Notiziario Tecnico

Archivio

1/1995

TIM

Global System for Mobile communications (GSM): strumenti e procedure per il controllo della funzionalità di Roaming Internazionale

G. Muratore, M. Tealdo (*)

L'introduzione e la rapida crescita in tutti i paesi europei (e ormai anche extraeuropei) del servizio radiomobile GSM consente agli utenti di utilizzare il proprio telefono anche all'estero.

Tale possibilità è detta di Roaming Internazionale e consiste nel fatto di poter generare e ricevere chiamate dal proprio telefonino anche fuori dal proprio paese, utilizzando la copertura della rete GSM straniera visitata. Usufruire di questa prestazione è estremamente semplice, anche se a ciò si contrappone una notevole complessità nell'interlavoro tra le reti GSM del paese visitato e del paese di appartenenza dell'utente.

Tale complessità unita al fatto che l'Italia, anche per ragioni turistiche, è un paese frequentemente visitato dagli stranieri, ha comportato la necessità di istituire procedure e costruire strumenti "ad hoc" che siano di ausilio all'effettuazione dei controlli periodici relativi al Roaming Internazionale.

1. Introduzione

La prestazione di Roaming Internazionale consente ad un utente di una rete GSM(1) di operare nell'ambito dell'area di servizio della rete di un altro Operatore estero (rete estera "visitata" o Visited-PLMN), utilizzandone le risorse come se si trovasse nell'area di servizio della rete di "casa" (Home-PLMN).

Ciò è possibile grazie alle informazioni d'utente memorizzate in apposite basi di dati (HLR) ed alla "lingua comune" (protocollo MAP) adoperata dalle reti GSM per accedere in tempo reale alle informazioni medesime.

La rete GSM visitata deve infatti riconoscere l'utente straniero e deve pertanto consultare le informazioni depositate nel HLR della Home-PLMN. La Home-PLMN viceversa deve individuare la posizione del proprio utente e deve memorizzarla in HLR. Tale informazione sarà necessaria per indirizzare in modo corretto eventuali chiamate (da Rete Radiomobile o Fissa) dirette verso quell'utente.

È quindi necessario, per garantire la funzionalità di roaming, che l'interlavoro tra reti GSM avvenga senza interruzioni, controllando con periodicità che il "cordone ombelicale" che lega un utente alla propria Home-PLMN non sia interrotto.

2. Il Roaming Internazionale

2.1 *Aspetti tecnici*

Nelle reti radiomobili non esiste corrispondenza tra il numero d'abbonato e la località in cui esso si trova quando debba effettuare o ricevere una chiamata. Di conseguenza una rete radiomobile deve gestire dinamicamente appositi registri che contengono le informazioni necessarie per localizzare i terminali mobili.

In particolare, i dati che individuano l'utente e che ne definiscono il profilo di servizio sono permanentemente

(*) Ing. Giuliano Muratore, ing. Marco Tealdo -Telecom Italia DG- Roma

(1) Il significato degli acronimi è riportato nella lista in coda all'articolo.

memorizzati in un registro di uno degli HLR della Home-PLMN. Inoltre, ogni volta che è riscontrata la presenza dell'utente in una specifica area di localizzazione, tali dati sono temporaneamente trasferiti in un registro della base di dati (VLR) che nella Visited-PLMN gestisce quell'area.

Caratteristica saliente del sistema GSM è quella di non limitare entro i confini nazionali la possibilità di utilizzare il terminale mobile, ma di permettere ciò anche nell'ambito di altre reti ed in altri Paesi. Tale caratteristica comporta uno scambio di informazioni tra registri (degli HLR e dei VLR) di reti GSM diverse, che va ad interessare i collegamenti internazionali di segnalazione instaurati secondo lo standard CCITT n.7.

Nel seguito sono brevemente esaminate le procedure che regolano l'accesso alla rete (o "registrazione"), la chiamata originata e la chiamata terminata per un terminale GSM (apparato terminale + carta SIM), mettendo in rilievo gli aspetti caratteristici connessi al Roaming Internazionale.

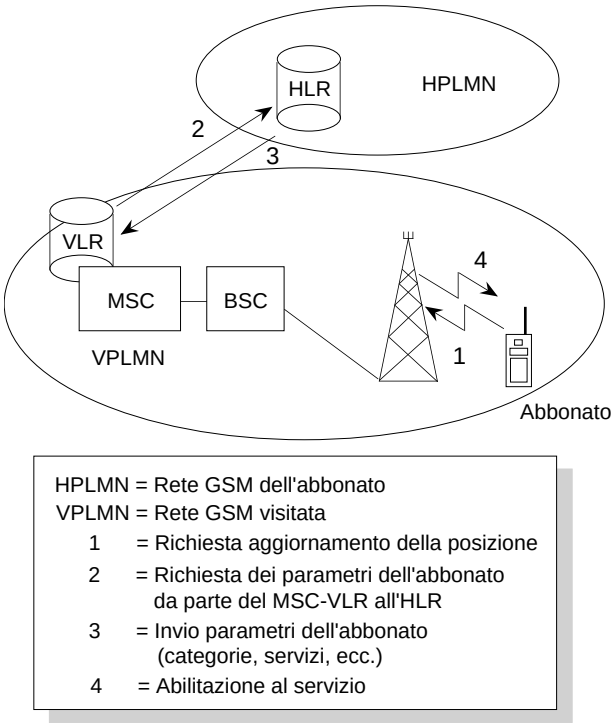


Figura 1

2.1.1 Accesso alla rete (o "registrazione")

Una stazione mobile, proveniente dalla propria rete GSM di casa, ed attivata nell'area servita da una rete GSM straniera, individua tale rete in base alle indicazioni ricevute dai canali radio di controllo e chiede l'autorizzazione per accedere al servizio inviando via radio il proprio identificativo, il cosiddetto IMSI, memorizzato all'interno della carta SIM (fig. 1).

In tale circostanza è instaurato un colloquio in segnalazione MAP tra il VLR della rete visitata e l'HLR della rete di casa, per lo scambio di informazioni relative al riconoscimento ed alla localizzazione dell'utente.

Il VLR richiede infatti all'HLR i parametri che sono necessari per identificare l'utente e quelli che ne caratterizzano il profilo di servizio (messaggi MAP: SEND PARAMETERS e UPDATE LOCATION REQUEST).

Una volta stabilito che l'utente è "in regola" per usufruire delle risorse appartenenti alla rete visitata, il VLR lo registra (stato "attached") ed invia alla stazione mobile il messaggio di abilitazione al servizio (messaggio: LOCATION UPDATE CONFIRM).

Completata con successo questa procedura il VLR sollecita l'HLR ad aggiornare lo stato di localizzazione dell'utente in esame (messaggio MAP: UPDATE LOCATION).

La principale distinzione tra la procedura di registrazione nella rete GSM di casa e la procedura di registrazione in una rete GSM straniera, consiste nella modalità di indirizzamento reciproco tra i registri degli HLR e dei VLR.

Nel caso di scambio informativo tra VLR ed HLR all'interno della stessa rete sono normalmente utilizzati indirizzi della rete di segnalazione (SPC) che appartengono ad un piano di numerazione di livello nazionale. Quando invece la transazione coinvolge nodi di reti GSM straniere, occorre utilizzare un indirizzo internazionale (MGT), che deve essere ricavato dall'IMSI sulla base di tabelle di corrispondenza "ad hoc".

In questo secondo caso particolari nodi della rete di segnalazione (nodi SCCP) contribuiscono ad instaurare il collegamento tra VLR ed HLR, estrapolando dall'indirizzo MGT le informazioni necessarie per realizzare il transito dei dati attraverso le reti fisse internazionali e per gestire il ritorno al sistema di indirizzamento nazionale una volta giunti nella rete dell'HLR o del VLR di destinazione.

2.1.2 Chiamata originata

La chiamata originata da un utente straniero in una Visited-PLMN è trattata dalla rete visitata in modo sostanzialmente analogo a quella originata da un proprio utente. In entrambi i casi, infatti, il trattamento della chiamata è operato sulla base delle informazioni acquisite nel VLR durante la fase di registrazione. Tali informazioni governano ad esempio gli eventuali sbarramenti a particolari tipologie di traffico uscente.

Anche la gestione dei dati di tassazione non subisce sostanziali modifiche nel caso di utente straniero se non per il fatto che tali dati devono

essere trasferiti dall'Operatore della Visited-PLMN all'Operatore estero.

2.1.3 Chiamata terminata

Come già evidenziato in precedenza, nell'ambito della rete radiomobile GSM l'utente è univocamente individuato dal numero IMSI.

Tuttavia nel contesto delle numerazioni telefoniche internazionali l'utente radiomobile GSM è caratterizzato da un numero (MSISDN) che è associato all'IMSI ed è composto dal codice di paese, dall'indicativo distrettuale e dalle cifre d'abbonato.

La chiamata verso un determinato MSISDN, indipendentemente da dove sia originata e dalla posizione dell'utente chiamato, è pertanto instradata in ogni caso verso il paese identificato dal relativo codice (es. 39 per gli utenti GSM italiani) e verso un GMSC (vedi fig. 2) della rete GSM corrispondente all'indicativo distrettuale (335, 338 o 339 per gli utenti GSM Telecom Italia).

Il GMSC, per conoscere lo stato di localizzazione dell'utente ed instradare quindi la chiamata, interroga l'HLR fornendogli il numero MSISDN del chiamato (messaggio MAP: SEND ROUTING INFO).

L'HLR, in base alle informazioni di identificazione (IMSI) e localizzazione associate all'MSISDN in esame, interroga il VLR estero specificando l'IMSI dell'utente chiamato (messaggio MAP: PROVIDE ROAMING NUMBER). Il VLR risponde all'HLR consegnandogli un numero telefonico (MSRN), temporaneamente associato all'IMSI dell'utente, che contiene il codice di paese e l'identificativo distrettuale della rete visitata.

L'MSRN è poi trasferito dall'HLR al GMSC che lo utilizza per completare (tratta GMSC-MSC delle rete visitata) l'instradamento della chiamata.

La caratteristica saliente dell'instaurazione di una chiamata terminata quando l'utente è roaming all'estero consiste, oltre che nel dialogo in segnalazione tra HLR e VLR di reti diverse, nel dover comunque instradare la chiamata passando attraverso la rete GSM di origine del chiamato. Tale situazione comporta la presenza costante di almeno una connessione internazionale tra la rete GSM di origine e quella visitata dall'utente.

Va osservato inoltre che il rispetto rigoroso delle regole di instradamento della fonia, applicate a livello internazionale, determina un caso particolare, il cosiddetto "tromboning", quando un utente in visita presso una rete GSM estera è chiamato dall'interno del paese che sta visitando. In questo caso la chiamata utilizza due connessioni internazionali, prima verso il paese di appartenenza dell'utente chiamato e poi da qui nuovamente verso il paese visitato.

Tale situazione, evidentemente non ottimizzata ed attualmente oggetto di studi e sperimentazioni migliorative da parte degli Operatori GSM, sarà definitivamente superata con il coinvolgimento degli Operatori di rete fissa.

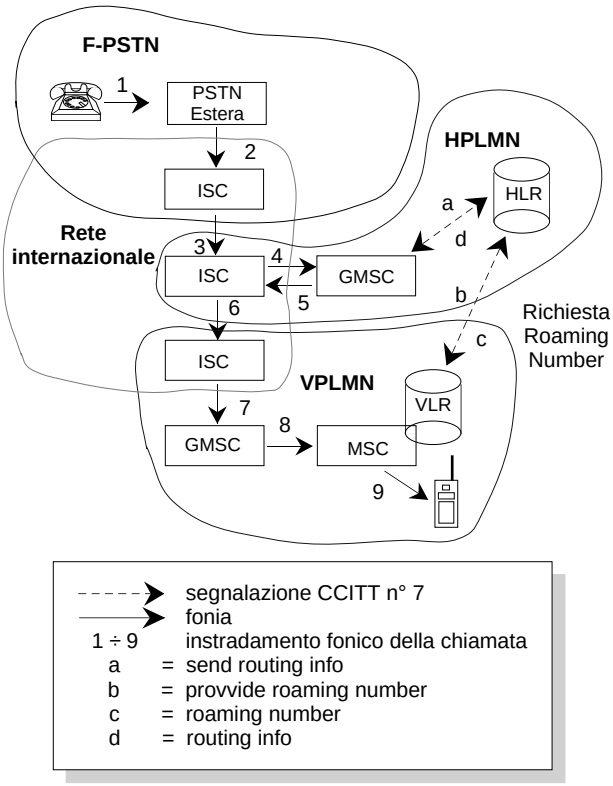


Figura 2

2.2 S*ituazione attuale*

È ormai possibile per un abbonato GSM Telecom Italia utilizzare il proprio telefonino praticamente ovunque in Europa e già in alcuni paesi extraeuropei (vedi tab. 1). Analogamente gli utenti degli Operatori con cui Telecom Italia ha sottoscritto accordi di Roaming Internazionale, possono utilizzare in Italia il proprio telefonino.

La situazione attuale, aggiornata ad Aprile 1994, è riportata nella tab. 1 e sarà comunque soggetta ad un'evoluzione molto rapida.

Tenendo conto infatti che nel mondo il numero degli Operatori radiomobili che hanno adottato o sono in procinto di adottare lo standard GSM ha ormai superato il centinaio (118 al 30/04/95), è prevedibile nel breve termine un ulteriore consistente aumento del numero di Paesi nei quali sarà possibile utilizzare il proprio telefonino GSM.

In particolare per il 1995 Telecom Italia potrebbe estendere il Roaming Internazionale agli Operatori GSM elencati nella tab. 2.

Paese dove si può utilizzare il proprio telefonino GSM	Operatore sotto la cui rete è possibile registrarsi
Australia	OPTUS
Australia	VODAFONE PTY
Austria	AUSTRIA PTT
Belgio	BELGACOM
Danimarca	SONOFON
Danimarca	TELEDANMARK
Finlandia	TELECOM FINLAND
Finlandia	RADIOLINJA
Francia	FRANCE TELECOM
Francia	SFR
Germania	DETEMOBIL
Germania	MANNESMANN
Grecia	PANAFON
Grecia	STET HELLAS
Hong Kong	HONG KONG TLC
Hong Kong	SMARTONE
Inghilterra	CELLNET
Inghilterra	VODAFONE
Irlanda	EIRCELL
Lussemburgo	LUX P&T
Norvegia	NETCOM
Norvegia	TELEMOBIL
Olanda	PTT NETHERLANDS
Portogallo	TMN
Portogallo	TELECEL
Singapore	SINGAPORE TLC.
Spagna	TELEFONICA
Svezia	TELIA MOBITEL
Svezia	COMVIQ
Svezia	EUROPOLITAN
Svizzera	SWISS PTT
Sud Africa	VODACOM
Turchia	TURKCELL
Turchia	TELSIM
Ungheria	WESTEL 900
Ungheria	PANNON

Tabella 1 Roaming Internazionale di Telecom Italia

Paese	Operatore
Andorra	S.T.A.
Australia	MOBILE NET
Cipro	Cyprus Tlc
Emirati Arabi Uniti	ETISALAT
Estonia	Radiolinja eesti as
Islanda	Postur og simi
Jersey	Jersey Tlc
Lettonia	Latvia Mobitel
Libano	Libancell
Malesia	MRCB Tlc SDN BHD
Nuova Zelanda	Bell South
Quatar	QTEL
Russia	North West
Sud Africa	MTN
Tailandia	AIS

Tabella 2 Possibile evoluzione del Roaming Internazionale di Telecom Italia

3. Strumenti di Test di Roaming Internazionale

Alla luce delle considerazioni fin qui svolte è facilmente intuibile quanto possa essere impegnativo il garantire la continuità della prestazione di Roaming Internazionale.

Un possibile incremento in termini di affidabilità può essere ottenuto eseguendo prove "periodiche" che consentano di supervisionare le funzionalità di base del servizio in modo continuativo nel corso dell'anno con riferimento a tutti gli Operatori interessati.

Inoltre, dall'esperienza maturata in campo è emersa anche la necessità di effettuare prove "sequenziali". Tali prove consistono in ripetizioni mirate di un medesimo test (es. chiamata terminata a Mobile), in giorni prestabiliti, con riferimento a particolari Operatori e con modalità rigorosamente definite (es. orario di esecuzione, numero di ripetizioni, ecc.). I risultati ottenuti sono poi elaborati statisticamente per esprimere valutazioni in merito alla qualità del servizio.

Sia nel caso delle prove periodiche che in quello delle prove sequenziali la quantità e la ripetitività dei test che devono essere effettuati ha determinato l'esigenza di ricorrere al supporto di strumenti automatici.

Tali strumenti sono finalizzati a guidare l'operatore nell'attività di esecuzione dei test, e devono essere caratterizzati da un elevato grado di flessibilità per seguire l'evoluzione del servizio, sia in termini di incremento del numero di Operatori interessati, sia in termini di nuove prestazioni.

Nel corso del 1993, a fronte della indisponibilità di prodotti che rispondessero alle esigenze descritte, nell'ambito di Telecom Italia è stato intrapreso lo sviluppo di un tool software (denominato ROAMING) orientato ad una facile ed immediata utilizzazione da parte del personale che opera presso le centrali.

Il programma ROAMING consente all'operatore di "navigare" all'interno del software selezionando da menù le operazioni elementari che compongono l'attività di interesse.

Il programma include le basi di dati che memorizzano le informazioni relative ai risultati delle prove, ai dati caratteristici delle SIM card di prova, ed agli instradamenti di segnalazione CCITT n.7 verso le reti degli Operatori stranieri.

A titolo di esempio, in fig. 3 è riprodotta la schermata di consultazione inerente ai dati di una specifica SIM card, quali il codice di accesso alla SIM (PIN), il numero telefonico (MSISDN), i servizi supplementari abilitati, gli eventuali sbarramenti verso particolari direttrici di traffico telefonico, i limiti di validità temporale ed i recapiti telefonici e facsimile per l'invio di eventuali comunicazioni all'Operatore in esame.

