, sono state collocate in strati diversi;
- c) l'interfaccia tra due strati è stabilita in modo da minimizzare la complessità ed il numero di interazioni tra gli strati stessi.

L'applicazione di questi principi ha portato alla definizione di sette strati funzionali: lo strato di *applicazione* (strato 7); lo strato di *presentazione* (strato 6); lo strato di *sessione* (strato 5); lo strato di *trasporto* (strato 4); lo strato di *rete* (strato 3); lo strato di *collegamento* (strato 2); lo strato *fisico* (strato 1).

Nel seguito del paragrafo uno strato potrà essere individuato sia con il proprio numero d'ordine sia con le iniziali del termine inglese corrispondente: Application (A); Presentation (P); Session (S); Transport (T); Network (N); Data Link (DL) e Physical (PH).

L'insieme degli strati ora definiti e la loro struttura gerarchica sono mostrati in Fig. III-9.

Lo strato più alto è quello di applicazione; gli strati inferiori forniscono l'insieme dei servizi tramite il quale le entità di applicazione di sistemi diversi possono cooperare. Ogni strato, da quello fisico a quello di presentazione, arricchisce il servizio offertogli dallo strato inferiore. La frontiera tra due strati identifica un particolare stadio del processo di arricchimento. Le funzionalità di uno strato sono governate dallo specifico protocollo utilizzato.

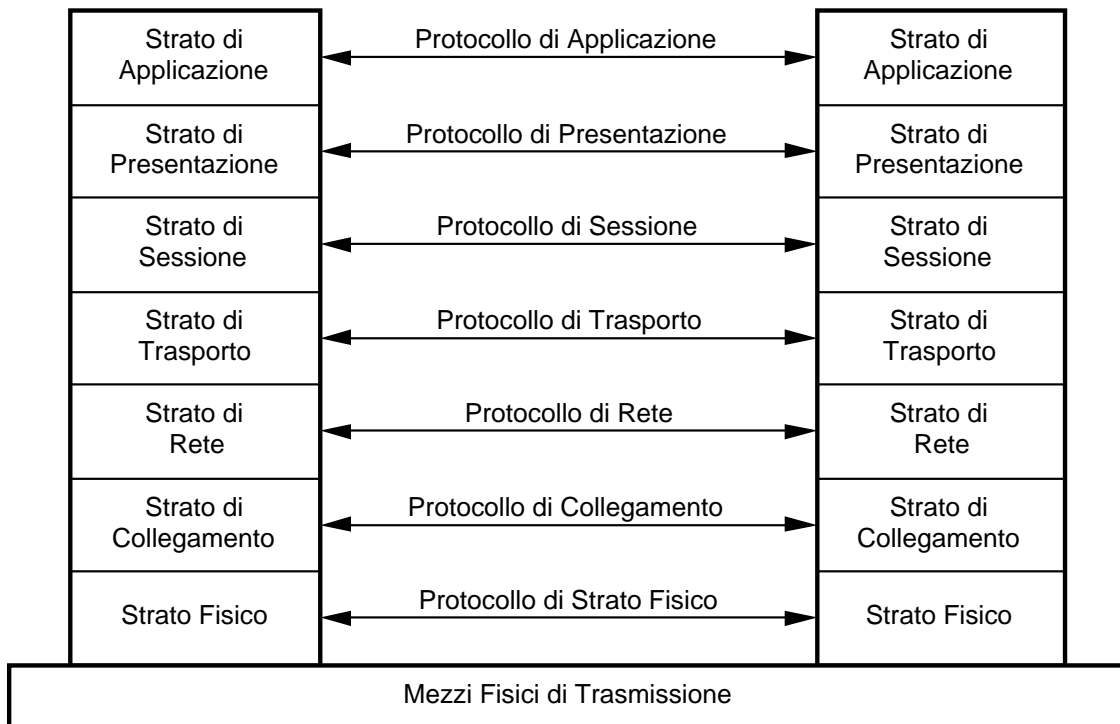


Fig. III-9 - Architettura del modello OSI.

E' il caso di sottolineare che, in un ambiente di comunicazione di dati, l'architettura a sette strati del modello OSI non è l'unica che possa essere derivata dai criteri generali precedentemente esposti. A criteri analoghi sono infatti ispirate altre architetture sviluppate in precedenza dai singoli costruttori, quali ad esempio la *System Network Architecture* (SNA) della IBM e la *Digital Network Architecture* (DNA) della DEC. Allo sviluppo dell'architettura OSI hanno d'altra parte contribuito anche altri aspetti quali:

- il soddisfacimento di esigenze di semplicità, che corrisponde all'avere un numero di strati non troppo elevato;
- l'applicazione delle esperienze maturate nel progetto di sistemi reali, che ha portato ad individuare le interfacce tra gli strati in base ad esempi di cui erano già dimostrate la correttezza e l'efficienza.

Nel seguito saranno descritte le funzionalità dei singoli strati dell'architettura.

III.3.3 Strato di applicazione

Lo strato di applicazione ha lo scopo di fornire ai processi applicativi residenti in un sistema i mezzi per accedere all'ambiente OSI. Poichè lo strato di applicazione è lo strato più elevato dell'architettura, esso non si interfaccia

superiormente con nessun altro strato e, quindi, deve contenere tutti i mezzi per accedere agli strati inferiori.

Appartengono allo strato di applicazione tutte le funzioni che sono utilizzate dalla comunicazione tra sistemi e non sono espletate dagli strati inferiori. Lo strato di applicazione fa sì che l'ambiente OSI costituisca una macchina virtuale in grado di associare un processo applicativo residente in un sistema con qualsiasi altro processo applicativo residente in un qualsiasi altro sistema remoto.

Ogni processo applicativo è associato, all'interno dello strato di applicazione, ad una specifica *entità di applicazione*. I processi applicativi residenti in sistemi diversi scambiano informazioni per mezzo di queste entità, utilizzando il protocollo di applicazione e sfruttando il servizio messo a disposizione dallo strato di presentazione.

Quando si presenta una necessità di comunicazione tra due processi applicativi residenti in sistemi remoti, il processo chiamante richiede, all'entità di applicazione associata, di instaurare, all'interno dello strato di applicazione, una connessione con l'entità di applicazione associata al processo applicativo con cui si vuole colloquiare. A sua volta l'entità di applicazione chiamante richiederà allo strato inferiore (strato di presentazione) l'instaurazione di una connessione, e così via sino a che non sia instaurata anche la connessione fisica tra i due sistemi. Quando le connessioni tra tutte le entità di pari livello sono state attivate, i processi applicativi possono iniziare lo scambio informativo.

Poiché gli utenti dei servizi offerti dallo strato di applicazione possono essere processi applicativi di qualsiasi natura, risulta arduo individuare un insieme di funzioni sufficientemente completo e generale che si adatti a tutte le possibili applicazioni. Per la descrizione delle funzioni interne allo strato di applicazione si segue quindi il criterio di identificare raggruppamenti funzionali omogenei. In altre parole si segue la strada di individuare, dapprima, elementi di servizio comuni a tutte le applicazioni e, in seguito, moduli disgiunti comprendenti elementi di servizio orientati a determinate classi di applicazioni.

Questa impostazione del problema porta alla definizione della seguente partizione dei possibili elementi di servizio dello strato di applicazione (fig. II.3.2):

- *elementi di servizio comuni a tutte le applicazioni* (Common Service Application Elements - CASE);
- *elementi di servizio specifici per classi di applicazioni* (Specific Service Application Elements - SASE);

- *elementi di servizio specifici di particolari applicazioni* (User Specific Service Elements - USSE).

Elementi di servizio specifici di particolari applicazioni (USSE)
Elementi di servizio specifici per classi di applicazioni (SASE)
Elementi di servizio comuni a tutte le applicazioni (CASE)

ENTITÀ DI APPLICAZIONE

Fig. III-10 - Definizione degli elementi di servizio dello strato di applicazione.

Gli elementi di servizio della classe CASE forniscono le potenzialità che sono generalmente utilizzate da tutte le applicazioni. Appartengono a tale classe i seguenti elementi di servizio:

- l'instaurazione di connessioni tra processi applicativi e la loro identificazione;
- la verifica della disponibilità delle risorse necessarie alla comunicazione;
- la determinazione della qualità di servizio richiesta dalle applicazioni;
- la determinazione di una sintassi di trasferimento;
- il trasferimento delle informazioni;
- l'abbattimento delle connessioni di cui in a).

Occorre tener presente che l'elenco ora esposto rappresenta solo un esempio e certamente non comprende tutti i possibili elementi di servizio comuni.

Gli elementi di servizio appartenenti alla classe SASE si concretizzano negli elementi aggiuntivi a quelli appartenenti alla classe CASE per la definizione di servizi particolari.

L'insieme USSE fa riferimento invece alle funzioni di interazione tipiche dell'utente finale.

Esempio III.3-1 - Un tipico esempio di servizio di strato di applicazione è quello di trasferimento e di manipolazione di archivi (File Transfer and Manipulation - FTAM), che consente l'accesso ad archivi remoti per scrivere o per leggere informazioni. Tutti gli elementi che sono utilizzati in questo servizio, diversi da quelli citati per l'insieme CASE, fanno parte

dell'insieme SASE.

Esempio III.3-2 - Le procedure di interrogazione di una banca di dati remota rappresentano un particolare caso di elementi di servizio appartenenti all'insieme USSE.

III.3.4 Strato di presentazione

Lo strato di presentazione ha due compiti fondamentali:

- rendere direttamente disponibili alle entità di applicazione i servizi dello strato di sessione relativi alla strutturazione, alla sincronizzazione e alla gestione del loro dialogo;
- risolvere i problemi di compatibilità tra due qualsiasi entità di applicazione per quanto riguarda la *rappresentazione dei dati* da trasferire e il modo di riferirsi a strutture di dati del sistema remoto.

In altre parole lo strato di presentazione ha lo scopo di sollevare le entità di applicazione da qualsiasi compito relativo alla *trasformazione della sintassi dei dati*.

In generale, all'interno dello strato di presentazione esistono tre tipi di sintassi: due *sintassi locali*, utilizzate dalle singole entità di applicazione, e una *sintassi di trasferimento* utilizzata per lo scambio informativo vero e proprio. La sintassi di trasferimento è negoziata dalle entità di applicazione, mentre la conversione tra sintassi locali e sintassi di trasferimento è effettuata dallo strato di presentazione.

La trasformazione di sintassi riguarda sia la conversione dei codici e degli insiemi di caratteri, sia l'adattamento delle azioni eseguite dalle entità di applicazione sulle strutture di dati.

Esempio III.3-3 - Un esempio di trasformazione di sintassi riguarda la conversione di alfabeti. Nel caso in cui due sistemi utilizzino una diversa codifica dell'informazione, lo strato di presentazione deve transcodificare i messaggi in modo che essi siano riportati nel tipo di codifica utilizzata da ogni sistema.

Esempio III.3-4 - Una funzione particolare che il servizio di presentazione può offrire è quella di *cifratura* dell'informazione da trasferire in modo che essa sia inintelligibile da un eventuale utilizzatore non autorizzato. Il codice crittografico rappresenta in questo caso una particolare sintassi di trasferimento.

III.3.5 Strato di sessione

Lo scopo dello strato di sessione è quello di assicurare alle entità di presentazione una connessione di sessione, e cioè una risorsa logica per organizzare il colloquio tra le entità di presentazione. Ciò ha lo scopo di

strutturare e di sincronizzare lo scambio dati in modo da poterlo sospendere, riprendere e terminare ordinatamente.

Una connessione di sessione è sempre in corrispondenza biunivoca con una connessione tra processi applicativi; in altre parole non è prevista in questo strato, come in quello di presentazione, nessuna funzione di moltiplicazione o di suddivisione.

Lo strato di sessione presuppone che siano risolti tutti i problemi relativi al trasferimento dell'informazione tra i sistemi. Il suo compito consiste invece nel mascherare, per quanto possibile, eventuali interruzioni del servizio di trasporto e nel mantenere una continuità logica nell'evoluzione della connessione di sessione.

Esempio III.3-5 -. Il servizio di sessione deve fare in modo che l'utente non si accorga, nel corso della comunicazione, di eventuali periodi di mancanza del servizio di trasporto (cadute della linea, riattivazioni, ecc.). Nel caso in cui l'utente stia lavorando in modo interattivo con un calcolatore, il presentarsi di questi eventi deve essere percepito al più come un semplice ritardo nella risposta del sistema remoto.

Il colloquio tra due entità di presentazione è strutturato in *attività* che a loro volta sono formate da una o più *unità di colloquio* (fig. II.3.3). Un'attività è definita in base alle finalità dello scambio informativo che deve essere effettuato. L'uso delle attività prevede una fase di inizializzazione e una fase di terminazione. Un'attività può essere interrotta, cioè terminata provvisoriamente, ed essere successivamente ripresa senza conseguenze sulla significatività delle informazioni scambiate; inoltre può essere scartata ed in questo caso viene persa ogni traccia del lavoro effettuato.

Esempio III.3-6 - Spesso un utente ha la necessità di gestire dal proprio terminale più processi contemporaneamente, come l'esecuzione di un programma, la stesura di un testo e l'accesso ad un archivio. In questo caso il passaggio da un processo all'altro, rappresentato sul terminale dalla visualizzazione di una finestra diversa dalla precedente, è gestito dal livello di sessione attraverso la definizione di attività successive.

Un'unità di colloquio è invece definita con lo scopo di creare, all'interno di una attività, punti di sincronizzazione che possono servire come riferimento nel caso in cui si presenti la necessità di riinizializzare il colloquio in seguito a malfunzionamenti della connessione di trasporto.

Un'unità di colloquio è delimitata da punti di sincronizzazione primari, che devono essere riscontrati esplicitamente, e contiene punti di sincronizzazione secondari che non sono soggetti ad esplicita conferma.

Un punto di sincronizzazione primario costituisce un punto da cui il colloquio può essere riinizializzato in caso di guasti o di malfunzionamenti degli strati inferiori.

Esempio III.3-7 - Nell'invio di un blocco di dati particolarmente voluminoso, come nel caso di un testo di diverse pagine, è necessario inserire punti di riferimento in corrispondenza, ad esempio, dell'inizio di ogni pagina. Se dovessero accadere malfunzionamenti nel corso del trasferimento, lo strato di sessione riprenderebbe il trasferimento a partire dalla pagina successiva all'ultima correttamente ricevuta prima dell'errore.

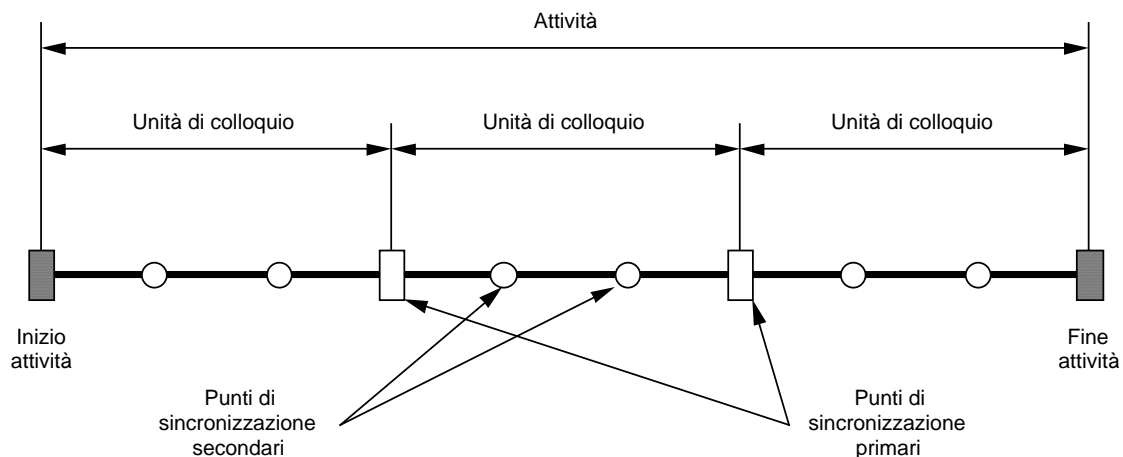


Fig. III-11 - Definizione di attività e di unità di colloquio nello strato di sessione.

Lo scambio di dati attraverso una connessione di sessione può avvenire secondo tre modalità:

- *bidirezionale simultanea*, quando entrambe le entità di presentazione possono emettere unità di dati in qualsiasi momento;
- *bidirezionale alternata*, quando può emettere unità di dati solo un'entità di presentazione alla volta sulla base dello scambio periodico di un diritto alla emissione;
- *unidirezionale*, quando in tutto il colloquio solo una entità di presentazione può emettere unità di dati.

III.3.6 Strato di trasporto

Lo strato di trasporto fornisce alle entità di sessione una connessione di trasporto, che corrisponde ad una risorsa virtuale per il trasferimento trasparente delle unità di dati. Le caratteristiche della connessione di trasporto, stabilite nel momento in cui questa viene instaurata, rimangono costanti sino all'istante in cui viene richiesto il suo abbattimento.

Lo strato di trasporto deve quindi, da un lato, colmare eventuali deficienze e fluttuazioni della qualità di servizio delle connessioni di rete disponibili, e

dall'altro, ottimizzare l'uso del servizio di rete in modo da poter garantire, al costo minimo, le prestazioni richieste.

Esempio III.3-8 - Se il tasso di errore residuo offerto dalla rete è insufficiente, lo strato di trasporto deve prevedere le funzionalità per un nuovo controllo d'errore, addizionale a quello previsto tra le funzioni dello strato di collegamento (cfr. § II.3.8).

Procedendo dal basso verso l'alto nell'esame dell'architettura OSI, lo strato di trasporto è il primo che abbia esclusivamente un significato da estremo ad estremo. In altre parole, mentre le funzionalità dei primi tre strati dell'architettura possono essere presenti sia nei sistemi terminali che nei nodi della rete di comunicazione, le procedure relative allo strato di trasporto sono eseguite solo dai sistemi terminali. Infatti, tutti i problemi relativi all'accesso dei sistemi terminali alla rete di comunicazione, all'instradamento e all'utilizzazione di eventuali reti di transito si considerano completamente risolti dallo strato di rete.

La qualità del servizio offerto da una connessione di trasporto dipende dalle richieste presentate dalle entità di sessione al momento dell'instaurazione della connessione. In particolare, se la qualità di servizio resa disponibile dallo strato di rete è insufficiente, lo strato di trasporto ha il compito di eseguire tutte le funzioni necessarie a portare tale qualità di servizio al livello richiesto dallo strato di sessione. Per fare questo le funzioni dello strato di trasporto sono organizzate in classi di servizio caratterizzate dal soddisfacimento di determinati parametri di qualità di servizio come: portata, ritardo di trasferimento, tasso d'errore residuo, disponibilità della connessione, ecc..

La classe di servizio è scelta al momento dell'instaurazione della connessione di trasporto ed è mantenuta sino al suo abbattimento. Se nel corso della connessione non può essere mantenuta una fissata qualità di servizio, la connessione di trasporto è abbattuta e tale evento viene comunicato allo strato di sessione.

Esempio III.3-9 - Se una rete non fornisce un servizio adeguato dal punto di vista del tasso d'errore, lo strato di trasporto dovrà scegliere una classe di servizio in cui è presente la funzione di controllo addizionale d'errore. Viceversa, se la rete offre un servizio accettabile, la classe di servizio scelta potrà non contenere quella funzione.

Come accennato in precedenza, un compito dello strato di trasporto è quello di ottimizzare l'utilizzazione delle risorse di rete in modo da offrire la qualità di servizio richiesta al costo minimo.

L'ottimizzazione dipende dall'ammontare delle richieste di connessione provenienti dallo strato di sessione e dalla qualità complessiva del servizio di rete. I mezzi che lo strato di trasporto ha per eseguire tale compito consistono nelle funzioni di moltiplicazione e di suddivisione.

Nel primo caso più connessioni di trasporto sono supportate dalla stessa connessione di rete; nel secondo caso più connessioni di rete sono a disposizione della stessa connessione di trasporto.

III.3.7 Strato di rete

Lo strato di rete fornisce i mezzi per instaurare, mantenere ed abbattere le connessioni di rete tra le entità dello strato di trasporto; inoltre fornisce i mezzi funzionali e procedurali per lo scambio di unità di dati tra tali entità.

Ciò significa che lo strato di rete rende invisibile allo strato di trasporto il modo in cui sono utilizzate le risorse di rete, in particolare se sono utilizzate diverse sottoreti in cascata. L'unico elemento che è visibile all'interfaccia con lo strato di trasporto è la qualità di servizio complessiva della connessione.

Se la connessione di rete interessa due sottoreti che offrono differenti qualità di servizio, l'interconnessione tra le sottoreti può avvenire secondo due diverse modalità:

- le reti sono interconnesse senza l'esecuzione di nessuna funzione addizionale; in questo caso la qualità di servizio risultante sarà sicuramente non superiore a quella corrispondente alla rete con minore qualità di servizio; tale modalità è mostrata in Fig. III-12;
- sulla rete con qualità di servizio inferiore sono eseguite funzioni che hanno l'obiettivo di innalzare tale qualità di servizio allo stesso livello della seconda rete; in questo caso, mostrato in Fig. III-13, la qualità complessiva può essere uguale a quella della sottorete a più alta qualità.

La scelta tra queste due alternative dipende dalla differenza di prestazioni delle due sottoreti e da fattori di costo.

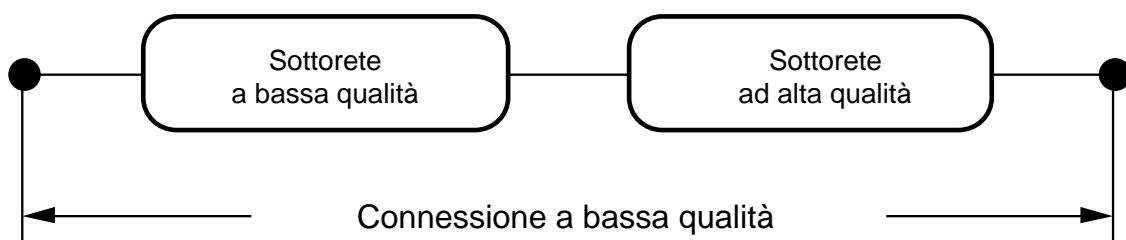


Fig. III-12 - Primo esempio di interconnessione di sottoreti.

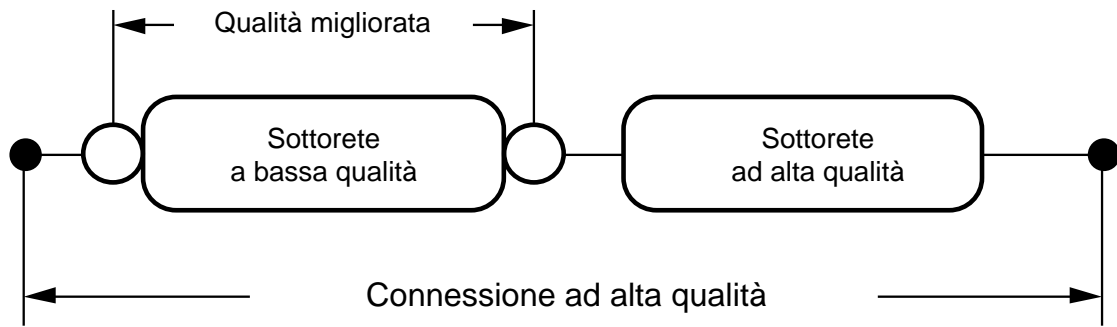


Fig. III-13 - Secondo esempio di interconnessione di sottoreti.

Tra le funzioni che lo strato di rete deve eseguire, particolarmente importante è quella relativa al *controllo del flusso delle unità di dati* sulla connessione di rete. Può accadere infatti che in un certo intervallo di tempo le richieste di trasferimento delle unità di dati provenienti dalle entità di trasporto superino la capacità della connessione di rete, causando la saturazione delle risorse di trasferimento, di memorizzazione e di elaborazione a disposizione dello strato. In questo caso il servizio di rete blocca il trasferimento delle unità di dati dallo strato superiore attraverso l'interfaccia tra i due strati. Il regolare flusso di dati sarà ripristinato nel momento in cui lo stato di sovraccarico è superato.

III.3.8 Strato di collegamento

Lo strato di collegamento fornisce i mezzi funzionali e procedurali per il trasferimento delle unità di dati tra entità di rete e per fronteggiare eventuali malfunzionamenti che dovessero verificarsi all'interno dello strato fisico.

La funzione fondamentale eseguita dallo strato di collegamento è quella di *rivelazione e di recupero degli errori trasmissivi* verificatisi durante il trasferimento sulla connessione fisica. Tale funzione dovrebbe garantire allo strato di rete la fornitura di un servizio di trasferimento dell'informazione esente da errori. Nel caso in cui tali errori siano rivelati, ma risultino non recuperabili, lo strato di collegamento provvede ad abbattere la connessione ed a comunicare l'evento allo strato di rete.

III.3.9 Strato fisico

Lo strato fisico fornisce i mezzi meccanici, elettrici, funzionali e procedurali per attivare, mantenere e disattivare le connessioni fisiche; inoltre ha il compito di effettuare il trasferimento fisico delle cifre binarie scambiate dalle entità di collegamento.

Lo strato fisico è lo strato più basso dell'architettura OSI e si interfaccia direttamente con i mezzi fisici di trasmissione. Le funzioni eseguite al suo interno e i servizi forniti allo strato superiore hanno lo scopo di assicurare l'indipendenza della comunicazione dalle caratteristiche del particolare mezzo trasmissivo utilizzato.

Esempio III.3-10 - Una connessione fisica tra sistemi può essere realizzata mediante l'unione di diverse tratte, ognuna delle quali può essere costituita da un mezzo fisico diverso; lo strato fisico ha lo scopo di rendere invisibile agli strati superiori il tipo di mezzo utilizzato. Se si utilizzano, ad esempio, tratte in fibra ottica che sono in cascata con tratte in cavo coassiale, lo strato fisico esegue, tra le altre funzioni, quella di conversione elettro-ottica dei segnali da trasmettere.

I servizi che devono essere forniti allo strato di collegamento sono:
 la gestione della connessione fisica;
 l'identificazione della connessione fisica;
 la trasmissione delle unità di dati.

Per quanto riguarda la gestione delle connessioni fisiche, occorre sottolineare che una connessione fisica può essere costituita da uno o più circuiti in cascata. Nel secondo caso le entità di strato fisico interne ai sistemi attraversati hanno il compito di eseguire le opportune funzioni di interconnessione tra i diversi circuiti. La Fig. III-14 mostra il modello generale di una connessione fisica.

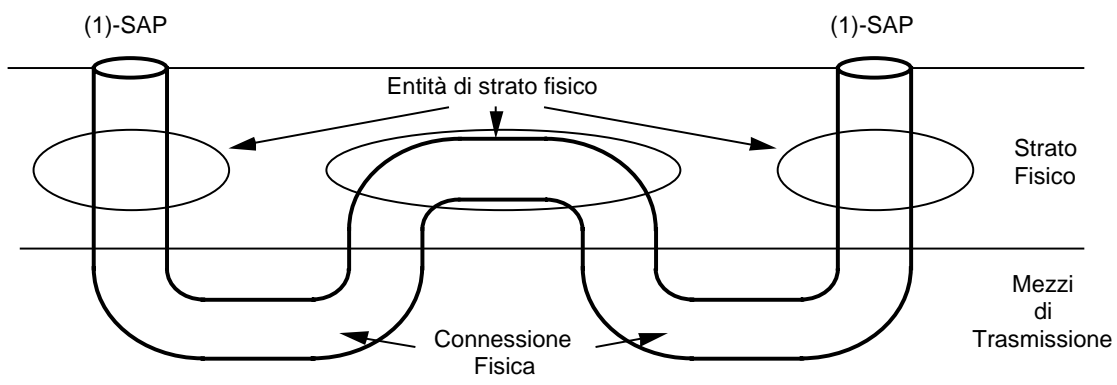


Fig. III-14 - Modello generale di una connessione fisica.