Nel caso di prove periodiche il programma conduce l'attività dell'operatore fornendo tutte le informazioni relative al tipo di test da eseguire, alla SIM card da

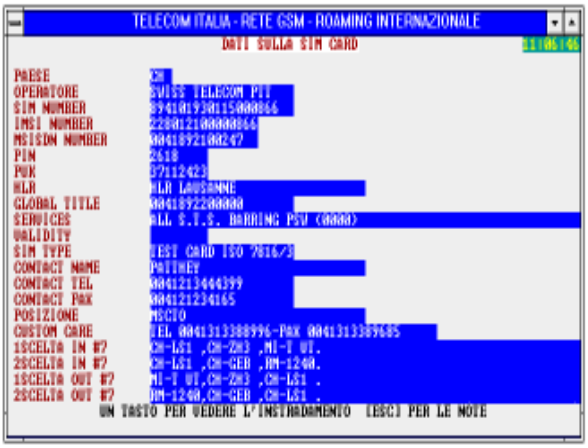


Figura 3

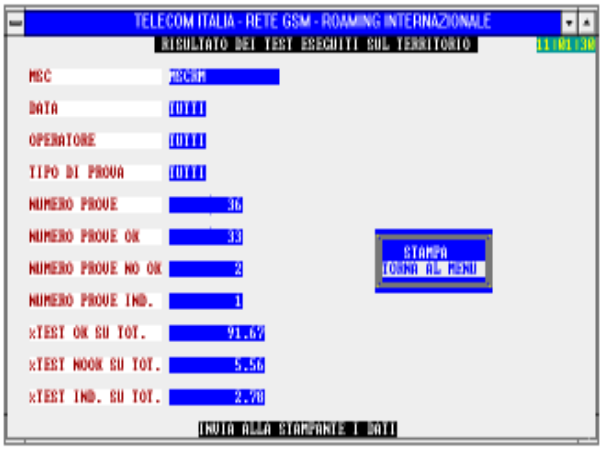


Figura 5

utilizzare per la verifica ed alle modalità di esecuzione, in funzione del giorno della settimana e del sito nel quale si svolgono le prove.

Anche la consultazione delle basi di dati che immagazzinano i risultati dei test è guidata dai menù del programma, in modo tale da poter selezionare gruppi di prove secondo il criterio di interesse.

Nella schermata riprodotta in fig. 4, ad esempio, si riconosce la possibilità di accedere alla consultazione dei risultati ottenuti nel corso di prove periodiche in uno specifico MSC (Roma), con riferimento ad un arco temporale definito e ad uno o più Operatori.

Nelle figg. 5 e 6 sono riprodotte le schermate di presentazione dei risultati relativi a prove rispettivamente periodiche e sequenziali.

Tenuto conto del fatto che il contenuto delle basi di dati del programma ROAMING è oggetto di frequenti scambi tra Direzione Generale e Territorio (rif. par. 4 fig. 7), si è reso opportuno inserire nelle funzionalità del software alcuni meccanismi per la protezione dei dati e per la gestione degli accessi.

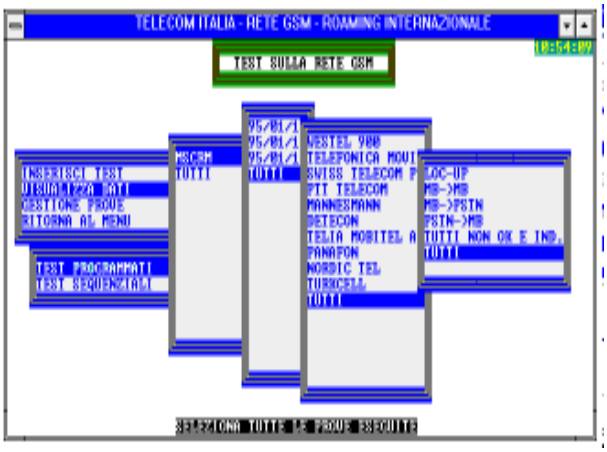


Figura 4

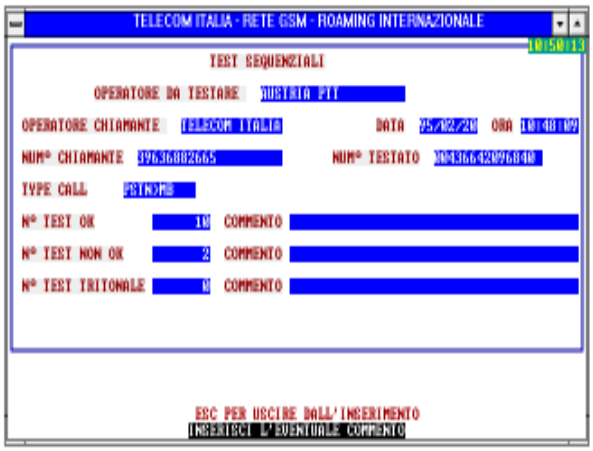


Figura 6

In particolare le informazioni sono automaticamente memorizzate all'interno del computer in forma criptata. Inoltre l'accesso in lettura o scrittura a particolari campi degli archivi elettronici è disciplinato mediante l'impostazione di password che corrispondono a differenti profili di autorità.

Considerando ad esempio l'archivio che memorizza i dati caratteristici delle SIM card sono stati definiti un profilo "di Direzione Generale" ed un profilo "di Territorio". Il primo consente libero accesso in lettura e scrittura a tutti i dati, mentre il secondo permette l'accesso in lettura a tutte le informazioni e l'accesso in scrittura ad un solo campo descrittivo, che è normalmente utilizzato per commenti circa lo stato operativo della SIM card in esame.

Il programma ROAMING è stato sviluppato in ambiente DOS su pacchetto applicativo DBIV. La piattaforma hardware necessaria per garantire il corretto funzionamento del software è costituita da un Personal Computer 386 (o superiore) con disponibilità di spazio su Hard Disk pari ad almeno 10 Mbyte.

4. **Aspetti connessi all'esercizio di rete**

L'esperienza di esercizio relativa al Roaming Internazionale GSM ha evidenziato che è necessario affrontare i problemi di affidabilità adottando adeguate misure in termini di prevenzione.

Ciò ha determinato, ad esempio, l'esigenza di consolidare i rapporti di reciproca collaborazione con gli Operatori di rete fissa al fine di armonizzare gli interventi di aggiornamento hardware/software degli impianti e le attività di modifica degli instradamenti.

Inoltre gli Operatori GSM hanno concordato sulla necessità di effettuare prove telefoniche riproducendo le situazioni operative dell'utente che effettua Roaming Internazionale, ed hanno attivato lo scambio vicendevole di consistenti quantitativi di SIM card.

Ad oggi sono state distribuite tra i 12 MSC/VLR in esercizio approssimativamente 400 SIMcard di circa 40 Operatori stranieri. La gestione di queste SIM card e di quelle (circa 300) che Telecom Italia ha consegnato agli altri Operatori si traduce in una notevole mole di lavoro, sia per la quantità di prove da eseguire, sia per le attività di aggiornamento, manutenzione e custodia dei "set" di carte.

In quest'ambito il programma ROAMING costituisce un ausilio indispensabile per guidare l'operatore nella corretta esecuzione delle prove, siano esse periodiche o sequenziali, e nell'organica archiviazione dei risultati ottenuti.

4.1 *Prove periodiche*

Le prove sono coordinate a livello nazionale, in modo tale che, nell'arco di una settimana (tab. 3), possa essere verificato il corretto funzionamento del Roaming su tutto il Territorio e per tutti gli Operatori di interesse.

Giorno → Sito ↓	Lun	Mar	Mer	Gio	Ven
MSC TO MSC FI MSC BA	Gruppo operat. A	Gruppo operat. E	Gruppo operat. D	Gruppo operat. C	Gruppo operat. B
MSC MI MSC AN MSC PA	Gruppo operat. B	Gruppo operat. A	Gruppo operat. E	Gruppo operat. D	Gruppo operat. C
MSC VE MSC RM	Gruppo operat. C	Gruppo operat. B	Gruppo operat. A	Gruppo operat. E	Gruppo operat. D
MSC GE MSC NA	Gruppo operat. D	Gruppo operat. C	Gruppo operat. B	Gruppo operat. A	Gruppo operat. E
MSC BO MSC CA	Gruppo operat. E	Gruppo operat. D	Gruppo operat. C	Gruppo operat. B	Gruppo operat. A

Tabella 3 Programmazione delle prove periodiche sul Territorio

Ogni sito di MSC/VLR è stato dotato di quanto riportato nel seguito:

a) un "set " di SIMcard degli Operatori esteri;
b) il pacchetto software ROAMING;
c) due apparati GSM.

Nell'arco di una settimana, per ciascun Operatore e presso ogni sito di MSC/VLR, sono effettuate le prove seguenti:

a) registrazione in rete;
b) chiamata originata da telefono GSM e terminata su telefono di rete fissa;
c) chiamata originata da telefono di rete fissa e terminata su telefono GSM;
d) chiamata originata da telefono GSM e terminata su telefono GSM.

Ciascuna prova è definita per ciò che concerne le modalità di esecuzione e gli esiti che debbono essere rilevati. In particolare la prova (1) deve iniziare con la deregistrazione della SIM card in esame per riprodurre le condizioni di accesso alla rete che comportano l'impegno dei collegamenti in segnalazione internazionali (rif. par. 2).

Inoltre, per quanto attiene alle prove (2), (3) e (4), il risultato del test è ritenuto positivo se, entro tre tentativi consecutivi, l'operatore percepisce il tono di libero dal telefono chiamante e almeno due squilli di suoneria dal terminale chiamato. Questo criterio è stato adottato per conciliare le opposte esigenze di snellire la procedura di test e di ottenere risultati che non si prestino ad ambiguità di interpretazione.

Periodicamente o con riferimento a casi "sospetti", l'operatore risponde alla chiamata di prova verificando la presenza della fonia in entrambi i versi. Eventuali anomalie riscontrate nel corso delle prove sono tempestivamente comunicate alla Direzione Generale che provvede, in collaborazione con le equivalenti strutture presso gli Operatori esteri, al coordinamento delle attività di analisi ed alla validazione degli interventi correttivi.

In presenza di un'accertata anomalia di rete occorre normalmente eseguire prove mirate che esulano dal calendario prestabilito. In questo caso, qualora sia necessario effettuare test con la SIM card di uno specifico nodo HLR è possibile, consultando il programma "ROAMING", valutare in modo rapido se e dove la carta di interesse sia disponibile.

I file che documentano i risultati di tutte le prove eseguite sono inviati dal Territorio alla Direzione Generale con cadenza settimanale. Il trasferimento, che inizialmente era operato mediante invio postale di floppy-disk, è ora effettuato via modem su rete commutata (fig. 7).

Dall'esame delle informazioni pervenute in Direzione Generale si evince che i problemi di instradamento in fonia e segnalazione sui collegamenti internazionali sono annoverabili tra le cause di disservizio rilevate con maggior frequenza. Tale situazione rispecchia verosimilmente il fatto che, nonostante siano già stati

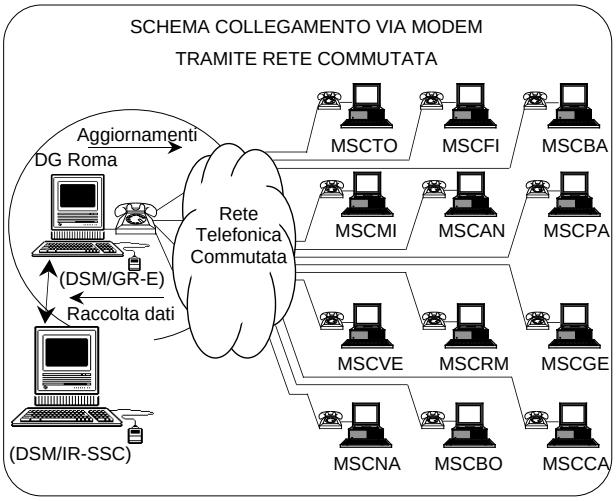


Figura 7

profusi notevoli sforzi in questo senso, molto ancora può essere migliorato in tema di coordinamento delle attività tra gli Operatori di PLMN e PSTN.

4.2 *Prove sequenziali*

Ad oggi, nell'ambito della rete GSM Telecom Italia, sono stati eseguiti due cicli di prove sequenziali, nei mesi di giugno '94 e dicembre '94, secondo le modalità riportate nel seguito:

a) *periodo*: una settimana lavorativa;
b) *numero di Operatori*: quattro, di cui due appartenenti alla stessa Nazione;
c) *tipo di prova*: chiamata terminata a Mobile (caso peggiore per il fenomeno del "tromboning", rif. par. 2);
d) *fasce orarie di prova*: 09:00-10:30, 10:30-12:00, 14:30-16:00;
e) *procedura operativa*: esecuzione di un ciclo di prove al giorno, per ciascun Operatore in esame, presso ogni MSC/VLR (per ciclo di prove si intende l'esecuzione in sequenza di 20 chiamate terminate a Mobile, di cui 10 originate da PSTN e 10 da apparato BL(2)).

Le campagne di test sono state attivate con lo scopo di ottenere riscontri in merito al funzionamento del Roaming Internazionale, per gli Operatori che attualmente generano la maggior quantità di traffico nell'ambito della rete Telecom.

In particolare, l'obiettivo di conoscere le correlazioni tra le criticità di gestione del Roaming Internazionale e la qualità del servizio percepita dagli utenti stranieri, ha determinato la necessità di riferirsi a condizioni di test particolarmente critiche (punti (c) e (d) delle modalità di prova).

Ciascuna sessione di prove sequenziali ha comportato l'esecuzione di circa 5000 prove di chiamata, uniformemente distribuiti nell'arco temporale delle cinque giornate su tutto il Territorio. Questa mole di test è stata ritenuta la minima sufficiente per dare la necessaria consistenza alle indicazioni attese.

5. **Evoluzione futura**

Come già evidenziato (par. 2) il numero degli accordi di Roaming e la quantità di prestazioni e servizi offerti dalle reti GSM sono in rapida crescita e vanno ad incrementare sensibilmente il livello di complessità delle procedure di test.

Alcuni Costruttori hanno recepito l'importanza di queste problematiche e stanno avviando lo sviluppo di strumenti di test caratterizzati da un elevato grado di automazione. Soluzioni di questo tipo consentono tra l'altro di aumentare la frequenza di esecuzione dei test senza impiego aggiuntivo di risorse umane con evidenti vantaggi sul fronte della prevenzione delle anomalie.

In generale (fig. 8) è prevista la realizzazione di una "Test Unit" dotata di appositi alloggiamenti multipli per le SIMcard che devono essere provate. L'unità di test si dovrebbe interfacciare direttamente con il nodo MSC simulando la stazione Mobile e la catena degli apparati radio. In fig. 8 è inoltre rappresentata una stazione di lavoro, collegata alla Test Unit, che fornisce il supporto alle attività di operatore. Tali attività consistono principalmente nella programmazione dei test, nella consultazione dei

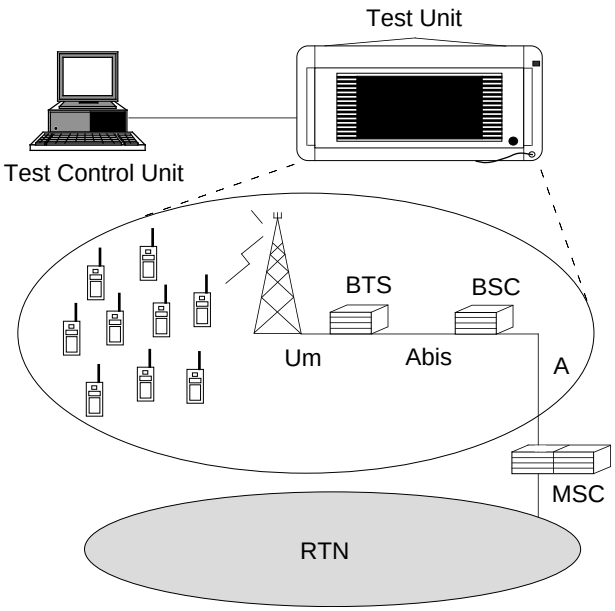


Figura 8

(2) Con il termine BL ci si riferisce ad un comune apparato telefonico connesso direttamente allo stadio di commutazione dell'impianto MSC.

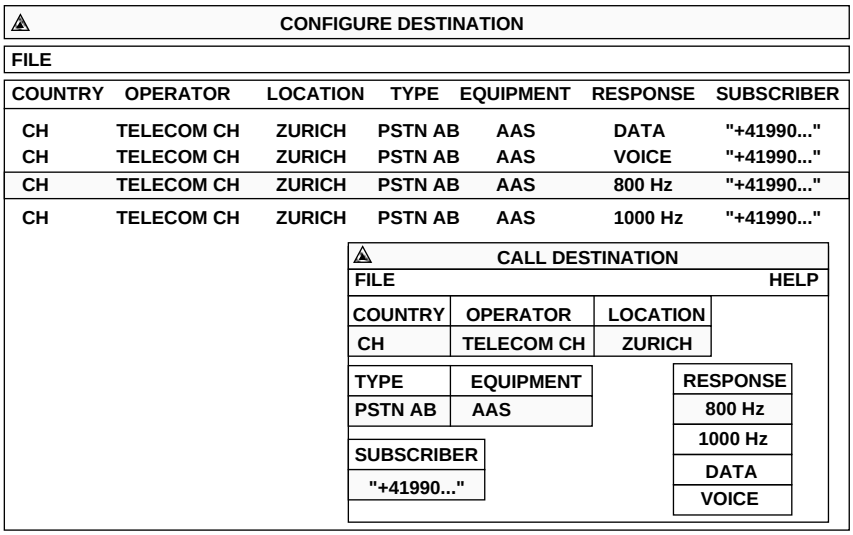


Figura 9

risultati e nell'elaborazione di report statistici.

In futuro, in aggiunta a queste funzionalità di base, il "Roaming Tester" potrebbe consentire l'invio automatico di "trouble-report" per comunicare agli Operatori interessati eventuali anomalie.

Allo stato attuale per il conseguimento di una completa automazione dei test di Roaming è necessario superare alcune criticità.

Attualmente il costo di un apparato di automazione dei test, considerato l'elevato livello di specializzazione, è ipotizzabile nell'ordine delle centinaia di milioni di lire. Ciò porterà inevitabilmente a limitare la collocazione delle macchine di prova in un numero ristretto di siti, riducendo le potenzialità diagnostiche su anomalie localizzate in particolari MSC.

Inoltre, per alcuni tipi di prova, è in genere previsto che il "Roaming Tester" verifichi l'esito della chiamata sulla base di una conferma (es. tono di risposta a frequenza predeterminata, o combinazione di toni in multifrequenza) proveniente da un dispositivo automatico collocato presso la rete GSM di casa della SIM card in esame. Escludendo allora l'eventualità che tutti gli Operatori GSM dispongano della medesima apparecchiatura di test, è comunque necessario un accurato coordinamento dei test e la predisposizione di apparati risponditori che forniscano risposte preregistrate (es. annuncio contenente il nome dell'Operatore o altre informazioni utili per i test). Tali risposte dovrebbero essere concordate tra gli Operatori e configurate nella Test Unit (fig. 9).