In accordo alla funzione di identificazione della connessione fisica, sono forniti alle entità di collegamento i mezzi per identificare la particolare connessione fisica utilizzata; una connessione fisica può essere sia punto-punto che multipunto.

Infine lo strato fisico assicura il trasferimento sulla connessione fisica delle unità di dati emesse dalle entità di collegamento. Una unità di dati di strato fisico corrisponde ad un bit nel caso di trasmissione seriale, oppure a n bit nel caso di trasmissione parallela.

La trasmissione può essere effettuata con modalità duplice, semi-duplice o semplice. Nella modalità duplice la trasmissione può avvenire contemporaneamente nelle due direzioni; in quella semi-duplice può essere attiva solo una direzione di trasmissione alla volta; nella modalità semplice la trasmissione è invece unidirezionale.

La qualità di servizio di una connessione fisica è caratterizzata dal tasso d'errore, dalla disponibilità del servizio, dal ritmo binario di trasmissione e dal ritardo di trasferimento. Il valore di questi parametri, ad esclusione del ritmo binario di trasmissione, può subire variazioni nel corso della connessione.

IV Tecniche di trasferimento dell'informazione

Attualmente, l'utenza dei mezzi di comunicazione, con riferimento specifico a quella degli ambienti di lavoro, richiede, in misura sempre più pressante, di allargare le proprie possibilità di scambi di informazione. Si è già passati da un quadro di esigenze che possono essere soddisfatte dalla disponibilità di servizi singoli (*utenza mono-servizio*) a un altro in cui la richiesta riguarda un'ampia gamma di possibilità di comunicazione, sotto forma di servizi da fruire in modo integrato (*utenza multi-servizio*).

La risposta a questa esigenza specifica si è manifestata sotto vari aspetti. In primo luogo, si è già da tempo riconosciuta la convenienza di una nuova forma di *integrazione delle tecniche*. Cioè, dai mezzi di comunicazione realizzati con tecnologia completamente analogica, si sta passando a un impiego generalizzato di tecnologie numeriche, basate oggi su componentistica di tipo elettronico e, nel prossimo futuro, anche di tipo ottico. La convenienza di questa trasformazione di natura tecnica risiede in una migliore qualità dei servizi offerti e in ridotti costi di fornitura.

In secondo luogo, con accordo ormai totale nel mondo della tecnica, è giunta ad una prima fase attuativa l'idea di soddisfare le nuove esigenze dell'utenza con risorse di telecomunicazione che prevedano:

- un accesso comune in forma numerica per tutti i servizi di base, relativi allo scambio di voce, di dati e di immagini (*accesso integrato*);
- un mezzo di controllo, preposto a legare le risorse di elaborazione condivise che sono presenti nei mezzi di comunicazione e quelle indivise che si trovano nel sito di utenza, e quindi a mettere a disposizione quanto necessario per ottenere un arricchimento dei servizi di base (*servizi a valore aggiunto*);
- un trasferimento di informazione attuato con modalità esclusivamente numeriche (*connettività numerica da estremo a estremo*), e quindi senza i limiti di qualità insiti nell'impiego delle tecnologie analogiche.

Queste caratteristiche qualificano un obiettivo di sviluppo delle reti di telecomunicazione, che viene chiamato *rete numerica integrata nei servizi* (ISDN - Integrated Services Digital Network) e che prevede l'integrazione dei servizi offerti in un ambiente con impiego generalizzato di tecnologie realizzative di tipo numerico.

Tra gli svariati problemi che sono stati affrontati e che sono tuttora oggetto di studio per perseguire questo obiettivo, il presente capitolo ne focalizza uno di

particolare rilevanza concettuale, e cioè' quello riguardante la definizione dei cosiddetti modi di trasferimento dell'informazione. A questo argomento sono dedicati i sei paragrafi in cui e' articolata la trattazione; i primi cinque (parr. III.1-III.5) riguardano l'informazione d'utente, mentre l'ultimo (par. III.6) si riferisce a quella di segnalazione.

IV.1 I modi di trasferimento

Lo scenario a breve termine delle reti di telecomunicazioni e' caratterizzato dallo sviluppo della *ISDN a banda stretta* (N-ISDN), che integrera' gradualmente le esistenti reti dedicate a un servizio per consentire, in comunicazioni prevalentemente monomediali su base chiamata e con configurazione punto-punto, il trasferimento di voce, di dati e di immagini fisse o con limitate esigenze di movimento; cio' nell'ambito di servizi che, orientativamente, appartengono alle classi a bassa e a media velocita'.

In questo sviluppo si assume, come punto di partenza, che il trasferimento dell'informazione attraverso la rete avvenga con modalita' completamente numeriche sia nella sezione interna che in quella di accesso (*connettivita' numerica da estremo a estremo*). La rete da cui la N-ISDN evolvera' deve essere quindi di tipo *integrato nelle tecniche numeriche*. Nella sua sezione interna deve cioe' sussistere comunanza delle tecniche numeriche impiegate per svolgere le funzioni di trasmissione e di commutazione, mentre e' prevista l'utilizzazione di canali di accesso solo numerici.

Questo risultato implica, secondo gli attuali orientamenti della tecnica, che le informazioni di utente e di segnalazione siano presentate alla rete e vengano da questa trasportate in una opportuna forma di *multiplazione a divisione di tempo* (TDM - Time Division Multiplexing).

Per cio' che riguarda poi gli sviluppi della N-ISDN, e' oggi riconosciuto che, in una prospettiva a medio e lungo termine, le potenzialita' di questa rete debbano essere ampliate per soddisfare le esigenze di comunicazione attualmente emergenti negli ambienti domestico e affari. L'obiettivo finale di questo ulteriore processo evolutivo e' la *ISDN a larga banda* (B-ISDN).

Tali nuove esigenze dovrebbero comprendere, in aggiunta a quelle gia' previste nella N-ISDN:

- comunicazioni su base chiamata con possibilita' di trasferire anche informazioni di dati e di immagine con movimento completo, nell'ambito di servizi ad alta velocita';
- comunicazioni senza connessione per trasferire dati tra reti in area locale e stazioni di lavoro ad alta velocita' elaborativa;
- comunicazioni multimediali con trasferimento, in forma correlata, di flussi informativi composti di voce, di dati e di immagini fisse o in movimento;
- comunicazioni con configurazione multipunto per servizi sia interattivi che distributivi.

Il conseguimento di questi nuovi obiettivi deve tuttavia tenere conto delle caratteristiche dei flussi informativi che la B-ISDN dovrebbe essere in grado di trattare per soddisfare le esigenze di comunicazione sopra elencate. Questi flussi hanno origine da sorgenti numeriche caratterizzate da ritmi binari di picco compresi in una dinamica di almeno sei decadi, e cioe' tra un limite inferiore minore di 1 kbit/s e un limite superiore di circa 1 Gbit/s. Le sorgenti possono essere CBR o VBR e, in questo secondo caso, il grado di intermittenza puo' assumere valori che vanno da qualche unita' ad alcune decine.

Il trattamento di flussi con queste caratteristiche richiede che la rete renda disponibile, come risorsa condivisa, una potenzialita' di trasferimento ad alta capacita' in corrispondenza dell'interfaccia utente-rete e delle apparecchiature della sezione interna. Tale capacita', a cui nel seguito verra' talvolta sostituito, come sinonimo, il termine *banda disponibile*, dovrebbe infatti essere superiore di almeno due ordini di grandezza a quella prevista per una N-ISDN.

In questo ambiente un ruolo chiave ha la scelta del *modo di trasferimento*, e cioe' della modalita' operativa per trasferire l'informazione attraverso la rete. Un primo requisito di questa scelta e' il conseguimento di un adeguato *grado di flessibilita' di accesso* alla banda disponibile. Con questo termine si fa riferimento alla adattabilita' del modo di trasferimento:

- a) nel trattare flussi informativi aventi origine da sorgenti con capacita' di emissione e con caratteristiche di attivita' tra loro anche molto diverse;
- b) nell'assicurare, in modo indipendente dalle caratteristiche del flusso, un elevato rendimento di utilizzazione della banda disponibile.

Un secondo requisito e' il conseguimento di un'accettabile prestazione di qualita' per l'intera gamma dei servizi offerti.

Per chiarire allora le implicazioni delle scelte dei modi di trasferimento, che verranno dapprima operate per la realizzazione della N-ISDN e che

successivamente dovranno essere effettuate per l'evoluzione verso la B-ISDN, e' importante analizzare piu' in profondita' come possa essere caratterizzato un modo di trasferimento e quali siano le conseguenze di tale caratterizzazione sulle prestazioni conseguibili. Cio' consentira', tra l'altro, di riesaminare in modo organico le scelte sistemiche che presiedono il funzionamento delle attuali reti dedicate a un servizio.

In quanto segue verranno dapprima (§ III.1.1) identificate e definite le componenti di un modo di trasferimento. Successivamente (§ III.1.2) si preciseranno le prestazioni del trasferimento che hanno maggiore rilevanza ai fini della qualita' di servizio. Infine (§ III.1.3) verra' sottolineata l'importanza degli studi svolti negli ultimi anni per la definizione di nuovi modi di trasferimento, in vista della realizzazione di una B-ISDN.

IV.1.1 Le componenti di un modo di trasferimento

Le componenti che caratterizzano un modo di trasferimento sono lo *schema di moltiplicazione*, il *principio di commutazione* e l'*architettura protocollare*.

Lo schema di moltiplicazione identifica le modalita' logiche adottate per utilizzare la capacita' di trasferimento dei rami della rete, sia nella sezione di accesso che in quella interna, e cioe' i modi in cui la banda disponibile di questi rami viene condivisa logicamente dai flussi informativi che li attraversano.

Il principio di commutazione riguarda i concetti generali sui quali e' basato il funzionamento logico dei nodi di rete, e cioe' i modi secondo cui l'informazione e' trattata in un nodo per essere guidata verso la destinazione desiderata. In particolare questo principio descrive le modalita' logiche adottate per attraversare i nodi e per utilizzarne la relativa capacita' di elaborazione.

Infine, l'architettura protocollare definisce la stratificazione delle funzioni di comunicazione, sia nell'ambito degli apparecchi terminali che in quello delle apparecchiature di rete (nodi di accesso o di transito). In particolare questa architettura individua le funzioni che ogni nodo deve svolgere sull'informazione in esso entrante e da esso uscente. Si tratta quindi dell'organizzazione delle funzioni che sono espletate, nelle sezioni di accesso e interna della rete, per assicurare il trasferimento entro fissati obiettivi prestazionali.

Dal punto di vista di una architettura a strati, la funzione di commutazione puo' essere vista come il risultato di una demoltiplicazione dei flussi entranti nel nodo e di una loro successiva moltiplicazione in accordo agli indirizzi (impliciti o espliciti) delle unita' di dati che compongono tali flussi. In altri termini, la funzione di commutazione puo' essere espletata sulla base delle stesse

informazioni di protocollo, che sono usate per identificare le varie unita' di dati in un flusso multiplato. Conseguentemente essa puo' essere considerata come appartenente allo strato piu' basso in cui viene effettuata la funzione di multiplazione. Nel seguito le unita' di dati che sono scambiate tra le entita' di questo strato saranno chiamate *unita' informative* o, piu' sinteticamente, *UI*.

Un modo di trasferimento e' completamente definito solo se le sue tre componenti sono adeguatamente descritte. In questo senso ogni componente puo' essere considerata una variabile indipendente, il cui valore puo' essere opportunamente stabilito ai fini del conseguimento di un dato scopo (ad esempio, di fissati obiettivi prestazionali).

IV.1.2 Caratteristiche prestazionali di un modo di trasferimento

Tra i vari tipi di requisiti prestazionali che caratterizzano la qualita' di un servizio di telecomunicazione ne esistono due di particolare rilevanza, e cioe' quelli riguardanti l'*integrita' informativa* e la *trasparenza temporale* dell'operazione di trasferimento.

Con questi due termini ci riferiamo alle diversita' che si possono manifestare, in relazione a un particolare modo di trasferimento, tra:

- 1) la sequenza di cifre binarie emesse dalla sorgente nell'ambito del servizio considerato (*sequenza di emissione*);
- 2) la corrispondente sequenza di cifre binarie ricevute, e cioe' a valle dell'operazione di trasferimento dell'informazione dalla sua origine alla sua destinazione (*sequenza di ricezione*).

L'*integrita' informativa* fa riferimento alle diversita', che si possono manifestare, in modo aleatorio, sia tra le singole cifre binarie emesse e quelle corrispondenti ricevute (*integrita' di cifra binaria*), sia tra le intere sequenze di emissione e ricezione (*integrita' di sequenza di cifre binarie*). Nel primo caso le diversita' sono dovute ad *errori*, normalmente di tipo trasmissivo; nel secondo sono addebitabili solitamente a errori procedurali. Il *grado di integrita' informativa* e' quindi tanto piu' elevato quanto minore e' la frequenza media di errore, misurata, ad esempio nel primo caso, dal *tasso di errore binario*.

Un elevato grado di integrita' informativa e' un requisito particolarmente importante nel caso di comunicazione di dati, mentre, nel caso del servizio telefonico, sono per esso accettabili valori di livello decisamente inferiore senza alterare in modo apprezzabile la qualita' della voce ricevuta. Cio' spiega il motivo per cui, nei servizi di comunicazione di dati, occorre normalmente mettere in atto procedure protettive, che siano in grado di recuperare gli errori

quando questi vengano rivelati e che consentano, in definitiva, di contenere il *tasso di errore binario residuo* a valori orientativamente inferiori a 10^{-10} .

La trasparenza temporale riguarda invece i *ritardi di transito* che differenti UI della sequenza di ricezione possono presentare rispetto alle corrispondenti UI della sequenza di emissione e che sono dovuti a cause di natura varia agenti, in generale, in modo aleatorio. Il *grado di trasparenza temporale* puo' quindi essere valutato quantitativamente con un parametro che qualifichi la variabilita' dei ritardi di transito ed e' tanto piu' elevato quanto minore e' tale variabilita'.

Un modo di trasferimento e' *temporalmente trasparente* quando la variabilita' dei ritardi di transito e' nulla o di entita' trascurabile, cioe' quando i ritardi di transito sono di valore praticamente costante. Non lo e' quando tale condizione non e' soddisfatta. La distinzione tra queste due caratteristiche prestazionali e' chiarita graficamente in Fig. IV-1.

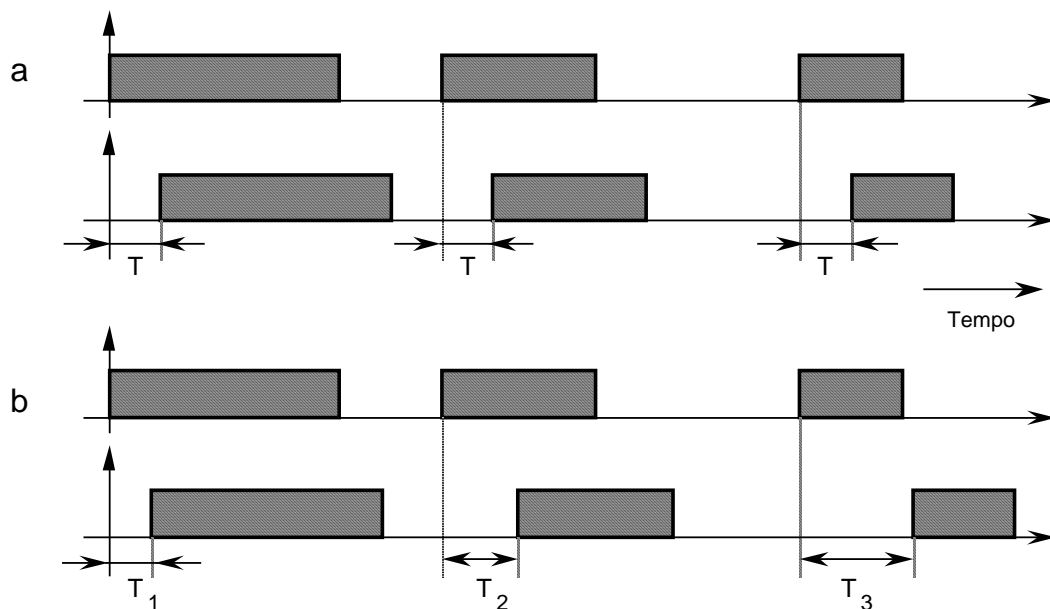


Fig. IV-1 - Sequenze di emissione e di ricezione nei due casi di modo di trasferimento (a) temporalmente trasparente e (b) non temporalmente trasparente.

Esistono servizi, detti *isocroni*, per i quali una corretta interpretazione dell'informazione in ricezione richiede che il trasferimento sia temporalmente trasparente. Sono invece chiamati *anisocroni* i servizi che non hanno particolari esigenze di un elevato grado di trasparenza temporale.

Esempi significativi di servizi isocroni si incontrano in tutti quei casi in cui l'informazione da trasferire sia il risultato di una operazione di conversione analogico-numerica, e cioe' in cui dal segnale analogico originario si estrae una sequenza di campioni, che occorre restituire in ricezione con la stessa

periodicità dell'operazione di campionamento. Pertanto, un esempio tipico di servizio isocrono è offerto dal servizio telefonico. Tra i servizi anisocroni rientrano invece le comunicazioni di dati.

Un servizio isocrono può utilizzare un trasferimento anche non temporalmente trasparente. Ciò può avvenire solo se la distribuzione dei ritardi di transito è tale da presentare un valore medio sufficientemente basso e deviazioni intorno a questo sufficientemente contenute. In tal caso, in ricezione, si può operare una *equalizzazione dei ritardi*, che eleva il grado di trasparenza temporale senza pregiudicare quello di integrità informativa.

Ai fini della qualità di servizio è d'interesse anche il valore di picco dei ritardi di transito. Questo deve essere inferiore a un limite che dipende dal servizio considerato e che è tanto più basso quanto maggiori sono le esigenze di interattività della comunicazione, come nel caso dei servizi di conversazione. Nel caso del servizio telefonico questo limite deve essere inferiore a circa 300 ms.

IV.1.3 Esigenza di nuovi modi di trasferimento

I modi di trasferimento, che sono impiegati nelle reti dedicate, sono il *modo di trasferimento a circuito* e quello *a pacchetto* con le due alternative di servizio a *chiamata virtuale* e a *datagramma*. Il primo modo è adottato nelle *reti a circuito*, il secondo nelle *reti a pacchetto*. Questi modi presentano caratteristiche complementari dai punti di vista dei gradi di flessibilità di accesso e di trasparenza temporale conseguibili.

In particolare, il modo di trasferimento a circuito opera con una quantizzazione della banda disponibile, in cui ogni quanto ha capacità di trasferimento idonea a trasportare un flusso informativo di fissato ritmo binario di picco. Come conseguenza, questo modo di trasferimento fornisce una completa trasparenza temporale, ma a spese di un basso rendimento di utilizzazione della banda disponibile, quando deve trattare flussi emessi da sorgenti VBR con elevato grado di intermittenza; tale rendimento è invece del tutto soddisfacente in presenza di sole sorgenti CBR, purché il loro ritmo binario di picco si armonizzi con i vincoli posti dalla quantizzazione della banda disponibile. Questa limitata adattabilità alle caratteristiche dei flussi di sorgente equivale a bassi valori del grado di flessibilità di accesso.

Il modo di trasferimento a pacchetto utilizza, invece, l'intera banda disponibile senza quantizzazioni. La conseguenza prestazionale è un elevato grado di flessibilità di accesso nei confronti di sorgenti con capacità di

emissione e con caratteristiche di attività anche fortemente diverse; cioè, però, a spese di un basso grado di trasparenza temporale, che è solitamente insufficiente per le esigenze dei servizi isocroni.

Questi limiti prestazionali sono accettabili per una N-ISDN; non lo sono invece per una B-ISDN. Negli ultimi anni è stato allora prodotto un ingente sforzo di ricerca al fine di individuare *nuovi* modi di trasferimento che meglio si prestino alle esigenze di una B-ISDN. Le nuove soluzioni, unitamente a quelle tradizionali, possono essere suddivise in due categorie: i *modi di trasferimento orientati al circuito* e quelli *orientati al pacchetto*. I primi, che includono il modo di trasferimento a circuito, presentano, rispetto ai secondi, un miglior grado di trasparenza temporale e un peggior grado di flessibilità di accesso. I secondi, che includono il modo di trasferimento a pacchetto nelle sue due versioni a chiamata virtuale e a datagramma, hanno prestazioni complementari alle precedenti.

Per chiarire, almeno in prima approssimazione, la distinzione tra modi di trasferimento nuovi e tradizionali, disponiamo questi su un asse come in Fig. IV-2 e con la seguente regola: spostando l'osservazione dall'estremità sinistra dell'asse verso quella destra, si incontrano i modi con crescente grado di flessibilità di accesso, mentre, nel verso contrario, sono collocati ordinatamente i modi con crescente grado di trasparenza temporale. All'estremità sinistra si trova il modo di trasferimento a circuito, mentre a quella destra è posto il modo di trasferimento a pacchetto nella versione a datagramma, preceduto da quello nella versione a chiamata virtuale. In posizione intermedia sono collocati i nuovi modi di trasferimento, che quindi sono il risultato di un compromesso tra i due requisiti prestazionali considerati in Fig. IV-2.

Questo compromesso è stato guidato da due elementi-chiave: da un lato, la possibilità oggi offerta di realizzare nodi di commutazione a pacchetto con portata media decisamente maggiore di quella conseguibile in passato; dall'altro, la disponibilità di canali trasmissivi ad alta velocità e a basso tasso d'errore, quali possono essere supportati da mezzi in fibra ottica.

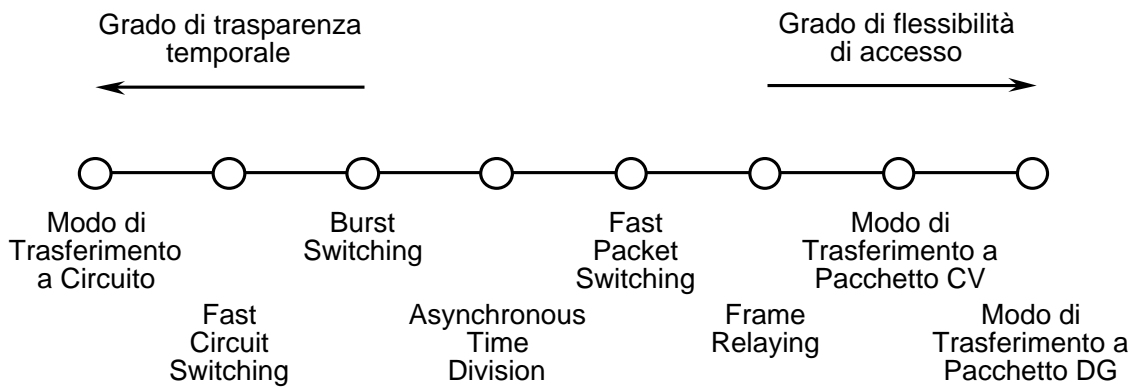


Fig. IV-2 - Relazione tra modi di trasferimento e relative prestazioni, con riferimento ai gradi di trasparenza temporale e di flessibilità di accesso.

Si può subito osservare che la Fig. IV-2 presenta, solo in termini qualitativi, il comportamento prestazionale dei modi di trasferimento ivi rappresentati, senza entrare nel merito delle loro differenze intrinseche. Per superare questo limite e per motivare le scelte, che stanno guidando nella realizzazione della N-ISDN e che sono state effettuate per il nuovo modo di trasferimento da impiegare in una B-ISDN, è necessario effettuare una descrizione delle possibili alternative di realizzazione per ognuna delle tre componenti menzionate nella § III.1.1. E' quanto verrà fatto nei prossimi tre paragrafi IV.2, IV.3 e IV.4, permettendo così di individuare (par. IV.5) le caratteristiche specifiche dei principali modi di trasferimento già impiegati nelle reti dedicate a un servizio e di prevista utilizzazione in quelle integrate nei servizi a banda stretta o larga.

IV.2 Gli schemi di moltiplicazione

Per un ramo di rete, lo schema di moltiplicazione definisce il modo secondo cui le UI entranti condividono logicamente la banda disponibile del ramo (*canale moltiplicato*), per essere trasferite verso l'estremità di uscita. Lo scopo principale di questo schema è quello di rendere efficiente, e quindi economicamente conveniente, l'uso della connessione fisica in corrispondenza dell'interfaccia utente-rete ovvero tra due nodi della sezione interna. Il conseguimento di questo scopo può essere tuttavia subordinato alla esigenza di soddisfare requisiti prestazionali specifici, come, ad esempio, un elevato grado di trasparenza temporale per i servizi isocroni.

Per questa ragione uno schema di multiplazione ideale, per un ambiente di servizi integrati, dovrebbe essere in grado di gestire la banda disponibile di ramo sia su base *dinamica* che *statica*, e cioe' di adottare, come strategie di assegnazione della risorsa, sia, da un lato, una pre-assegnazione collettiva o una assegnazione a domanda (gestione dinamica), sia, dall'altro, una pre-assegnazione individuale (gestione statica).

Occorre tuttavia osservare che, accanto al problema della gestione della banda disponibile, e' presente anche quello delle differenti modalita' di formazione del flusso multiplato, a seconda che l'accesso al canale multiplato sia centralizzato o distribuito. Nel caso di accesso centralizzato, esiste un organo (*multiplatore*), che riceve, dalle sorgenti tributarie, le UI da moltiplicare e che provvede direttamente a formare il flusso multiplato. Cio' avviene assegnando alle sorgenti, in modo statico o dinamico, la banda richiesta e risolvendo le eventuali situazioni di contesa. E' questa la soluzione che normalmente si adotta per la gestione della capacita' di trasferimento dei rami nella sezione interna di una rete.

Nel caso, invece, di accesso distribuito (chiamato solitamente *accesso multiplo*), sono le stesse sorgenti tributarie a dover gestire collettivamente la formazione del flusso multiplato. Questa gestione dovra' avvenire nel rispetto di regole di comportamento, che consentano una efficiente utilizzazione della banda disponibile e un limitato ritardo di attesa. Le regole a cui si allude sono dedicate al *controllo di accesso al mezzo* (MAC - Medium Access Control) e cioe' sono finalizzate allo svolgimento delle funzioni di assegnazione di banda e di risoluzione delle eventuali situazioni di contesa, con l'obiettivo di assicurare equita' di accesso per tutti i richiedenti. Tali regole sono definite nei *protocolli MAC*.

Nel seguito del paragrafo ci riferiremo esclusivamente al caso di accesso centralizzato.

In Fig. IV-3 e' mostrata una possibile classificazione degli schemi di multiplazione a divisione di tempo con accesso centralizzato che sono realizzabili. I criteri seguiti per tale classificazione sono due. Il primo considera il modo secondo cui le entita' emittente e ricevente trattano l'asse dei tempi nello svolgimento delle funzioni di multiplazione e di demultiplazione. Il secondo riguarda le strategie di assegnazione di banda adottate dal multiplatore e i conseguenti compiti del demultiplatore.