Va comunque osservato che, nella fase di apertura del Roaming con un nuovo Operatore, in concomitanza dell'introduzione di nuovi servizi, e nella fase di analisi e risoluzione di un'anomalia di rete, l'intervento manuale specialistico non può attualmente essere sostituito dall'automazione.

Acronimi

BSC	Base Station Controller
BL	Both way Line
GMSC	Gateway MSC
GSM	Global System for Mobile Communications
HLR	Home Location Register
IMSI	International Mobile Subscriber Identity
ISC	International Switching Center
MAP	Mobile Application Part
MGT	Mobile Global Title
MSC	Mobile Switching Center
MSISDN	Mobile Station ISDN number
MSRN	Mobile Station Roaming Number
PIN	Personal Identification Number
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
SCCP	Signalling Connection Control Part
SIM	Subscriber Identity Module
SPC	Signalling Point Code
TUP	Telephone User Part
VLR	Visited Location Register

Bibliografia

[1] *GSM Technical Specification.* ETSI.

[2] Mouly, M.; Pautet, M-B.: *The GSM System for Mobile Communications.*

[3] Giordani, M.; Grimaldi, F.; Santinelli, M.: *Global System for Mobile communications (GSM): le caratteristiche e l'applicazione nella rete SIP*. «Notiziario Tecnico SIP» Vol. 3, n. 1, Aprile 1994.

Impiego della tecnica di multiplazione a sottoportante per il trasporto di canali analogici e numerici su fibra ottica in rete d'accesso

G. Aureli, S. Betti, V. C. Di Biase (*)

L'impiego della fibra ottica nella rete di accesso ha ricevuto recentemente un forte impulso, determinato soprattutto dalla prospettiva di fornire servizi a larga banda all'utenza residenziale. La possibilità di realizzare reti integrate allo scopo di fornire servizi voce, dati e video è sempre stato l'obiettivo dei gestori di telecomunicazioni: una rete integrata risulta infatti potenzialmente più economica e flessibile rispetto ad un'infrastruttura caratterizzata da reti distinte, dedicate ai diversi servizi. Negli ultimi anni, da un lato il processo di evoluzione tecnologica, dall'altro la tendenza ad una diffusa "deregolamentazione" nell'ambito dei servizi e delle infrastrutture di rete, hanno portato ad una "convergenza" e ad una progressiva "integrazione" dei servizi che trovano nel "servizio basato su immagini" (sia multimediale sia, più semplicemente, televisivo) l'elemento di riferimento. Nell'ambito di tale contesto, in questo articolo vengono esaminate alcune importanti tematiche tecniche relative al trasporto di segnali per la fornitura di servizi nella rete di accesso. In particolare, viene considerata la tecnologia ibrida fibra ottica/cavo coassiale, che risulta molto promettente, per quanto riguarda flessibilità d'impiego ed economicità d'installazione, e permette di fornire servizi sia di tipo diffusivo che interattivo. Sono inoltre messi in risalto i principali problemi relativi al trasporto sulla sezione ottica della rete di accesso, considerando la tecnica di multiplazione a sottoportante (Sub-Carrier Multiplexing, SCM) e la modulazione dei segnali sia in formato analogico che numerico.

1. Introduzione

La possibilità di fornire servizi voce, dati e video su una rete integrata è sempre stato l'obiettivo di ogni gestore di telecomunicazioni. Una rete integrata risulta potenzialmente più economica e flessibile rispetto ad un'infrastruttura caratterizzata da reti distinte, dedicate ai diversi servizi. Tuttavia, il processo di evoluzione tecnologica e le regolamentazioni finora vigenti hanno determinato una diversificazione dei servizi e delle reti per la loro offerta, creando uno scenario di reti multiple (distribuzione video, telecomunicazioni). Le architetture di rete si sono quindi sviluppate seguendo indirizzi differenziati e risultano ottimizzate per la tipologia di servizio a cui sono rivolte: è sufficiente confrontare, a questo proposito, la semplice architettura di rete per la diffusione di programmi televisivi con la complessa struttura della rete di telecomunicazione.

Questa situazione ha determinato un consolidamento delle diverse funzioni degli operatori di rete con differenti regolamentazioni (o deregolamentazioni), definite per i servizi specifici. In sostanza, si sono creati due contesti separati come filosofia di rapporto con l'utenza (interattiva/diffusiva), come struttura di rete (punto-punto/punto-multipunto), come mezzi trasmissivi utilizzati (doppino in rame, cavo coassiale, portante radio o fibra ottica).

L'ottimizzazione delle reti nei due contesti ha portato a due diverse topologie di rete. Mentre per la telefonia la topologia di riferimento è risultata quella a "stella", sufficientemente economica per servizi a banda stretta, diversa è la situazione per quanto riguarda i servizi diffusivi a larga banda, per i quali tale struttura sarebbe risultata estremamente onerosa. In questo caso sono state realizzate strutture alternative per collegamenti punto-multipunto come quelle per la radiodiffusione di

(*) Ing. Silvello Betti -Fondazione Ugo Bordoni- Roma; ing. Guglielmo Aureli, sig. Valerio Claudio Di Biase -Telecom Italia DG- Roma

segnali video e quelle di tipo "tree-and-branch" per la televisione via cavo (Community Antenna TV, CATV).

Nelle strategie di evoluzione della rete hanno avuto un ruolo di rilievo anche fattori di natura politico-legislativa e di scelte di politica industriale che hanno spesso influenzato scelte di sviluppo negli specifici ambiti nazionali. Come conseguenza si sono delineati scenari piuttosto eterogenei, con elevate specificità nazionali: a questo proposito, basti considerare la differenza di sviluppo dei servizi video tra i paesi in cui la diffusione dei segnali televisivi è stata regolamentata o meno.

Ciò nonostante sono stati spesso effettuati tentativi di sviluppo di servizi congiunti di comunicazione e diffusivi come, ad esempio, nel caso dei servizi televideo (funzione di distribuzione di informazioni contemporaneamente ai segnali televisivi) e dei servizi videotel (funzione di trasmissione dati con possibilità di interattività) con risultati piuttosto limitati poiché le diverse reti si sono sviluppate autonomamente, con riferimento a precise tipologie di servizio.

Negli ultimi anni, da un lato l'evoluzione tecnologica, dall'altro la tendenza ad una diffusa deregolamentazione nell'ambito dei servizi e delle infrastrutture di rete hanno portato ad una "convergenza" dei servizi e ad una loro progressiva integrazione, trovando nel servizio "basato su immagini" (sia multimediale sia, più semplicemente, televisivo) l'elemento di riferimento. Tale passaggio potrà richiedere tempi non brevi dato che gli operatori del settore, siano essi gestori delle reti di telecomunicazione o di reti CATV, dovranno rivedere la struttura della propria architettura di rete.

Solo negli ultimi anni si è verificata una spinta da parte di entrambe le categorie di gestori verso una convergenza dei due tipi di servizi. Le modalità con cui, in base a questa operazione di convergenza, gli aspetti di competizione-collaborazione tra i servizi, e quindi tra i gestori, potranno evolvere, dipende dallo sviluppo che si avrà nei settori economico, legislativo, politico e tecnologico. Quest'ultimo ha fornito gli elementi principali per una possibile convergenza tra i diversi tipi di servizi; in particolare, vanno menzionati progressi concernenti:

- la codifica e compressione di segnali video a bassa frequenza di cifra;
- la capacità di memorizzazione ed elaborazione dei dati;
- la capacità di multiplazione e commutazione a pacchetto (ATM);
- la tecnologia per la rete di accesso (fibra ottica, cavo coassiale, doppino in rame o portante radio).

Lo sviluppo di tali tematiche consentirà di realizzare reti con elevate prestazioni in termini di capacità di trasporto e flessibilità per la gestione dei nuovi servizi a larga banda. Le potenzialità tecnologiche ormai disponibili hanno sensibilmente ridotto il divario esistente tra le industrie legate rispettivamente alla distribuzione del segnale televisivo ed alla rete di telecomunicazione: in particolare, il massiccio impiego delle fibre ottiche nella rete di accesso consentirà l'offerta di una gamma molto ampia di servizi, sia di tipo diffusivo che interattivo.

Anche se sembra unanimemente riconosciuto che, in futuro, sarà conveniente integrare le reti di telecomunicazione in una rete numerica commutata basata sull'impiego di fibre ottiche, differenze permangono tra i diversi operatori del settore riguardo alle modalità da seguire ed alla gradualità necessaria per la realizzazione di tale prospettiva, anche in relazione ai rischi connessi agli ingenti investimenti richiesti, ed all'incertezza di un mercato non ancora definitivamente consolidato. Si stanno pertanto affermando, per il breve-medio termine, soluzioni tecniche di tipo ibrido, di carattere transitorio ma in continua evoluzione in uno scenario in cui, relativamente ai servizi di tipo televisivo, si assiste ad una competizione tra possibili opzioni che riguardano la scelta tra modulazione analogica e digitale, la qualità e la quantità di canali da fornire all'utenza, la scelta del portante fisico di trasmissione (fibra ottica, cavo in rame o radiopropagazione), la scelta di fornire servizi su base commutata o su base diffusiva.

Le soluzioni principali che si stanno delineando prevedono la trasmissione sia di canali analogici che numerici, eventualmente sottoposti a compressione, l'uso di fibre ottiche per i collegamenti di giunzione e la parte primaria della distribuzione, con collegamenti in cavo coassiale per la parte secondaria della distribuzione.

La realizzazione di reti sovrapposte (overlay network) alla rete di telecomunicazione, di facile implementazione, permette di ottenere i vantaggi economici connessi alla condivisione delle risorse, limitati inizialmente all'uso delle stesse infrastrutture ma che potranno essere più significativi in una seconda fase, con il consolidamento delle tecnologie ed una più precisa definizione da un lato dello scenario di regolamentazione, dall'altro delle condizioni di mercato.

Nell'ambito della rete di accesso, al di là di soluzioni intermedie e temporanee quali l'ADSL (Asymmetrical Digital Subscriber Loop), indicate per l'utilizzo in particolari nicchie di mercato, la tendenza è quella di sviluppare sistemi in fibra ottica.

Per le architetture di rete l'orientamento più diffuso è verso una soluzione di rete in fibra ottica con nodi periferici, condivisi da gruppi di 100-400 clienti, in cui i segnali ottici vengono convertiti in segnali elettrici che sono trasmessi verso l'utente o tramite cavo coassiale (con struttura "tree-and-branch") e/o tramite doppino (con struttura punto-punto). Tipicamente, la capacità dei nodi per le reti CATV è più elevata (500-2000 clienti per nodo) ma la prospettiva di inserire servizi di tipo telefonico e video ad alta interattività, con richiesta quindi di una significativa banda di ritorno verso la centrale, ha determinato una riduzione del numero di utenti serviti dal nodo.

Per quanto riguarda le tecniche di multiplazione sono stati individuati due possibili approcci:

- *tecniche in banda-base (Baseband)*: i canali numerici associati ai servizi interattivi sono multiplati a divisione di tempo (TDM), così da generare un unico flusso ad elevata frequenza di cifra che modula la portante ottica; il segnale così ottenuto viene trasportato in rete analogamente a quanto avviene per i tradizionali sistemi numerici operanti nella rete di telecomunicazione;
- *tecniche passa-banda (Passband)*: sia i canali analogici sia quelli numerici vengono multiplati mediante l'uso di sottoportanti a radiofrequenza (RF), opportunamente separate in frequenza; il segnale risultante modula la portante ottica che è trasmessa ai nodi periferici della rete. Una delle principali tecniche impiegate in tal caso è quella della multiplazione a sottoportante (Sub-Carrier Multiplexing, SCM).

Le tecniche in banda-base sono in genere preferite dagli operatori delle reti di telecomunicazione in quanto si integrano efficacemente con i sistemi già operanti nella rete. Gli ostacoli maggiori all'impiego di tali tecniche sono connessi ad una minore flessibilità rispetto alla tecnica SCM: inferiore efficienza spettrale e necessità che tutti i canali associati ai servizi abbiano formato numerico. Quest'ultimo aspetto riveste particolare importanza e fa si che tale approccio possa risultare efficace solo quando le tecniche di codifica e decodifica video siano consolidate sia dal punto di vista della standardizzazione che dei costi. L'impossibilità di un trasporto simultaneo di canali analogici e numerici sulla stessa portante ottica, come invece può avvenire con la tecnica SCM, non consente un'applicazione immediata delle tecniche in banda-base per servizi già disponibili come, ad esempio, quelli video analogici.

Le tecniche SCM sono generalmente preferite dagli operatori CATV per diversi motivi:

- analogia con le trasmissioni via etere;
- tecnologia consolidata a costi contenuti;
- compatibilità con i ricevitori domestici per i segnali in formato analogico;
- compatibilità con le strutture CATV esistenti;
- evoluzione flessibile dei formati di modulazione (migrazione da canali analogici a canali numerici; canali con differenti formati di modulazione numerica);
- elevata efficienza spettrale (l'uso di tecniche di modulazione multilivello per i segnali numerici permette di ottenere agevolmente efficienze spettrali di 4 bit/s/Hz);
- capacità di trasporto flessibile in funzione dello sviluppo della rete e delle esigenze dell'utenza.

Recentemente, diversi importanti gestori di reti di telecomunicazione del Nord America hanno optato per la tecnica SCM: in particolare, negli Stati Uniti, tutte le Regional Bell Operating Company (RBOC), alcune Local Exchange Carrier (LEC), tra le quali la GTE, e tutte le compagnie CATV hanno commissionato importanti progetti per la realizzazione, o il rinnovamento, nel caso delle compagnie CATV, di reti ibride fibra ottica/cavo coassiale in cui l'impiego della tecnica SCM consente l'implementazione di strutture "Full Service Network", termine coniato da Time Warner, secondo operatore CATV degli Stati Uniti, per identificare una rete in grado di fornire tutti i servizi di telecomunicazione.

Questo lavoro si propone di approfondire, prescindendo dalla definizione delle architetture di rete e dai servizi forniti, le principali tematiche relative al trasporto in rete di accesso di segnali analogici e numerici con la tecnica SCM.

Dopo aver introdotto, nel paragrafo 2, le tecniche di multiplazione a sottoportante, sono brevemente presentate, nel paragrafo 3, le principali caratteristiche dei segnali video, sia per quanto riguarda i formati di modulazione (analogica e numerica) che le tecniche di compressione (MPEG).

Nel paragrafo 4 sono analizzate le prestazioni dei sistemi di trasmissione ottica SCM sia nel caso di canali analogici che numerici: sono considerate le sorgenti di rumore, gli effetti di distorsione nonlineare ed infine è valutato l'impiego di amplificatori ottici in fibra.

Il paragrafo 5 introduce le principali problematiche relative ai sistemi SCM nel caso di trasmissione simultanea di segnali video AM-VSB e multilivello QAM.

Viene infine riportata una rassegna delle principali realizzazioni sperimentali di sistemi ottici SCM nel caso di trasmissione numerica o ibrida analogico/numerica.

2. Tecniche di multiplazione a sottoportante (Sub-Carrier Multiplexing, SCM)

La distribuzione via cavo del segnale televisivo, originariamente denominata CATV (Community Antenna TV), è nata inizialmente per estendere la portata dei segnali televisivi diffusi via etere nelle aree in cui la ricezione era difficoltosa e la qualità inaccettabile. Le difficoltà di ricezione potevano essere dovute sia alla distanza dal trasmettitore, sia all'ambiente particolarmente ostile. Successivamente, in alcuni paesi (ad esempio, Stati Uniti, Nord Europa), questa tecnica di distribuzione basata sul cavo coassiale si è diffusa parallelamente alla tecnica convenzionale via etere. I sistemi basati sull'uso del cavo coassiale con architettura ad albero (tree-and-branch) si sono infatti dimostrati efficienti anche per la distribuzione su larga scala dei segnali video.

Negli ultimi anni un interesse crescente è stato rivolto verso l'impiego delle fibre ottiche anche perché sono state ormai raggiunte le prestazioni limite della

"tecnologia coassiale". Risulta, infatti, particolarmente attraente l'impiego delle fibre ottiche singolo modo le cui potenzialità, in termini di bassa attenuazione (≈0.2 dB/km alla lunghezza d'onda di 1.5 μm) e larghezza di banda (≈20 THz) possono essere vantaggiosamente sfruttate sia per incrementare le prestazioni delle reti esistenti, sia per la realizzazione di reti di distribuzione di elevatissima capacità [1-2].

L'evoluzione verso architetture di rete di accesso "interamente" o "quasi interamente" ottiche è consentita anche dal progressivo miglioramento delle prestazioni dei componenti ottici (laser, modulatori elettro-ottici, amplificatori ottici, fotodiodi), parallelamente ad una graduale riduzione dei costi.

Una tappa fondamentale verso una rete d'accesso ottica è rappresentata dalla configurazione FTTC (Fibre-To-The-Curb), realizzabile con l'impiego della fibra ottica nella rete primaria. Il collegamento ottico risulta in questo caso condiviso da più utenti fino ad un determinato punto della rete di accesso; da tale punto fino all'utente, la connessione avviene tramite cavi in rame. Passo successivo potrebbe essere rappresentato dall'evoluzione di questa rete verso la configurazione FTTB (Fibre-To-The-Building) con la fibra impiegata fino al condominio e, in una prospettiva a più lungo termine, verso la configurazione FTTH (Fibre-To-The-Home) con la fibra fino a casa del cliente.

In questo contesto, la tecnica SCM permette di distribuire segnali a larga banda sfruttando vantaggiosamente la larghezza di banda delle fibre ottiche singolo modo mediante l'impiego delle consolidate tecnologie a RF, richiedendo dispositivi ottici con requisiti di elevata linearità e basso rumore [3-7].

Inoltre, essa è compatibile con la possibile evoluzione verso reti ottiche strutturate in maniera "gerarchica", in cui la tecnica di multiplazione a divisione di lunghezza d'onda (Wavelength Division Multiplexing, WDM), con spaziatura tra i canali nel dominio ottico dell'ordine dei nanometri, venga sovrapposta all'SCM in modo tale da sfruttare in maniera ancora più efficiente la larghezza di banda delle fibre ottiche singolo modo.

In linea di principio, la tecnica SCM può essere considerata come una naturale evoluzione nel dominio ottico delle tecniche a radiofrequenza cui solitamente si fa riferimento con il nome generico di Modulazione Multiportante (Multicarrier Modulation, MCM) [8]. Nel caso di trasmissione in fibra ottica di segnali SCM, un'unica portante ottica, ad una data lunghezza d'onda, è modulata da più sottoportanti a RF, ognuna delle quali è a sua volta modulata dal messaggio associato ad un determinato canale.

Questa tecnica di multiplazione permette la trasmissione su fibra ottica di canali analogici e numerici per dati, segnali audio, video, video ad alta definizione e, in linea di principio, qualunque combinazione di servizi mediante un adeguato posizionamento delle portanti nella banda disponibile. La flessibilità offerta rende questa tecnica estremamente interessante come "piattaforma fisica" per applicazioni a larga banda, anche nel caso in cui i servizi distribuiti provengano da sorgenti distinte ed adottino formati di modulazione diversi e differenti larghezze di banda [9]. Tali possibilità fanno sì che i sistemi basati sulla tecnica SCM possano seguire l'evoluzione della tecnologia video e delle tecniche di compressione dei segnali numerici, anche in relazione ai mutamenti delle condizioni del mercato.

Per la rivelazione del segnale ottico può essere impiegato un ricevitore a rivelazione diretta (basato su fotodiodi PIN o APD) oppure un ricevitore ottico coerente nel caso in cui sia necessaria una sensibilità più elevata (fino a circa 10-15 dB). Tale opportunità, in aggiunta alle potenzialità offerte dall'impiego di amplificatori in fibra drogata all'erbio, sia come booster in trasmissione sia come amplificatore di linea, rendono la tecnica SCM estremamente flessibile nel caso in cui sia necessario aumentare le prestazioni della rete. A fronte di una sensibilità più elevata del ricevitore è comunque da tenere presente la maggiore complessità di un ricevitore basato su tecniche di rivelazione coerente rispetto al caso di rivelazione diretta, complessità principalmente dovuta alle caratteristiche più stringenti richieste ai dispositivi ottici (purezza spettrale ed elevata sintonizzabilità dei laser, elevata larghezza di banda e bassa perdita d'inserzione del modulatore elettro-ottico nel caso di modulazione angolare della portante ottica), con conseguente sensibile incremento dei costi.

Nel seguito del lavoro verranno trattati esclusivamente i problemi relativi ai sistemi SCM con rivelazione diretta, di interesse applicativo più immediato.

Lo schema a blocchi di riferimento di un sistema SCM con rivelazione diretta è mostrato in fig. 1: i segnali in banda base analogici o numerici modulano sottoportanti a RF; tali segnali risultano pertanto traslati intorno alle frequenze delle sottoportanti. Successivamente, le sottoportanti vengono sommate, ad esempio mediante combinatori a RF, in modo tale da ottenere un unico segnale che modula d'intensità la portante ottica, o mediante modulazione diretta di una sorgente laser a semiconduttore, oppure tramite un modulatore elettro-ottico d'intensità esterno. In ricezione, dopo la rivelazione del segnale ottico mediante un fotodiodo, la demultiplazione è effettuata, a livello elettrico, mediante un oscillatore locale sintonizzabile, un mixer ed un opportuno filtro passa-banda; ciò permette di selezionare il canale SCM da inviare al demodulatore.

Uno degli svantaggi principali della tecnica SCM è rappresentato dalla non-linearità dei componenti, principalmente delle sorgenti laser che, essendo modulate da più sottoportanti, possono dare luogo ad effetti di distorsione armonica e a prodotti di

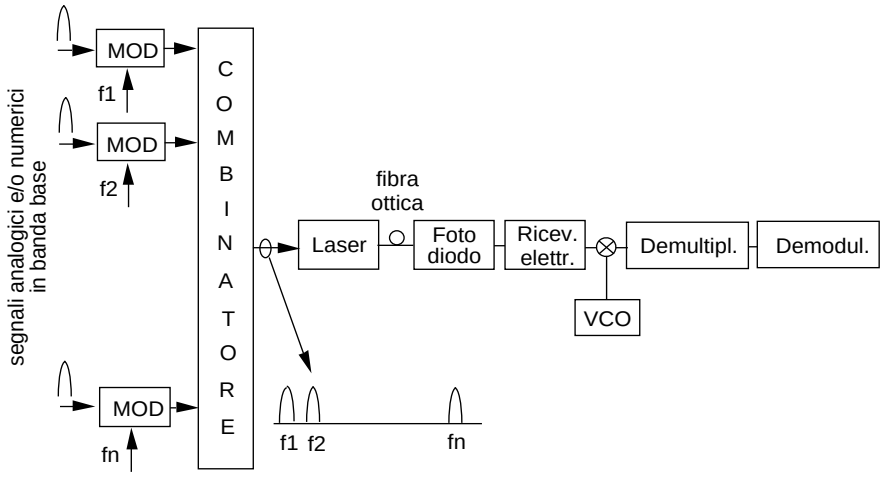


Figura 1 Schema a blocchi di un sistema SCM del tipo modulazione d'intensità - rivelazione diretta

intermodulazione tra i segnali. Questo problema risulta particolarmente rilevante nel caso in cui la sorgente laser sia modulata da canali analogici, così da limitare l'indice di modulazione di ciascun canale, ovvero il numero dei canali. Il problema delle distorsioni di non linearità dei laser a semiconduttore è stato ampiamente studiato e sono stati sviluppati laser a semiconduttore di tipo DFB (Distributed Feedback Laser) con elevate caratteristiche di linearità della risposta potenza ottica/ corrente di polarizzazione.

3. **Tecniche di modulazione del segnale video**

3.1 *Formati di modulazione analogica*

Il formato di modulazione dei segnali video distribuiti via cavo all'utenza domestica è attualmente di tipo AM-VSB (Amplitude Modulation-Vestigial Side Band): in fig. 2 è mostrato lo spettro di potenza di un segnale video PAL (norma G): la larghezza di banda nominale del segnale è pari a 8 MHz (norma B, 7 MHz) ed il limite inferiore della banda è a 1.25 MHz al di sotto della portante video; la portante audio, a 5.5 MHz al di sopra della portante video, è modulata di frequenza con deviazione pari a ±50 kHz, mentre la sottoportante di colore si trova a 4.43 MHz al di sopra della portante video.

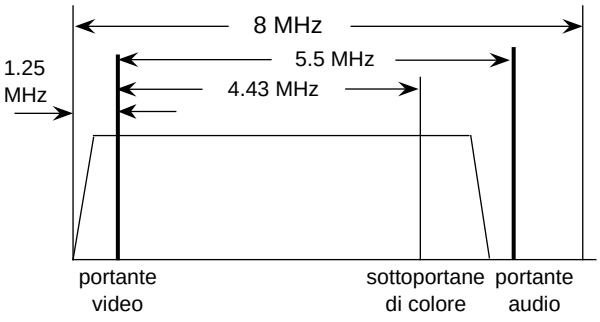


Figura 2 Spettro di potenza di un segnale video PAL (norma G)

Sebbene un canale video PAL standard occupi complessivamente una banda di 8 MHz, il rumore è specificato relativamente ad una larghezza di banda di 5 MHz, centrata entro il canale a 8 MHz (contro i 4 MHz nel caso di un canale video standard NTSC, con larghezza di banda nominale pari a 6 MHz).

Il rapporto segnale-disturbo pesato (weighted S/N) si ottiene sommando al rapporto portante-disturbo il guadagno di demodulazione: nel caso di un segnale video PAL G/B, per un canale AM-VSB si ottiene

$$S/N\ (dB) = C/N\ (dB) + 1 \div 1.2\ dB. \qquad (1)$$

Tipicamente, in rete di accesso a livello della connessione di utente viene richiesto un valore di C/N pari a 48 dB.

Un altro formato di modulazione analogico che è diffusamente impiegato per trasmissioni di segnali video da satellite ed è stato adottato per collegamenti di elevata qualità (ad esempio, nell'ambito di reti di contributo tra studi televisivi) è la modulazione di frequenza (FM). Con questo formato di modulazione si ottiene un sensibile aumento del rapporto segnale-disturbo a prezzo di un incremento della larghezza di banda richiesta dal canale. Come è noto, in base alla regola di Carson, un valore approssimato per la larghezza di banda richiesta B è dato da $B \approx \Delta f_{pp} + 2f_m$, essendo Δf_{pp} il valore picco-picco della deviazione di frequenza del modulatore e f_m la frequenza della sottoportante audio. Il rapporto segnale-disturbo pesato S/N risulta in questo caso espresso dalla relazione

$$S/N(dB) = C/N(dB) + 10\log\left[\frac{3B}{2f_v}\left(\frac{\Delta f_{pp}}{f_v}\right)^2\right] + W + PE \quad (2)$$

in cui:

- f_v rappresenta la banda video;
- W è un fattore peso che tiene conto della risposta non uniforme dell'occhio al rumore bianco nella banda video;
- PE è il fattore di preenfasi.

L'incremento dell'S/N rispetto al C/N varia per i diversi sistemi ma è compreso, tipicamente nell'intervallo 39÷44 dB. Ad esempio, per un segnale NTSC trasmesso in banda C da satellite (3.7÷4.2 GHz), si hanno i seguenti valori dei parametri

$$\begin{cases} \Delta f_{pp} = 22.5 \text{ MHz} \\ f_m = 6.8 \text{ MHz} \\ f_v = 4.2 \text{ MHz} \\ W = 13.8 \text{ dB} \\ PE = 0 \div 5 \text{ dB} \end{cases}$$

tali da ottenere (cfr. (2)):

$$B \approx \Delta f_{pp} + 2f_m = 36.1 \text{ MHz} \quad (3)$$

$$S/N \text{ (dB)} = C/N \text{ (dB)} + 39 \div 44 \text{ dB}. \quad (4)$$

Come può essere dedotto confrontando la (1) con la (4), l'impiego di un segnale FM permette di ottenere un sensibile guadagno di demodulazione (ad esempio, 39 dB) rispetto al caso di segnali AM-VSB, a scapito di un'occupazione di banda sensibilmente maggiore (36 MHz per l'FM contro 8 MHz per l'AM-VSB). Ad esempio, un valore del C/N di 17 dB permette di ottenere un S/N di qualità studio pari a 56 dB, anche in assenza di preenfasi.

3.2 *Formati di modulazione numerica*

In linea di principio, per la trasmissione su fibra ottica di segnali video può essere usato qualunque formato di modulazione numerica, sia di tipo binario (Amplitude-Shift-Keying, ASK, Frequency-Shift-Keying, FSK, Phase-Shift-Keying, PSK), che multilivello (M-FSK, M-PSK, M-Quadrature Amplitude Modulation, M-QAM, M-Vestigial Side Band, M-VSB) [10-11]. Tuttavia, le caratteristiche dei dispositivi optoelettronici (laser, modulatori elettro-ottici, fotodiodi, amplificatori ottici) ed elettronici (modulatori, combinatori, VCO ed amplificatori) e la struttura del sistema ottico SCM condizionano in maniera sensibile la scelta del formato di modulazione. Infatti, nel caso binario la tecnica SCM può essere vantaggiosamente sfruttata mediante l'impiego di formati di modulazione angolare (tipicamente FSK e BPSK) piuttosto che di ampiezza. In caso di modulazione multilivello vengono tipicamente adottati i formati QPSK, M-QAM e M-VSB, in relazione all'esigenza di utilizzare in maniera ottimale la larghezza di banda disponibile, limitata dai circuiti elettronici di elaborazione del segnale.

3.3 *Tecniche di compressione del segnale video*

La flessibilità offerta dalla tecnica SCM è tale da permettere la realizzazione di reti d'accesso basate sulle tecnologie ottiche in cui, oltre a canali analogici, siano offerti all'utenza anche canali numerici sia per la trasmissione di segnali video, eventualmente a diversi livelli di compressione, sia per la trasmissione di dati.

Un segnale video PAL digitalizzato e non compresso richiede una velocità di trasmissione di circa 100 Mbit/s. Tecniche di co/decodifica di tipo DPCM, basate sulla riduzione della ridondanza pixel-to-pixel, permettono di operare ad una velocità di trasmissione di circa 45 Mbit/s.

Il Motion Picture Experts Group (MPEG) ha definito standard per segnali video e audio numerici altamente compressi. Gli standard di interesse per l'impiego in sistemi CATV in fibra ottica sono MPEG1 ed MPEG2.

Lo standard MPEG1 è stato sviluppato diversi anni fa ed approvato formalmente nel Novembre 1992 come standard ISO (International Organization for Standardization) per la compressione del segnale video e del relativo audio ad una velocità di trasmissione di circa 1.5 Mbit/s, mantenendo una qualità comparabile a quella della registrazione VHS: la risoluzione spaziale del segnale video all'ingresso del codificatore MPEG1 è infatti pari a 360 campioni per riga, 240 righe per quadro a 30 Hz (Source Input Format, SIF).

Lo standard MPEG1, sviluppato sostanzialmente per applicazioni di "storage & retrieval" non soddisfa le esigenze degli operatori di broadcasting che intendano fornire programmi in formato numerico con codifica in tempo reale e con standard di qualità comparabili con le attuali trasmissioni. Questi problemi sono stati affrontati e risolti mediante lo sviluppo e la definizione dello standard MPEG2.

Uno degli obiettivi dello standard MPEG2 è stato quello di fornire una scala di possibili rapporti di compressione da adottare in funzione delle specifiche applicazioni (ad esempio, "storage & retrieval" a 1.5 Mbit/s, CCIR 601 in tempo reale a 4÷8 Mbit/s, HDTV per distribuzione e contributo a 20÷60 Mbit/s), utilizzando gli stessi moduli base con diverse configurazioni per la realizzazione dei co/decodificatori. Nel contempo, lo standard MPEG2 contiene elementi relativi al trasporto (ad esempio, sincronismi, correzione errori) e alla multiplazione di diverse sorgenti audio e video (di particolare interesse per applicazioni multimediali).

Nel Marzo 1993 venne presentato il documento di lavoro MPEG2 "Main Profile, Main Level" che definiva i requisiti per il raggiungimento della qualità del segnale video in base alla raccomandazione CCIR 601. I diversi requisiti che MPEG2 deve soddisfare sono stati raccolti nel documento "Agreements on Profile/Level" redatto nel Luglio 1993 [12]. A tale riguardo si può suddividere MPEG2 in tre "profili" (Simple Profile, Main Profile e Next Profile) e ogni "profilo" in quattro "livelli" (Low Level, Main Level, High Level-1440, High Level).

In termini di definizione dell'immagine, Main Level corrisponde allo standard CCIR-601, High Level alla risoluzione HDTV e Low Level alla risoluzione SIF.

Dei tre profili, il più consolidato è Main Profile in grado di gestire immagini dal più basso livello, con qualità MPEG1, al più alto con qualità HDTV. Il Main Level di questo "profilo" è in grado di fornire una soluzione generale per la distribuzione del segnale video nei diversi ambiti di trasmissione (ad esempio, via cavo, in fibra ottica o in rame, diretta via satellite, VCR). Nella tab. 1 sono riportati alcuni dei principali parametri relativi ai "livelli" di Main Profile.

High Level (da 2 a 60 Mbit/s)	Pixel/linea Linee/quadro Quadri/s	1920 1152 60
High Level-1440 (da 2 a 60 Mbit/s)	Pixel/linea Linee/quadro Quadri/s	1440 1152 60
Main Level (da 2 a 15 Mbit/s)	Pixel/linea Linee/quadro Quadri/s	720 576 30
Low Level (da 2 a 4 Mbit/s)	Pixel/linea Linee/quadro Quadri/s	352 288 30

Tabella 1 Principali parametri relativi ai "livelli" di Main Profile di MPEG2

Per comodità di confronto, sono riassunte alcune delle tipiche corrispondenze tra velocità di trasmissione richiesta e qualità del segnale video per gli standards MPEG1 ed MPEG2 (tab. 2).

MPEG1	$\approx$ 1.5 Mbit/s $\longrightarrow$ qualità VHS
MPEG2	$\approx$ 3.1 Mbit/s $\longrightarrow$ qualità PAL/CATV $\approx$ 6.2 Mbit/s $\longrightarrow$ qualità studio (CCIR 601) $\approx$ 2 $\div$ 10 Mbit/s $\longrightarrow$ PAL (CCIR 601) $\approx$ 10 $\div$ 20 Mbit/s $\longrightarrow$ HDTV

Tabella 2 Ritmi binari tipici per varie applicazioni video con codifiche MPEG1 e MPEG2

4. Prestazioni dei sistemi di trasmissione con multiplazione a sottoportante

4.1 Sistemi analogici

Le prestazioni di un sistema SCM sono valutate a partire dal rapporto portante-disturbo (Carrier-to-Noise Ratio, C/N) che determina, per i segnali analogici, la qualità del segnale video in termini di rapporto segnale-disturbo pesato (weighted Signal-to-Noise Ratio, weighted S/N). Se si considera un sistema SCM, nel caso di modulazione diretta di una sorgente laser a semiconduttore e di rivelazione diretta del segnale ottico tramite un fotodiodo PIN o un APD, il C/N può essere espresso dalla seguente relazione

$$\mathrm{C/N} = \frac{I_c^2}{\sigma_{SN}^2 + \sigma_{TN}^2 + \sigma_{RIN}^2 + \sigma_{MPI}^2 + \sigma_{NLD}^2} \qquad (5)$$

dove I_c rappresenta la fotocorrente relativa al generico canale trasmesso, espressa da

$$I_c = \frac{mMI_s}{\sqrt{2}} = \frac{mMRP\exp(-\alpha L/10)}{\sqrt{2}} \qquad (6)$$

in cui:

- I_s rappresenta il valore medio della fotocorrente associata al segnale ottico ricevuto;
- m è la profondità di modulazione (valore di picco) relativa al canale considerato;
- M è il guadagno di moltiplicazione nel caso di un APD (M=1 per un fotodiodo PIN);
- $R=e\eta/h\nu$ è la responsività del fotodiodo, essendo η l'efficienza quantica del fotodiodo, e la carica dell'elettrone ($\approx 1.6\cdot10^{-19}$ C), h la costante di Planck ($\approx 6.6\cdot10^{-34}$ J·s), $\nu=c/\lambda$ la frequenza del segnale ottico di lunghezza d'onda λ;
- P rappresenta la potenza del segnale ottico trasmesso in fibra;
- α (dB/km) è l'attenuazione della fibra ottica;
- L (km) è la lunghezza della fibra ottica.