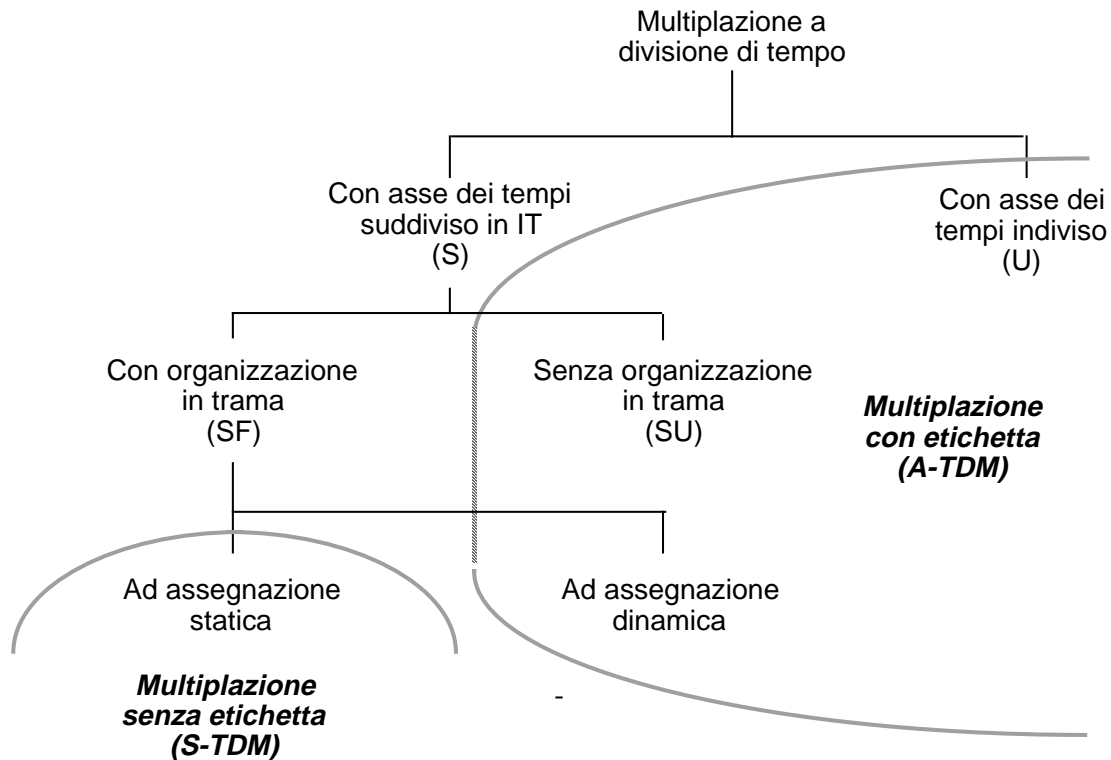


Fig. IV-3 - Classificazione degli schemi di moltiplicazione a divisione di tempo.

IV.2.1 Modalità di gestione dell'asse dei tempi

Secondo il primo criterio di classificazione utilizzato nella rappresentazione di Fig. IV-3, la moltiplicazione a divisione di tempo utilizza il flusso continuo di cifre binarie emesse dal moltiplicatore come un vettore di trasporto in cui le UI provenienti da sorgenti diverse (*sorgenti tributarie*) vengono caricate in un opportuno ordine temporale. Ciò equivale a dire che la moltiplicazione viene effettuata seguendo un principio di separazione a divisione di tempo, che prevede la emissione di UI diverse in intervalli di tempo diversi.

Un primo problema che si pone allora è quello di consentire al demoltiplicatore, nel flusso continuo di cifre binarie ricevute, di individuare ogni singola UI, distinguendo il raggruppamento di cifre binarie che la compongono dalle altre cifre binarie che la precedono e la seguono. Si tratta cioè di operare, nel processo di moltiplicazione, una delimitazione, *implicita* o *esplicita*, di ogni singola UI, in modo che la individuazione di questa in demoltiplicazione possa avvenire in modo univoco.

Un secondo problema che deve essere risolto deriva dalla necessità di indirizzare ogni singola UI, in modo che questa, quando perviene al demoltiplicatore (ad esempio, all'ingresso di un nodo di rete), possa essere

inoltrata verso la corretta destinazione. Anche questa operazione di indirizzamento può essere effettuata in modo *implicito* o *esplicito* e deve essere armonizzata con il principio di commutazione che regola il funzionamento della rete (cfr. § III.2.4).

Nel caso in cui la delimitazione e/o l'indirizzamento siano effettuati in modo esplicito, secondo le modalità che verranno chiarite nel seguito, ad ogni UI deve essere aggiunta, nella sua intestazione, un'*extra-informazione*, che riduce l'efficienza di trasferimento.

Circa il modo in cui le due apparecchiature emittente e ricevente trattano l'asse dei tempi, possono essere identificate due alternative di base (Fig. IV-4): nella prima, chiamata qui *alternativa S* (slotted), l'asse dei tempi è *suddiviso* in *intervalli temporali* (IT) tutti di ugual durata, mentre nella seconda, denominata *alternativa U* (unslotted), tale asse è *indiviso*.

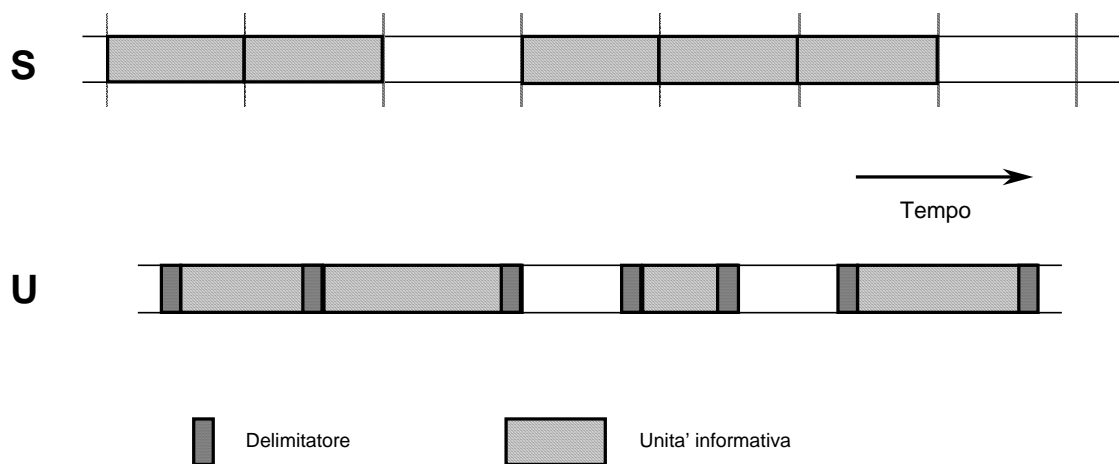


Fig. IV-4 - Multiplazione a divisione di tempo: alternative di attuazione con suddivisione dell'asse dei tempi in intervalli temporali (S) e senza suddivisione (U).

L'alternativa S tratta UI aventi lunghezza, che è costante e commisurata alla durata di un singolo IT. Quest'ultimo è quindi utilizzato per contenere e trasportare una singola UI. Ogni UI risulta così delimitata in modo implicito se esiste una *sincronizzazione di IT* tra moltiplicatore e demoltiplicatore. Cioè la sequenza degli istanti di inizio di ogni IT deve essere a disposizione di ambedue tali apparati terminali. Inoltre, per fronteggiare i casi di perdita di questa condizione di sincronizzazione, deve essere prevista una *procedura di recupero* veloce. Ciò ha lo scopo di ridurre nei limiti del possibile il numero di UI che, durante la condizione di fuori-sincronismo, risultano inevitabilmente perdute, in quanto non delimitabili e quindi non riconoscibili.

Nell'alternativa U, le UI possono essere di lunghezza variabile e vengono emesse senza vincoli sul loro istante di inizio, compatibilmente con le situazioni di contesa che si presentano nell'accesso alla risorsa di trasferimento. Deve allora essere adottata una delimitazione esplicita delle singole UI. Tale operazione può essere effettuata con gruppi di cifre binarie, chiamate *delimitatori* (flags), posti all'inizio e alla fine di ogni UI, e distinguibili in modo univoco da parte del demultiplicatore, nel senso che un delimitatore non deve poter essere simulato da analoghi gruppi di cifre binarie appartenenti ad altre UI.

Esempio IV.2-1 - Un delimitatore non simulabile è quello impiegato per delimitare le trame dei protocolli dello strato di collegamento orientati al bit, del tipo HDLC. Tale delimitatore è costituito da un ottetto di cifre binarie, comprendente uno ZERO, seguito da sei UNO e concluso da uno ZERO. Per evitare la possibilità di simulazione, nel contenuto di una UI in emissione si inserisce uno ZERO (detto di *riempimento*) dopo una sequenza di cinque UNO consecutivi. Il bit di riempimento viene rimosso in ricezione.

L'alternativa S è realizzabile con due possibili soluzioni (Fig. IV-5): una *con organizzazione in trama* (framed), chiamata qui *alternativa SF* e l'altra *senza organizzazione in trama* (unframed), denominata *alternativa SU*.

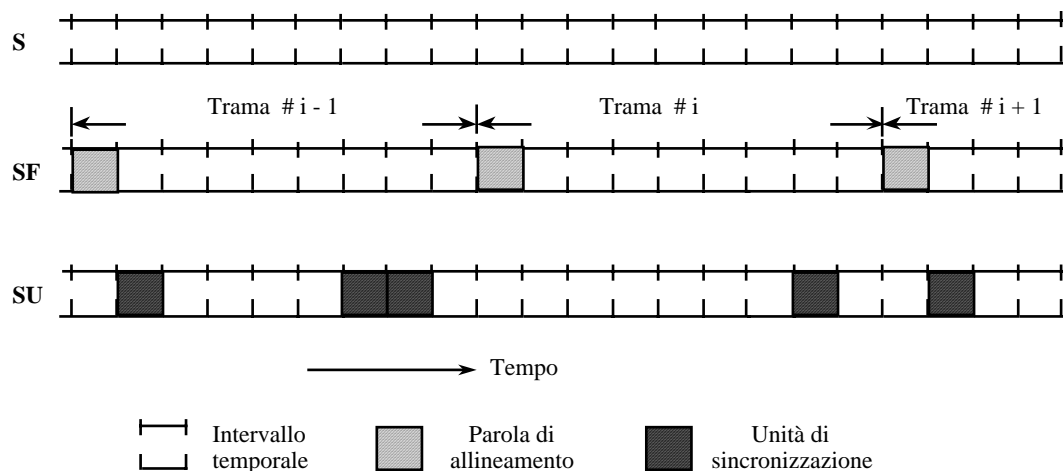


Fig. IV-5 - Soluzioni per l'alternativa con suddivisione dell'asse dei tempi in intervalli temporali (S): con organizzazione in trama (SF) e senza organizzazione in trama (SU).

Nell'alternativa SF, gli IT sono strutturati in *trame*, e cioè in intervalli, tutti di ugual durata, che comprendono un numero intero di IT e che contengono, in posizione fissa (ad esempio, all'inizio della trama), anche una configurazione di cifre binarie, chiamata *parola di allineamento*. È proprio questa che controlla la condizione di *sincronizzazione di trama* e che consente quindi di ottenere

implicitamente anche la sincronizzazione di IT. Per verificare la condizione di sincronizzazione di trama e per il suo recupero una volta che sia stata persa, viene seguita una opportuna strategia (*strategia di allineamento*), che consente di operare anche in presenza di errori trasmissivi nel riconoscimento delle cifre della parola di allineamento e di interruzione del canale multiplato.

Nell'alternativa SU, gli IT si susseguono senza alcuna struttura sovrapposta. Di conseguenza deve essere previsto un opportuno meccanismo di sincronizzazione di IT, quale quello che prevede di inserire, in IT non utilizzati (*IT vuoti*), particolari gruppi di cifre binarie, chiamati *unita' di sincronizzazione*. Il demultipatore riconosce tali unita' e puo' quindi verificare lo stato di allineamento o anche recuperarlo alla prima opportunita' offerta dalla ricezione di una unita' di sincronizzazione. Bisogna tuttavia osservare che il successo di questo meccanismo dipende dalla possibilita' di emettere unita' di sincronizzazione con frequenza media sufficientemente elevata. La tecnica e' quindi tanto piu' efficiente quanto meno il canale multiplato e' caricato. Infatti, man mano che il carico medio del canale aumenta, cresce l'intervallo medio tra due consecutive unita' di sincronizzazione e, conseguentemente, aumentano la probabilita' di perdere il sincronismo e il tempo di recupero dello stesso.

Un differente meccanismo di sincronizzazione di IT, che non presenta questo inconveniente, e' basato su una *auto-sincronizzazione*, ottenuta con una accorta utilizzazione delle intestazioni di UI ovvero con l'impiego di codici di linea a blocchi.

Applicando poi il secondo criterio di classificazione utilizzato in Fig. IV-3, la assegnazione della capacita' del canale multiplato puo' essere *statica* o *dinamica*. Le relative strategie di attuazione seguono i principi generali introdotti in § I.3.3 e verranno descritte in § IV.2.2 e § IV.2.3.

IV.2.2 *Multiplazione a divisione di tempo sincrona*

La assegnazione statica prevede che la capacita' complessiva del canale sia suddivisa in *sub-canali*, aventi capacita' che assumono valori in un insieme discreto. All'instaurazione di una chiamata, un sub-canale e' pre-assegnato individualmente alle necessita' di trasferimento di questa chiamata. La durata dell'assegnazione e' uguale quella della chiamata (*pre-assegnazione individuale su base chiamata*).

Una assegnazione statica puo' essere effettuata solo se viene adottata l'alternativa SF. Infatti la strutturazione in trame crea una associazione logica tra

gli IT che sono collocati nella stessa posizione entro trame differenti. Ciò significa che una sequenza di IT, scelti uno per trama in modo da rispettare tale associazione logica, costituisce un *sub-canale di base*, la cui capacità C_s è facilmente valutabile.

Infatti, indichiamo con L_s e L_f le lunghezze (in cifre binarie) di un IT e di una trama rispettivamente e con C_m la capacità del canale multiplo. Allora

$$(III.2.1) \quad C_s = C_m L_s / L_f .$$

In altre parole la capacità C_m può essere considerata suddivisa in tanti sub-canali di base quanto è il numero N di IT in una trama. Si noti che, in generale, $L_f > NL_s$ in quanto una trama comprende, oltre agli N IT, anche altri gruppi di cifre binarie per scopi di servizio.

Nella modalità di moltiplicazione statica, si possono effettuare assegnazioni *a singolo IT* o *a IT multiplo*. Nel primo caso (*moltiplicazione di base*) le UI generate nell'ambito di una singola chiamata sono trasferite su un singolo sub-canale di base. Nel secondo caso (*sovra-moltiplicazione*), a una chiamata viene reso disponibile un sub-canale di capacità, che è un multiplo di quella del sub-canale di base; l'ordine di molteplicità è uguale al numero di IT assegnati in una trama e non può, ovviamente, essere maggiore di N .

Esempio IV.2-2 - Se la capacità del sub-canale di base è di 64 kbit/s e se la capacità disponibile del canale multiplo è di 1.920 kbit/s (cfr. § IV.4.3), la sovra-moltiplicazione consente di affasciare cinque sub-canali di capacità uguale a 384 kbit/s (6×64 kbit/s) ovvero un insieme di sub-canali con questa capacità e di sub-canali di base, in numero tale che la somma delle relative capacità non sia superiore alla capacità disponibile del canale multiplo.

Secondo la modalità di moltiplicazione statica, le UI generate da una singola sorgente tributaria sono trasferite come una sequenza periodica e con un ritardo di transito costante. Ogni periodo è uguale alla durata della trama e può includere una o più UI in accordo all'ordine di molteplicità dell'assegnazione di IT. In relazione a questa periodicità di trasferimento, tale schema di moltiplicazione è anche chiamato *moltiplicazione a divisione di tempo sincrona* (S-TDM - Synchronous TDM).

Come conseguenza delle prestazioni di trasferimento offerte dalla modalità S-TDM con particolare riferimento alla sua completa trasparenza temporale, questo schema di moltiplicazione è la componente obbligatoria dei modi di trasferimento orientati al circuito e ben si presta alle esigenze prestazionali dei servizi isocroni.

Cio' guida nel dimensionamento della lunghezza L_s di un IT. Assumiamo, infatti, che la capacita' C_s del sub-canale di base sia utilizzata per il trasferimento di un flusso tributario CBR, risultante da un'operazione di conversione analogico-numerica di un segnale analogico (ad esempio, di voce), che e' campionato con periodo T_c e che e' codificato con un numero b fisso di cifre binarie per ogni campione; conseguentemente $C_s = b/T_c$. In queste condizioni, e' conveniente assumere la durata T_f della trama uguale a un multiplo di ordine m ($m = 1, 2, \dots$) di T_c . Con questa scelta, dato che $L_s/C_s = T_f$, la lunghezza di un IT ha un valore uguale a m volte il numero b ($L_s = mb$).

Esempio IV.2-3 - Nella moltiplicazione PCM il sub-canale di base deve trasferire voce di qualita' telefonica con codifica PCM; si assume allora $T_f = T_c = 125 \mu s$, $C_s = 64 \text{ kbit/s}$ (cfr. § I.2.3) e $b = 8 \text{ bit}$; conseguentemente $L_s = 8 \text{ bit}$. Se, invece (cfr. § IV.6.1), si assume $T_f = 250 \mu s$, il ritmo binario risultante da una codifica PCM puo' essere trasferito con un IT di lunghezza uguale a due ottetti di cifre binarie.

Oltre alla moltiplicazione di base e alla sovra-moltiplicazione si puo' anche effettuare una *sotto-moltiplicazione*, che consente di ottenere una capacita' di sub-canale, che e' sottomultipla di quella del sub-canale di base. Si possono a questo scopo adottare due modalita', che chiameremo *a trama singola* o *a multi-trama*.

Nella modalita' a trama singola, il minimo sotto-multiplo ottenibile della capacita' del sub-canale di base e' uguale a C_s/L_s e corrisponde ad assegnare, a un singolo flusso tributario, una sola cifra binaria tra le L_s che compongono l'IT. Conseguentemente, un singolo IT, assegnato con periodicita' di trama, puo' essere il supporto di (al massimo) L_s sub-canali di capacita' uguale a C_s/L_s ovvero di un opportuno insieme di sub-canali con capacita' multiple di C_s/L_s fino a saturare la capacita' C_s .

Esempio IV.2-4 - Se $C_s = 64 \text{ kbit/s}$ e $L_s = 8 \text{ bit}$, la minima capacita' assegnabile e' uguale a 8 kbit/s e il sub-canale di base puo' ospitare un numero di flussi tributari con ritmi binari di 8, 16 o 32 kbit/s, con il vincolo che la somma di questi ritmi non superi il valore di 64 kbit/s.

Per *adattare* poi un flusso tributario con ritmo binario minore di C_s/L_s (o di un suo multiplo) ad una capacita' uguale a C_s/L_s (o a un suo multiplo), si puo' ricorrere al *riempimento* di tale flusso con cifre addizionali e alla *ripetizione* di ogni cifra binaria del tributario.

Esempio IV.2-5 - Un flusso con ritmo di 4,8 kbit/s puo' essere adattato ad una capacita' di 8 kbit/s, riempiendo con 32 cifre binarie addizionali (con scopi di servizio) una sequenza di

48 cifre binarie del tributario; in tal modo la capacita' di 8 kbit/s viene utilizzata per 48/80 della sua piena potenzialita', secondo quanto necessario. Invece, un corrispondente adattamento di flussi tributari con ritmi di 0,6 o 1,2 o 2,4 kbit/s richiede, oltre al provvedimento precedente, anche la ripetizione di ogni cifra binaria di questi tributari per 8 o 4 o 2 volte, rispettivamente. Cio' pero' comporta, per la banda disponibile, un rendimento di utilizzazione che e' tanto piu' basso quanto minore e' il ritmo binario da adattare.

Per cio' che riguarda poi la modalita' a multi-trama, se si richiede che il minimo sotto-multiplo della capacita' C_s sia uguale a C_s/K ($K = 2, 3, \dots$), occorre definire una multi-trama di lunghezza (in cifre binarie) uguale a KL_f e operare su questa come nella moltiplicazione di base o nella sovra-moltiplicazione.

Osserviamo che l'ordine K di sotto-moltiplicazione non puo' essere aumentato al di sopra di un opportuno limite, che e' legato all'esigenza di contenere il *ritardo di riempimento* D di un IT. Tale ritardo, che e' il tempo necessario a riempire un IT con un flusso tributario CBR di ritmo binario uguale a C_s/K , e' infatti uguale a

$$(III.2.2) \quad D = K L_s / C_s .$$

Esso deve essere limitato, in quanto costituisce una componente additiva del ritardo di transito nell'operazione di trasferimento, se, ad esempio, detto flusso tributario risulta da un'operazione di conversione analogico-numerica di un segnale analogico, che e' campionato con un periodo T_c uguale alla durata T_f della trama di base (cioe' $T_c = L_s/C_s$) e che e' codificato con $b = L_s/K$ cifre binarie per campione.

Esempio IV.2-6 - Se la capacita' C_s e' di 2,048 Mbit/s con una durata di trama di base uguale a 125 μ sec, la lunghezza L_s di un IT e' uguale a 256 bit. In queste condizioni, se il tributario CBR ha minor ritmo binario fosse l'uscita di un codificatore PCM e richiedesse quindi una capacita' di 64 kbit/s (cioe' 1/32 di C_s), la multi-trama di moltiplicazione dovrebbe avere una lunghezza uguale a 32 trame di base e il ritardo di riempimento D sarebbe uguale a $32 \times 125 \mu s = 4 \text{ ms}$.

IV.2.3 *Moltiplicazione a divisione di tempo asincrona*

L'assegnazione dinamica puo' essere realizzata con ambedue le alternative S ed U. La capacita' globale del canale moltiplicato e' trattata come una risorsa singola, a cui si puo' accedere in accordo a una opportuna procedura di controllo, dato che la banda e' assegnata a domanda o pre-assegnata collettivamente. Questo secondo tipo di strategia di assegnazione e' normalmente attuato con autorizzazioni di accesso aventi validita' per l'intera durata di una chiamata (*pre-assegnazione collettiva su base chiamata*).

A seguito di queste strategie di assegnazione, e' ogni sorgente tributaria a condizionare l'effettivo ritmo binario di emissione da parte del moltiplicatore, qualunque sia questo ritmo possibilmente variabile durante la comunicazione. Percio', sul canale moltiplicato, la ricorrenza di UI provenienti da una singola sorgente non e' necessariamente periodica. In accordo a tale caratteristica, uno schema di moltiplicazione basato su una assegnazione dinamica e' anche chiamato *moltiplicazione a divisione di tempo asincrona* (A-TDM - Asynchronous TDM).

L'attuazione di una A-TDM richiede sempre la soluzione di contese di utilizzazione e, nel caso di adozione della strategia di pre-assegnazione collettiva, anche di contese di pre-assegnazione. La soluzione di questi due tipi di contese puo' essere attuata con le modalita' descritte in generale in § I.3.4 e I.3.5, rispettivamente.

Per cio' che riguarda la soluzione delle contese di utilizzazione, nella modalita' *orientata al ritardo* una UI entrante nel moltiplicatore puo' subire due diversi trattamenti. In assenza di contesa, la UI viene inoltrata in uscita senza ritardo. In presenza di contesa, occorre invece operare una memorizzazione della UI, in attesa che il canale moltiplicato si renda per lei accessibile. Conseguentemente le UI subiscono ritardi di transito variabili e il grado di trasparenza temporale che ne deriva puo' risultare insufficiente per i servizi isocroni. Per questi ultimi, si impone allora la procedura di equalizzazione dei ritardi, a cui si e' accennato in § III.1.2 e che deve essere attuata o nell'apparato terminale ricevente o in corrispondenza dell'interfaccia utente-rete.

La modalita' *orientata alla perdita* e' basata sullo scarto di porzioni di UI corrispondenti al tempo di attesa affinche' il canale moltiplicato si renda accessibile. Non si ha quindi un deterioramento del grado di trasparenza temporale, ma, corrispondentemente, data la perdita di cifre binarie, si verifica, rispetto alla modalita' orientata al ritardo, una diminuzione del grado di integrita' informativa. Per queste ragioni, la applicabilita' di questo modo dipende strettamente dai requisiti di qualita' di servizio dell'informazione di utente da trasferire ed e' ragionevole solo se il grado di integrita' informativa puo' essere di basso valore.

Infine, la soluzione mista *orientata al ritardo con perdita* puo' avere il vantaggio di distribuire le degradazioni di qualita' di servizio tra le varie comunicazioni in corso.

Passando poi alla soluzione delle contese di pre-assegnazione, occorre una procedura (*controllo di accettazione di chiamata*) per definire il numero massimo di chiamate che possono coesistere sul canale moltiplicato. A tale riguardo esistono le due alternative principali gia' considerate in § I.3.5. Cioe' la

banda disponibile del canale multiplato puo' essere condivisa da una numero di sorgenti tributarie (e quindi di chiamate a queste corrispondenti) tale che si verifichi, in alternativa, uno dei seguenti due casi:

- a) la somma dei ritmi binari *medi* delle sorgenti multiplate sia minore di una opportuna porzione della capacita' di trasferimento del canale multiplato;
- b) la somma dei ritmi binari di picco delle sorgenti non superi detta capacita'.

L'alternativa di cui in a), che e' chiamata *assegnazione a domanda di banda media* e' largamente impiegata nelle reti per dati a pacchetto, ove si parla di *multiplazione statistica*. Essa consente di conseguire la massima efficienza nell'utilizzazione della banda disponibile. Tuttavia, dato che il suo principio e' basato su una descrizione statistica delle caratteristiche di emissione delle sorgenti, possono verificarsi transitoriamente contese di utilizzazione. Al verificarsi di tali condizioni, la domanda istantanea di banda e' maggiore della capacita' di trasferimento disponibile e cio' determina una corrispondente degradazione delle prestazioni relative al trasferimento delle UI; ad esempio, incrementi del ritardo di attesa e/o del grado di perdita, a seconda della modalita' adottata per la soluzione delle contese di utilizzazione. Si nota comunque che tali prestazioni sono assicurate *in media*; ad esempio, nel caso di soluzione orientata al ritardo, viene garantito che il valor medio del ritardo di attesa non superi un massimo prefissato.

Per cio' che riguarda poi l'alternativa di cui in b), e cioe' l'*assegnazione a domanda di banda di picco*, il numero massimo di chiamate che possono essere contemporaneamente abilitate all'accesso e' minore del caso precedente. L'efficienza di utilizzazione della banda e' quindi inferiore, ma con il vantaggio di poter garantire un dato livello di qualita' di servizio. Ad esempio, nel caso in cui si adotti una soluzione orientata al ritardo per la soluzione delle contese di utilizzazione e se le UI sono di lunghezza costante, l'applicazione di questa alternativa assicura che il valore di picco del ritardo di attesa non superi un massimo prefissato.

Per concludere, si puo' affermare che, in base alle caratteristiche delle prestazioni di trasferimento dell'A-TDM, questo schema di multiplazione e' tipico dei modi di trasferimento orientati al pacchetto.

IV.2.4 *Multipolazione con o senza etichetta*

Circa poi le modalita' di indirizzamento, queste debbono essere esplicite nel caso di una A-TDM e implicite quando si adotta una S-TDM. In particolare, nell'A-TDM, l'informazione di indirizzo e' contenuta in una *etichetta*, che viene aggiunta, nell'operazione di multipolazione, ad ogni UI, entrandone a far parte integrante fino allo svolgimento dell'operazione di demultipolazione. Circa l'informazione di indirizzo, questa dipende dal principio secondo cui operano i nodi di rete e puo' corrispondere, in alternativa, all'indirizzo completo della destinazione o a un riferimento della chiamata nell'ambito della quale la UI viene trasferita. Su queste due possibilita' ritorneremo in § III.3.4. In base alla modalita' di indirizzamento in essa impiegata, la A-TDM e' anche denominata *multipolazione con etichetta*.

Nel caso di S-TDM, le UI contenute in un IT sono quelle generate nell'ambito della chiamata a cui tale IT e' stato assegnato; percio' esse sono implicitamente identificate e quindi indirizzate per mezzo della posizione temporale dell'IT nella trama. Non e' quindi necessario un indirizzamento esplicito e si parla allora di *multipolazione senza etichetta*. Tuttavia l'associazione logica tra una chiamata e un IT richiede che, durante la fase di instaurazione, tutte le unita' funzionali coinvolte nella chiamata siano informate di tale associazione. Cio' puo' essere ottenuto tramite lo scambio di informazioni di segnalazione.

IV.2.5 *Multipolazione ibrida*

Per completare il quadro delle tecniche di multipolazione a divisione di tempo, deve essere menzionata una ulteriore alternativa, che e' chiamata *multipolazione ibrida*.

Questa utilizza una struttura a trama, divisa in due *regioni*: una suddivisa in IT, l'altra indivisa. La capacita' associata ad ognuna di queste due regioni puo' essere trattata applicando gli stessi schemi descritti in § III.2.2, l'altra, che puo' essere suddivisa in IT oppure indivisa, la cui capacita' puo' essere trattata applicando gli schemi discussi in § III.2.3. La frontiera tra le due parti della trama puo' essere *fissa* o *mobile*; il caso di frontiera mobile si applica per adattare le capacita' delle due regioni alle richieste di trattamenti secondo le due tecniche S-TDM e A-TDM.

IV.3 Principi di commutazione

La commutazione e' la funzione svolta da un nodo della rete logica. Nell'ambito di comunicazioni punto-punto, essa consente di associare logicamente un qualunque ingresso del nodo (*ramo di ingresso*) con una sua qualunque uscita (*ramo di uscita*). Tale associazione ha una durata temporale quale e' richiesta per trasferire, tra questi rami, le UI scambiate nell'utilizzazione di un servizio di telecomunicazione. Se si opera invece nell'ambito di comunicazioni multipunto, l'associazione di cui sopra deve essere stabilita tra un ramo di ingresso e un numero di rami di uscita uguale all'ordine di molteplicita' della comunicazione.

La funzione di commutazione e' attuata per mezzo delle funzioni di *instradamento* e di *attraversamento*.

L'instradamento e' una *funzione decisionale*, avente lo scopo di stabilire il ramo di uscita verso cui deve essere inoltrata una UI che perviene da un ramo di ingresso. Percio' la scelta risultante dalla funzione di instradamento e' il mezzo che consente ai nodi di rete di guidare le UI verso la loro corretta destinazione.

Tale scelta non e' in generale univoca e deve quindi essere operata in base a opportune regole (*algoritmi di instradamento*). Per fare in modo che la scelta rispetti i requisiti di qualita' del servizio desiderato, gli algoritmi di instradamento piu' evoluti prevedono che questa funzione sia espletata da ogni nodo in modo coordinato con gli altri nodi della rete. Cio' puo' consentire di tenere conto di possibili guasti di nodo o di ramo e di modificare le scelte quando certi segmenti di rete entrano in uno stato di congestione.

L'attraversamento e' una *funzione attuativa* e riguarda il trasferimento delle UI da un ramo di ingresso a uno di uscita attraverso un nodo di rete. Esso pertanto realizza quanto deciso dalla funzione di instradamento e percio' puo' attuarsi solo se la funzione di instradamento e' stata gia' espletata.

La funzione di attraversamento si effettua, individualmente, su ogni UI che perviene al nodo. D'altra parte ogni UI entra nel nodo e ne esce all'interno di flussi multiplati su un ramo entrante e su uno uscente, rispettivamente. Conseguentemente, prima di svolgere la funzione di attraversamento, occorre, per ogni ramo entrante, demultiplexare il relativo flusso, estraendo cosi' le UI che quel ramo trasferisce al nodo. Dopo aver svolto la funzione di attraversamento, le UI, che sono state inoltrate verso il pertinente ramo uscente, debbono essere multiplate per costituire il flusso uscente su questo secondo ramo.

L'attraversamento del nodo avviene seguendo un *percorso interno ingresso-uscita*, per la cui formazione possono essere impiegate due differenti tecniche:

- quella della *connessione diretta*, in cui il transito delle UI avviene in modo temporalmente trasparente;
- quella della *connessione a immagazzinamento e rilancio*, in cui le UI vengono completamente memorizzate allo scopo di poterle elaborare e di consentire loro l'accesso a una operazione di moltiplicazione dinamica; successivamente sono rilanciate in avanti.

Da un punto di vista prestazionale, queste due tecniche si distinguono in base al *ritardo di attraversamento* che viene imposto alle UI che transitano attraverso un nodo di rete da un suo ramo di ingresso a uno di uscita. Tale ritardo coincide con l'intervallo di tempo che intercorre tra l'istante in cui una UI entra completamente nel nodo e quello in cui inizia ad uscirne.

Nel caso della connessione diretta il ritardo di attraversamento e' di valore costante per tutte le UI originate dallo stesso utente. Nel caso invece della connessione ad immagazzinamento e rilancio, tale ritardo assume valori che variano aleatoriamente per ogni UI.

Le possibili alternative per attuare la funzione di commutazione sono descritte nel diagramma ad albero di Fig. IV-6, ove si assumono come riferimenti:

- 1) i possibili servizi di trasferimento, in relazione alle modalita' di attuazione della funzione di instradamento;
- 2) le possibili strategie di assegnazione delle risorse;
- 3) le relazioni temporali tra gli orologi che governano la temporizzazione dei nodi di rete.

Il primo di questi argomenti e' trattato in § IV.3.1, mentre il secondo e' sviluppato in § IV.3.2÷IV.3.4. Infine il terzo e' oggetto di § IV.3.5.

Piu' in particolare, in § IV.3.1, § IV.3.2 e § IV.3.3 a completamento dei contenuti di Fig. IV-6, si stabiliranno le possibili corrispondenze tra le alternative relative ai punti 1) e 2) di cui sopra, le tecniche per l'attuazione della funzione di attraversamento e gli schemi di moltiplicazione sui rami in uscita. Sara' allora possibile identificare varie modalita' per attuare la funzione di commutazione e attribuire ognuna di queste come componente di uno specifico modo di trasferimento.

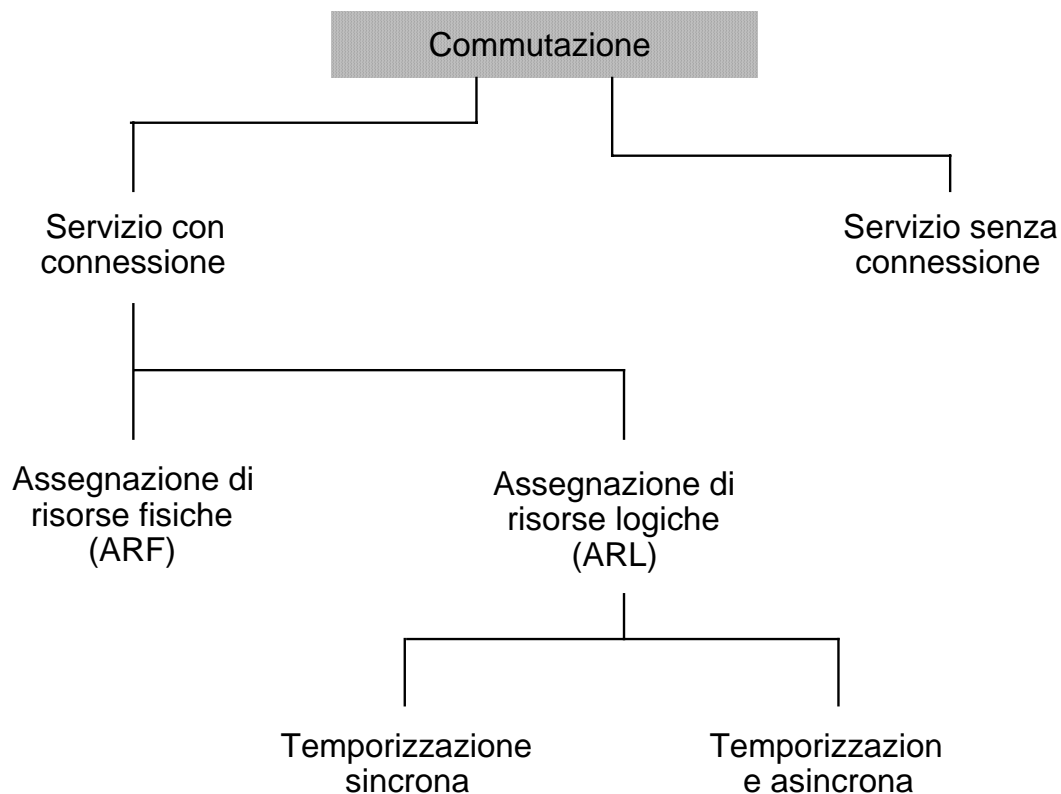


Fig. IV-6 - Classificazione dei principi di commutazione.

IV.3.1 Servizi di trasferimento

In base ai modi per realizzare la funzione di instradamento e con riferimento allo strato architetturale in cui tale funzione viene svolta (strato che coincide con quello MT in cui sono svolte le funzioni di commutazione e di multiplazione), si possono distinguere, come è noto, un servizio di strato *con connessione* e uno *senza connessione*. Nel seguito, per questi due tipi di servizi di strato viene utilizzata, con ovvia motivazione, la denominazione di servizio di trasferimento.

Nel caso di comunicazione punto-punto e di servizio di trasferimento con connessione, viene preliminarmente instaurato un *percorso di rete da estremo a estremo*, e cioè tra gli apparecchi terminali dei due utenti da porre in corrispondenza. Questo percorso di rete è un'associazione temporanea di connessioni da nodo a nodo ed è identificato dalla sequenza di rami e di nodi che occorre attraversare per passare da un estremo all'altro della rete.

Una volta che detto percorso sia stato instaurato, il meccanismo di controllo della rete tratta tutte le UI originate dagli utenti così connessi, conservando l'identità di tali UI tra gli estremi del percorso. Quando questo

deve essere abbattuto, vengono liberate le risorse fisiche o virtuali, che sono state utilizzate per connettere i due utenti.

In un servizio di trasferimento senza connessione, una comunicazione non comporta l'instaurazione preventiva di un percorso di rete da estremo a estremo. Gli utenti consegnano alla rete le loro UI e queste vengono trattate come unità indipendenti, e cioè senza alcun legame con la comunicazione nell'ambito della quale vengono scambiate. La rete garantisce l'accettazione di tali UI nel punto di accesso di origine e, entro certi limiti, la loro consegna nel punto di accesso a destinazione.

Un servizio di trasferimento con connessione è offerto da un nodo di commutazione sotto due forme alternative: una *connessione commutata* o una *connessione semi-permanente*. Nella prima di queste, l'instaurazione può avvenire in qualunque istante a seguito di una richiesta da parte di un utente, mentre l'abbattimento si verifica quando la connessione non è più necessaria. Nella seconda, la connessione è instaurata a seguito di un accordo contrattuale tra cliente e gestore e ha durata che, in relazione a tale accordo, può essere indefinita ovvero definita nell'ambito di un opportuno intervallo di tempo (un giorno, una settimana, ecc.).

Si nota che il caso usuale di un servizio di trasferimento con connessione commutata si verifica nell'ambito di una comunicazione su base chiamata. In casi di questo tipo il nodo opera sulla base di richieste di servizio che sono strutturate secondo le fasi di attività descritte nella § I.2.1. Tali richieste possono essere presentate dagli utenti sia direttamente (come avviene nel caso dei nodi di accesso), che indirettamente (come si verifica per i nodi di transito). I nodi di accesso a cui sono connessi direttamente l'utente chiamante e quello chiamato sono detti *di origine* e *di destinazione*, rispettivamente.

Un servizio di trasferimento con connessione commutata e offerto nell'ambito di una comunicazione su base chiamata presenta quindi tre fasi temporalmente distinte: l'*instaurazione* del percorso di rete da estremo a estremo, il *trasferimento dell'informazione* e l'*abbattimento* del percorso. L'instaurazione avviene per iniziativa dell'utente chiamante, mentre l'abbattimento si attua su richiesta di uno qualunque dei due utenti in comunicazione. Il trasferimento dell'informazione ha luogo secondo le esigenze del servizio di telecomunicazione di cui gli utenti usufruiscono.

Una rete può anche offrire *connessioni permanenti*, che sono instaurate senza l'impiego di organi di commutazione, e cioè, ad esempio tramite una associazione fisica non temporanea di rami di rete. In questo caso, per gli utenti

così connessi, si ha una disponibilità indivisa della connessione in un qualunque istante del periodo stabilito contrattualmente tra cliente e gestore.

Si nota che, in ambedue i casi di connessioni semi-permanenti e permanenti, gli utenti che ne usufruiscono possono procedere direttamente allo scambio di informazione che è di loro interesse, senza preoccuparsi di presentare richieste alla rete affinché questa provveda a una instaurazione o a un abbattimento. Tuttavia i due casi sono sostanzialmente diversi per ciò che riguarda le caratteristiche del modo di trasferimento che ne consegue.

Infatti, mentre nel caso di connessione permanente non si ha alcun coinvolgimento dei nodi di commutazione e quindi delle funzionalità della rete logica, nel caso di connessione semi-permanente questo coinvolgimento è del tutto analogo a quello che si verifica per le connessioni commutate, almeno per ciò che riguarda la fase di trasferimento dell'informazione.

Per connessioni commutate e semi-permanenti si applicano quindi in gran parte tutte le considerazioni che verranno svolte nel seguito di questo paragrafo. Tuttavia, per semplicità espositiva, nel seguito supporremo che, nei servizi di trasferimento orientati alla connessione, gli utenti usufruiscano di connessioni commutate nell'ambito di comunicazioni su base chiamata.

Servizio di strato	Strategia di assegnazione delle risorse di trasferimento	Attraversamento	Multiplazione (TDM)	Modo di trasferimento	
con connessione	pre-assegnazione su base chiamata	individuale (ARF)	connessione diretta	senza etichetta	orientato al circuito
			connessione diretta	senza etichetta	orientato al circuito
		collettiva (ARL)			
senza connessione	assegnazione a domanda		connessione ad immagazzinamento e rilancio	con etichetta	orientato al pacchetto
			connessione ad immagazzinamento e rilancio	con etichetta	orientato al pacchetto

Fig. IV-7 - Possibili modi di attuazione di una funzione di commutazione.

IV.3.2 *Modalita' di utilizzazione delle risorse*

Per espletare la funzione di instradamento, ogni nodo di rete coinvolto in una comunicazione tra due utenti utilizza le proprie risorse di elaborazione. Cio' vale sia per servizi di trasferimento con connessione, sia per quelli senza connessione. Differenti risultano invece, per questi due servizi, le strategie di assegnazione delle risorse e quindi i momenti in cui viene presa la decisione relativa all'instradamento.

Nel caso di servizio di trasferimento senza connessione si opera una assegnazione a domanda, coerentemente con il fatto che ogni UI viene trattata in modo indipendente dalle altre che la precedono o la seguono nell'attraversamento del nodo. In particolare la decisione di instradamento viene presa per ogni UI. In tal modo due UI successive, anche se scambiate nell'ambito della stessa comunicazione, possono essere trasferite lungo percorsi di rete diversi. Le risorse di trasferimento sono utilizzate secondo lo schema di una moltiplicazione con etichetta, mentre l'attraversamento dei nodi e' attuato con la tecnica dell'immagazzinamento e rilancio. Pertanto, quando la commutazione e' attuata secondo questa modalita', essa e' una componente di un modo di trasferimento orientato al pacchetto (Fig. IV-7).

Nel caso invece di servizio di trasferimento con connessione commutata, si adotta una strategia di pre-assegnazione applicata sulla base di una chiamata. Conseguentemente la decisione relativa all'instradamento e' presa, chiamata per chiamata, durante la fase di instaurazione e riguarda la assegnazione delle risorse di trasferimento ed eventualmente di quelle di elaborazione. Tali risorse sono necessarie per costituire, in cooperazione con gli altri nodi della rete, un percorso da estremo a estremo che e' seguito da tutte le UI appartenenti alla stessa chiamata.

Si possono pero' distinguere le due diverse strategie di pre-assegnazione definite in § I.5.3, e cioe' le pre-assegnazioni individuale e collettiva, che in questo contesto sono chiamate *assegnazione di risorse fisiche* (ARF) e *assegnazione di risorse logiche* (ARL), rispettivamente.

IV.3.3 *Assegnazione di risorse fisiche*

Nella strategia ARF le funzioni di elaborazione sono svolte solo durante le fasi di instaurazione e di abbattimento, mentre sono praticamente assenti durante la fase di trasferimento dell'informazione. Per cio' che riguarda poi il trattamento delle risorse di trasferimento, dato che la strategia ARF e' sempre accoppiata con una moltiplicazione senza etichetta (Fig. IV-7), viene effettuata una

quantizzazione della banda con quanti fissi corrispondenti a una assegnazione statica di IT. Inoltre la funzione di attraversamento e' attuata per mezzo di una connessione diretta che opera, in alternativa, con la tecnica *a divisione di tempo* o *a divisione di spazio* e che e' svolta da una struttura di organi di commutazione, chiamata *rete di connessione*.

Nella tecnica a divisione di tempo, che e' usualmente impiegata nelle reti di connessione realizzate in tecnologia numerica, la connessione diretta tra un ingresso e una uscita e' attuata identificando questo ingresso e questa uscita con gli IT che trasportano le UI di una chiamata, rispettivamente, su una trama entrante ed una uscente nell'ambito di uno schema di moltiplicazione statica. Conseguentemente ogni connessione viene attuata trasferendo, con periodicit  uguale alla durata della trama di moltiplicazione, il contenuto dell'IT corrispondente all'ingresso in quello dell'IT corrispondente all'uscita (Fig. IV-8). Questo trasferimento puo' avvenire con uno o piu' scambi di IT .

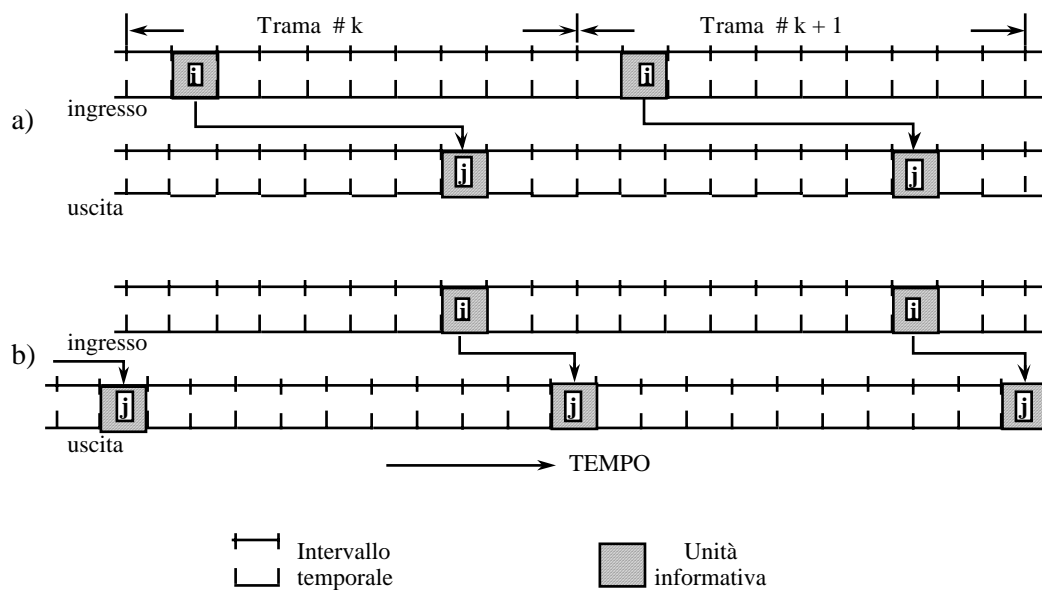


Fig. IV-8 - Tecnica di commutazione a divisione di tempo: scambio di IT
(a) nella stessa trama o (b) in due trame successive.

Con la tecnica a divisione di tempo, il ritardo di attraversamento, che, come in tutte le connessioni dirette, e' di valore costante, e' costituito da due componenti. Una, di valore normalmente trascurabile, corrisponde al ritardo di propagazione tra l'ingresso e l'uscita del nodo. L'altra e' un multiplo della durata di un IT. Se, nell'ambito di una trama con N IT, i e' il numero d'ordine dell'IT corrispondente all'ingresso ($i = 1, 2, \dots, N$) e j quello corrispondente all'uscita (j

= 1, 2, ..., N), tale multiplo e' almeno uguale a $j - i$ o a $N + j - i$ a seconda che sia $j > i$ o $j < i$, e cioe' a seconda che lo scambio di IT avvenga nella stessa trama o in due trame successive. Puo' essere maggiore, se, all'interno della rete di connessione, e' effettuato piu' di uno scambio di IT.

Nella tecnica a divisione di spazio, che e' stata largamente impiegata nelle realizzazioni in tecnologia analogica e che puo' essere utilizzata anche nel caso di realizzazioni in forma numerica, l'informazione di utente, scambiata nell'ambito di una comunicazione, attraversa il nodo su un mezzo fisico individuale. Conseguentemente ogni connessione tra un ingresso e una uscita della rete di connessione e' spazialmente distinta da tutte le altre che contemporaneamente sussistono attraverso il nodo.

Nel caso della tecnica a divisione di spazio, il ritardo di attraversamento e' determinato dal solo ritardo di propagazione tra l'ingresso e l'uscita del nodo ed e' quindi di valore normalmente trascurabile.

Un commutatore a divisione di tempo, realizzato in tecnologia numerica, opera quindi, con la durata di una chiamata, la connessione diretta tra un sub-canale di base del flusso multiplato in ingresso e un corrispondente sub-canale di base del flusso multiplato in uscita. Ad esempio, se la multiplazione statica e' di tipo PCM e se quindi i suoi sub-canali di base hanno capacita' di 64 kbit/s, tale e' anche la capacita' della connessione che viene messa a disposizione della chiamata per tutta la sua durata.

Almeno in linea di principio, possono anche essere attuate connessioni con capacita' che siano multiple o sotto-multiple di quella del sub-canale di base della multiplazione statica. Nel primo caso si parla di *commutazione a multi-IT*, nel secondo di *commutazione a sotto-ritmo*. Le relative modalita' di attuazione sono immediatamente deducibili, ove si tenga conto di quanto detto, in § III.2.2, a proposito della sovra- e della sotto-multiplazione, rispettivamente.

Come conseguenza di queste caratteristiche, un nodo operante secondo la strategia ARF puo' presentare solo situazioni di contesa di pre-assegnazione, che vengono normalmente risolte con modalita' orientata alla perdita.

La strategia ARF e' impiegata esclusivamente nel modo di trasferimento a circuito, di cui costituisce un elemento distintivo. Infatti, con essa, durante la fase di trasferimento dell'informazione in una comunicazione punto-punto, le due parti hanno a disposizione una connessione fisica, che e' caratterizzata da una portata di trasferimento e da un ritardo di transito, che sono costanti per tutta la durata della chiamata.

Questa condizione di trasparenza temporale per l'informazione di utente ha due implicazioni. Da un lato essa offre un canale di trasferimento individuale,

che viene reso disponibile per tutta la durata della chiamata e in maniera esclusiva per le parti in comunicazione. Dall'altro essa obbliga i due utenti alle estremità di questo canale a verificare, prima dell'instaurazione della connessione, la loro compatibilità in termini di ritmo binario emesso e ricevuto, di procedure di controllo, ecc..

IV.3.4 Assegnazione di risorse logiche

La strategia ARL ha lo scopo di ottenere una assegnazione dinamica della banda disponibile. A questo fine le funzioni di elaborazione sono attivate, oltre che nelle fasi di instaurazione e di abbattimento, anche in quella di trasferimento dell'informazione.

Questo obiettivo è normalmente conseguito adottando una moltiplicazione con etichetta e, in tal caso, la strategia ARL è utilizzata dai modi di trasferimento orientati al pacchetto (cfr. fig. III.3.2). Tuttavia la stessa alternativa di assegnazione può essere attuata insieme a una moltiplicazione senza etichetta, se è disponibile una tecnica di segnalazione veloce, quale è quella a canale comune (cfr. par. III.6). In tal caso la strategia ARL può essere utilizzata anche da modi di trasferimento orientati al circuito (cfr. fig. III.3.2).

Quando la strategia ARL è impiegata in unione alla moltiplicazione con etichetta, la funzione di attraversamento è attuata con la tecnica dell'immagazzinamento e rilancio. In questo caso il ritardo di attraversamento è costituito da due componenti:

- una è costante ed è data dalla somma della durata di elaborazione di una UI e del ritardo di propagazione tra l'ingresso e l'uscita del nodo;
- l'altra è variabile ed è determinata dalla durata di memorizzazione di ogni UI all'interno del nodo;

dipende quindi dal carico di lavoro richiesto sia all'unità di elaborazione, che al moltiplicatore dinamico di uscita. Poiché questo carico è una variabile aleatoria (cfr. § I.3.1), ne consegue che anche il ritardo di attraversamento è, nel caso in esame, descrivibile solo in termini probabilistici.

Inoltre, sempre con riferimento alla strategia ARL in unione alla moltiplicazione con etichetta, è opportuno precisare il contenuto informativo dell'etichetta di ogni UI uscente dal nodo di rete coinvolto in una chiamata e trasferita nell'ambito di questa.

La moltiplicazione è riferita a un particolare ramo uscente dal nodo e appartenente al percorso di rete da estremo a estremo, sul quale la chiamata considerata è stata instradata. Per questo ramo viene stabilita una associazione

logica tra le sue estremità, che è chiamata *canale virtuale* (o canale logico) e che permette il trasferimento unidirezionale delle UI inoltrate su quel ramo durante la fase di trasferimento dell'informazione nell'ambito della chiamata. Il criterio di pre-assegnazione di un canale virtuale dipende dai requisiti di qualità di servizio; in particolare, nelle reti per dati a pacchetto, è adottata usualmente l'assegnazione a domanda di banda media (cfr. § III.2.3).

La concatenazione di canali virtuali, lungo il percorso di rete da estremo a estremo assegnato alla chiamata, dà luogo a una connessione logica, che è detta *circuito virtuale*. In queste condizioni l'etichetta associata ad ogni UI può contenere solo una *indicazione di riferimento di chiamata* (IRC), a differenza di quanto accade in un servizio di trasferimento senza connessione, per il quale l'etichetta deve contenere l'indirizzo completo di destinazione.

Il valore dell'IRC può essere specificato in vari modi. Ad esempio, nelle reti per dati a pacchetto è fornito tramite un numero d'ordine del canale virtuale. Il valore dell'IRC è fissato durante la fase di instaurazione. In ogni nodo attraversato dal percorso di rete da estremo a estremo su cui la chiamata è stata instradata, si stabilisce una corrispondenza tra gli IRC riguardanti i canali virtuali entrante e uscente che sono assegnati a quella chiamata. Durante la fase di trasferimento dell'informazione, il valore dell'IRC in ogni UI entrante è trasformato nel valore dell'IRC che identifica il canale virtuale uscente. Questa trasformazione è chiamata *traduzione di IRC*. In § III.5.2 verrà dato un esempio del meccanismo ora descritto con riferimento al modo di trasferimento a pacchetto, quale è impiegato nelle reti per dati.

Nel caso in cui si adotti la strategia ARL in unione a una moltiplicazione con etichetta, possono essere conseguite alcune importanti proprietà. La prima riguarda l'indipendenza tra utente e rete nell'ambito della loro interfaccia, con alcune significative conseguenze: ad esempio, non è necessaria una sincronizzazione tra rete e apparecchio terminale. La seconda proprietà si riferisce alla sezione interna di rete; per questa, una riduzione del ritmo binario di una sorgente comporta un aumento del quantitativo di risorse virtuali che sono rese disponibili per altre sorgenti. La terza proprietà è relativa alla portata di rete e al ritardo di transito: questi parametri prestazionali sono variabili in relazione al carico di rete.

A tale ultimo riguardo, il trasferimento di ogni unità informativa avviene attraverso una successione di *salti* che consentono a questa di passare da un nodo a quello successivo dall'origine alla destinazione. Conseguentemente il ritardo di transito di ogni UI dipende da:

- a) le capacita' di trasferimento dei rami di rete attraversati dal circuito virtuale;
- b) i ritardi di propagazione su questi rami;
- c) la lunghezza (in cifre binarie) della UI;
- d) il numero di salti intermedi;
- e) il ritardo di attraversamento di ogni nodo coinvolto nel trasferimento.

Il rapporto tra la lunghezza di cui in c) e la capacita' di cui in a) riferita a un ramo di rete e' il *tempo di trasmissione* della UI su questo ramo. Inoltre, ogni salto contribuisce alla determinazione del ritardo di transito con la somma del ritardo di attraversamento del nodo a monte e del ritardo di trasferimento della UI sul ramo uscente da questo nodo. Il primo contributo ha le caratteristiche di variabilita' precedentemente sottolineate. Il secondo e' la somma del ritardo di propagazione e del tempo di trasmissione della UI sul ramo in questione.