Con riferimento alla fig. 3, nel caso in cui si consideri un solo segnale sinusoidale che moduli d'ampiezza un laser a semiconduttore, l'indice di modulazione m può essere definito dall'espressione

$$m = \frac{P_{max} - P_0}{P_0} \qquad (7)$$

essendo P_{max} e P_0 rispettivamente il valore di picco ed il valore medio della potenza ottica emessa dal laser. Se più canali multiplati SCM modulano la sorgente laser, la stessa definizione può essere direttamente adottata per l'indice di modulazione ottico totale m_{tot} (Optical Modulation Index, OMI), nel caso in cui si operi entro l'intervallo di linearità della caratteristica potenza ottica-

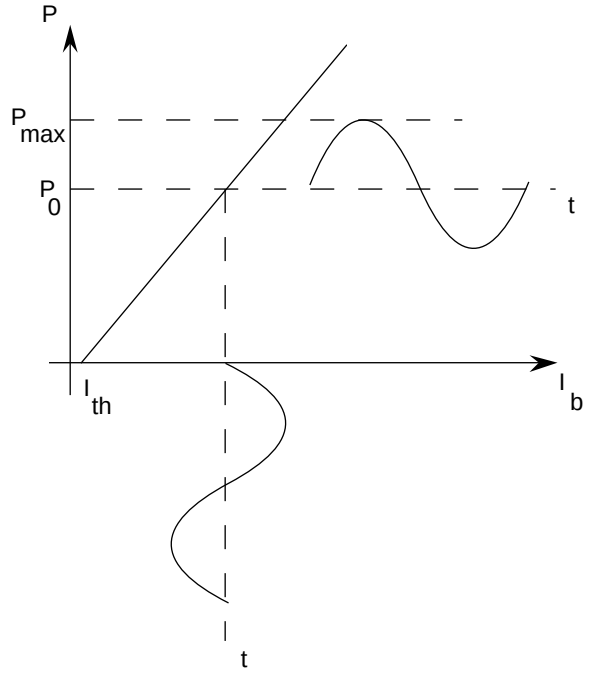


Figura 3 Modulazione diretta di un laser a semiconduttore: caso di un singolo segnale sinusoidale modulante

corrente del laser in modo tale che, essendo generalmente l'indice di modulazione lo stesso per ognuno dei canali, risulti $m_{tot}=mN$. In genere, nel caso di segnali analogici, si effettua una sovramodulazione del laser a semiconduttore (m_{tot}>100%), operando in un intervallo più ampio della regione lineare e tollerando un incremento delle distorsioni di non linearità.

Infatti, per definire la sovramodulazione si può fare riferimento alla fig. 4 in cui si considera il caso di N canali, di ampiezza A, che modulano direttamente il laser polarizzato al centro della regione lineare della caratteristica la quale potrebbe accogliere soltanto $L<N$ canali. Poiché l'indice di modulazione per canale è pari ad $m=1/L$, l'indice di modulazione ottico totale m_{tot} risulta espresso da $m_{tot}=mN=N/L>1$. In un sistema SCM le diverse sottoportanti risultano, in generale, scorrelate e quindi, per un numero di canali N sufficientemente elevato (N>10), la distribuzione delle loro ampiezze tende ad essere gaussiana cosicché risulta più opportuno far riferimento ad un indice di modulazione efficace, definito come $\mu=m\sqrt{(N/2)}$ [13, 14]. Valori di m_{tot} pari a 150÷300% (μ=15÷30%) nel caso di modulazione AM-VSB e 300÷500% (μ=30÷90%) per la modulazione FM sono tipici di realizzazioni sperimentali.

In (5) il rumore complessivo è dovuto a diversi processi, supposti indipendenti e gaussiani, il cui effetto dipende principalmente dal formato di modulazione, dalla sorgente laser e dalla struttura del ricevitore: il rumore di rivelazione (Shot-Noise) σ^2_{SN}, il rumore termico σ^2_{TN}, il rumore d'intensità del laser a

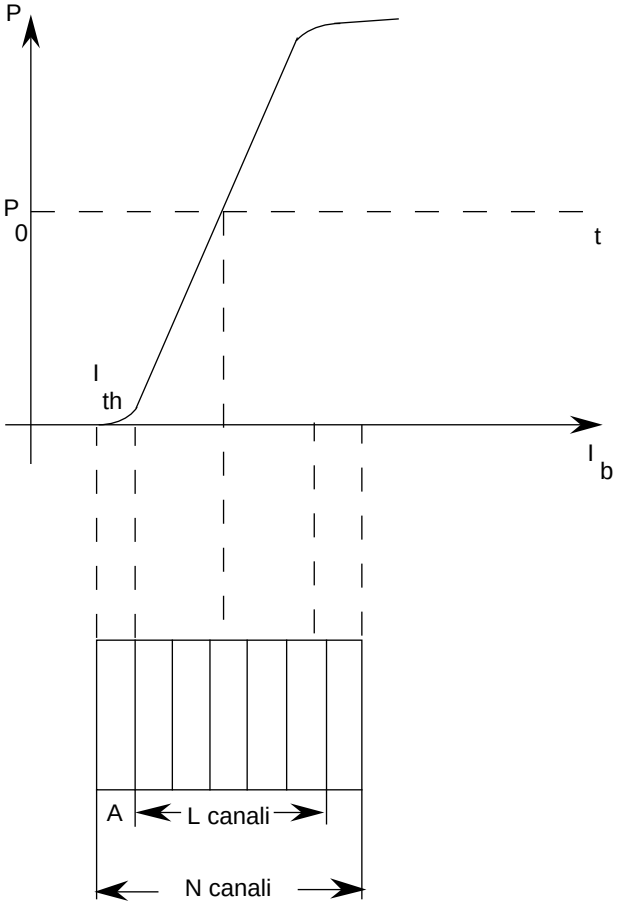


Figura 4 Modulazione diretta di un laser a semiconduttore: caso di N segnali sinusoidali modulanti di ampiezza A

semiconduttore σ^2_{RIN}, il rumore dovuto alle riflessioni multiple in fibra (Multi-Path Interference, MPI) σ^2_{MPI} ed il rumore associato alle distorsioni di nonlinearità (Nonlinear Distorsion) σ^2_{NLD}, dovute principalmente alla nonlinearità della caratteristica potenza ottica/corrente di polarizzazione del laser, fenomeno che per un numero di canali sufficientemente elevato (N>10) può essere considerato con buona approssimazione gaussiano.

4.1.1 Rumore di rivelazione

Per quanto riguarda il rumore di rivelazione, dovuto al processo statistico di conversione del segnale ottico in segnale elettrico operata dal fotodiodo, la varianza σ^2_{SN} può essere espressa dalle seguenti relazioni nel caso in cui venga impiegato un fotodiodo PIN o un APD:

$$\sigma^2_{SN} = 2eI_sB \quad \text{(PIN)} \qquad (8a)$$

$$\sigma^2_{SN} = 2e\left(I_s + I_d\right)M^2F(M)B \quad \text{(APD)} \qquad (8b)$$

in cui B rappresenta la larghezza di banda equivalente di

rumore (nel caso di modulazione AM-VSB, pari a 5 MHz per il segnale PAL), I_d la corrente di buio dell'APD, M il fattore di moltiplicazione dell'APD, $F(M) \approx M^x$ il fattore di rumore in eccesso (Excess Noise Factor) dell'APD ($x \approx 0.5$ per Si e GaAs).

4.1.2 Rumore termico

Il rumore termico è dovuto principalmente all'operazione di amplificazione del segnale elettrico generato dal fotodiodo; il valore quadratico medio della corrente associata al rumore termico è espresso dalla relazione

$$\sigma_{TN}^2 = \frac{4kTFB}{R_L} \qquad (9)$$

in cui k è la costante di Boltzmann ($\approx 1.38 \cdot 10^{-23}$ J/K), T la temperatura assoluta, F la figura di rumore dell'amplificatore ed R_L la resistenza di carico.

4.1.3 Rumore d'intensità

Il rumore d'intensità della sorgente laser è quantificato tramite il Relative Intensity Noise (RIN), definito come rapporto tra la densità spettrale delle fluttuazioni di potenza del laser $S_P(\omega)$ ed il quadrato della potenza media emessa P_0

$$\mathrm{RIN\ (dB/Hz)} = 10 \log_{10}\left(\frac{S_P(\omega)}{P_0^2}\right). \qquad (10)$$

Pertanto il valore quadratico medio della corrente di rumore associata a tale effetto risulta espresso dalla relazione

$$\sigma_{RIN}^2 = I_s^2 B 10^{RIN/10}. \qquad (11)$$

In un laser a semiconduttore il RIN ha una dipendenza dalla corrente di polarizzazione I_b del tipo $\mathrm{RIN} \propto (I_b/I_{th} - 1)^{-3}$, essendo I_{th} la corrente di soglia del laser. Con strutture laser a semiconduttore di tipo DFB (Distributed Feedback) è possibile ottenere valori del RIN dell'ordine di $-140 \div -150$ dB/Hz.

4.1.4 Riflessioni multiple

Le prestazioni del sistema risultano ulteriormente degradate a causa di riflessioni dovute a discontinuità presenti nel collegamento (connettori, giunti), tali da reiniettare una parte del segnale ottico nella cavità laser oppure a causa di riflessioni multiple tra discontinuità della fibra. Nel primo caso, l'effetto delle riflessioni sul laser può essere efficacemente limitato mediante l'impiego di isolatori ottici. L'effetto delle riflessioni multiple genera invece un rumore d'intensità di tipo interferometrico dovuto all'interferenza di più segnali ottici che incidono sul fotodiodo con un certo ritardo. La distribuzione spettrale di tale rumore dipende dallo spettro ottico della sorgente laser. Nel caso di modulazione diretta di un laser a semiconduttore lo spettro del segnale ottico dipende in maniera sensibile dall'effetto del chirping che provoca una conversione AM-FM. Infatti il chirping è legato alle fluttuazioni dinamiche della frequenza ottica di emissione, in funzione delle variazioni della corrente di modulazione del laser [15].

Se un laser a semiconduttore è modulato direttamente da un numero di canali AM-VSB sufficientemente elevato ($N > 10$), lo spettro ottico della sorgente modulata risulta praticamente gaussiano con larghezza a metà altezza pari a $B_{1/2} = \sqrt{2}\mu K I_b$, essendo K (GHz/mA) l'efficienza di modulazione di frequenza del laser e I_b la corrente di polarizzazione del laser. In questo caso, anche la distribuzione spettrale del rumore dovuto agli effetti di interferenza multicammino risulta gaussiana e tale fenomeno può essere considerato equivalente ad un rumore d'intensità relativo $\mathrm{RIN_{MPI}}$. Nell'ipotesi in cui il ritardo tra segnale diretto e riflesso sia sufficientemente elevato da potere considerare tali segnali scorrelati (condizione che, nel caso di segnali video, corrisponde alla presenza di centri di riflessione lungo la fibra distanti almeno alcuni metri), il rumore d'intensità relativo $\mathrm{RIN_{MPI}}$ risulta espresso dalla relazione [16]

$$\mathrm{RIN_{MPI}\ (dB/Hz)} = 10 \log_{10}\left(\frac{4R_1R_2}{\sqrt{2\pi}B_{1/2}}\right) \qquad (12)$$

essendo R_1 ed R_2 le riflettività in potenza di due centri di riflessione. Pertanto, il valore quadratico medio della corrente di rumore associato a tale effetto può essere espresso dalla relazione

$$\sigma_{MPI}^2 = I_s^2 B 10^{RIN_{MPI}/10}. \qquad (13)$$

Per coefficienti di riflessione R_1 ed R_2 dell'ordine di -30 dB si ottiene tipicamente, $\mathrm{RIN_{MPI}} \approx -150$ dB/Hz.

Nel caso di modulazione FM o di formati di modulazione numerica, lo spettro del segnale ottico è in genere più complesso e, se le frequenze di modulazione si estendono su alcuni GHz, l'allargamento spettrale dovuto alla modulazione si aggiunge a quello dovuto al chirping, così da distribuire la potenza ottica su una banda piuttosto ampia. Pertanto, per questi formati di modulazione l'effetto del rumore d'intensità di tipo interferometrico sul singolo canale risulta modesto rispetto al caso di modulazione AM-VSB.

Se viene invece impiegato un modulatore esterno lo spettro ottico non è alterato dal chirping: in questo caso l'effetto del rumore d'intensità interferometrico dipende in maniera sensibile dalle relazioni di fase tra il segnale utile ed i segnali riflessi.

4.1.5 Distorsioni di nonlinearità

Il fattore principale che determina effetti di distorsione di nonlinerità in un sistema CATV è rappresentato dalla nonlinearità della caratteristica potenza ottica-corrente di polarizzazione del laser a semiconduttore, la quale pone un limite alla frazione di potenza ottica associata ad ogni canale.

Nel caso di trasmissione multicanale, la risposta non lineare di un dispositivo come il laser a semiconduttore genera componenti di distorsione non lineare: le più importanti risultano essere quelle del second'ordine ($\omega_A+\omega_B$ e $\omega_A-\omega_B$, essendo ω_A e ω_B le frequenze angolari associate a due generici canali) ed i termini di battimento a tre onde (principalmente del tipo $\omega_A+\omega_B-\omega_C$, che cadono più frequentemente all'interno della banda dei canali). Il numero delle componenti di distorsione del terz'ordine, del tipo $2\omega_A-\omega_B$ aumenta proporzionalmente a $N(N-1)$ mentre i termini di battimento a tre onde aumentano proporzionalmente a $N(N-1)(N-2)/2$, pertanto questi ultimi risultano dominanti per un numero di canali sufficientemente elevato ($N>10$). Nel caso in cui i canali occupino una larghezza di banda entro un'ottava, devono essere considerati soltanto i termini di distorsione non lineare del terz'ordine mentre, se viene occupata una banda maggiore di un'ottava, anche i termini del second'ordine devono essere tenuti in conto. Un altro effetto di distorsione, dovuto ad una risposta non lineare di tipo cubico [3], è rappresentato dalla modulazione incrociata (cross modulation) che consiste nel trasferimento della modulazione da uno o più canali ad altri presenti nel segnale multicanale: ad esempio, può verificarsi come risultato del battimento tra la sottoportante di un canale, una sottoportante interferente e le bande laterali di quest'ultima. Nel caso di modulazione AM-VSB si considera tipicamente l'effetto delle componenti di distorsione del second'ordine (Composite Second Order, CSO) e quello dovuto ai battimenti a tre onde (Composite Triple Beat, CTB).

Se si considerano, per semplicità di analisi, due canali modulanti a frequenza angolare ω_1 ed ω_2 ed aventi lo stesso indice di modulazione m, in regime lineare la potenza ottica istantanea emessa dal laser a semiconduttore risulta espressa dalla relazione

$$P(t) = P_b\,[1 + m\cos\omega_1 t + m\cos\omega_2 t], \qquad (14)$$

in cui P_b rappresenta la potenza ottica associata al punto di lavoro del diodo laser. Considerando invece per il laser una risposta nonlineare di tipo cubico, si ottiene, per questo caso

$$\begin{aligned} P(t) = P_b\,\{&[1 + m\cos\omega_1 t + m\cos\omega_2 t]+ \\ &+C_2[m\cos\omega_1 t + m\cos\omega_2 t]^2+ \\ &+C_3[m\cos\omega_1 t + m\cos\omega_2 t]^3\}, \end{aligned} \qquad (15)$$

essendo C_2 e C_3 fattori di proporzionalità correlati ai coefficienti di distorsione armonica. Dalla precedente espressione si possono dedurre i valori dei rapporti tra i segnali di battimento del secondo e del terz'ordine e la portante, le cui frequenze angolari e potenze medie sono riportate in tab. 3 [2]:

Frequenza angolare	Potenza media	
$2\omega_1$, $2\omega_2$	$1/8(C_2m)^2$	(armoniche del second'ordine)
$\omega_1 \pm \omega_2$	$1/2(C_2m)^2$	(CSO)
$3\omega_1$, $3\omega_2$	$1/32(C_3m^2)^2$	(armoniche del terz'ordine)
$2\omega_1 \pm \omega_2$, $\omega_1 \pm 2\omega_2$	$9/32(C_3m^2)^2$	(CTB)

Tabella 3 Potenze medie associate ai segnali di battimento prodotti da una non linearità di ordine 3

Anche se la caratteristica potenza ottica-corrente di polarizzazione del laser fosse perfettamente lineare al di sopra della corrente di soglia, si avrebbe comunque un effetto di limitazione (clipping) dovuto alla soglia dato che, ovviamente, la potenza ottica emessa dal laser non può essere negativa. Infatti, se l'ampiezza del segnale modulante sovrapposto alla corrente di polarizzazione del laser è tale da portare la corrente al di sotto del valore di soglia, il segnale d'uscita risulta limitato in maniera asimmetrica, con conseguente distorsione. Sono stati proposti diversi modelli per valutare l'effetto della distorsione dovuta al clipping sulle prestazioni del sistema [13, 14, 17]. Nel caso più semplice, per un numero di canali N sufficientemente elevato ($N>10$), il processo associato al fenomeno del clipping può essere supposto gaussiano [13], con valor medio proporzionale alla corrente di polarizzazione del laser I_{bias} e varianza

$$\sigma^2_{NLD} \approx \sqrt{\frac{2}{\pi}} I^2_{bias}\mu^5 \exp\left(-\frac{1}{2\mu^2}\right), \qquad (16)$$

espressione da considerare nella (5) per la valutazione del rapporto portante-disturbo C/N.

Nella tab. 4 sono riportati valori tipici dei principali parametri richiesti a livello di utente per la trasmissione in una rete di distribuzione via cavo di segnali video PAL nel caso di modulazione AM-VSB.

4.1.6 Impiego di amplificatori ottici in fibra drogata all'erbio

L'impiego di amplificatori in fibra drogata all'erbio (Erbium Doped Fibre Amplifier, EDFA) in reti ottiche a larga banda permette di incrementare in maniera sensibile il bilancio di potenza del sistema [18-21].

Spaziatura dei canali	7-8 MHz
Larghezza di banda di rumore	5 MHz
CSO (Composite Second Order)	- 53 dBc
CTB (Composite Triple Beat)	- 53 dBc

Tabella 4 Parametri tipici relativi alla trasmissione di segnali PAL nel caso di modulazione AM-VSB, nell'ambito di una rete di distribuzione via cavo

L'elevato guadagno (≈20÷30 dB), la linearità di funzionamento entro una banda ottica di alcuni nanometri ed il limitato contributo di rumore dovuto all'emissione stimolata amplificata (Amplified Spontaneous Emission, ASE) rendono questo dispositivo particolarmente adatto all'impiego in reti ottiche di accesso sia per aumentare la distanza del collegamento sia per compensare le perdite d'inserzione dovute all'inserimento di splitter ottici, così da incrementare il numero di utenti della rete afferenti ad un singolo dispositivo. Particolarmente interessante è il caso in cui l'EDFA sia impiegato come amplificatore ottico di potenza in trasmissione (booster): indicando con P_s^{in} e G rispettivamente, la potenza ottica d'ingresso ed il guadagno dell'amplificatore, che dipende dalla potenza ottica di pompa, e considerando un fotodiodo PIN (M=1), per un segnale AM-VSB multiplato a sottoportante, il C/N può essere espresso dalla relazione

$$\mathrm{C/N} = \frac{I_c^2}{\sigma_{SN}^2 + \sigma_{TN}^2 + \sigma_{RIN}^2 + \sigma_{MPI}^2 + \sigma_{NLD}^2 + \sigma_{sp-sp}^2 + \sigma_{sig-sp}^2} \quad (17)$$

in cui $I_c = mI_s/\sqrt{2}$ essendo $I_s = RG\,P_s^{in}\,\exp(-\alpha L/10)$ e

$$\sigma_{sig-sp}^2 = 4Re\eta n_{sp}(G-1)\exp(-2\alpha L/10)GP_s^{in}B \quad (18a)$$

$$\sigma_{sp-sp}^2 = 4R^2\left[h\nu n_{sp}(G-1)\exp(-\alpha L/10)\right]^2 BB_0 \quad (18b)$$

essendo B_0 la banda ottica, eventualmente fissata da un filtro ottico posto di fronte al fotodiodo, n_{sp} il coefficiente di inversione di popolazione, legato alla figura di rumore F del dispositivo e dipendente dalla potenza ottica d'ingresso, dalla potenza di pompa e dalla lunghezza dell'amplificatore. Per valori sufficientemente elevati del guadagno G e della potenza ottica d'ingresso P_s^{in}, condizioni che rendono σ_{sig-sp}^2 predominante rispetto a σ_{sp-sp}^2 è possibile ottenere, per la figura di rumore F la seguente relazione approssimata

$$F \approx 2n_{sp}\frac{G-1}{G} \approx 2n_{sp} \quad (19)$$

Operando in regime lineare, con opportuni valori di G e P_s^{in} ed una adeguata potenza di pompa, è possibile avvicinarsi al valore minimo del coefficiente di inversione di popolazione n_{sp}, uguale ad uno, in corrispondenza del quale si ottiene il limite teorico per la figura di rumore dell'EDFA, pari a

$$F \approx 2 \quad (F = 3\ \mathrm{dB}), \quad (20)$$

limite dovuto al fenomeno dell'emissione spontanea nella fibra attiva. Valori di F pari a 4÷5 dB sono tipici per amplificatori drogati all'erbio di tipo commerciale, con pompaggio a 980 nm.

Sono ormai disponibili sul mercato numerosi modelli di amplificatori ottici all'erbio per le diverse possibili configurazioni (booster, amplificatori in linea, preamplificatori), sia con pompaggio a 980 nm che a 1480 nm, ed un'analisi più dettagliata delle caratteristiche di questo componente ottico è al di là degli scopi di questo lavoro.

4.2 *Sistemi numerici*

Le prestazioni dei sistemi basati sulla modulazione numerica sono espresse in termini di probabilità d'errore P_e (tipicamente, per sistemi di trasmissione in fibra ottica, P_e=10^{-9}). La probabilità d'errore è funzione del rapporto portante-disturbo C/N che viene normalmente calcolato considerando la larghezza di banda di Nyquist bilatera. Talvolta, invece di specificare la funzione P_e=f(C/N), le prestazioni del sistema sono espresse tramite la relazione P_e=$f(E_b/N_0)$ in cui

$E_b = CT_b = C/R_b$ = energia media associata al bit,

N_0 = densità spettrale di potenza di rumore (potenza di rumore entro una banda di 1 Hz), avendo assunto pari a C la potenza media del segnale ed indicato con T_b ed R_b, rispettivamente il tempo di bit e la velocità di trasmissione. Tenendo conto che $N_0 = N/B$, la seguente espressione mette in relazione E_b/N_0 con C/N

$$E_b/N_0 = \mathrm{C/N}\cdot B/R_b. \quad (21)$$

Il parametro E_b/N_0 è una grandezza normalizzata spesso adottata per l'analisi teorica delle prestazioni dei sistemi di trasmissione in quanto risulta indipendente dalla larghezza di banda del ricevitore. Nel caso in cui le prestazioni siano espresse in termini di P_e=f(C/N) è necessario specificare la larghezza di banda di rumore del ricevitore. Ovviamente, nel caso di trasmissione binaria, risulta E_b/N_0 = C/N se si considera, per il ricevitore, una larghezza di banda bilatera pari a quella di Nyquist.

4.2.1 Modulazione FSK

Il sistema di trasmissione ottico SCM-FSK binario presenta il vantaggio di poter essere implementato in maniera semplice dato che la modulazione è ottenuta

variando la capacità di un diodo varactor in un VCO (Voltage Controlled Oscillator). I limiti principali di questo sistema sono dovuti alla modesta stabilità in frequenza del VCO ed all'effetto passa-basso del diodo varactor che limita la velocità di trasmissione ad alcune centinaia di Mbit/s (tipicamente, < 200 Mbit/s).

Nel caso di demodulazione sincrona di segnali FSK binari, la probabilità d'errore risulta espressa dalla seguente relazione

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{2N_0}}\right), \qquad (22)$$

da cui si deduce un valore teorico E_b/N_0 = C/N = 15.6 dB per una probabilità d'errore P_e=10^{-9}. Nell'ambito delle comunicazioni ottiche è però più frequente il caso di demodulazione asincrona di segnali FSK binari per il quale vale la seguente espressione della probabilità d'errore

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right), \qquad (23)$$

che fornisce, per una probabilità d'errore P_e=10^{-9}, un valore teorico E_b/N_0 = C/N = 16 dB.

4.2.2 Modulazione PSK

Il formato di modulazione PSK, sia binario (BPSK) che multilivello (M-PSK) è invece adottato a velocità di trasmissione più elevate (fino ad alcuni Gbit/s, limitata dalla larghezza di banda a frequenza intermedia - Intermediate Frequency, IF- del mixer). Nel caso di modulazione PSK binaria con demodulazione coerente la probabilità d'errore risulta espressa dalla relazione

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right), \qquad (24)$$

in base alla quale una probabilità d'errore di 10^{-9} richiede un valore del rapporto E_b/N_0 (uguale in questo caso al valore del rapporto portante-disturbo C/N) pari a 12.6 dB. Con questo formato di modulazione è spesso adottata la tecnica di demodulazione differenziale (DPSK, Differential Phase-Shift-Keying), sostanzialmente basata su un demodulatore di fase a linea di ritardo. In questo caso la probabilità d'errore risulta espressa dalla relazione

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right), \qquad (25)$$

da cui risulta un valore E_b/N_0 = C/N = 13 dB in corrispondenza ad una probabilità d'errore di 10^{-9}.

Di solito, in sistemi ottici SCM basati sulla modulazione numerica di fase, vengono impiegati i formati multilivello QPSK o DQPSK (Differential QPSK), caratterizzati da una efficienza spettrale teorica di 2 bit/s/Hz (efficienza spettrale effettiva 1.2÷2 bit/s/Hz), a scapito di una riduzione di circa 3÷3.5 dB del C/N rispetto al PSK binario (circa 2 dB di penalità alla probabilità d'errore di 10^{-9} per il DQPSK rispetto al DPSK).

Nel caso di segnali PSK multilivello con demodulazione coerente, una espressione della probabilità d'errore di simbolo P_M approssimata ma piuttosto accurata per valori sufficientemente bassi di P_M ($P_M \leq 10^{-4}$) è data da

$$P_M \approx erfc\left(sin\left(\frac{\pi}{M}\right)\sqrt{\frac{E_b}{N_0}\log_2 M}\right), \qquad (26)$$

in cui M rappresenta il numero di livelli. Se si considera una codifica di Gray, la probabilità d'errore di bit P_e per segnali M-PSK è bene approssimata dalla relazione [10]

$$P_e \approx \frac{P_M}{\log_2 M}. \qquad (27)$$

Pertanto, nel caso particolarmente importante di modulazione QPSK (M=4), si ottiene per la probabilità d'errore di bit la seguente espressione

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right), \qquad (28)$$

identica a quella relativa al formato di modulazione BPSK. Tale uguaglianza deriva dal fatto che le prestazioni sono valutate in termini di E_b/N_0; nel caso in cui si consideri il rapporto portante-disturbo C/N, le prestazioni dei due sistemi differiscono di 3 dB (12.6 dB per il BPSK contro 15.6 dB per il QPSK, per una probabilità d'errore di 10^{-9}), condizione consistente con la definizione di E_b/N_0 in cui il rapporto B/R_b vale 1 per il formato BPSK e 0.5 per il QPSK. Rispetto al QPSK, il formato di modulazione multilivello DQPSK presenta il vantaggio di una struttura più semplice del ricevitore, basato su un demodulatore differenziale. Se si considera anche in questo caso una codifica di Gray, l'espressione della probabilità d'errore per il formato DQPSK risulta piuttosto complessa anche se riferibile a funzioni note, pertanto si rimanda a testi classici di comunicazione numerica [10].

In fig. 5 è mostrato l'andamento della probabilità d'errore di bit P_e in funzione di E_b/N_0 per i formati di modulazione BPSK e QPSK con codifica di Gray, DPSK e DQPSK: per elevati valori del rapporto segnale-disturbo quest'ultima modulazione presenta una penalità di circa 2.3 dB rispetto al QPSK.

Il limite principale dei formati multilivello QPSK e DQPSK è rappresentato dalla modesta efficienza spettrale. Sebbene la larghezza di banda messa a disposizione dalla fibra ottica singolo-modo sia praticamente illimitata (≈13 THz alla lunghezza d'onda di 1550 nm), nei sistemi SCM la larghezza di banda disponibile è sostanzialmente limitata dal ricevitore e dai circuiti di elaborazione elettronica del segnale. Pertanto risulta conveniente l'impiego di formati di modulazione multilivello con una buona efficienza spettrale al fine di incrementare, nella banda disponibile, il numero di canali video trasmessi.

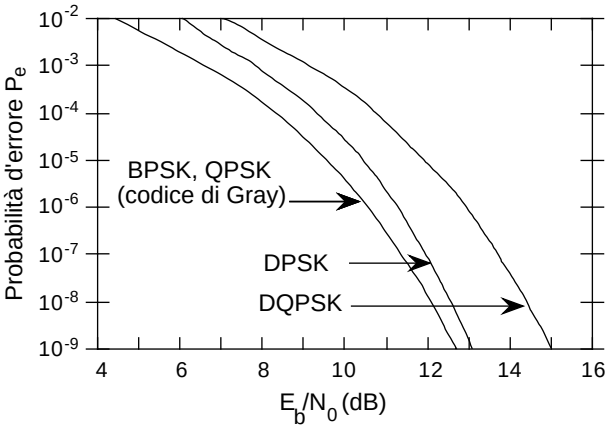


Figura 5 Probabilità d'errore di bit P_e in funzione del rapporto E_b/N_0 per i formati di modulazione BPSK e QPSK con codifica di Gray, DPSK e DQPSK

4.2.3 Modulazione M-QAM

Nell'ambito di reti ottiche di accesso, risulta particolarmente vantaggioso l'uso delle tecniche multilivello QAM (M-QAM), finora diffusamente impiegate in canali di trasmissione limitati in banda (ad esempio, ponti radio a radiofrequenza e a microonde). Oltre ad una efficienza spettrale sensibilmente superiore rispetto al QPSK, la modulazione QAM presenta l'indubbio vantaggio di essere una tecnica consolidata nell'ambito delle trasmissioni a RF, con riflessi indubbiamente positivi per quanto riguarda la disponibilità ed il costo degli apparati. Infine, come recentemente dimostrato da realizzazioni sperimentali, questo formato di modulazione risulta particolarmente adatto alla realizzazione di sistemi ottici SCM di distribuzione di tipo ibrido AM/QAM, caratterizzati da un notevole grado di flessibilità.

Per un sistema QAM ad M livelli (M-QAM), se l'efficienza spettrale $\log_2 M$ è un numero pari, la probabilità d'errore di simbolo P_M risulta espressa dalla relazione

$$P_M = 2\left(1-\frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{3\log_2 M}{2(M-1)}\frac{E_b}{N_0}}\right)\times \\ \times\left[1-\frac{1}{2}\left(1-\frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{3\log_2 M}{2(M-1)}\frac{E_b}{N_0}}\right)\right] \quad (29)$$

in cui E_b è valutato in termini di valor medio sull'insieme delle M ampiezze. Per un arbitrario valore $\log_2 M \geq 1$, la probabilità d'errore di simbolo può essere stimata con notevole precisione per valori di P_M sufficientemente bassi ($P_M \leq 10^{-4}$), tramite la relazione

$$P_M \leq 2erfc\left(\sqrt{\frac{3\log_2 M}{2(M-1)}\frac{E_b}{N_0}}\right). \quad (30)$$

Se si considera anche in questo caso una codifica di Gray, la probabilità d'errore di bit P_e può essere espressa in funzione della probabilità d'errore di simbolo P_M mediante la relazione (27).

In fig. 6 è mostrato l'andamento della probabilità d'errore di simbolo P_M in funzione del rapporto portante-disturbo C/N per segnali multilivello PSK e QAM: le prestazioni sono ottenute considerando, per ogni formato, un rumore gaussiano bianco entro una banda bilatera pari a quella di Nyquist.

4.2.4 Modulazione M-VSB

Recentemente la Zenith ha proposto, come possibile alternativa al QAM, il formato di modulazione multilivello M-VSB (M-Vestigial Sideband), caratterizzato da una efficienza superiore rispetto al QAM ma non ancora standardizzato in ambito internazionale [22].

Un confronto tra i formati QAM e M-VSB può essere fatto considerando un segnale a frequenza f modulato da una sequenza di simboli con periodo T; esprimendo tale segnale mediante le componenti in fase ed in quadratura intorno alla frequenza f si ottiene

$$s(t) = A_c\, u_c(t)\cos(2\pi ft) - A_s\, u_s(t)\sin(2\pi ft), \quad (31)$$

in cui gli impulsi in banda base $u_c(t)$ ed $u_s(t)$ sono normalizzati rispetto alle ampiezze dei simboli A_c ed A_s [10]. Nel formato QAM la modulazione è associata al valore delle ampiezze A_c ed A_s mentre gli impulsi in banda base $u_c(t)$ ed $u_s(t)$ sono identici. Ciò rende questa modulazione di tipo banda laterale doppia, portante soppressa. Nel caso di modulazione M-VSB, le ampiezze A_c ed A_s risultano uguali mentre gli impulsi in banda

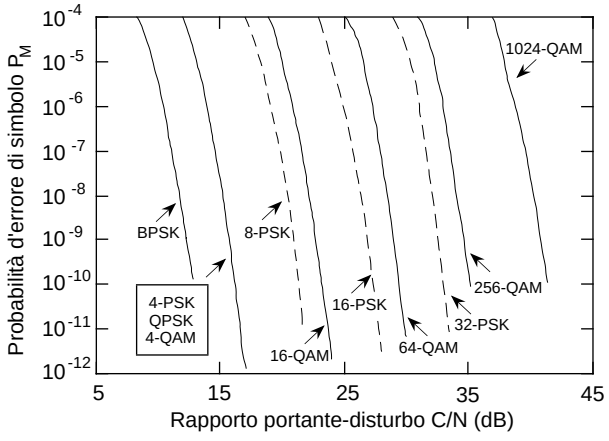


Figura 6 Probabilità d'errore di simbolo P_M in funzione del rapporto C/N per formati di modulazione multilivello PSK e QAM

base $u_c(t)$ ed $u_s(t)$ sono approssimativamente l'uno la trasformata di Hilbert dell'altro, così da permettere, anche se in maniera approssimata, la cancellazione di una delle bande laterali nello spettro del segnale [23], con conseguente migliore sfruttamento della banda disponibile. Nella tab. 5 sono riportati alcuni dei parametri principali relativi al formato di modulazione 16-VSB.

Velocità di trasmissione (larghezza di banda di 6 MHz) (larghezza di banda di 8 MHz)	 43 Mbit/s 57 Mbit/s
Efficienza spettrale teorica (Nyquist) Efficienza spettrale effettiva (assenza di codifica di correzione d'errore)	8 bit/s/Hz 7.2 bit/s/Hz
C/N (P_e=10^{-8}, assenza di codifica di correzione d'errore) C/N (P_e=10^{-8}, codifica di correzione d'errore Reed-Solomon (167-147))	34.5 dB 29 dB
Numero di canali video MPEG1(@1.5Mbit/s) Numero di canali video MPEG2 (@4 Mbit/s)	23 9

Tabella 5 Parametri principali del formato di modulazione 16-VSB in base alla proposta Zenith

Nella tab. 6 sono riassunte alcune caratteristiche dei principali formati di modulazione multilivello precedentemente esaminati. E' ovvio che la scelta del formato di modulazione è condizionata non solo da parametri tecnici ma soprattutto dal costo, dall'affidabilità e dalla disponibilità degli apparati. Se da un lato le tecniche M-VSB offrono generalmente migliori prestazioni in termini di efficienza spettrale effettiva rispetto al QAM, questa seconda tecnica di modulazione presenta il vantaggio di essere ormai standardizzata e di aver trovato diffuse applicazioni nell'ambito delle trasmissioni a RF, specialmente con sistemi 64-QAM e 256-QAM.