Infine, un nodo operante secondo la strategia ARL, in unione a una moltiplicazione con etichetta, puo' presentare situazioni di contesa sia di pre-assegnazione che di utilizzazione. Le prime vengono normalmente risolte con una modalita' orientata alla perdita, mentre le seconde, almeno nelle reti per dati a pacchetto, sono affrontate con una modalita' orientata al ritardo e con meccanismi di controllo di flusso.

IV.3.5 Relazioni tra le temporizzazioni negli orologi di nodo

Esaminiamo infine quali debbano essere le relazioni tra le temporizzazioni degli orologi di nodo in ognuno dei principi di commutazione precedentemente descritti.

Nella strategia ARF a divisione di tempo, i flussi di cifre binarie entranti e uscenti in ogni nodo di rete, cosi' come quelli emessi e ricevuti da ogni apparecchio terminale, sono posti in una corrispondenza temporalmente trasparente. Essi debbono quindi essere supportati da sincro-segnali aventi tutti uguale frequenza (almeno in termini medi) e uguale fase.

Questo vincolo puo' essere soddisfatto, innanzitutto, assicurando condizioni di *sincronizzazione di rete*, e cioe' mettendo in atto quanto e' necessario affinche' le frequenze emesse dagli orologi di nodo siano almeno mediamente uguali (*condizione di mesocronismo*) o con scarti relativi che debbono essere contenuti entro ristrettissimi margini di tolleranza (*condizione di plesiocronismo*). Queste condizioni possono essere realizzate mediante opportune *strategie di sincronizzazione*.

Inoltre, occorre impiegare organi di interfaccia tra nodo e rete, che garantiscano, per tutti i flussi numerici entranti, condizioni di *sincronismo di cifra* e di *allineamento di trama*. Infine, la temporizzazione degli apparecchi terminali e di altri apparati (multiplatori, concentratori, ecc.) nella sezione di accesso deve essere asservita alla temporizzazione di rete.

Nella strategia ARL, in unione a una multiplazione con etichetta, possono invece essere impiegate due relazioni di temporizzazione:

- a) quella *sincrona*, in cui, come nel caso della strategia ARF a divisione di tempo, e' richiesta una completa sincronizzazione di rete;
- b) quella *asincrona*, in cui non c'e' necessita' di sincronizzazione di rete, in quanto la temporizzazione dei flussi di cifre binarie entranti puo' essere adattata all'orologio del nodo senza perdita di informazione.

Il funzionamento asincrono puo' essere utilizzato se le UI sono completamente memorizzate entro il nodo, elaborate e successivamente inoltrate verso la loro destinazione. In questo caso, le memorie di ingresso e/o di uscita, se hanno capacita' adeguata, costituiscono una separazione tra le temporizzazioni dei flussi entranti e di quelli uscenti.

L'impiego del funzionamento sincrono puo' essere richiesto allo scopo di ridurre il ritardo di attraversamento. Un esempio significativo e' offerto dall'applicazione del principio del "*cut-through*". Questo deriva dall'applicazione di ambedue le strategie ARF a divisione di tempo e ARL. Se una UI trova una condizione di occupazione sul ramo di uscita di un nodo di rete, il trattamento della UI e' lo stesso di quello della strategia ARL. Se invece il ramo e' libero, il risultato e' simile a quello riscontrabile nell'applicazione della strategia ARF a divisione di tempo. Questo modo di funzionamento e' allora possibile solo se sono soddisfatte le condizioni di sincronizzazione di rete richieste dalla strategia ARF a divisione di tempo.

IV.4 Architetture protocollari

L'architettura protocollare riguarda il modello di riferimento che viene adottato per realizzare uno specifico modo di trasferimento. Essa descrive la stratificazione delle funzioni di comunicazione e stabilisce l'attribuzione di tali funzioni alle apparecchiature di rete. In questo paragrafo ci riferiremo, per semplicita', ai modi di trasferimento riguardanti le informazioni di utente; il caso dell'informazione di segnalazione verra' considerato nel par. III.6.

IV.4.1 Protocolli di accesso e di rete interna

Per analizzare le interazioni tra le sezioni di accesso e interna di una rete di telecomunicazioni, consideriamo l'evoluzione di un processo di comunicazione che coinvolge due apparecchi terminali. Ogni interazione riguarda coppie di entità residenti in questi ultimi o nelle apparecchiature di rete. Le modalità organizzative ed esecutive di queste interazioni, per consentire a queste una attuazione coordinata, sono normalmente descritte tramite un modello di riferimento, che applica i principi dell'architettura stratificata.

Nel modello generale di riferimento in Fig. IV-9 vengono distinti due apparecchi terminali, due nodi di accesso e un nodo di transito. Le due parti della sezione di accesso consentono ad ogni apparecchio terminale di interagire con il pertinente autocommutatore locale attraverso l'interfaccia utente-rete. La sezione interna interconnette i nodi di accesso attraverso il nodo di transito e le relative interfacce nodo-nodo.

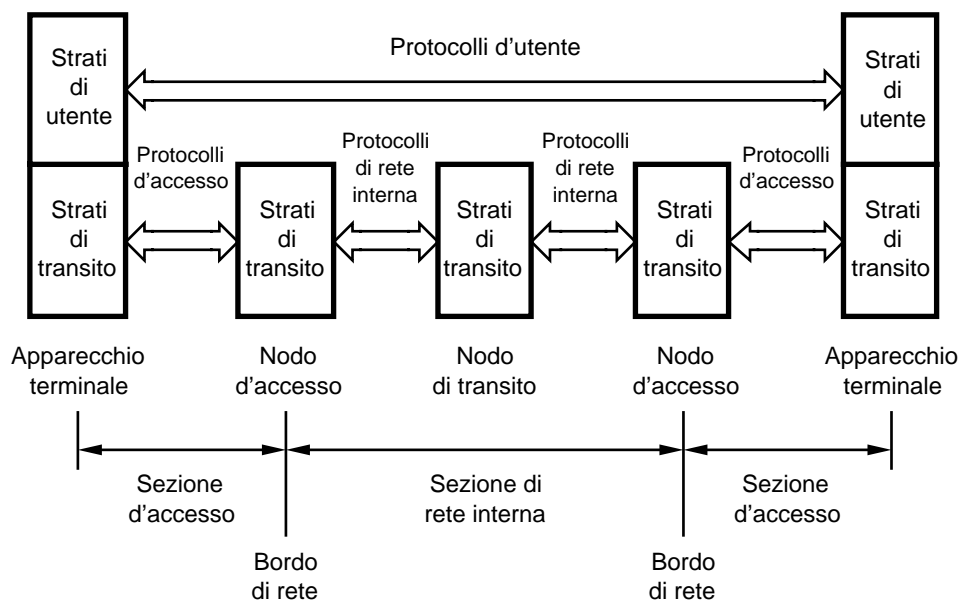


Fig. IV-9 - Modello di riferimento generale di un processo di comunicazione.

Secondo la terminologia del modello OSI, oggi impiegata anche in contesti diversi da quello delle comunicazioni di dati in cui detto modello è stato sviluppato, gli apparecchi terminali del modello di Fig. IV-9 possono essere considerati come sistemi terminali e possono quindi essere visti come organizzati in tanti strati funzionali quanti ne sono previsti nell'architettura di comunicazione considerata. I nodi possono invece essere trattati come sistemi di rilegamento; ciascuno di essi comprenderà quindi un sottoinsieme di detti strati,

che, partendo da quello di ordine piu' basso, arriva fino allo strato piu' alto che include una funzione di rilegamento. Questo sottoinsieme e' indicato in Fig. IV-9 come quello degli strati di transito, mentre lo strato piu' alto nel sottoinsieme verra' indicato qui di seguito, per univocita' di riferimento, come strato S. I rimanenti strati dell'architettura sono invece chiamati strati di utente.

L'identificazione degli strati di transito e, in particolare, dello strato S qualifica il modo di trasferimento utilizzato per trasportare l'informazione attraverso la rete e puo' condurre a scelte diverse a seconda che si considerino i nodi di accesso da un lato e quelli di transito dall'altro.

L'architettura protocollare di un processo di comunicazione che si svolge attraverso una rete geografica prevede sempre la definizione di protocolli di accesso, di protocolli di rete interna e di protocolli di utente.

I protocolli di accesso regolano le interazioni tra un apparecchio terminale e la relativa terminazione di rete; essi appartengono agli strati di transito fino a quello di ordine piu' elevato tra questi, che e' presente nell'interfaccia utente-rete.

I protocolli di rete interna definiscono invece le regole di interazione tra un nodo di accesso e un nodo di transito, ovvero tra due nodi di transito, ovvero direttamente fra i due nodi di accesso; essi appartengono agli strati di transito fino a quello di ordine piu' elevato, che e' presente nelle interfacce nodo-nodo.

Infine i protocolli di utente riguardano gli strati omonimi e vengono gestiti da estremo a estremo, nel senso che le entita' alla pari interagenti per il loro tramite risiedono negli apparecchi terminali.

In Fig. IV-9 e' considerato il caso in cui gli strati di transito sono gli stessi sia nei nodi di accesso che in quelli di transito. In questo caso, ogni nodo, che deve interagire con tutte le apparecchiature di utente e di rete con cui si interfaccia, gestisce, per ognuno degli strati di transito, il protocollo che e' pertinente ad ognuna delle interfacce di interazione. La gestione di questi protocolli e' allora effettuata sezione per sezione.

Conseguentemente, in un processo di comunicazione quale e' ipotizzato nel modello di riferimento in Fig. IV-9, un nodo di accesso deve operare una conversione tra il protocollo di accesso e quello di rete interna, mentre un nodo di transito deve effettuare una rigenerazione del protocollo di rete interna. In ambedue i casi devono essere effettuate una chiusura del protocollo relativo all'interfaccia a monte e una apertura di quello relativo all'interfaccia a valle.

Quando invece, a differenza di quanto ipotizzato in Fig. IV-9, uno o piu' tra gli strati di transito di ordine gerarchico piu' elevato siano presenti nei soli nodi

di accesso e non in quelli di transito, i protocolli di rete interna sono di due tipi come in Fig. IV-10:

- 1) quelli appartenenti agli strati di transito piu' bassi, che sono presenti sia nei nodi di accesso, che in quelli di transito;
- 2) quelli appartenenti agli strati di transito piu' alti, che sono presenti nei soli nodi di accesso.

I primi rientrano nel caso gia' considerato in Fig. IV-9 e sono quindi gestiti sezione per sezione. I secondi riguardano invece l'interazione diretta tra entita' alla pari residenti nei nodi di accesso e sono quindi gestiti da bordo a bordo.

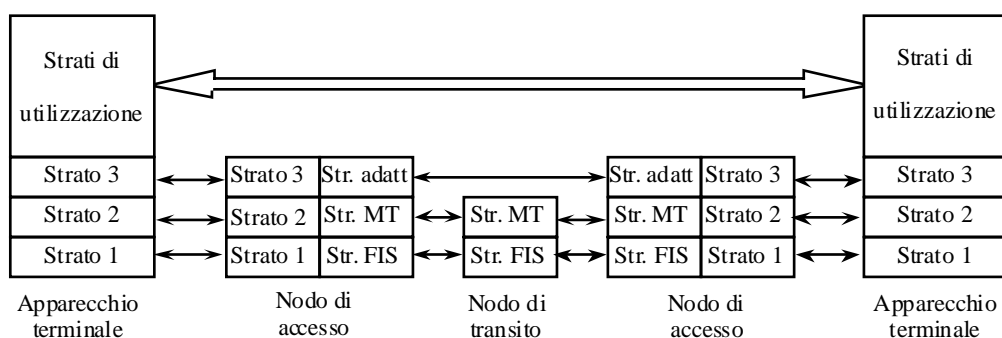


Fig. IV-10 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con un nuovo modo di trasferimento orientato al pacchetto.

IV.4.2 Possibili alternative architetturali

Per analizzare ora alcune possibili alternative di architettura protocollare, assumeremo, per fissare le idee, che la stratificazione delle funzioni di comunicazione sia quella a sette strati adottata nel modello OSI. Inoltre ci limiteremo dapprima a considerare particolarizzazioni del modello generale di riferimento mostrato in Fig. IV-9.

A questo scopo e' sufficiente individuare lo strato S . Questo, come gia' detto, e' lo strato gerarchicamente piu' elevato che interessa, oltre agli apparecchi terminali, anche i nodi di rete e che include una funzione di rilegamento. Almeno in linea di principio e nei limiti della stratificazione adottata dal modello OSI, lo strato S puo' essere quello fisico, ovvero quello di collegamento, ovvero infine quello di rete.

La individuazione dello strato S e' legata a considerazioni varie. Ne presentiamo qui alcune per chiarire, seppure sommariamente, i termini della questione.

Innanzitutto, come e' facile convincersi, la funzione di rilegamento svolta da un nodo ha, tra le sue componenti essenziali, la funzione di commutazione.

Inoltre, chiamiamo *unita' informative* (UI) le unita' di dati che sono scambiate tra le entita' dello strato a cui sono attribuite le funzioni di commutazione e di multiplazione. Osserviamo poi che la funzione di commutazione:

- e' il risultato di una demultiplazione dei flussi informativi entranti in un nodo e di una loro successiva multiplazione in accordo agli indirizzi (impliciti ed espliciti) delle UI che compongono questo flussi;
- puo' essere espletata sulla base delle stesse informazioni di protocollo di strato, che sono utilizzate per identificare le UI in un flusso multiplato.

Si puo' quindi concludere che lo strato S e' quello in cui sono svolte le funzioni di commutazione e di multiplazione.

D'altra parte la individuazione dello strato S ha implicazioni dirette sulle prestazioni del servizio di trasferimento e, in particolare, sui gradi di trasparenza temporale e di integrita' informativa. Cio' e' chiarito nei due successivi esempi, ove viene messo in evidenza che gli effetti di una scelta dello strato S sono tra loro contrastanti per cio' che riguarda il conseguimento di fissati obiettivi prestazionali.

Se ci riferiamo al trattamento protocollare del flusso informativo entrante in un nodo di rete e se teniamo conto dei principi di una architettura stratificata, lo svolgimento di una funzione appartenente a un dato strato richiede, da parte del nodo, la preventiva chiusura dei protocolli degli strati inferiori. Cio' equivale a dire che, se la scelta in esame fosse indirizzata verso uno strato S piu' alto, aumenterebbe la complessita' delle procedure che debbono essere gestite da un nodo di rete e quindi peggiorerebbe la prestazione di trasparenza temporale. Ovviamente si otterrebbe un effetto opposto indirizzando la scelta verso uno strato S piu' basso.

Se lo strato S fosse ad esempio quello di rete o quello di collegamento, il protocollo dello strato di collegamento (che e' tra l'altro preposto alla funzione di recupero degli errori trasmissivi) potrebbe essere gestito sezione per sezione, in accordo al modello di riferimento di Fig. IV-9. Cio' consentirebbe di fronteggiare con maggior successo i casi di canali trasmissivi con tasso di errore binario relativamente elevato, quali si incontrano con l'impiego di mezzi trasmissivi in rame. Conseguentemente la scelta di uno strato S di ordine piu' elevato dello strato fisico comporta migliori prestazioni di integrita' informativa, almeno nell'ambito di reti con impiego di mezzi trasmissivi in rame.

Alla luce delle considerazioni ora svolte, possiamo commentare le due alternative di architettura protocollare mostrate nelle Fig. IV-11 e Fig. IV-12,

che costituiscono una particolareggiata del modello generale di riferimento in Fig. IV-9.

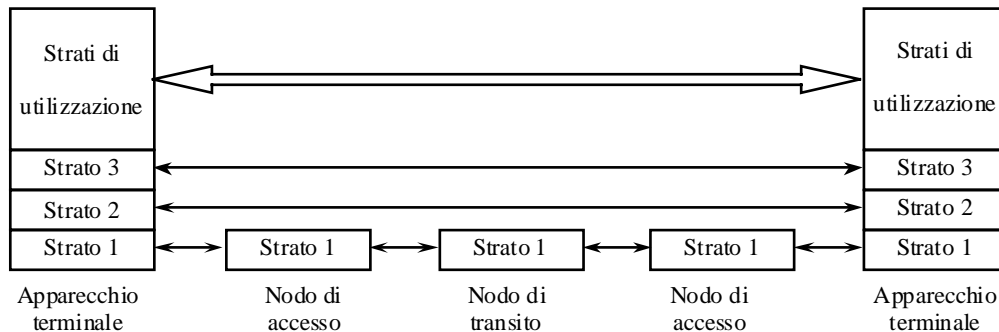


Fig. IV-11 – Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con modo di trasferimento a circuito.

Nella Fig. IV-11 lo strato S è quello fisico e quindi in questo strato sono svolte le funzioni di commutazione e di multiplazione. Inoltre tutte le funzioni logiche preposte al trattamento protocollare dell'informazione (a cominciare da quelle tipiche dello strato di collegamento) sono espletate con interazioni da estremo a estremo. Conseguentemente i nodi di accesso e di transito sono completamente trasparenti nei confronti dei flussi informativi che li attraversano. Si può quindi affermare che questa architettura protocollare è particolarmente adatta per la fornitura di un servizio di trasferimento con spiccate esigenze di un elevato grado di trasparenza temporale e con requisiti meno stringenti di integrità informativa.

In Fig. IV-12 si considera invece il caso in cui la multiplazione e la commutazione sono funzioni dello strato di rete. Ne consegue che i protocolli dei primi tre strati del modello OSI debbono essere aperti e chiusi in ogni nodo della sezione interna. Questa architettura protocollare risponde quindi bene ai requisiti prestazionali delle comunicazioni di dati, e cioè alle esigenze di un elevato grado di integrità informativa, ma con un modesto grado di trasparenza temporale.

Le architetture protocollari dei nuovi modi di trasferimento orientati al pacchetto sono intermedie tra i due casi estremi ora considerati, in quanto sono state definite in modo da consentire un trasferimento di informazione con prestazioni indipendenti dalle caratteristiche del corrispondente servizio di telecomunicazione.

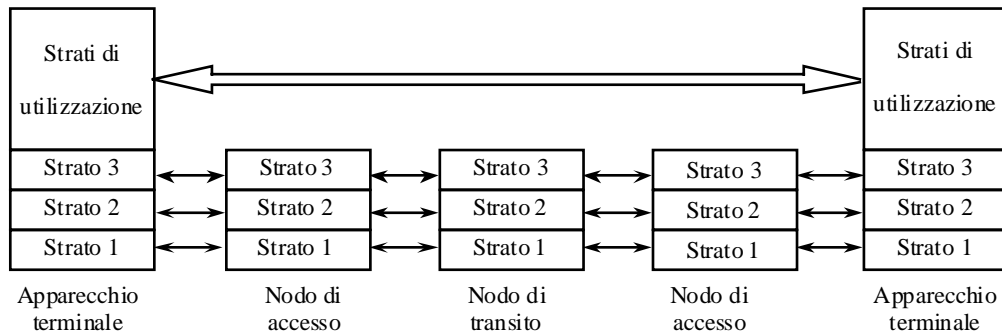


Fig. IV-12 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con modo di trasferimento a pacchetto.

Per conseguire questo scopo, sono stati spostati verso i bordi della rete i siti di svolgimento delle funzioni di protocollo, che dipendono dalle caratteristiche del servizio e che sono onerose in termini di tempo richiesto per il loro svolgimento; si allude in particolare al controllo di flusso, al recupero d'errore e alla sequenzializzazione delle UI. Ciò ha comportato una drastica semplificazione delle procedure svolte nei nodi di transito e, conseguentemente, sia una diminuzione dei ritardi di attraversamento, che un aumento delle portate di nodo.

Come risultato, le funzioni di multiplazione e di commutazione sono state collocate via via negli strati più bassi, in modo che la relativa architettura protocollare si avvicinasse corrispondentemente al modello di riferimento delle reti a circuito. Ciò ha richiesto una completa ridefinizione degli strati più bassi del modello OSI e, in particolare, l'introduzione di due nuovi strati immediatamente al di sopra dello strato fisico. Questi sono lo strato di *adattamento* e lo strato di *modo di trasferimento*, in ordine gerarchico decrescente.

Lo strato di adattamento contiene le funzioni dipendenti dal servizio, che hanno lo scopo di adattare lo specifico flusso informativo del servizio alle caratteristiche di trasferimento offerte dalla rete. Nello strato di modo di trasferimento sono incluse le funzioni di multiplazione e di commutazione, mentre lo strato fisico garantisce una efficiente utilizzazione del mezzo trasmissivo e un trasferimento affidabile della sequenza di cifre binarie.

L'architettura protocollare che ne risulta è mostrata in Fig. IV-13, dove si assume che gli apparecchi terminali abbiano una struttura protocollare del tipo OSI. I nodi di transito effettuano solo funzioni di strato fisico e di strato di modo di trasferimento, mentre le funzioni di strato di adattamento sono localizzate nei nodi di accesso. In tal modo, il nucleo della sezione interna di rete effettua solo

le funzioni comuni a tutti i servizi, mentre, in corrispondenza dei bordi di questa sezione, sono svolte le funzioni dipendenti dal servizio.

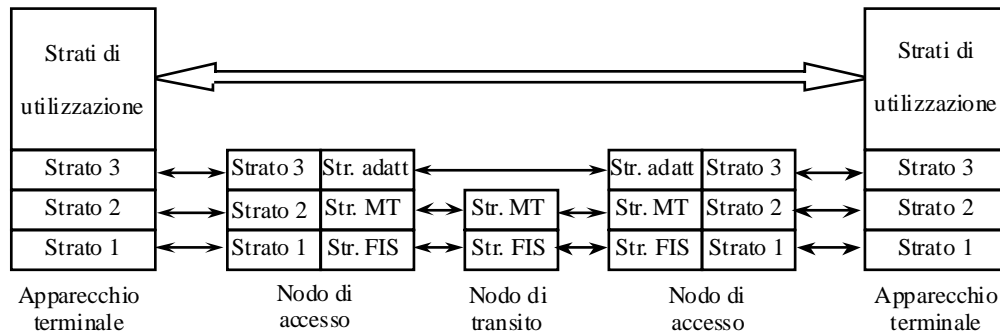


Fig. IV-13 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con un nuovo modo di trasferimento orientato al pacchetto.

IV.5 Esempi di modi di trasferimento

Riassumiamo dapprima (§ III.5.1 e III.5.2) gli aspetti specifici che caratterizzano i modi di trasferimento tradizionali (cfr. par. III.1), che trovano utilizzazione anche in una N-ISDN. Successivamente si illustreranno le caratteristiche dei nuovi modi di trasferimento (§ III.5.3) e, in particolare (§ III.5.4), del *modo di trasferimento asincrono* (ATM - Asynchronous Transfer Mode), che, secondo gli attuali orientamenti, appare il più serio candidato per la realizzazione della futura B-ISDN. Si conclude con il modo di trasferimento a rilegamento di trama (§ III.5.5).

Per ognuno di questi modi, verranno precisati lo schema di multiplazione, il principio di commutazione e l'architettura protocollare, in accordo ai criteri di classificazione introdotti nei paragrafi precedenti. In particolare, per ciò che riguarda l'architettura protocollare, viene individuato lo strato a cui appartengono le funzioni di multiplazione e di commutazione. Verranno infine distinti i modi di trasferimento orientati al circuito e quelli orientati al pacchetto.

IV.5.1 Modo di trasferimento a circuito

Un modo di trasferimento a circuito utilizza:

- 1) uno schema di multiplazione senza etichetta (S-TDM), di cui l'esempio più largamente diffuso nelle reti telefoniche numeriche è costituito dalla multiplazione PCM, con una UI costituita da 8 cifre binarie e con

- un sub-canale di base avente capacità di trasferimento uguale a 64 kbit/s;
- 2) un principio di commutazione, detto *a circuito*, che offre un servizio di trasferimento con connessione e che applica la strategia di assegnazione fisica delle risorse, con un attraversamento attuato per mezzo di una connessione diretta a divisione di tempo; per le relazioni tra le temporizzazioni degli orologi di nodo e' richiesto un funzionamento sincrono;
 - 3) una architettura protocollare, in cui la multiplazione e la commutazione sono funzioni di strato 1 (Fig. IV-11).

Nel caso di una comunicazione punto-punto su base chiamata, il trasferimento dell'informazione d'utente avviene su una connessione fisica commutata (*circuito fisico*) da utente a utente. Per instaurare tale connessione all'inizio della chiamata e per abbatterla al termine della comunicazione, avviene uno scambio di informazione di segnalazione tra i due utenti chiamante e chiamato da un lato e la rete dall'altro. Nell'ambito della sezione interna si ha anche un trasferimento dell'informazione di segnalazione tra i nodi coinvolti nella chiamata.

La durata richiesta per instaurare la connessione e' il *ritardo di instaurazione* e puo' variare, a seconda delle caratteristiche della rete e dei nodi che la compongono, da alcune decine di millisecondi ad alcuni secondi. Questo ritardo e' il risultato dei tempi richiesti per elaborare l'informazione di segnalazione in ognuno dei nodi coinvolti nella chiamata e per trasferire tale informazione dal chiamante al chiamato (*segnalazione in avanti*) e viceversa (*segnalazione all'indietro*). Esso dipende anche dai tipi di protocolli di segnalazione che vengono impiegati.

In relazione poi al carico della rete, possono verificarsi condizioni di *congestione*, che impediscono la instaurazione del circuito fisico da utente a utente. Tali condizioni possono corrispondere a temporanea insufficienza di risorse di trasferimento sia all'interno di uno tra i nodi coinvolti nella chiamata, sia all'esterno di questi. Nel primo caso si ha l'impossibilita' di formare una connessione diretta ingresso-uscita e si parla allora di *congestione interna*; il secondo caso corrisponde invece a una condizione di *congestione esterna*, che si manifesta con la impossibilita' di pre-assegnare individualmente alla chiamata una delle linee di giunzione uscenti dal nodo.

IV.5.2 *Modo di trasferimento a pacchetto*

Un modo di trasferimento a pacchetto, quale e' stato realizzato nelle reti per dati, e' invece caratterizzato da:

- 1) uno schema di moltiplicazione con etichetta (A-TDM), in cui le UI (*pacchetti*) sono, in generale, di lunghezza variabile e sono trasferite su canali con asse dei tempi indiviso; il trattamento delle contese di utilizzazione e' orientato al ritardo e, nel caso di servizio a chiamata virtuale, il controllo di accettazione di chiamata e' attuato con l'assegnazione a domanda di banda media;
- 2) un principio di commutazione, detto *a pacchetto*, in cui il servizio di trasferimento puo' essere di due tipi:
 - a) e' senza connessione e con strategia di assegnazione a domanda nel caso di servizio a datagramma;
 - b) e' invece con connessione e con strategia di assegnazione di risorse logiche nel caso di servizio a chiamata virtuale;
 in ambedue i casi l'attraversamento e' attuato con connessione ad immagazzinamento e rilancio, mentre, per le relazioni tra le temporizzazioni degli orologi di nodo, e' applicato un funzionamento asincrono;
- 3) una architettura protocollare, in cui la moltiplicazione e la commutazione sono funzioni di strato 3 (Fig. IV-12).

Con riferimento a una chiamata virtuale per una comunicazione di dati, la rete, a seguito di scambio di informazione di segnalazione, mette a disposizione della comunicazione un circuito virtuale, e cioe' (cfr. § III.3.4) un'associazione logica tra l'origine e la destinazione. Tale associazione e' definita nella fase di instaurazione ed e' attuata, in ogni nodo coinvolto nella chiamata, tramite una corrispondenza tra un canale virtuale entrante ed uno uscente.