5. Sistemi ibridi AM/QAM

Le tecniche SCM di tipo ibrido analogico/numerico (AM/QAM) sono attualmente considerate tra le più promettenti per la distribuzione su fibra ottica sia di canali video convenzionali che di altri tipi di servizi, tipicamente di video interattivo e di telecomunicazione (telefonia, videoconferenza, ecc.). Tali sistemi presentano il vantaggio di una notevole flessibilità di gestione dei servizi e compatibilità con le attuali reti CATV, potendo essere configurati in maniera tale da ottimizzare lo sfruttamento della banda disponibile nei collegamenti esistenti tra terminazione ottica ed utenti. Tra i formati di modulazione numerica, la scelta del formato multilivello QAM per i canali numerici di un sistema ibrido è principalmente motivata dalla buona

Formato di modulazione	Efficienza spettrale teorica (bit/s/Hz)	Efficienza spettrale effettiva (bit/s/Hz)	Velocità di trasmissione per un canale video (Mbit/s)	Rapporto C/N (P_e=10^{-8}) (dB)	Rapporto C/N (P_e=10^{-8}, codice RS 167,147) (dB)
QPSK (4-QAM)	2	1.2-2	9.6-16	15	-
8-PSK	3	2.5-3	20-24	20.5	-
16-QAM	4	2.5-3.5	20-28	22.5	17
4-VSB	4	-	-	22.5	17
64-QAM	6	4.5-5	36-40	28.5	23
8-VSB	6	-	-	28.5	23
128-QAM	7	4.5-5.5	36-44	31.5	-
256-QAM	8	5-7	40-56	34.5	29
16-VSB	8	7.2	57	34.5	29

Tabella 6 Efficienza spettrale e rapporto C/N per i principali formati di modulazione multilivello per sistemi ottici CATV

efficienza spettrale e dalla disponibilità, offerta dalle tecniche di trasmissione a RF, di apparati affidabili, ormai standardizzati a livello internazionale.

In fig. 7 è mostrato un esempio di canalizzazione, proposto da Telecom Italia, nel caso d'impiego della tecnica di multiplazione SCM: esso comprende canali telefonici upstream e downstream, canali di controllo e segnalazione, canali analogici e numerici sia per servizi di tipo diffusivo (broadcasting) che per servizi di tipo interattivo (multimedia).

Il problema principale dei sistemi di trasmissione ottica SCM di tipo ibrido AM/QAM è rappresentato dalla degradazione della probabilità d'errore nei canali QAM dovuta all'effetto di "limitazione" (clipping) del diodo laser a semiconduttore. Infatti, se da un lato tale fenomeno causa direttamente effetti di distorsione sui canali AM, le caratteristiche di "rumore impulsivo" non-gaussiano del clipping condizionano in modo sensibile le prestazioni dei canali QAM fino a causare, per valori eccessivamente elevati dell'indice di modulazione dei canali QAM, un plateau o addirittura un andamento crescente della probabilità d'errore.

La probabilità d'errore dei canali QAM può essere determinata assumendo l'effetto del clipping come un rumore impulsivo non-gaussiano [24]. Con questa ipotesi, la probabilità d'errore di bit P_e risulta espressa dalla relazione

$$P_e \approx 2\left(1-\frac{1}{\sqrt{M}}\right)\exp(-A)\cdot\sum_{j=0}^{\infty}\frac{A^j}{j!}erfc\left[\frac{\sqrt{C/N_G}}{\sqrt{2}\left(\sqrt{M}-1\right)\sigma_j}\right] \quad (32)$$

in cui:

- C/N_G rappresenta il rapporto portante-disturbo riferito alla sola componente di rumore gaussiano;
- $\sigma_j^2=(j/A+G)/(1+G)$, essendo G uguale al rapporto tra la varianza del rumore gaussiano σ_G^2 e quella relativa al rumore impulsivo σ_I^2;
- A rappresenta l'indice di rumore impulsivo.

L'indice A è definito come il prodotto del numero medio di impulsi ricevuti nell'unità di tempo per la durata dell'impulso, pertanto risulta equivalente alla probabilità del fenomeno di clipping per unità di tempo [25].

In fig. 8 è mostrato l'andamento della probabilità d'errore per un canale 64-QAM in funzione dell'indice di modulazione dei canali QAM, al variare del numero di canali 64-QAM, per un sistema ibrido AM/64-QAM costituito da 70 canali AM (con indice di modulazione pari al 4.2%, tale da garantire, in assenza di canali QAM, un C/N=53 dB [26]). La probabilità d'errore in funzione dell'indice di modulazione dei canali QAM presenta un valore minimo il quale, fissati i parametri del sistema e la potenza ottica ricevuta (pari ad 1 mW nel caso in esame), aumenta al crescere del numero dei canali QAM. Il valore ottimo dell'indice di modulazione dei canali QAM diminuisce invece al crescere dei canali. Tale andamento deve essere opportunamente considerato nel progetto di sistemi ottici SCM di tipo ibrido.

Infatti, fissata la probabilità d'errore di riferimento (ad esempio, $P_e=10^{-9}$), una volta determinati il numero dei canali QAM ed il valore ottimo dell'indice di modulazione ottico per tali canali, è necessario valutare il corrispondente livello di potenza e confrontarlo con quello dei canali analogici AM. Nel caso in cui il livello di potenza dei canali QAM sia confrontabile a quello dei canali analogici, si determinerebbe una condizione di

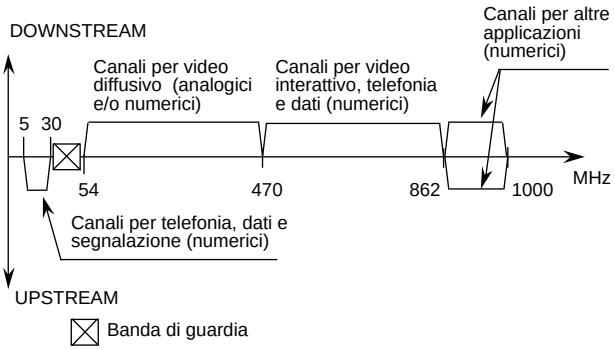


Figura 7 Esempio di possibile canalizzazione di una banda di ampiezza pari ad 1 GHz nel caso d'impiego della tecnica di multiplazione SCM di tipo ibrido AM/QAM. La banda relativa ai canali per "altre applicazioni" (862-1000 MHz) non può essere contemporaneamente utilizzata nei due versi DOWNSTREAM ed UPSTREAM

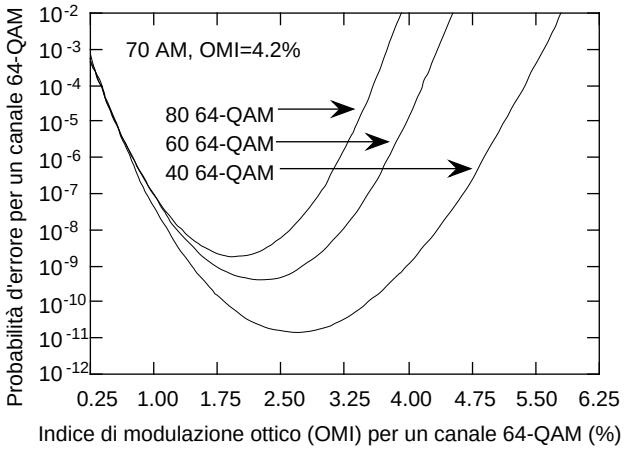


Figura 8 Probabilità d'errore P_e per un canale 64-QAM, in un sistema ibrido AM/64-QAM, in funzione dell'indice di modulazione ottico del segnale 64-QAM al variare del numero di canali 64-QAM (70 canali AM con indice di modulazione ottico pari al 4.2%; potenza ottica ricevuta uguale ad 1 mW; larghezza di banda del canale video pari a 6 MHz; R=0.9 A/W; RIN=–155 dB/Hz) [26]

interferenza sui canali AM tale da pregiudicarne la qualità. E' possibile ovviare a questo problema o diminuendo direttamente il numero dei canali QAM oppure riducendo l'indice di modulazione ottico dei canali AM, così da variare indirettamente sia il valore ottimo dell'indice di modulazione che il numero dei canali QAM. Nel primo caso è necessario verificare se la condizione di valore ottimo, o perlomeno subottimo, dell'indice di modulazione ottico dei canali QAM è mantenuta. Nel secondo caso occorre valutare se la corrispondente riduzione del rapporto portante-disturbo C/N dei canali AM può essere tollerata ai fini di una accettabile qualità del segnale [26].

6. **Realizzazioni sperimentali**

Negli ultimi anni sono stati effettuati numerosi esperimenti di trasmissione su fibra ottica di segnali video basati sulla tecnica SCM [27÷29], sia utilizzando formati di modulazione analogica (AM, FM), che numerica (FSK, BPSK, QPSK, QAM). In questa sede ci si limiterà all'analisi dei risultati relativi ad esperimenti di trasmissione ottica SCM particolarmente significativi per la realizzazione di sistemi ibridi analogici/numerici.

È stato sperimentato un sistema di trasmissione ottica SCM con 94 sottoportanti nell'intervallo 139÷697 MHz (≈6 MHz per canale), operante alla lunghezza d'onda di 1560 nm su 20 km di fibra ottica singolo-modo [30, 31]. Impiegando un laser a semiconduttore di tipo DFB come trasmettitore ed in linea un amplificatore in fibra drogata all'erbio (G=20 dB per una potenza ottica d'ingresso di –13 dBm), per un indice di modulazione per canale pari al 6.7%, è stato ottenuto un C/N≈20 dB (corrispondente ad una probabilità d'errore di 10^{-5} per il sistema multilivello 16-QAM), per una potenza del segnale ottico ricevuto pari a –23 dBm. In queste condizioni, modulando tutte le sottoportanti con il formato multilivello 16-QAM si potrebbe ottenere la trasmissione di circa 400 canali video MPEG2 (ad una velocità di cifra di 4÷5 Mbit/s), ovvero di circa 1300 canali video MPEG1 (ad 1.2÷1.5 Mbit/s), con un bilancio del collegamento ottico di 23÷26 dB, in grado di servire fino a 128 nodi ottici. Adottando invece il formato multilivello 64-QAM, circa 570 canali video MPEG2 (ovvero circa 1900 canali MPEG1) potrebbero essere trasmessi con un valore C/N≈27 dB (corrispondente, anche in questo caso, ad una probabilità d'errore di 10^{-5}) ed un bilancio del collegamento ottico di 17÷20 dB.

Recentemente sono stati presentati i risultati di due esperimenti di sistemi SCM di tipo ibrido analogico/numerico. Nel primo [32] è stata effettuata la trasmissione di 60 canali AM-VSB (nella banda 55.25÷439.25 MHz) e di 10 canali 64-QAM (90 Mbit/s/canale, con una larghezza di banda di 22.5 MHz, nella banda 525÷750 MHz), mediante un solo laser a semiconduttore di tipo DFB. Per un indice di modulazione dei canali AM-VSB pari al 3.7%/canale (CNR≈52 dB, CSO, CTB≈–65 dBc), l'effetto dovuto al clipping è risultato trascurabile e tale da ottenere una probabilità d'errore di circa 10^{-9} per i canali QAM in corrispondenza ad un indice di modulazione pari a circa l'1%/canale.

Nel secondo esperimento [33], 60 canali AM/VSB (indice di modulazione 4%/canale, CNR≥51 dB, CSO≤–60 dBc, CTB≤–65 dBc, con una larghezza di banda di 6 MHz nella banda 55.25÷415.25 MHz) e 90 canali 16-QAM (indice di modulazione 0.7%/canale, con una larghezza di banda di 6 MHz nella banda 433.25÷967.25 MHz, $P_e \leq 10^{-6}$) sono distribuiti a 128 nodi ottici mediante l'impiego di un laser MQW-DFB, operante alla lunghezza d'onda di 1561 nm, con elevate caratteristiche di linearità, e di due amplificatori in fibra drogata all'erbio, uno dei quali usato come booster, l'altro come amplificatore di linea.

7. **Conclusioni**

Negli ultimi anni l'impiego della fibra ottica nella rete di accesso sta ricevendo un forte impulso, determinato soprattutto dalla prospettiva di fornire servizi a larga banda alla clientela residenziale.

Le direttive comunitarie stimolano al riguardo la competizione tra le aziende del settore e la convergenza verso forniture "globali", che comprendano sia i servizi di telecomunicazione convenzionali che i servizi a larga banda, in particolare di tipo video. Ciò determina una vivace attività da parte degli operatori del settore che si sta concretizzando in sperimentazioni, anche di notevole ampiezza, ed in ambiziosi programmi per la realizzazione di nuove infrastrutture di rete.

Con riferimento a tale scenario, questo documento ha trattato alcune importanti tematiche tecniche relative al trasporto dei segnali per la fornitura di servizi nella rete di accesso. In particolare, è stata considerata la tecnologia ibrida fibra ottica/cavo coassiale che risulta assai promettente per quanto riguarda flessibilità d'impiego ed economicità di installazione e, allo stesso tempo, è in grado di fornire servizi anche eterogenei (diffusivi ed interattivi, analogici e digitali), almeno fino a quando le necessità di banda (dal fornitore di servizi al cliente e viceversa) non risultino tali da giustificare l'impiego di ulteriori risorse.

Con riferimento a tale architettura di rete, sono stati messi in risalto i principali problemi relativi al trasporto sulla sezione ottica della rete di accesso, considerando la tecnica di multiplazione SCM e la modulazione dei segnali sia in formato analogico che digitale, valutandone, in prima approssimazione, le potenzialità.

Nel lavoro sono stati considerati anche gli aspetti connessi all'introduzione di amplificatori ottici nella catena di distribuzione. L'analisi mantiene la sua validità anche nel caso di architetture di rete differenti, purché vengano adottati gli stessi formati di modulazione.

Ulteriori studi in corso sono rivolti alla definizione di un modello matematico in grado di fornire indicazioni relative all'ottimizzazione dei parametri del sistema di trasporto ottico in rete di accesso, in modo tale da costituire un supporto teorico sia per il confronto e la interpretazione dei risultati delle sperimentazioni in campo che per la valutazione delle prestazioni di sistemi disponibili commercialmente.

Bibliografia

[1] Special Issue *Optical Fiber Video Delivery Systems of the Future*. «IEEE-LCS, The Magazine of Lightwave Communication Systems», Vol.1, n.1, February 1990.

[2] Chiddix, J.A.; Laor, H.; Pangrac, D.M.; Williamson, L.D.; Wolfe, R.W.: *AM Video on Fiber in CATV Systems: Need and Implementation*. «IEEE-Journal on Select. Area in Comm.», Vol.8, n.7, September 1990, pp. 1229-1239.

[3] Way, W.I.: *Subcarrier Multiplexed Lightwave System Design Considerations For Subscriber Loop Applications.* «IEEE-Journal of Lightwave Technol.», Vol.7, n.11, November 1989, pp. 1806-1818.

[4] Darcie, T.E.; Iannone, P.P.; Kasper, B.L.; Talman, J.R.; Burrus, C.A.; Baker, T.A.: *Wide-Band Lightwave Distribution System Using Subcarrier Multiplexing*. «IEEE-Journal of Lightwave Technol.», Vol.7, n.6, June 1989, pp. 997-1004.

[5] Olshansky, R.; Lanzisera, V.A.; Hill, P.M.: *Subcarrier Multiplexed Lightwave Systems for Broad-Band Distribution.* «IEEE-Journal of Lightwave Technol.», Vol.7, n.9, September 1989, pp. 1329-1341.

[6] Darcie, T.E.; Bodeep, G.E.: *Lightwave Subcarrier CATV Transmission Systems*. «IEEE-Transactions on Microwave Theory and Techniques», Vol.38, n.5, May 1990, pp. 524-533.

[7] Darcie, T.E.: *Subcarrier Multiplexing for Lightwave Networks and Video Distribution Systems*. «IEEE-Journal on Select. Area in Comm.», Vol.8, n.7, September 1990, pp. 1240-1248.

[8] Bingham, J.A.C.: *Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come*. «IEEE Communication Magazine», May 1990, pp. 5-14.

[9] Olshansky, R. et al.: *Subcarrier Multiplexed Broad-Band Service Network: A Flexible Platform for Broad-Band Subscriber Services.* «IEEE-Journal of Lightwave Technol.», Vol.11, n.1, January 1993, pp. 60-69.

[10] Proakis, J.G.: *Digital Communications*. II Edition, Mc-Graw-Hill, 1989.

[11] Feher, K.: *Advanced Digital Communications*. Prentice-Hall, Inc., 1987.

[12] Lockwood, L.W.: *MPEG-2: A wide ranging standard*. «International Cable», December 1993, pp. 56-61.

[13] Saleh, A.A.M.: *Fundamental Limit on Number of Channels in Subcarrier-Multiplexed Lightwave CATV System.* «Electronics Letters», Vol.25, n.12, 1989, pp. 776-777.

[14] Alameh, K.; Minasian, R.A.: *Ultimate Limits of Subcarrier-Multiplexed Lightwave Transmission*. «Electronics Letters», Vol.27, n.14, 1991, pp. 1260-1262.

[15] Koch, T.L.; Bowers, J.E.: *Nature of wavelength chirping in directly modulated semiconductor lasers.* «Electronics Letters», Vol.20, 1984, pp. 1038-1039. Linke, R.A.: *Modulation induced transient chirping in single frequency lasers*. «IEEE-Journal Quantum Electronics», Vol.21, 1985, pp. 593-597.

[16] Darcie, T.E.; Bodeep, G.E.; Saleh, A.A.M.: *Fiber-Reflection-Induced Impairments in Lightwave AM-VSB CATV Systems*. «IEEE-Journal of Lightwave Technol.», Vol.9, n.8, August 1991, pp. 991-995.

[17] Frigo, N.J.; Bodeep, G.E.: *Clipping Distortion in AM-VSB CATV Subcarrier Multiplexed Lightwave Systems.* «IEEE-Photonics Technol. Letters», Vol.4, n.7, July 1992, pp. 781-784.

[18] Yoneda, E.; Kikushima, K.; Tsuchiya, T.; Suto, K.: *Erbium-Doped Fiber Amplifier for Video Distribution Networks*. «IEEE-Journal on Select. Area in Comm.», Vol.8, n.7, September 1990, pp. 1249-1267.

[19] Way, W.I.; Choy, M.M.; Yi-Yan, A.; Andrejco, M.; Saifi, M.; Lin, C.: *Multichannel AM-VSB Television Signal Transmission Using an Erbium-Doped Optical Fiber Power Amplifier*. «IEEE-Photonics Technol. Letters», Vol.1, n.10, October 1989, pp. 343-345.

[20] Habbab, I.M.I.; Cimini, L.J.: *Optimized Performance of Erbium-Doped Fiber Amplifiers in Subcarrier Multiplexed Lightwave AM-VSB CATV Systems*. «IEEE-Journal of Lightwave Technol.», Vol.9, n.10, October 1991, pp. 1321-1329.

[21] Habbab, I.M.I.; Saleh, A.A.M.: *Fundamental Limitations in EDFA-Based Subcarrier-Multiplexed AM-VSB CATV Systems.* «IEEE-Journal of Lightwave Technol.», Vol.11, n.1, January 1993, pp. 42-48.