Esempio IV.5-1 - Per chiarire questo punto e per esemplificare quanto detto in § III.3.4, indichiamo con Y_i il nodo a cui ci riferiamo per descrivere le funzioni di trattamento di chiamata e con Y_{i-1} l'apparecchiatura (terminale di origine o altro nodo), che e' collocata a monte di Y_i sul circuito virtuale in corso di instaurazione. Per la chiamata considerata, il canale virtuale entrante in Y_i e' preassegnato da Y_{i-1} e fa parte dell'insieme di canali virtuali che sono pre-assegnabili per la moltiplicazione statistica sul ramo di rete che connette Y_{i-1} con Y_i (ramo entrante in Y_i).

Il canale virtuale uscente da Y_i e' pre-assegnato da quest'ultimo a seguito della ricezione di un messaggio di segnalazione, che richiede l'instaurazione del circuito virtuale. Tale pre-assegnazione e' attuata attraverso i seguenti passi: a) conoscendo il nodo di destinazione (precisato dall'informazione di segnalazione), viene applicato un algoritmo di instradamento e, sulla base dei risultati da questo forniti, viene individuato un ramo uscente da Y_i ; tale ramo

concorrerà a far parte del percorso di rete da utente a utente; b) nell'insieme dei canali virtuali pre-assegnabili per la moltiplicazione statistica su questo ramo, ne viene prescelto uno tra quelli liberi da precedenti impegni.

Come già detto in § III.3.4, i due canali virtuali, quello entrante e quello uscente, sono identificati da due numeri d'ordine, scelti in due insiemi di numerazione univoca. Il *numero di canale virtuale entrante* è univoco nell'insieme utilizzato da Y_{i-1} per il ramo entrante in Y_i ed è comunicato a quest'ultimo tramite il messaggio di segnalazione, che richiede l'instaurazione del circuito virtuale. Il *numero di canale virtuale uscente* è univoco nell'insieme di numerazione utilizzato da Y_i per individuare i canali virtuali attivabili su tutti i rami da esso uscenti.

In tal modo, a conclusione della fase di instaurazione di un particolare circuito virtuale, ogni nodo, che appartiene al corrispondente percorso di rete, può memorizzare, in un'apposita *tabella di attraversamento* e in corrispondenza al ramo entrante che compone quel percorso, una coppia di numeri. Di questi, il primo identifica il canale virtuale pre-assegnato per il trasferimento dei pacchetti di dati scambiati nell'ambito di quella chiamata e entranti nel nodo, mentre il secondo identifica il canale virtuale pre-assegnato per il rilancio dei pacchetti di dati verso la loro destinazione.

Cio' equivale a dire che il circuito virtuale instaurato per una particolare chiamata è descritto da una *sequenza di numeri di canale virtuale*. Gli elementi, che compongono questa sequenza, sono memorizzati nelle apparecchiature di rete per tutta la durata della chiamata fino all'espletamento della fase di abbattimento.

Nella fase di trasferimento dell'informazione relativa a una particolare chiamata, un pacchetto di dati uscente da un nodo viene trasferito sul ramo appartenente al percorso di rete prescelto per quella chiamata e contiene, nella sua intestazione, il numero di canale virtuale che è associato a questo ramo nell'ambito della chiamata.

Pertanto, in corrispondenza della ricezione di un pacchetto di dati, il processore preposto al trattamento dei pacchetti in ogni nodo di rete deve provvedere a svolgere le seguenti funzioni:

- deve leggere l'intestazione del pacchetto e, in particolare, il numero di canale virtuale entrante che è in essa contenuto;
- in base alla consultazione della tabella di attraversamento deve trovare il numero di canale virtuale uscente, che è in corrispondenza con quello entrante;
- deve modificare, nell'intestazione del pacchetto di dati, il numero di canale virtuale entrante in quello uscente;
- deve inoltrare il pacchetto di dati così modificato verso il ramo di uscita, che è in corrispondenza con il numero di canale virtuale uscente.

IV.5.3 Nuovi modi di trasferimento

La Fig. IV-14 riassume gli aspetti specifici che caratterizzano i modi di trasferimento tradizionali descritti nelle due sezioni precedenti e quelli considerati per nuove applicazioni. Questi ultimi sono già stati menzionati in Fig. IV-2 e quindi a questa si rinvia per il significato dei relativi acronimi.

	Modo di trasferimento	Schema di multiplazione	Principio di commutazione	Architettura protocollare
MTOC	a circuito	S - TDM	ARF	Strato 1
	FCS		ARL con temporiz zazione sincrona	
MTOC	BS	A - TDM con alternativa SF		ARL con temporiz zazione asincrona
	ATD	A - TDM con alternativa SU		
	FPS	A - TDM con alternativa U		
	FR			
	a pacchetto con servizio VC			
	a pacchetto con servizio DG		senza connessione	Strato 3

MTOC - Modo di trasferimento orientato al circuito

MTOC - Modo di trasferimento orientato al pacchetto

Fig. IV-14 - Schemi di multiplazione, principi di commutazione e architetture protocollari nei modi di trasferimento orientati al circuito (MTOC) e orientati al pacchetto (MTOC).

Il contenuto della figura si spiega in modo autonomo; conseguentemente, ci limitiamo qui a esaminare i modi di trasferimento ATD e FPS, in quanto questi hanno condotto alla definizione dell'ATM. Per ciò che riguarda poi il modo FR (Frame Relaying), questo verrà analizzato in § III.5.5.

I modi ATD (Asynchronous Time Division) e FPS (Fast Packet Switching) adottano lo stesso principio di commutazione (servizio di trasferimento con connessione, assegnazione di risorse logiche, connessione ad immagazzinamento e rilancio, temporizzazione asincrona) e sono basati sulla stessa architettura protocollare in accordo con la Fig. IV-13-. Essi differiscono per lo schema di multiplazione, che è con etichetta in ambedue i casi, ma segue l'alternativa U nel caso dell'FPS e quella SU nel caso dell'ATD.

Per questi due modi, il ritardo di transito ha le componenti già definite nel caso generale di strategia ARL (cfr. § III.3.4); tuttavia il suo valor medio e la sua varianza sono tali da assicurare un trasferimento con un elevato grado di trasparenza temporale. Ciò è dovuto alla riduzione delle funzionalità di protocollo espletate dai nodi di transito (cfr. par. III.4) e alle possibilità oggi offerte di realizzare, da un lato, nodi con una drastica riduzione dei ritardi di attraversamento e, dall'altro, sistemi di trasmissione ad alta velocità.

Questo obiettivo è conseguibile, oltre che a seguito della utilizzazione di protocolli a bassa funzionalità, anche per effetto di quanto oggi consente la tecnologia dei sistemi di elaborazione e di quelli di trasmissione. È infatti possibile realizzare elaboratori di UI con alto parallelismo e quindi con ridottissimi tempi di elaborazione per ogni UI. Inoltre l'impiego di canali trasmissivi ad alta capacità, quali sono realizzabili su fibra ottica, consente di ridurre fortemente i tempi di trasmissione di ogni UI.

Occorre tuttavia tenere presente che l'utilizzazione di uno di questi modi in un ambiente di comunicazioni integrate implica due ulteriori componenti del ritardo di transito, che assumono, in un contesto di alta velocità di trasferimento, valori non trascurabili e, in alcuni casi, addirittura dominanti. Si allude al *ritardo di pacchettizzazione* e a quello di *de-pacchettizzazione*, che sono introdotti, rispettivamente, all'inizio e alla fine di ogni segmento di rete in cui è applicato il modo di trasferimento in esame.

Il ritardo di pacchettizzazione è il tempo necessario a riempire una UI ed è dato dal rapporto fra la lunghezza (in cifre binarie) di una UI e il ritmo binario di emissione della sorgente. Il ritardo di de-pacchettizzazione è invece il tempo necessario a equalizzare i ritardi di transito (cfr. § III.1.2), se questa operazione è richiesta come nel caso dei servizi isocroni, e a permettere la ricostruzione del flusso informativo CBR.

In ambedue i modi ATD e FPS, l'informazione è trasferita in piccole UI aventi lunghezza *fissa* nel caso dell'ATD e *variabile* in quello dell'FPS. Le UI entranti in un nodo sono trattate da un protocollo, che è in grado di rivelare eventuali errori entro l'etichetta e di scartare la UI entro cui si verifica l'errore.

I vantaggi potenziali di UI con lunghezza variabile risiedono in considerazioni di flessibilità, di efficienza e di qualità di servizio. Di questi vantaggi possono essere dati due esempi, con riferimento alle applicazioni riguardanti la voce e i dati.

Le sorgenti di dati emettono tratti informativi (cfr. § I.2.3) composti da un numero di ottetti (di cifre binarie) che può variare da 1 o 2 unità fino a varie centinaia di migliaia. In questo caso l'uso di UI con lunghezza fissa sarebbe

inefficiente per il trasferimento di tratti informativi sia piu' brevi della lunghezza fissa (in quanto parte della UI non verrebbe utilizzata), sia piu' lunghi (in quanto sarebbe necessario l'impiego di una molteplicita' di UI, ognuna con la sua informazione aggiuntiva di protocollo).

Per cio' che riguarda le applicazioni vocali, queste sono basate (cfr. § I.2.3) su un vasto campo di possibili ritmi binari di codifica (da una decina di kbit/s fino ad alcune centinaia di kbit/s in corrispondenza di differenti qualita' dell'informazione vocale trasferita), ma comportano in ogni caso analoghi vincoli sul valore di picco del ritardo di transito. Il mancato rispetto di questi vincoli comporta l'introduzione di dispositivi per il controllo dell'eco (*cancellatori d'eco*), con conseguenti costi aggiuntivi. D'altra parte, in queste applicazioni, il ritardo di transito e' largamente dipendente dal ritardo di pacchettizzazione, che, come si e' visto, e' direttamente proporzionale alla lunghezza della UI e inversamente proporzionale al ritmo binario di emissione della sorgente. Conseguentemente, per ogni valore di questo ritmo, esiste un valore massimo della lunghezza di UI.

I vantaggi di una UI con lunghezza fissa sono principalmente legati alla semplificazione delle funzionalita' di commutazione nei nodi di rete. Infatti questa soluzione facilita il dimensionamento degli organi di nodo (buffer) preposti all'immagazzinamento delle UI e assicura un flusso continuo di informazione di etichetta negli elaboratori di UI. Inoltre e' consentita una elaborazione piu' veloce del flusso asincrono di UI, dato che, nell'ambito dell'alternativa S dello schema di moltiplicazione (cfr. § III.2.1), la delimitazione delle UI e' identificabile in modo agevole, quando siano assicurate condizioni di sincronizzazione di IT.

Se viene adottata una UI con lunghezza fissa, questa deve essere la piu' corta possibile. Tale scelta e' legata a ragioni di efficienza per cio' che riguarda la funzione di commutazione e a motivi prestazionali per quanto attiene la riduzione del ritardo di pacchettizzazione. Inoltre, l'impiego di UI di piccola lunghezza rende piu' tollerabile il loro eventuale scarto a seguito di rivelazione di errore sulle loro etichette.

Per concludere, e' poi da considerare che la portata media di un nodo che applica i modi di trasferimento ATD e FPS deve essere aumentata di oltre due ordini di grandezza rispetto a quella che e' oggi necessaria nei nodi operanti con modo di trasferimento a pacchetto nelle reti per dati.

Esempio IV.5-2 - Supponiamo infatti che l'informazione da trasferire sia quella emessa da una sorgente vocale in una conversazione telefonica e assumiamo che: 1) la codifica della voce sia effettuata con rivelazione dei tratti di voce attiva e il grado di intermittenza della

sorgente sia uguale a 2,5; 2) la lunghezza, supposta fissa, di una UI sia tale da contenere un segmento di voce attiva con durata di 10 ms; 3) ogni chiamata telefonica abbia una durata media di 180 s. In queste condizioni, sulla base delle ipotesi 1) e 2), in ognuna delle due direzioni di trasferimento vengono emesse in media $1/(10 \times 10^{-3} \times 2,5) = 40$ UI/s e quindi $40 \times 180 = 7.200$ UI nella durata di una chiamata.

Cio' si riflette sul valore di portata media richiesto per un nodo. A tale riguardo consideriamo un autocommutatore locale telefonico, che sia in grado, secondo gli attuali orientamenti, di servire fino a 100.000 utenti. Allora, se le chiamate hanno la durata media che si e' sopra ipotizzata e se l'intensita' media di traffico offerto da ogni utente e' di 0,1 Erl, l'autocommutatore deve essere in grado di trattare in media fino a circa 55 chiamate/s. Conseguentemente, un autocommutatore locale, che operi con uno dei due modi ATD e FPS e che abbia una capacita' di trattamento di chiamata avente valore comparabile con quella dell'autocommutatore telefonico, dovrebbe avere una portata media di circa 800.000 UI/s, in accordo con quanto si e' anticipato circa il confronto con le portate medie dei nodi a pacchetto per applicazioni di dati.

Valori cosi' elevati di portata media di nodo sono oggi realizzabili attraverso i mezzi gia' menzionati ai fini della riduzione del ritardo di attraversamento.

IV.5.4 *Modo di trasferimento asincrono*

L'ATM e' un modo di trasferimento orientato al pacchetto, in cui:

- 1) lo schema di multiplazione e' con etichetta; le UI trattate sono chiamate *celle* e hanno un formato di lunghezza fissa; l'asse dei tempi e' suddiviso in IT senza organizzazione in trama; la gestione delle contese di utilizzazione e' orientata al ritardo, mentre il controllo di accettazione di chiamata e' tuttora oggetto di attivita' di ricerca;
- 2) il principio di commutazione offre un servizio orientato alla connessione, con strategia di assegnazione di risorse logiche; l'attraversamento e' con connessione ad immagazzinamento e rilancio, mentre le relazioni tra le temporizzazioni degli orologi di nodo sono basate su un funzionamento asincrono;
- 3) l'architettura protocollare segue le linee guida descritte nel modello di Fig. IV-13, dove lo strato di modo di trasferimento e' chiamato *strato ATM*.

Piu' in particolare una cella e' formata da un *campo informativo* e da una *intestazione*, con una lunghezza complessiva di 53 ottetti di cui 5 sono dedicati all'intestazione. Lo scopo primario di quest'ultima e' di portare l'etichetta, che identifica le celle appartenenti allo stesso canale virtuale per mezzo di una opportuna identificazione di riferimento di chiamata (cfr. § III.3.4). L'informazione di indirizzamento deve essere protetta contro possibili errori;

conseguentemente al contenuto dell'instestazione e' applicata una procedura di controllo di errore.

Una connessione virtuale e' identificata da una sequenza di IRC. Ogni valore di questa e' assegnato, per ciascun segmento di rete, durante la fase di instaurazione e viene restituito quando non e' piu' necessario. La connessione, cosi' identificata, rimane invariata per tutta la durata della chiamata.

Nel corso della chiamata, la capacita' di trasferimento e' assegnata a domanda in relazione all'attivita' della sorgente e alle risorse disponibili. L'integrita' della sequenza di celle su un canale virtuale e' garantita dallo strato ATM.

L'ATM offre una potenzialita' di trasferimento flessibile, comune a tutti i servizi, inclusi quelli senza esigenze di connessione. A tale scopo provvedono le funzionalita' dello strato di adattamento (cfr. par.III.4). Il limite tra lo strato ATM e gli strati superiori corrisponde a quello tra le funzioni supportate dai contenuti dell'instestazione e quelle legate all'informazione contenuta nel campo informativo della cella.

IV.5.5 Applicazione del LAPD nel Rilegamento di trama

Consideriamo il modo di trasferimento FR (Frame Relaying), che e' stato citato nelle figg. III.1.2 e III.5.1 e che viene qui descritto come soluzione alternativa di utilizzazione del canale D per trasferire dati strutturati a pacchetto attraverso una ISDN. Tale soluzione e' prevista nella Racc. I.122. Le motivazioni che ne hanno guidato la definizione sono quelle di carattere sistemistico e tecnologico riportate in § III.5.3 a proposito dei nuovi modi di trasferimento.

Nel modo FR, lo schema di moltiplicazione e' con etichetta e con unita' informative di lunghezza variabile; queste sono trasferite su canali con asse dei tempi indiviso; il trattamento delle contese di utilizzazione e' orientato al ritardo. Il principio di commutazione prevede un servizio di trasferimento con connessione e una strategia di assegnazione di risorse logiche; l'attraversamento e' attuato con connessione a immagazzinamento e rilancio. Nell'architettura protocollare, la moltiplicazione e la commutazione sono funzioni di strato 2.

A tale ultimo scopo viene utilizzata, per l'accesso a una ISDN, la moltiplicazione effettuata nello strato 2 dal LAPD. Per cio' che riguarda le funzioni di strato 2 presenti nei nodi della rete, queste coincidono con il cosiddetto *nucleo della Racc. Q.921*, e cioe' con:

- 1) la delimitazione, l'allineamento e la trasparenza di trama;

- 2) la multiplazione/ demultiplazione delle trame, utilizzando il loro campo di indirizzo;
- 3) la verifica della trama, per assicurare che essa sia costituita da un numero intero di ottetti, prima dell'inserimento degli ZERO di riempimento (cfr. § III.2.1) e dopo l'estrazione di questi;
- 4) la verifica della trama, per assicurare che essa non sia troppo lunga o troppo corta;
- 5) la rivelazione di errori trasmissivi.

La rete quindi non offre controllo di flusso, ne' recupero d'errore: le trame errate sono semplicemente scartate senza notifiche ai suoi utenti. Il controllo di flusso e il recupero d'errore, se desiderati, possono essere realizzati come funzioni di utente (e cioe' da estremo a estremo).

Conseguentemente un modo FR offre un servizio di trasferimento con le seguenti specificita':

- a) conservazione dell'ordine di trasferimento da un estremo all'altro della rete;
- b) assenza di duplicazione delle trame;
- c) probabilita' molto piccola di perdita delle trame.

IV.6 Il trattamento della segnalazione

In quanto precede ci siamo interessati solo dei modi di trasferimento relativi all'informazione d'utente. Verra' ora considerata anche l'informazione di segnalazione. Questa, come si e' gia' detto in § I.2.1, ha lo scopo di controllare una chiamata. E' quindi impiegata, oltre che per instaurare e per abbattere una connessione fisica o virtuale al servizio della chiamata, anche per modificarne le caratteristiche quando la chiamata e' stata gia' instaurata. Cio' al fine, ad esempio, di estendere una comunicazione, iniziata come punto-punto, a una configurazione multi-punto, quale e' richiesta da un servizio di tele-conferenza.

Nel seguito del paragrafo viene data, dapprima (§ III.6.1), una classificazione dei possibili modi di segnalazione adottati nelle reti per telefonia e per dati e di cui si prevede l'impiego nelle reti integrate a banda stretta o larga. Successivamente si parlera' brevemente della segnalazione a canale comune (§ III.6.2).

IV.6.1 Classificazione dei modi di segnalazione

E' opportuno distinguere preliminarmente due tipi di informazione di segnalazione, e cioe' quella che e' scambiata, nella sezione di accesso, tra apparato terminale e autocommutatore locale (*segnalazione di utente*) e quella che viene trasferita tra i nodi della sezione interna (*segnalazione inter-nodale*).

Consideriamo poi separatamente le modalita' di trattamento di questa informazione nelle reti a circuito, in quelle a pacchetto con servizio a chiamata virtuale e in quelle operanti con modo di trasferimento asincrono (*reti ATM*).

Nelle reti a circuito per telefonia, le segnalazioni di utente e inter-nodale sono state in passato, e sono tuttora, trattate con le modalita' piu' varie, che si differenziano per una molteplicita' di aspetti, quali la codifica di queste informazioni e il protocollo di comunicazione per l'interazione degli elementi di rete preposti al controllo della chiamata.

Un ulteriore aspetto distintivo riguarda l'associazione tra il modo di trasferimento dell'informazione di segnalazione preposta al trattamento di una data chiamata e quello dell'informazione d'utente scambiata nell'ambito della stessa chiamata. A tale riguardo esistono due alternative principali, e cioe' la *segnalazione associata al canale* (SAC) e la *segnalazione a canale comune* (SCC).

Nell'alternativa SAC, l'informazione di segnalazione preposta al controllo di una chiamata viene scambiata su un canale fisico distinto rispetto ad altri canali che trasferiscono informazioni dello stesso tipo, ma preposte al controllo di chiamate differenti. In queste condizioni si stabilisce una corrispondenza biunivoca tra due canali: quello su cui e' trasferita l'informazione di utente relativa alla chiamata considerata (*canale controllato*) e quello su cui viaggia l'informazione di segnalazione che tratta quella chiamata (*canale controllante*). Per ambedue i tipi di informazione si attua lo stesso modo di trasferimento a circuito.

I canali controllante e controllato (Fig. IV-15) possono coincidere (*soluzione a canale unico*) o essere distinti (*soluzione a canali separati*); nel primo caso si parla di *segnalazione in banda*, nel secondo di *segnalazione fuori banda*.

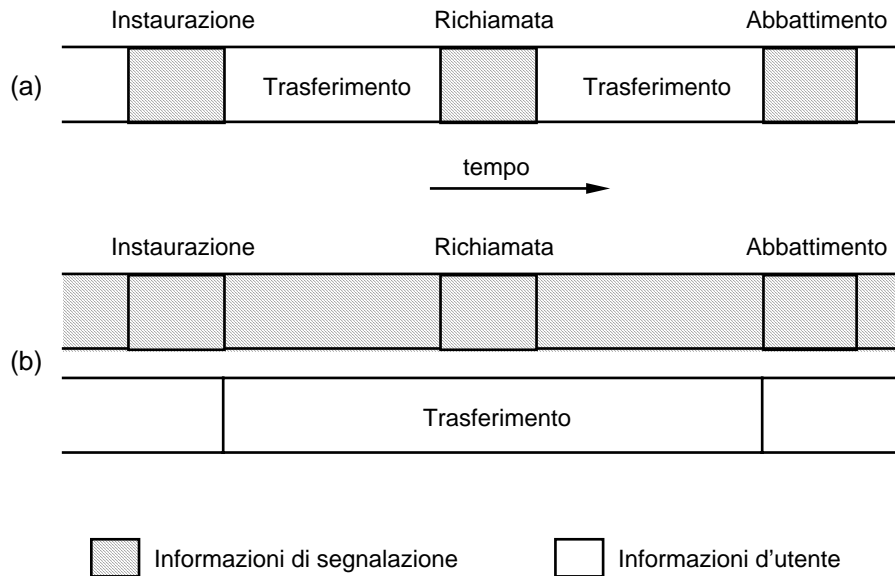


Fig. IV-15 - Modi di trasferimento dell'informazione di utente e di quella di segnalazione su un canale unico (a) e su canali separati (b).

Nella soluzione SAC a canale unico, le informazioni di utente e di segnalazione corrispondenti alla stessa chiamata condividono lo stesso canale con una utilizzazione in intervalli di tempo disgiunti. Le attuazioni di questa soluzione dipendono dal tipo di rete a circuito che si considera e dalle conseguenti modalità di codifica dell'informazione di segnalazione. Essa ha trovato impiego in ambedue le sezioni di rete, come nel caso delle reti telefoniche analogiche, ove la segnalazione è codificata da impulsi o da toni. Ma si incontra anche nella sezione di accesso, come nelle reti a circuito per dati in tecnica numerica, ove la segnalazione d'utente è presentata in un formato a caratteri di 8 cifre binarie, secondo l'*Alfabeto Internazionale no.5* (IA5) e come è specificato nella Racc. X.21.

La soluzione SAC a canali separati è utilizzata, nelle reti telefoniche, sia per la segnalazione d'utente che per quella inter-nodale. I modi di trasferimento sono a circuito per ambedue i canali. Un esempio tipico è offerto dalla modalità di trasferimento dell'informazione di segnalazione in uno schema di multiplazione statica PCM, attuata secondo la versione primaria europea. In questo caso l'informazione d'utente è trasportata da un canale scelto tra i 30 disponibili con capacità di 64 kbit/s, mentre l'informazione di segnalazione usufruisce di un canale, con capacità di 2 kbit/s, ottenuta sotto-multiplando l'IT no.16, con modalità a multitrama (cfr. § III.2.2).

Nell'alternativa SCC, come nella soluzione SAC a canali separati, il trasferimento dell'informazione di segnalazione avviene su un canale distinto

rispetto a quelli che trasportano l'informazione di utente, ma con due differenze sostanziali: il canale della segnalazione agisce come controllante per una pluralità di chiamate e il modo di trasferimento che viene impiegato è del tipo orientato al pacchetto, con servizio di trasferimento senza connessione.

L'alternativa SCC è oggi la soluzione più avanzata per il trasferimento della segnalazione inter-nodale in una rete a circuito in tecnica numerica. Ma viene impiegata anche nel caso della N-ISDN, sia nella sezione interna, che in quella di accesso.

Per concludere, rimangono da esaminare i casi delle reti a pacchetto con servizio a chiamata virtuale e di quelle ATM. Nel primo caso, se i protocolli di accesso e inter-nodali sono del tipo descritto dalle Racc.X.25 e X.75, l'informazione di segnalazione (sia di utente, che inter-nodale) è strutturata in UI specializzate (*pacchetti di segnalazione*), che condividono, in uno schema di multiplexazione dinamica, lo stesso canale virtuale utilizzato dalle UI contenenti l'informazione di utente (*pacchetti di dati*). Invece, nel caso delle reti ATM, le informazioni di utente e di segnalazione sono trasferite su canali virtuali separati. Anche in questi due casi si parla di segnalazione in banda e fuori banda, rispettivamente.

IV.6.2 *La segnalazione a canale comune*

Il trasferimento della segnalazione a canale comune nella sezione interna di una rete a circuito utilizza come supporto una rete apposita, che è sovrapposta alla precedente e che è chiamata *rete di segnalazione a canale comune* (rete SCC). Il trasferimento su questa rete avviene secondo le regole di protocolli, che, nel loro insieme, costituiscono il *sistema di segnalazione no.7* (SS no.7), che è normalizzato dal CCITT nelle Raccomandazioni della serie Q.700.

Nella rete SCC si distinguono i punti terminali e i nodi di commutazione. Un punto terminale, chiamato *punto di segnalazione* (SP - Signalling Point), rappresenta la sorgente e il collettore dell'informazione di segnalazione ed è in corrispondenza con l'unità di comando di un autocommutatore (AUT) della rete a circuito; nel seguito, per indicare questa corrispondenza, si dirà che SP e AUT costituiscono coppia.

Un nodo di commutazione della rete SCC è denominato *punto di trasferimento della segnalazione* (STP - Signalling Transfer Point), mentre i rami della rete SCC sono i *collegamenti di segnalazione*. Con riferimento a una specifica coppia (SP, AUT), un collegamento di segnalazione uscente dall'SP di

questa coppia e' al servizio di uno o piu' tra i fasci di linee di giunzione (*vie di giunzione*), che escono dall'AUT della stessa coppia verso altri autocommutatori della rete a circuito.

L'instaurazione di un circuito fisico nella sezione interna della rete a circuito fra un autocommutatore di origine e uno di destinazione avviene in accordo a una ricerca di percorso e a una operazione di impegno effettuate con il metodo *sezione per sezione*. Percio', in una connessione che coinvolge uno o piu' nodi della rete a circuito, le informazioni di segnalazione relative a tale connessione devono essere ricevute ed elaborate in ogni nodo intermedio tra quelli di origine e di destinazione.

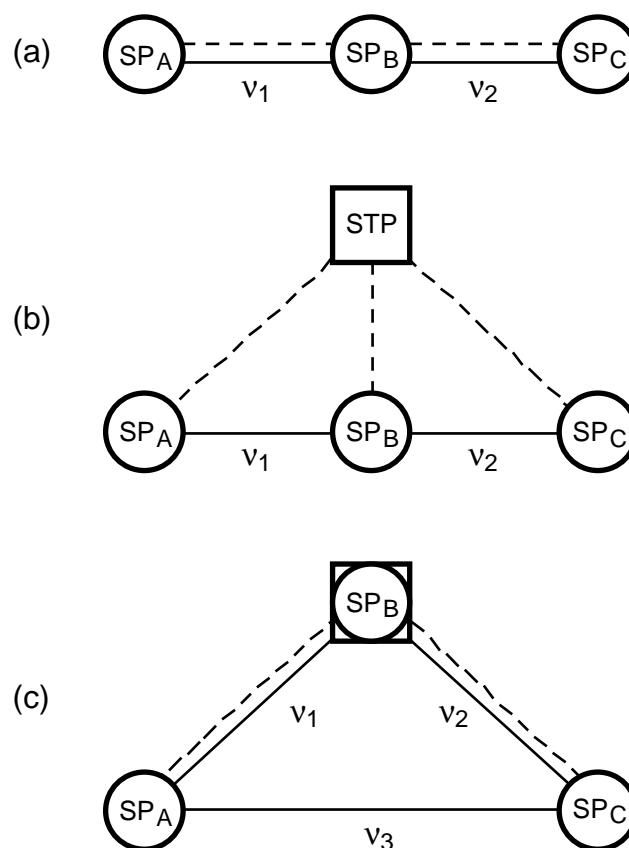


Fig. IV-16 - Modi di funzionamento in una rete SCC operante secondo l'SS no. 7: modo associato (a), modo non associato (b) e modo quasi associato (c).

Una rete SCC e' una rete di calcolatori tra gli elaboratori preposti al comando degli autocommutatori della rete a circuito. L'informazione e' scambiata sottoforma di messaggi. La metodologia dell'architettura a strati ben si presta a descrivere i protocolli della rete SCC. L'SS no.7 e' organizzato in quattro *livelli*. I tre livelli piu' bassi, con i numeri d'ordine 1, 2 e 3, sono, almeno orientativamente, in corrispondenza con i primi tre strati (1, 2 e 3) del modello

OSI e costituiscono, nel loro insieme, la cosiddetta *parte di trasferimento di messaggio* (MTP - Message Transfer Part).

Il livello 4 dell'SS no. 7 comprende funzioni, sempre orientativamente, in corrispondenza con gli ultimi quattro strati (4, 5, 6 e 7) del modello OSI. Tali funzioni sono mostrate in Fig. IV-17 e costituiscono le *parti di utilizzazione* (UP - User Part) nell'SS no.7; nella stessa figura e' sottolineata la corrispondenza tra gli strati OSI e i quattro livelli di questo sistema di segnalazione.

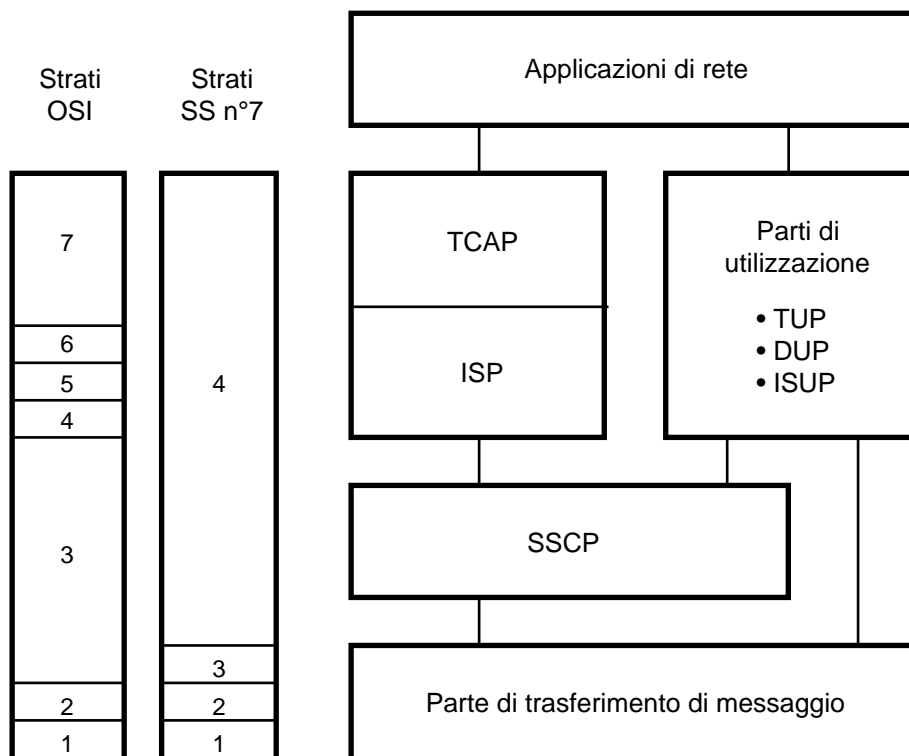


Fig. IV-17 - Struttura generale delle funzioni dell'SS no. 7.

Tra le parti di utilizzazione si distinguono innanzitutto la *Parte di controllo della connessione di segnalazione* (SSCP - Signalling Connection Control Part) e le *Potenzialita' di transazione* (Transaction capabilities). In particolare, la funzione SSCP fornisce i mezzi per controllare connessioni logiche in una rete SCC; essa quindi modifica il servizio offerto dall'MTP, a favore di quelle UP che richiedano un servizio orientato alla connessione, per trasferire segnalazione o altre informazioni a questa relative, quali sono utilizzate in applicazioni di esercizio e manutenzione o con lo scopo di controllare specifiche transazioni.

Le Potenzialita' di transazione sono composte da due elementi: la *Parte di applicazione delle potenzialita' di transazione* (TCAP - Transaction Capability Application Part) e la *Parte di servizio intermedio* (ISP - Intermediate Service

Part). Tali potenzialita' sono utilizzate per gestire l'interazione tra nodi di commutazione della rete a circuito e centri di servizio, nell'ambito di specifiche applicazioni, quali la gestione centralizzata dei gruppi chiusi di utente (cfr. § IV.3.3) e l'autenticazione di carte di credito.

Altre parti di utilizzazione, che utilizzano i servizi forniti dall'MTP (direttamente o indirettamente tramite l'SCCP) sono definite sia per trattare il controllo di specifiche connessioni, sia per supportare altre applicazioni legate, ad esempio, all'esercizio e alla manutenzione della rete SCC. Si hanno:

- la *UP per il servizio telefonico* (*TUP* - Telephony User Part), che e' preposta alla segnalazione necessaria per controllare le chiamate telefoniche;
- la *TUP migliorata* (*TUP-E* - TUP Enhanced), che e' stata definita per fronteggiare le prime realizzazioni di una ISDN;
- la *UP per i servizi di comunicazione di dati* in reti dedicate a circuito (*DUP* - Data User Part);
- la *UP*, che definisce le procedure di chiamata in una ISDN (*ISUP* - ISDN User Part);
- la *parte di applicazione per l'esercizio e la manutenzione* (*OMAP* - Operation and Maintenance Application Part), che e' preposta alle procedure di sorveglianza della rete SCC svolte da un SP preposto a funzioni di controllo (*CSP* - Control Signalling Point).

La Fig. IV-18 mostra poi in dettaglio le funzioni principali dell'MTP. Nel livello 1, abbiamo le funzioni del *circuito di dati di segnalazione* (Signalling Data Circuit), comprendendo in queste la trasmissione fino a 64 kbit/s e lo scambio di protezione di linea. Il livello 2 include le funzioni di *collegamento di dati di segnalazione* (Signalling Data Link), quali la strutturazione in opportune unita' informative, la sequenzializzazione e il recupero d'errore.

Nel livello 3, si distinguono le funzioni di *trattamento di messaggio* (Message Handling) e di *gestione di rete* (Network Management). La prima di queste comprende l'instradamento in accordo a un servizio di trasferimento senza connessione. La seconda riguarda le procedure aggiuntive per il funzionamento della rete SCC in condizioni anormali, quali guasti e sovraccarichi.

La Fig. IV-19 illustra infine i protocolli di comunicazione tra due SP per il tramite di un STP intermedio. I protocolli di livello 4 dell'SS no. 7 sono gestiti da estremo a estremo attraverso la rete SCC, dato che gli STP trattano funzioni solo fino al livello 3.

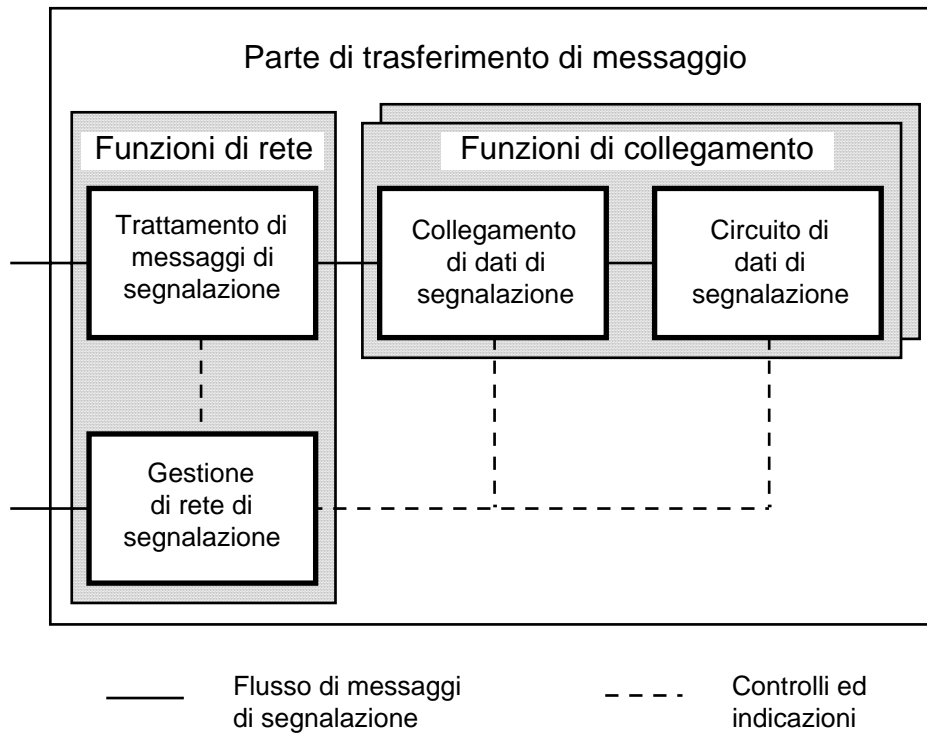


Fig. IV-18 - Struttura della parte di trasferimento di messaggio (MTP).

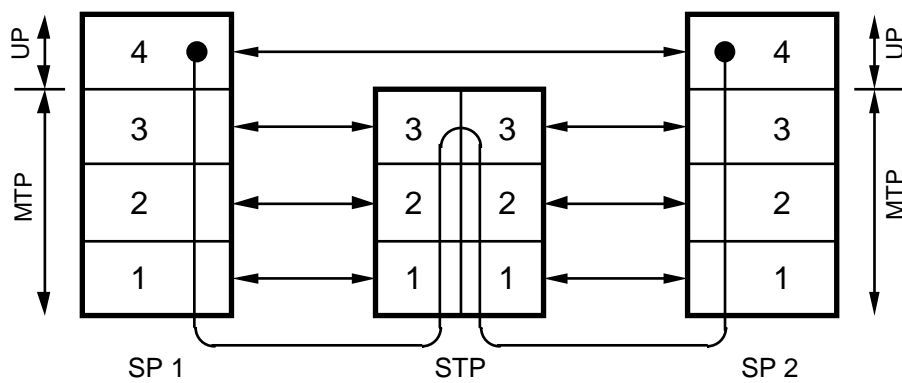
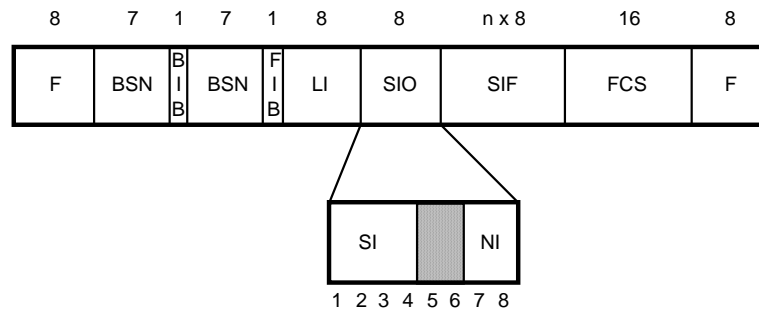
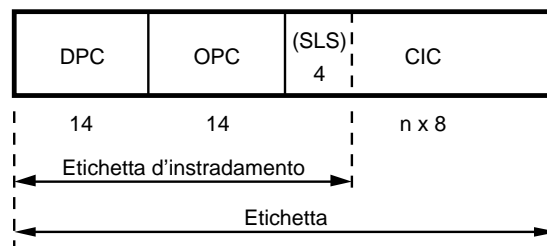


Fig. IV-19 - Architettura protocollare dell'SS no. 7.



F : Delimitatore
 FCS : Sequenza di controllo di trama
 SIF : Campo di informazione di segnalazione
 SIO : Ottetto di informazione di servizio
 LI : Indicatore di lunghezza
 SI : Indicatore di servizio
 FIB : Bit indicatore in avanti
 FSN : Numero di sequenza in avanti
 BIB : Bit indicatore all'indietro
 BSN : Numero di sequenza all'indietro
 NI : Indicatore nazionale

Fig. IV-20 - Formato della MSU nell'SS no. 7.



DPC : Codice del punto di destinazione
 OPC : Codice del punto di origine
 SLS : Selezione del collegamento di segnalazione
 CIC : Codice di identificazione del circuito

Fig. IV-21 - Struttura dell'etichetta dell'SS no. 7.

V RETI DEDICATE E RETI INTEGRATE

Gli argomenti sviluppati in precedenza sono di premessa essenziale per la trattazione che verrà svolta in questo capitolo, ove, con una articolazione in sei paragrafi, vengono introdotti gli elementi che riguardano l'evoluzione delle reti di telecomunicazione verso una ISDN e che caratterizzano questa secondo gli sviluppi più recenti della relativa attività di normalizzazione in sede internazionale.

V.1 Le reti dedicate a un servizio

Forniamo ora un quadro dello stato evolutivo delle reti di telecomunicazioni quale si presenta a monte degli sviluppi ISDN. Ci riferiamo in particolare allo stato delle reti dedicate alla telefonia (sez.IV.1.1) da un lato e alle applicazioni dei dati (sez.IV.1.2) dall'altro.

V.1.1 Rete telefonica

La tecnica di progettazione delle reti telefoniche, con riferimento specifico alle apparecchiature di commutazione e ai sistemi di trasmissione che ne fanno parte, si è sviluppata per oltre mezzo secolo in un ambiente di comunicazioni completamente *analogiche*, quale era consentito dalla tecnologia disponibile, e ha condotto a realizzazioni via via più perfezionate, che, ancora oggi, presentano esemplificazioni molto significative anche nelle reti telefoniche più evolute.

Dalla seconda metà degli anni '60, la rete telefonica ha iniziato l'evoluzione tecnologica e sistemistica, tuttora in atto, verso un impiego generalizzato delle tecniche *numeriche* in tutti gli apparati componenti. Questo sviluppo ha riguardato dapprima i sistemi di trasmissione e, successivamente (dopo circa un decennio), anche gli apparati di commutazione. L'obiettivo finale è la realizzazione (cfr. par.III.1) di una *rete numerica integrata nelle tecniche* (IDN - Integrated Digital Network).

Gli elementi distintivi di una IDN telefonica sono:

- 1) l'utilizzazione di tecnologie completamente elettroniche;
- 2) l'impiego di elaboratori per controllare, da un punto di vista decisionale, il funzionamento degli autocommutatori (*elaboratori di controllo*);

3) la graduale sostituzione, nella sezione interna, della segnalazione associata al canale con la segnalazione a canale comune.

Circa il punto 1), si puo' ricordare che, almeno per cio' che riguarda le apparecchiature di commutazione, la tecnologia utilizzata per circa 50 anni e' stata quella *elettromeccanica* e successivamente (dalla meta' degli anni '60) quella *semielettronica* (e cioe' in parte elettromeccanica e in parte elettronica). La transizione verso un impiego generalizzato di tecnologie realizzative elettroniche, dapprima con l'impiego di componentistica discreta a semiconduttori e, in un momento successivo, con l'utilizzazione di tecniche microcircuitali a vari livelli di integrazione, ha rivoluzionato le tecniche progettuali, le metodologie costruttive e l'organizzazione di esercizio degli impianti.

Relativamente poi al punto 2), vale la pena sottolineare che un elaboratore controlla un autocommutatore sulla base di un opportuno programma. Un autocommutatore *con controllo a programma memorizzato* (SPC - Stored Program Control) presenta vantaggi decisivi rispetto all'impiego di logiche di tipo cablato. Tra questi si possono citare: a) la *modularita'* e la *modificabilita'*, e cioe' la suddivisione del programma di controllo in moduli che possono essere modificati separatamente e indipendentemente, in modo relativamente agevole; b) la *adattabilita' al sito* e la *estensibilita'*, e cioe' la possibilita' di fronteggiare sia specifiche condizioni di sito all'atto della prima installazione, che mutate condizioni di esercizio dell'autocommutatore; c) la *autodiagnostica* e la *capacita' di adattamento* a condizioni di funzionamento anormali, e cioe', da un lato, la possibilita' offerta agli addetti di individuare e di eliminare rapidamente eventuali guasti e, dall'altro, la possibilita' di modificare la configurazione di sistema per consentirne la sopravvivenza.

Con riguardo infine al punto 3), nelle IDN telefoniche piu' evolute le informazioni di segnalazione vengono scambiate fra i nodi di commutazione per il tramite della rete SCC.

In Fig. V-1 e' mostrato un possibile modello di IDN telefonica. Vi si possono distinguere le sue sezioni di accesso e interna, oltre alla rete SCC. Descriviamo i principali elementi della IDN telefonica.

Gli *autocommutatori locali* sono composti da *stadi di utente*, con funzioni di interfaccia e di concentrazione/espansione, e da *stadi di gruppo*, con funzioni di distribuzione. Gli stadi di utente possono avere la stessa collocazione fisica degli stadi di gruppo ovvero possono essere posti in posizione a questi remota. Gli *autocommutatori di transito* sono invece composti da soli stadi di gruppo. Tali organi di commutazione operano secondo i principi del modo di

trasferimento a circuito e trattano informazioni foniche, numerizzate con ritmo binario uguale a 64 kbit/s, secondo la tecnica di codifica PCM (cfr. sez.I.2.3).

Un modello astratto di un autocommutatore a circuito con controllo a programma memorizzato e' mostrato in Fig. V-2.

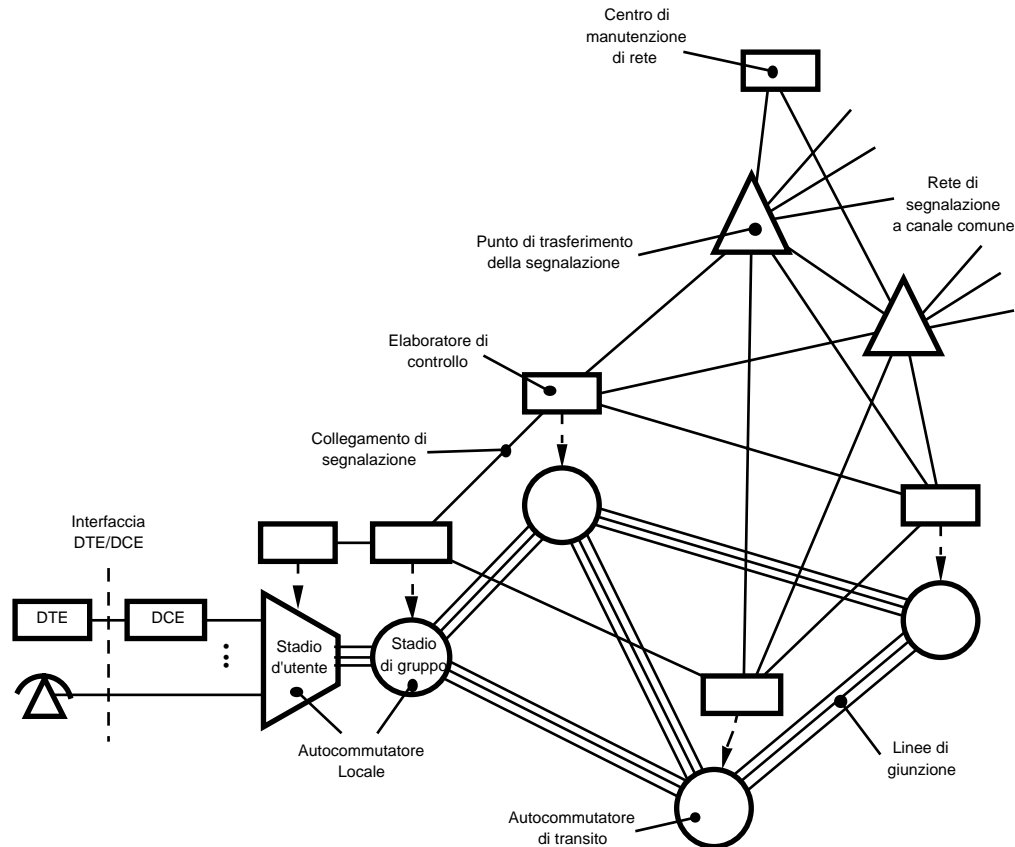


Fig. V-1 - Rete a circuito per telefonia e per dati sincroni, con l'impiego della segnalazione a canale comune.

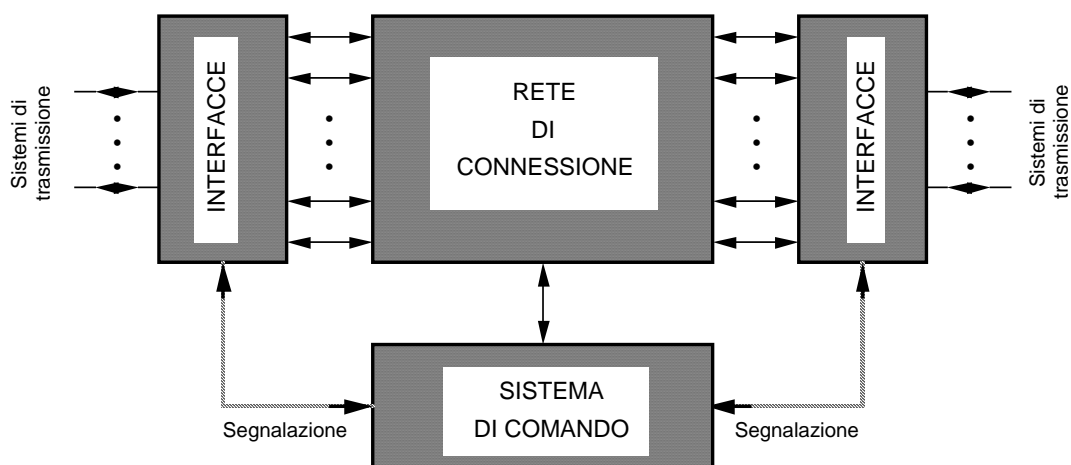


Fig. V-2 - Modello di un autocommutatore a circuito.

Tale modello comprende:

- 1) un insieme di interfacce tra autocommutatore e rete che include le terminazioni e le porte di ingresso e di uscita; da esso possono essere estratte o inserite le informazioni di segnalazione, nel caso in cui queste ultime siano trasferite con modo associato al canale;
- 2) una rete di connessione, attraverso la quale vengono effettuate le connessioni dirette ingresso-uscita;
- 3) un sistema di comando, che e' preposto al trattamento delle chiamate e alla gestione della rete (cfr. par.I.3); 4) un insieme di collegamenti di segnalazione che, da un lato, fanno capo al sistema di comando e, dall'altro, possono fare capo, in alternativa, alle interfacce di cui in 1) ovvero ai sistemi di comando di altri autocommutatori; mentre la prima possibilita' e' quella adottata per la segnalazione associata al canale, la seconda fa riferimento all'impiego della segnalazione a canale comune.

In particolare, un autocommutatore a circuito in tecnica numerica PCM (*autocommutatore PCM*) e' caratterizzato dal fatto che le informazioni entranti e uscenti sono strutturate in flussi multiplati in modo statico di tipo PCM e che la rete di connessione e' realizzata con tecnica a divisione di tempo numerica (cfr. sez.III.3.3).

Esempio V.1-1 - La potenzialita' di un autocommutatore a circuito e' descritta da due parametri prestazionali principali. Il primo di questi e' la massima intensita' media di traffico che puo' essere globalmente offerta all'autocommutatore con una prefissata *probabilita' di blocco* (cfr. esempio I.4.13) per le connessioni dirette ingresso-uscita e per ognuno dei fasci di giunzione uscenti. Questo parametro rende conto dei fenomeni di contesa di pre-assegnazione individuale relativi alle risorse di trasferimento interne ed esterne all'autocommutatore.

Il secondo parametro riguarda invece la capacita' del sistema di comando (e quindi delle risorse di elaborazione del nodo) a trattare, con vincoli di tempo reale (cfr. esempio I.3.7), i tentativi di chiamata che gli pervengono. Esso e' normalmente espresso mediante il numero massimo di tentativi di chiamata che possono essere trattati in un fissato intervallo temporale (ad esempio, in un'ora) con il rispetto di detti vincoli.

Le linee di giunzione sono il supporto di sub-canali di base, che hanno capacita' di trasferimento uguale a 64 kbit/s e che sono logicamente affasciati secondo lo schema di multiplazione statica PCM. Di questo schema sono state normalizzate (Racc. G.732 e G.733) due versioni *primarie*, una con ritmo binario di 2,048 Mbit/s adottata in Europa e l'altra di 1,544 Mbit/s adottata in Nord America.

Nella sezione interna della rete trasmissiva, sono ormai largamente impiegate, con vantaggi tecnico-economici decisivi, le fibre ottiche

monomodali, come mezzi alternativi a quelli tradizionali in rame (cavi a coppie simmetriche o a coppie coassiali) e in ponte radio.

Circa i sistemi di trasmissione, questi sono basati sull'impiego della *multiplazione numerica*. Questa opera un affasciamento di segnali tributari numerici in modo da generare un segnale multiplato con ritmo binario più elevato. La multiplazione numerica oggi impiegata è quella *asincrona*, con riferimento alla possibile diversità dei valori assunti dai ritmi binari tributari rispetto al loro valore nominale. La *gerarchia numerica asincrona* (Racc. G.702) definisce le sequenze di ritmi binari gerarchici riportate in Tab. V-1, secondo le normalizzazioni adottate in Europa e in Nord America.

Recentemente, in vista di un impiego sempre più diffuso delle fibre ottiche, è stata anche normalizzata una *gerarchia numerica sincrona* (Racc. G.707), che presenta un *modulo base* (STM - Synchronous Transport Module) con ritmo binario in linea di 155,520 Mbit. Gli altri ritmi di questa gerarchia (Tab. V-2) sono ottenibili come prodotto di questo ritmo di base con i termini di una progressione aritmetica con ragione uguale a 4 (cioè 4, 8, 12, 16 ...).

LIVELLO GERARCHICO NUMERICO	RITMI BINARI (in kbit/s) GERARCHICI PER RETI CON GERARCHIA NUMERICA BASATA SU UN RITMO BINARIO DI PRIMO LIVELLO UGUALE A	
	1.544 kbit/s (Nord America)	2.048 kbit/s (Europa)
1	64	64
2	1.544	2.048
3	6.312	8.448
4	32.064	34.368
	97.728	139.264

Tab. V-1 - Gerarchia numerica asincrona normalizzata in Nord America e in Europa.

LIVELLO GERARCHICO NUMERICO	RITMI BINARI GERARCHICI (in kbit/s)
1	155.520
4	622.080
8 (*)	1.244.160
12 (*)	1.866.240
16 (*)	1.488.320

Tab. V-2 - Gerarchia numerica sincrona.

(*) In corso di normalizzazione

Con riferimento poi ai vari orologi, che forniscono la temporizzazione delle apparecchiature della sezione interna, le frequenze da questi emesse debbono rispettare le condizioni di sincronizzazione di rete precisate nella sez.III.3.5 con riferimento alla strategia ARF a divisione di tempo.

Circa infine la sezione di accesso di una IDN telefonica, si osserva che ragioni di carattere economico non hanno giustificato finora l'impiego delle tecniche numeriche anche nella realizzazione delle linee d'utente e l'utilizzazione della segnalazione a canale comune anche al livello di interfaccia utente-rete, se queste soluzioni sono rapportate alle sole esigenze di base del *servizio telefonico ordinario* (POTS - Plain Old Telephone Service).

Vedremo nel seguito come le linee di utente numeriche e la segnalazione a canale comune al livello di interfaccia utente-rete sono elementi distintivi delle potenzialita' multiservizio insite nell'evoluzione verso l'integrazione dei servizi.

V.1.2 Reti per dati

Le reti per dati sono di concezione e di realizzazione molto piu' recente (anni '70) rispetto a quelle telefoniche. Esse quindi si sono sviluppate in un ambiente di comunicazione ormai fortemente orientato all'impiego delle tecniche numeriche e sono percio' inquadrabili come reti numeriche integrate.

Le IDN per dati sono state realizzate con modi di trasferimento sia a circuito, che a pacchetto. In Tab. V-3 sono riportati i ritmi binari rilevabili all'interfaccia tra DTE e DCE e normalizzati nella Racc.X.1 per il trasferimento di dati sincroni con modo a circuito o a pacchetto; sono ivi indicate anche le relative *classi di servizio*.

Le IDN a circuito per *dati sincroni* seguono gli stessi principi applicati nel caso delle IDN telefoniche.

CLASSI DI SERVIZIO CON TRASFERIMENTO		RITMO BINARIO (bit/s) DI EMISSIONE
A CIRCUITO	A PACCHETTO	
3		600
4	8	2.400
5	9	4.800
6	10	9.600
7	11	48.000
19	13	64.000

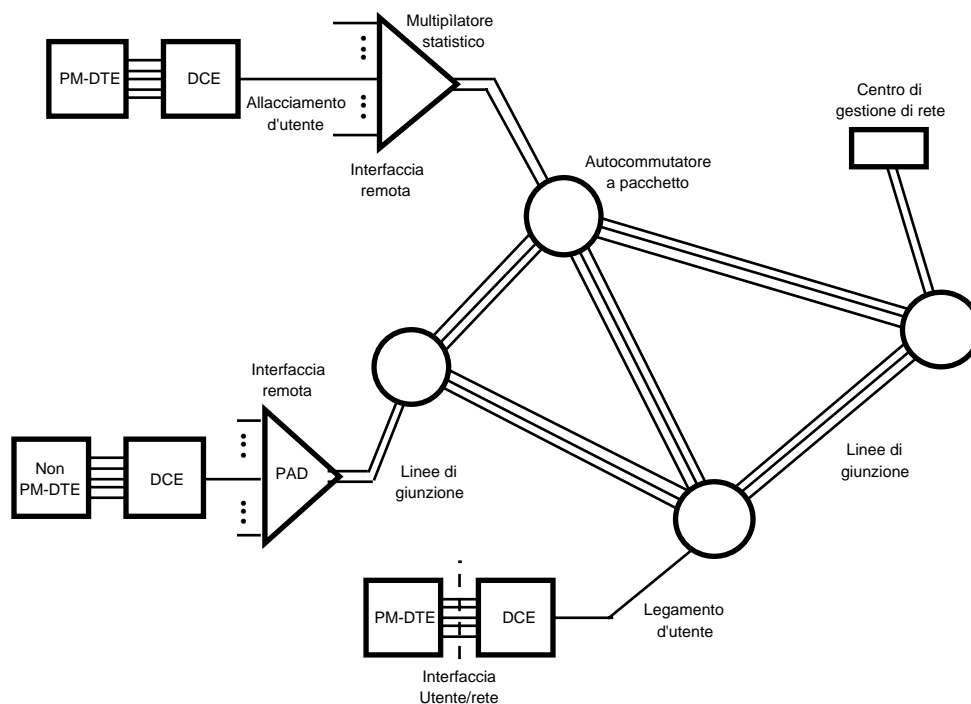
Tab. V-3 - Classi di servizio e relativi ritmi binari di emissione per il trasferimento di dati sincroni trattati a circuito o a pacchetto.

Tecniche opportune sono usate per sotto-multiplexare flussi di dati a ritmo binario inferiore a 64 kbit/s (*sotto-ritmo*) su un canale avente capacità di 64 kbit/s. Il modo di trasferimento a circuito per flussi di dati a sotto-ritmo può essere realizzato secondo due alternative principali. Si può, in primo luogo, utilizzare un pre-commutatore operante a sotto-ritmo. Come seconda alternativa si può impiegare una normale rete di connessione operante a 64 kbit/s; occorre allora fare ricorso a modalità di *riempimento/svuotamento* degli intervalli temporali al fine di trattare correttamente i flussi a sotto-ritmo.

Con queste precisazioni, il modello di IDN a circuito, mostrato in Fig. V-1, è applicabile anche al caso del trasferimento di dati sincroni.

Reti per dati a circuito sono state realizzate anche per comunicazioni di *dati asincroni (start-stop)*. Tali reti dedicate sono state progettate, almeno all'origine, per trattare il servizio telex o dati start-stop fino a 300 bit/s.

In Fig. V-3 è mostrato un modello di IDN per dati a pacchetto.



PM - DTE : DTE operante a pacchetto

Non PM - DTE : DTE non operante a pacchetto

Fig. V-3 - Rete per dati a pacchetto.

In esso si distinguono:

- 1) la sezione interna, costituita da nodi di commutazione, interconnessi da linee di giunzione numeriche oggi operanti con una capacità di 64 kbit/s;

- 2) la sezione di accesso, comprendente le apparecchiature di concentrazione, di multiplazione e di interfacciamento dei DTE;
- 3) il centro di gestione di rete, preposto a funzioni di supervisione per una efficiente utilizzazione delle risorse.

Le tecnologie attuali e la necessita' di assicurare elevata capacita' di traffico hanno portato per i nodi di commutazione a criteri di progetto di tipo modulare, secondo strutture distribuite in cui unita' specializzate per funzioni (interfaccia di linea, gestione dei protocolli) sono collegate tra loro da strutture di interconnessione ad alta velocita'.

Il funzionamento della sezione interna applica i principi del modo di trasferimento a pacchetto (cfr. sez.III.5.2) con servizio a chiamata virtuale ovvero a datagramma. Nel primo caso tutti i pacchetti scambiati nell'ambito della stessa chiamata seguono lo stesso percorso di rete; viene in tal modo garantita la sequenzialita' dei pacchetti e l'integrita' del messaggio a destinazione. Inoltre, all'atto della fase di instaurazione, viene accertata la disponibilita' dell'utente chiamato a stabilire una conversazione bidirezionale.

Nel caso invece di servizio di trasferimento a datagramma, ogni pacchetto (datagramma) e' provvisto di una intestazione che lo rende autonomo nel trasferimento attraverso la rete. Conseguentemente i percorsi di rete, seguiti dai datagrammi aventi origine da un DTE e aventi la stessa destinazione, possono, in generale, essere diversi l'uno dall'altro.

Esempio V.1-2 - Circa i parametri prestazionali di un autocommutatore a pacchetto operante nel modo a chiamata virtuale, vale quanto si e' detto nell'esempio I.4.12. Come e' ovvio, all'aumentare della portata media dell'autocommutatore, si acuiscono i fenomeni di contesa di utilizzazione e cresce quindi il ritardo di attraversamento. Di conseguenza, normalmente, si precisa quale debba essere il valor massimo della portata media entro cui detto ritardo si mantiene entro valori accettabili, per cio' che riguarda sia il suo valor medio, sia il valore che non e' superato, ad esempio, nel 95% dei casi (percentile 95°).

Gli algoritmi di instradamento (cfr. sez.III.3.1) applicati nelle reti a pacchetto provvedono a costruire, in ogni nodo, tabelle che pongono una corrispondenza biunivoca tra linee entranti e uscenti relative al nodo. Si possono avere instradamenti *statici* o *dinamici*, a seconda della cadenza di aggiornamento delle tabelle di instradamento.

Nel caso di instradamento statico, le tabelle di instradamento sono in genere fissate al momento dell'attivazione dell'impianto, memorizzate nei processori di controllo dei nodi e aggiornate con cadenza estremamente lenta. Nel caso invece di instradamento dinamico, l'algoritmo di instradamento e'

capace di adattarsi alle mutate condizioni di traffico nella rete e agli stati di funzionamento delle apparecchiature di rete. In questo ambito si distinguono soluzioni *centralizzate*, *distribuite* e *miste*.

Le tecniche di instradamento dinamico centralizzato si basano sulla disponibilit  di un centro di gestione di rete, che abbia una conoscenza, aggiornata con continuit , delle condizioni di traffico nella rete e che provveda, di conseguenza, a elaborare le tabelle di instradamento e a distribuirle a tutti i nodi di commutazione. Nel caso invece di instradamento dinamico distribuito, ogni nodo scambia informazioni con i nodi adiacenti in modo da poter effettuare una stima delle condizioni di traffico presente nella rete e da aggiornare conseguentemente le proprie tabelle di instradamento. Infine le tecniche miste adottano combinazioni opportune di algoritmi centralizzati e distribuiti.

Circa la scelta tra queste modalit  di instradamento, si pu  affermare che le reti operanti internamente con servizio a chiamata virtuale (ad esempio le reti TELENET, TRANSPAC e ITAPAC) tendono a usare soluzioni centralizzate. Invece, nelle reti che adottano internamente il servizio di trasferimento a datagramma (ad esempio nelle reti ARPANET e DATAPAC), la tecnica preferita   quella dell'instradamento distribuito.

Per cio' che riguarda i modi di accesso a una rete e pacchetto, in Fig. V-3 ne sono previsti tre casi possibili. Due di questi riguardano DTE che sono in grado di pacchettizzare l'informazione e cio  *DTE operanti a pacchetto* (PM-DTE, Packet Mode DTE), mentre il terzo si riferisce a DTE che, come i terminali start-stop, non hanno questa capacit  (*DTE non operanti a pacchetto*). Relativamente ai primi due casi, in fig.IV.1.3 vengono distinte due possibilit  di accesso: una prima con linea d'utente individuale; l'altra con accesso intermedio a un moltiplicatore statistico, su cui convergono altri DTE. Per cio' che riguarda poi il terzo caso, dispositivi per l'*assemblamento/ disassemblamento dei pacchetti* (PAD - Packet Assembling/Disassembling) sono impiegati per interfacciare DTE non operanti a pacchetto; l'accesso avviene tipicamente tramite connessioni commutate sulla rete telefonica.

Esempio V.1-3 - I primi due casi considerati in fig.IV.1.3 si riferiscono a DTE operanti a pacchetto che si interfacciano con il proprio DCE secondo la normativa della Racc. X.25. Apparecchi terminali dello stesso tipo possono anche accedere a una rete a pacchetto per il tramite di una rete a circuito (telefonica o per dati); in questo caso l'interfaccia DTE-DCE   definita nella Racc. X.32.

Con riferimento poi al terzo caso, l'interfaccia DTE-DCE per terminali start-stop che accedono a una rete a pacchetto tramite una unit  PAD   normalizzata nella Racc. X.28.

Infine, per consentire uno scambio di informazione tra due DTE, di cui uno operante a pacchetto e l'altro senza questa capacit  per il tramite di un PAD ovvero tra due PAD, sono

previste le procedure descritte nella Racc. X.29.

V.2 Definizione della ISDN

Secondo la definizione che ne fornisce il CCITT, la ISDN e' una rete che, evolvendo dalla IDN telefonica, e' in grado di offrire una completa connettivita' numerica da utente a utente per la fornitura di un'ampia gamma di servizi, vocali e non vocali, ai quali e' possibile accedere attraverso un insieme ristretto di interfacce utente-rete normalizzate.

Da questa definizione emergono i tre aspetti chiave che sono alla base del concetto di ISDN, e cioe':

- 1) la completa connettivita' numerica da utente a utente;
- 2) la capacita' di fornire una molteplicita' di servizi di telecomunicazione;
- 3) la normalizzazione di un ristretto numero di interfacce utente-rete.

Per cio' che riguarda i punti 1) e 2), questi sono gia' stati commentati in precedenza (cfr. par.III.1). D'altra parte, il punto 3) sottolinea che, per raggiungere una completa integrazione di trattamento dei vari servizi, il primo passo essenziale e' la normalizzazione di un numero limitato di interfacce utente-rete. Cio' consentira' di superare, con evidenti vantaggi economici, l'attuale situazione offerta dalle reti dedicate, in cui l'accesso a singoli servizi (o a limitati insiemi di questi) richiede interfacce diverse da rete a rete.

Una ISDN e' pienamente descritta dalle caratteristiche di servizio offerte alle sue interfacce di accesso, piuttosto che dalle soluzioni sistemiche e tecnologiche adottate nella sezione interna. Cio' comporta una completa indipendenza tra l'architettura della sezione interna e i servizi offerti agli utenti; l'evoluzione di questi due aspetti puo' avvenire in modo completamente separato. Conseguentemente una ISDN puo' essere realizzata in una varietas di configurazioni, a seconda dell'ambiente di comunicazione dal quale emerge, in termini dello stato iniziale della rete e della sua evoluzione successiva.

Secondo le indicazioni del CCITT, nella prima fase di sviluppo la ISDN sara' basata sulla IDN telefonica (cfr. sez.IV.1.1), introducendo in questa le potenzialita' connesse alla linea di utente numerica. Attraverso tale rete sara' possibile accedere ad altre strutture dedicate ad un servizio, quali quelle per dati a pacchetto.

Compatibilmente poi con quanto consentito dall'evoluzione tecnologica e dai vincoli economici, la ISDN incorporera' progressivamente funzioni addizionali e prestazioni migliorate, incluse quelle attualmente fornite dalle strutture dedicate ad un servizio. Chiariamo questo punto.

In primo luogo, i nuovi servizi introdotti nella ISDN saranno dapprima fondamentalmente basati su connessioni commutate o semi-permanenti con capacita' di trasferimento uguale a 64 kbit/s; in tempi successivi, potranno essere incluse anche connessioni commutate o semi-permanenti con capacita' di trasferimento superiore ed, eventualmente, inferiore a 64 kbit/s.

Inoltre, le funzionalita' di una ISDN non dovranno limitarsi al trasporto dell'informazione, ma potranno comprendere anche trattamenti di questa con lo scopo di fornire *servizi a valore aggiunto*, quali la memorizzazione e l'inoltro di messaggi (vocali o scritti), la conversione di protocolli, la raccolta e la elaborazione di dati. Le risorse di elaborazione necessarie per questo scopo potranno essere inserite direttamente nella ISDN come risorse condivise e, quando richiesto da specifici servizi, anche negli apparecchi terminali, curandone la compatibilita' con quanto offerto dalla rete.

In terzo luogo, essenziale per le funzioni di una ISDN e' l'integrazione delle sue funzionalita' per tutto cio' che concerne la gestione della rete, nei suoi vari aspetti coinvolti dai problemi di esercizio, di amministrazione e di manutenzione. Per la soluzione di questi problemi un ruolo chiave avranno ancora le risorse di elaborazione messe a disposizione dalla ISDN.

Circa l'evoluzione successiva della ISDN, una tendenza e' quella di procedere con l'integrazione, a partire dalle interfacce di accesso, anche verso le apparecchiature della sezione interna. In particolare, per migliorare l'economia del sistema, potra' essere conveniente incorporare le attuali risorse dedicate a un servizio in apparecchiature di rete comuni. Sotto questo punto di vista l'obiettivo finale potrebbe essere quello di ottenere una struttura di rete, che sia comune per tutti i servizi sia nella sezione di accesso, sia in quella interna.

Deve essere tuttavia sottolineato che su quest'ultimo punto non esiste unanimita' di consensi: cioe', mentre esiste completo accordo sull'integrazione degli accessi con l'impiego, almeno in futuro, di un unico modo di trasferimento (ad esempio, dell'ATM a cui si e' accennato nel par.III.6) per l'intera gamma dei servizi offerti, non altrettanto si puo' affermare per gli sviluppi relativi alla sezione interna.

Qualunque sara' la linea di evoluzione che prevarra', una ISDN sara' il risultato di attivita', che prevedibilmente, dureranno uno o piu' decenni e che prospettano, come primo passo, la realizzazione della ISDN a banda stretta e, come passo conclusivo, quella della ISDN a banda larga. Gli obiettivi conseguibili con queste realizzazioni sono gia' stati introdotti nel par.III.1.

VI Elenco degli acronimi

ACK	Positive Acknowledgement
ADPCM	Adaptive Differential PCM
ARF	Assegnazione di Risorse Fisiche
ARL	Assegnazione di Risorse Logiche
ATD	Asynchronous Time Division
A-TDM	Asynchronous TDM
ATM	Asynchronous Transfer Mode
AUT	Autocommutatore
BIB	Backward Indicator Bit
B-ISDN	Broadband ISDN
BSN	Backward Sequence Number
CBR	Constant Bit Rate
CCITT	Comite' Consultative Internationale pour la Telephonie
CEI	Connection End point Identifier
CES	Connection End point Suffix
CIC	Circuit Identification Code
CSMA-CR	Carrier Sense Multiple Access - Collision Resolution
CSP	Control Signalling Point
CP	Control Processor
CUG	Closed User Group
DCE	Data Circuit terminating Equipment
DL	Data Link
DLCI	Data Link Connection Identifier
DPC	Destination Point Code
DTE	Data Terminal Equipment
DUP	Data User Part
EA	Extended Address
ET	Exchange Termination
FCS	Frame Check Sequence
FIB	Forward Indicator Bit
FPS	Fast Packet Switching
FR	Frame Relaying
FSN	Forward Sequence Number
HDLC	High-level Data Link Control

IA5	International Alphabet no. 5
IDN	Integrated Digital Network
IRC	Indicazione di Riferimento di Chiamata
ISDN	Integrated Services Digital Network
ISDN-PRM	ISDN - Protocol Reference Model
ISO	International Organization for Standardization
ISUP	ISDN User Part
IT	Intervallo Temporale
IW	Interworking
LAN	Local Area Network
LAPD	Link Access Protocol-D channel
LFC	Local Functional Capabilities
LI	Length Indicator
LSSU	Link Status Signalling Unit
LT	Line Termination
MAC	Medium Access Control
MDL	Management Data Link
MSU	Message Signalling Unit
MTP	Message Transfer Part
NACK	Negative Acknowledgement
N-ISDN	Narrowband ISDN
NMC	Network Management Center
NT	Network Termination
OMAP	Operation - Maintenance Application Part
OPC	Originating Point Code
OSI	Open System Interconnection
PAD	Packet Assembling/Disassembling
PB	Protocol Block
PBX	Private Branch Exchange
PCM	Pulse Code Modulation
PH	Packet Handling
PM-DTE	Packet Mode DTE
PSE	Packet Switching Exchange
POTS	Plain Old Telephone Service
RR	Receive Ready
SAC	Segnalazione Associata al Canale
SAPI	Service Access Point Identifier
SCC	Segnalazione a Canale Comune

SCCP	Signalling Connection Control Part
SDU	Service Data Unit
SIF	Signalling Information Field
SIO	Service Information Octet
SLS	Signalling Link Selection
SM	Statistical Multiplexer
SP	Signalling Point
SPC	Stored Program Control
SS no. 7	CCITT Signalling System no. 7
S-TDM	Synchronous TDM
STM	Synchronous Transport Module
STP	Signalling Transfer Point
SU	Signalling Unit
TCAP	Transaction Capability Part
TDM	Time Division Multiplexing
TE	Terminal Equipment
TEI	Terminal End-point Identifier
TUP	Telephone User Part
TUP-E	Telephone User Part Enhanced
UI	Unnumbered Information
UP	User Part
VBR	Variable Bit Rate

VII Bibliografia

- [1] A. Roveri: “Reti di Telecomunicazioni”, Principi Generali, Scuola Superiore G. Reiss Romoli, L’Aquila, 1995.
- [2] R. Gallager, D. Bertsekas: "Data Networks", Prentice Hall, 1992.
- [3] D. Comer: “Internetworking with TCP/IP”, Volume I - Principles, Protocols and Architecture. Prentice Hall, 1991.
- [4] D. Comer, D. Stevens: “Internetworking with TCP/IP”, Volume II - Design, Implementation and Internals. Prentice Hall, 1994.
- [5] D. Comer, D. Stevens: “Internetworking with TCP/IP”, Volume III - Client-Server Programming and Applications for the AT&T TLI Version. Prentice Hall, 1994.
- [6] J. Cypser: "Communications for Cooperating Systems, OSI, SNA and TCP/IP, Addison-Wesley, 1992.
- [7] Ramteke: "Networks", Prentice Hall, 1994
- [8] W. Stallings: "ISDN and B-ISDN", Prentice Hall
- [9] A. Tanenbaum, "Computer Networks", Prentice Hall '89

VIII Lista delle figure, tabelle ed equazioni

VIII.1 Figure

Fig. I-1 - Alternative di grafo di rete.....	11
Fig. I-2 - Esempio di grafo logico (a) e di grafo fisico (c), relativi alla sezione interna di una rete di telecomunicazione; il grafo fisico risulta dalla scelta di percorso dei mezzi trasmissivi quale e' rappresentata in (b).	15
Fig. I-3 - Grafo logico relativo alla sezione interna di una rete di telecomunicazione: distinzione tra nodi di accesso (A) e di transito (T).	15
Fig. I-4 - Principali utilizzazioni delle risorse di elaborazione in un autocommutatore.	17
Fig. I-5 - Relazione tra le funzioni di trattamento di chiamata e le informazioni di segnalazione	17
Fig. I-6 - Esempio di sezione di accesso con linea di utente individuale.....	18
Fig. I-7 - Schema generale di una rete per dati.	20
Fig. I-8 - Modello base di un sistema di comunicazione di dati.	21
Fig. I-9 - Sequenza di emissione di una sorgente VBR "tutto o niente".	35
Fig. I-10 - Caratteristiche generali di una comunicazione e loro possibili attributi.....	46
Fig. I-11 - Classificazione dei servizi di telecomunicazione in base alle modalita' di comunicazione.....	47
Fig. I-12 - Sequenze di emissione e di ricezione nei due casi di modo di trasferimento (a) temporalmente trasparente e (b) non temporalmente trasparente.	68
Fig. II-1 - Struttura generale di un sistema di servizio.....	81
Fig. II-2 - Sequenza di istanti di richiesta di servizio (a) e rappresentazione di una realizzazione del processo di ingresso (b) e di quello di servizio (c).....	86
Fig. II-3 - Andamento della funzione di densita' di probabilita' di una distribuzione esponenziale negativa.....	90
Fig. II-4 - Per l'illustrazione dei regimi transitorio e permanente con differenti condizioni iniziali: caso a): $p_0(0) = 1$; $p_1(0) = 0$ caso b): $p_0(0) = 0$; $p_1(0) = 1$	105
Fig. II-5 - Decomposizione dei flussi di richiesta di servizio in un sistema ad attesa con perdita, operante in condizioni di equilibrio statistico.	108
Fig. II-6 - Andamento della probabilita' di rifiuto al variare dell'intensita' media di traffico offerto, per alcuni valori del numero m di serventi.	113
Fig. II-7 - Andamento della portata media normalizzata in un sistema di servizio ad attesa con perdita, al variare del carico medio offerto normalizzato.	114
Fig. II-8 - Andamento del valor medio del ritardo di attesa in funzione del rendimento di utilizzazione	115
Fig. II-9 - Corrispondenza tra la risorsa fisica R condivisa e le relative risorse virtuali R_i ($i = 1, 2, \dots, n$)	126
Fig. II-10 - Alternative di attuazione per le strategie di assegnazione delle risorse.	127
Fig. III-1 - Elementi fondamentali di una architettura di comunicazione.....	140
Fig. III-2 - Stratificazione funzionale di sistemi.	141
Fig. III-3 - Relazione tra gli elementi di una architettura di comunicazione con trasferimento nel modo con connessione.	144
Fig. III-4 - Esempi di connessioni di strato.....	148
Fig. III-5 - Esempio di relazione tra unita' di dati in strati adiacenti.	151

Fig. III-6 - Possibili relazioni tra unità di dati in strati adiacenti.	155
Fig. III-7 - Modello e primitive di servizio di un generico (N)-strato.	157
Fig. III-8 - Classificazione dei servizi di strato.	158
Fig. III-9 - Architettura del modello OSI.	161
Fig. III-10 - Definizione degli elementi di servizio dello strato di applicazione.	163
Fig. III-11 - Definizione di attività e di unità di colloquio nello strato di sessione.	166
Fig. III-12 - Primo esempio di interconnessione di sottoreti.	168
Fig. III-13 - Secondo esempio di interconnessione di sottoreti.	169
Fig. III-14 - Modello generale di una connessione fisica.	170
Fig. IV-1 - Sequenze di emissione e di ricezione nei due casi di modo di trasferimento (a) temporalmente trasparente e (b) non temporalmente trasparente.	177
Fig. IV-2 - Relazione tra modi di trasferimento e relative prestazioni, con riferimento ai gradi di trasparenza temporale e di flessibilità di accesso.	180
Fig. IV-3 - Classificazione degli schemi di moltiplicazione a divisione di tempo.	182
Fig. IV-4 - Moltiplicazione a divisione di tempo: alternative di attuazione con suddivisione dell'asse dei tempi in intervalli temporali (S) e senza suddivisione (U).	183
Fig. IV-5 - Soluzioni per l'alternativa con suddivisione dell'asse dei tempi in intervalli temporali (S): con organizzazione in trama (SF) e senza organizzazione in trama (SU).	184
Fig. IV-6 - Classificazione dei principi di commutazione.	194
Fig. IV-7 - Possibili modi di attuazione di una funzione di commutazione.	196
Fig. IV-8 - Tecnica di commutazione a divisione di tempo: scambio di IT (a) nella stessa trama o (b) in due trame successive.	198
Fig. IV-9 - Modello di riferimento generale di un processo di comunicazione.	204
Fig. IV-10 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con un nuovo modo di trasferimento orientato al pacchetto.	206
Fig. IV-11 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con modo di trasferimento a circuito.	208
Fig. IV-12 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con modo di trasferimento a pacchetto.	209
Fig. IV-13 - Architettura protocollare per il trasferimento dell'informazione di utente in una rete operante con un nuovo modo di trasferimento orientato al pacchetto.	210
Fig. IV-14 - Schemi di moltiplicazione, principi di commutazione e architetture protocollari nei modi di trasferimento orientati al circuito (MTOC) e orientati al pacchetto (MTOP). .	214
Fig. IV-15 - Modi di trasferimento dell'informazione di utente e di quella di segnalazione su un canale unico (a) e su canali separati (b).	221
Fig. IV-16 - Modi di funzionamento in una rete SCC operante secondo l'SS no. 7: modo associato (a), modo non associato (b) e modo quasi associato (c).	223
Fig. IV-17 - Struttura generale delle funzioni dell'SS no. 7.	224
Fig. IV-18 - Struttura della parte di trasferimento di messaggio (MTP).	226
Fig. IV-19 - Architettura protocollare dell'SS no. 7.	226
Fig. IV-20 - Formato della MSU nell'SS no. 7.	227
Fig. IV-21 - Struttura dell'etichetta dell'SS no. 7.	227
Fig. V-1 - Rete a circuito per telefonia e per dati sincroni, con l'impiego della segnalazione a canale comune.	230
Fig. V-2 - Modello di un autocommutatore a circuito.	230
Fig. V-3 - Rete per dati a pacchetto.	234

VIII.2 Tabelle

Tab. I-1 - Suddivisione dei servizi di telecomunicazione in classi di velocità.	34
Tab. II-1 – Simboli usati nella notazione sintetica.....	96
Tab. II-2.....	97
Tab. III-1 - Relazioni tra tipi di informazione e unita' di dati.....	150
Tab. V-1 - Gerarchia numerica asincrona normalizzata in Nord America e in Europa.....	232
Tab. V-2 - Gerarchia numerica sincrona.....	232
Tab. V-3 - Classi di servizio e relativi ritmi binari di emissione per il trasferimento di dati sincroni trattati a circuito o a pacchetto.	233

VIII.3 Equazioni

Errore. Non è stata trovata alcuna voce dell'indice delle figure.