[22] Olshansky, R.: *Video Distribution and the Evolution of Broadband Networks*. ECOC'94, 20th European Conference on Optical Communication, Firenze (Italy), September 25-29, 1994, Vol.3, pp. 95-142.

[23] Peroni, B.: *Comunicazioni Elettriche*, Siderea, 1973.

[24] Seo, J.-S.; Cho, S.-J.; Feher, K.: *Impact of Non-Gaussian Impulsive Noise on the Performance of High-Level QAM*. «IEEE-Trans. on Electromagnetic Compatibility», Vol.31, n.2, May 1989, pp. 177-180.

[25] Kinh Pham, et al.: *Performance of 64-QAM signals in a hybrid AM-VSB/QAM optical transmission system*. OFC'94, San Jose (California), February 20-25, 1994, WH2, pp. 107-108.

[26] Qun Shi: *Performance Limits on M-QAM Transmission in Hybrid Multichannel AM/QAM Fiber Optic Systems*. «IEEE-Photonics Technol. Letters», Vol.5, n.12, December 1993, pp. 1452-1455.

[27] Special Issue on *Applications of RF and Microwave Subcarriers to Optical Fiber Transmission in Present and Future Broadband Networks*. «IEEE-Journal on Select. Area in Comm.», Vol.8, n.7, September 1990.

[28] Special Issue on *Broad-Band Lightwave Video Transmission*. «IEEE-Journal of Lightwave Technol.», Vol.11, n.1, January 1993.

[29] Betti, S.; De Marchis, G.; Iannone, E.: *Coherent Optical Communications Systems*. J. Wiley & Sons, Inc., 1994.

[30] Joyce, G.R.; Olshansky, R.: *Subcarrier Transmission of Compressed Digital Video*. «IEEE-Photonics Technol. Letters», Vol.4, n.6, June 1992, pp. 665-667.

[31] Olshansky, R.; Joyce, G.R.: *Subscriber Distribution Networks Using Compressed Digital Video*. «IEEE-J. Lightwave Tech.», Vol.10, n.11, November 1992, pp. 1760-1765.

[32] Lu, X.; Bodeep, G.E.; Darcie, T.E.: *Single-Laser Broadband AM-VSB/64QAM Hybrid Cable TV System*. OFC'94, February 20-25, 1994, San Jose (California), PD25-1.

[33] Fuse, M. et al.: *128 Optical Distribution System of 150 ch AM/QAM Hybrid Signals*. ECOC'94, September 25-29, 1994, Florence (Italy), Vol.1, pp. 45-48.

Acronimi

ADSL	Asymmetric Digital Subscriber Loop
AM	Amplitude Modulation
APD	Avalanche Photo Diode
ASE	Amplified Spontaneous Emission
ASK	Amplitude Shift Keying
ATM	Asynchronous Transfer Mode
BER	Bit Error Rate
BPSK	Bipolar Phase Shift Keying
CATV	Community Antenna TV; CAble TV
CCIR	Comité Consultatif International des Radiocommunications
CSO	Composite Second Order
CTB	Composite Triple Beat
DFB	Distributed Feed-Back
DPCM	Differential Pulse Code Modulation
DPSK	Differential Phase Shift Keying
EDFA	Erbium Doped Fiber Amplifier
FM	Frequency Modulation
FSK	Frequency Shift Keying
FTTB	Fiber To The Building
FTTC	Fiber To The Curb
FTTH	Fiber To The Home
HDTV	High Definition TV
ISO	International Standard Organization
LEC	Local Exchange Carrier
MCM	Multi Carrier Modulation
MPEG	Motion Picture Expert Group
MPI	Multi Path Interference
MQW	Multiple Quantum Well
OMI	Optical Modulation Index
PAL	Phase Alternate Line
PIN	Positive Intrinsec Negative
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
RBOC	Regional Bell Operating Company
RF	Radio Frequency
RIN	Relative Intensity Noise
SCM	Sub-Carrier Multiplexing
TDM	Time Division Multiplexing
VCO	Voltage Controlled Oscillator
VSB	Vestigial Side Band
WDM	Wavelength Division Multiplexing

Problemi di sicurezza nei servizi e nei sistemi di telecomunicazione

Il termine "sicurezza", nella comune accezione di significato, evoca uno stato di certezza e di tranquillità, dovuto all'avvenuta adozione di opportune ed adeguate previdenze, capaci di scongiurare i pericoli in cui potrebbe incorrere un soggetto nell'esercizio di un suo diritto, o nell'impiego appropriato di un bene o di un servizio, e di evitargli possibili danni.

Questo concetto, unitario nella comune sensazione soggettiva, sottintende l'individuazione di un vasto insieme di problemi potenziali, "eventuali", e di soluzioni, per fronteggiarli efficacemente, quando tali problemi si dovessero concretamente presentare. E' intuibile che l'ampiezza e l'articolazione dell'insieme in questione siano correlate alla complessità del bene utilizzato e, nel caso di un servizio, a quella del sistema che ne consente l'offerta.

Ed è appunto dalla grande complessità del sistema delle telecomunicazioni civili che deriva l'elevato numero di problemi e di casi da considerare e da ipotizzare, nell'affrontare il tema della sicurezza nel suo insieme: un quadro che si delinea e si definisce *parzialmente* al concreto verificarsi di inconvenienti, e che tuttavia, per molti aspetti, può essere solo adombrato in base ad ipotesi e ad intuizioni.

Le previdenze messe a punto sulla base di dati in parte ipotetici, mentre impongono di attribuire realisticamente alla sicurezza conseguita un carattere *non assoluto*, inseriscono anche nello scenario elementi di tipo casuale: e ciò con effetti pratici dei quali non sfugge il contrasto con l'etimo della parola sicurezza.

La probabilità nella sicurezza

Nell'attribuire al concetto di sicurezza un significato relativo, si compie, di fatto, un passo di grande importanza per l'applicabilità di tale concetto alla realtà tecnica. Un passo *filosofico* verso una verità oggettiva, cui però l'uomo non è forse ancora preparato: pur avendo accolto ormai da oltre due secoli i metodi che fanno riferimento ai *processi aleatori* per descrivere molti fondamentali fenomeni fisici, stenta ad accettare l'approccio statistico nei fatti della vita quotidiana, soprattutto quando tali fatti lo interessano come soggetto individuale.

Sostenere che è *sicuro*, cioè *che non desta preoccupazione*, ciò che tuttavia presenta alcune probabilità di rischio, equivale ad addentrarsi nel paludoso terreno del "contrasto in termini": come non preoccuparsi se qualche rischio permane?

Come si sottolinea in un primo articolo di un ciclo dedicato alla sicurezza, che prende avvio in questo numero del Notiziario, nell'analisi dell'insieme degli eventi in cui si concretizza il problema da affrontare, l'attenzione degli studiosi si deve concentrare sui *casi sfavorevoli*, che sono appunto quelli da ridurre, adottando opportune misure. Appare per converso scarsamente rilevante l'aumento dei casi favorevoli: è di qualche consolazione per il malato sapere d'essere stato colpito da un morbo particolarmente raro?...

Ad un esame superficiale, l'attenzione rivolta ai casi sfavorevoli può sembrare solamente motivata dalla scelta di un criterio metodologico volto a rendere il problema

più maneggevole sotto il profilo matematico. Ma non è così.

Va infatti osservato, a questo proposito che, quanto più i casi sfavorevoli sono rari, tanto più sfuggono alle analisi statistiche, se non altro per l'esiguità del campione osservato; d'altra parte, al migliorare della sicurezza, cioè al ridursi dei casi da studiare, diviene di crescente difficoltà l'individuazione dei fattori casuali che, pur nell'evento causato da una violazione dolosa (della sicurezza) -evento *intrinsecamente determinato*- hanno *contribuito* a renderla possibile. Ma non basta: se è evidente che solo dalle indagini *sulle cause* possono emergere gli elementi per configurare le misure volte a ridurre la frequenza di eventi avversi, è anche chiaro che le misure da adottare dovranno interessare i sistemi ed i servizi che *potrebbero* essere oggetto di future violazioni.

Il costo della sicurezza

Si intuisce come, per prevenire le violazioni, si debba intervenire su tutti i sistemi e servizi dello stesso tipo di quelli che le hanno subite. Ed è anche da osservare che le misure dovranno essere tanto più sistematiche e generali quanto meno si conoscono le circostanze casuali che hanno contribuito a rendere possibili o a facilitare le azioni dolose.

Le conseguenze pratiche di queste riflessioni riguardano ovviamente il costo della sicurezza, cui è difficile contrapporre, per una razionale valutazione, il risparmio ottenuto evitando violazioni delle quali, sia il numero sia il peso economico, sono noti soltanto "a posteriori". Questa difficoltà pregiudiziale si proietta sul rapporto tra chi offre servizi di TLC e chi ne fruisce: l'utente, quando subisce un danno dovuto ad insufficiente sicurezza, è tendenzialmente portato a ritenere che chi offre il servizio abbia voluto risparmiare, a proprio esclusivo vantaggio, omettendo di utilizzare migliori e più costosi strumenti di protezione tecnicamente esistenti; la realtà è invece speculare, poiché i limiti all'adozione di previdenze più costose sono tipicamente posti dalla necessità di non penalizzare in modo generalizzato, attraverso i più alti costi di erogazione, tutti gli utilizzatori del servizio; anche quelli che, in materia di sicurezza, abbiano esigenze oggettivamente minori.

L'incertezza nella realtà tecnica

Come si è già accennato, l'etimologia del termine sicurezza richiama l'*assenza di preoccupazione* che, ovviamente, discende dalla certezza delle prestazioni e della inviolabilità del sistema di TLC di cui si fa uso. Un concetto che male si attaglia a descrivere le caratteristiche di un sistema di TLC: inadatto in generale a descrivere qualunque dispositivo progettato e realizzato all'uomo, ma particolarmente inappropriato quando tale dispositivo sia composto da numerose parti, ciascuna delle quali possiede caratteristiche che, in molti casi, possono essere definite solamente in termini statistici (MTBF, tolleranze dimensionali, ecc.). Nel sistema, infatti, tali caratteristiche si combinano secondo leggi che sono spesso così complesse da non prestarsi ad una rappresentazione con modelli matematici, anche se notevolmente "sofisticati".

L'umanità è ormai chiaramente incamminata verso l'accettazione di sistemi a sicurezza "relativa", cioè ad alta sicurezza statistica (comunque non assoluta); ma su un piano individuale e soggettivo l'uomo rifiuta istintivamente ogni richiamo ai

termini quantitativi del problema. E ciò tanto più quanto maggiore è l'oggettiva pericolosità del sistema, intesa non come frequenza degli eventi sfavorevoli ma come gravità degli effetti di tali eventi. Questo atteggiamento è probabilmente determinato dalla consapevolezza di non poter influire sull'entità del rischio che si accetta, e nell'intesa che i limiti di tale rischio siano stati fissati da esperti, che abbiano impiegato al meglio le possibilità tecnologiche disponibili.

Nel complesso la logica che guida gli esperti nel fissare i limiti del rischio accettabile non è nota agli utilizzatori che forse, in fondo, non la vogliono neppure conoscere.

Il dosaggio della sicurezza: il compito degli esperti

Nei sistemi ad alto rischio (nel senso non statistico ma "effettuale" sopra richiamato), si confrontano grandezze di elevata entità: il rischio, appunto, ed i costi da sostenere, rapidamente crescenti per ridurlo sia pure di pochi punti percentuali. Nel crescere rapidissimo di tali costi è contenuto "in nuce" un elemento che facilita le scelte degli esperti che, di fatto, si devono comunque arrestare prima di addentrarsi nell'area del "commercialmente inaccettabile"; cioè nella proposta di beni o servizi che sarebbero acquisiti solamente da una esigua minoranza di soggetti.

Sebbene non siano frequenti i casi in cui dalla insufficiente sicurezza di un sistema di TLC, nelle telecomunicazioni civili, possono derivare pericoli confrontabili con quelli conseguenti alla disfunzione, casuale o dolosa, di altri sistemi, paradossalmente la sicurezza dei servizi e dei sistemi di TLC costituisce per il progettista un problema di particolare difficoltà; non solo per la complessità, già ricordata, del sistema ma, soprattutto, per la grande varietà delle utilizzazioni che l'utente ne fa perfino nell'ambito di un singolo servizio.

A ciascuna delle utilizzazioni in questione, può infatti essere associato un diverso grado di rischio accettabile, con riferimento ai quattro fondamentali aspetti riguardanti l'informazione: la *confidenzialità*, l'*integrità*, l'*autenticità*, e la *certezza di tempestivo ed esclusivo recapito*. Per tale motivo, volendo contenere entro limiti ristretti i danni eventuali relativi agli impieghi più importanti, si vengono inevitabilmente a gravare di costi aggiuntivi gli impieghi meno critici. Di fatto il progettista deve mediare le varie esigenze, ma scontenta comunque molti degli utilizzatori, ognuno dei quali, tra l'altro, valuta di volta in volta la sicurezza che gli è necessaria in modo diverso.

Gli strumenti di dosaggio della sicurezza

Con riferimento alla individuazione delle misure adatte ai diversi servizi ed alle differenti modalità di impiego, si deve infine osservare un vincolo di natura strettamente tecnica derivante dall'impossibilità di modificare *nel continuo* la sicurezza adattandola alle varie esigenze. Il progettista dispone infatti di un numero limitato di alternative tecniche, caratterizzate per lo più da costi molto diversi, ed è frequente che debba scegliere tra una soluzione insufficiente, per alcuni degli utilizzatori, nelle prestazioni, ed una inaccettabile nei costi, per altri utenti.

p.r.

La sicurezza nei servizi di telecomunicazione

B. Degiovanni, C. Montechiarini, F. Riciniello, M. Volpe (*)

L'evoluzione dei servizi di telecomunicazione e di Information Technology nei diversi settori della società è connotata da richieste crescenti di sicurezza per garantire la privacy, proteggere la proprietà intellettuale, salvaguardarsi da frodi e manomissioni.

Questo contributo mira a fornire una descrizione delle problematiche connesse alle procedure, alle tecniche e alle valutazioni di efficacia ed usabilità della "security".

1. Cos'è la "sicurezza"?

Il termine "sicurezza" viene utilizzato in svariati contesti per indicare ciò che non presenta pericoli e dà la certezza di avvenire secondo le previsioni.

Il significato forse più immediato è però quello derivante dall'etimologia della parola (di origine latina):

sine cura = senza preoccupazione

Lo strumento "sicuro" è dunque quello che può essere utilizzato "senza preoccupazioni".

Generalmente i "pericoli" che attentano alla sicurezza si suddividono in due categorie: quelli intenzionali e quelli accidentali. Evidentemente, la protezione contro i pericoli intenzionali ingloba anche una protezione nei confronti di molti pericoli accidentali; infatti restano esclusi solo quelli che non sono "riproducibili" dall'uomo (ad esempio una tromba d'aria). Per tutti gli altri, la volontarietà rappresenta semplicemente un'aggravante; e quindi allestire delle contromisure contro l'evento intenzionale significa già proteggersi dal caso peggiore (una corda può spezzarsi per caso o, peggio ancora, essere tagliata da un attentatore).

In quest'ottica si colloca la sicurezza dei sistemi informatici e dei servizi di telecomunicazione. Un guasto accidentale o un errore di programmazione possono arrestare il servizio, ma ben altra cosa è l'intrusione volontaria di malintenzionati all'interno della rete di comunicazione. Un personal computer può bloccarsi per un malfunzionamento casuale oppure a causa di un virus informatico.

Nel seguito si parlerà di sicurezza sempre facendo riferimento al pericolo intenzionale: cioè si considererà la possibilità di attacco deliberato contro un servizio di telecomunicazione, dal più semplice (spiare una linea di comunicazione per "carpire" informazioni segrete) al più elaborato (penetrare, via rete, in un sistema informatico producendo danni irreversibili o sottraendo denaro).

2. Caratteristiche funzionali

La sicurezza di un sistema informatico o di un servizio di telecomunicazione viene tradizionalmente definita come la combinazione di:

- confidenzialità
- integrità
- disponibilità

La confidenzialità permette di gestire le informazioni in maniera da renderle leggibili solo a chi ne è autorizzato. È possibile, utilizzando opportuni algoritmi crittografici, trasformare le informazioni in modo da renderle incomprensibili a chiunque non sia in possesso della chiave di decifratura.

In questo modo, invece di riporre le informazioni riservate all'interno di casseforti o di trasmetterle per mezzo di fidati corrieri, si possono utilizzare le stesse procedure di archiviazione e trasmissione che si userebbero per dati non segreti. Naturalmente tutta la sicurezza del meccanismo viene cosi` trasferita sulla robustezza dell'algoritmo crittografico e sulla segretezza della chiave di decifratura.

Esistono due tecniche di crittografia che si distinguono per il fatto di usare chiavi private o pubbliche. Gli algoritmi a chiave privata vengono anche definiti simmetrici in quanto entrambe le parti che si scambiano messaggi provvedono a renderli sicuri codificandoli e decifrandoli con la stessa chiave segreta. Tale procedura è adatta per comunicazioni bilaterali, nelle quali lo scambio di

(*) ing. Bruno Degiovanni -CSELT- Torino; ing. Claudio Montechiarini, ing. Flavio Riciniello, ing. Michele Volpe -Telecom Italia DG- Roma

informazioni confidenziali coinvolge due soli interlocutori. Se invece si desidera mantenere un numero maggiore di relazioni sicure diventa più complesso gestire chiavi distinte per ciascuno dei canali che si vuole proteggere. Nei sistemi a chiave pubblica, detti anche asimmetrici, la codifica e la decodifica usano parametri diversi di cui almeno uno non è calcolabile a partire dall'altro (più precisamente il calcolo richiederebbe un impegno di risorse tale da risultare impraticabile). In questo modo una delle due trasformazioni può essere rivelata senza compromettere l'altra ed è quindi utilizzabile come algoritmo pubblico. Il processo di codifica è sostanzialmente simile ad una cassetta della posta nella quale tutti sono in condizione di inserire la corrispondenza che soltanto il proprietario può prelevare.

Chiunque intenda inviare un messaggio può cifrarlo con la chiave pubblica del destinatario, avendo la certezza che nessun altro possa leggerlo in quanto esclusivamente il legittimo ricevente possiede la chiave di decifrazione. In entrambi i casi, una delle maggiori criticità è rappresentata dagli aspetti di gestione: occorre infatti che le procedure di distribuzione, rinnovo e revoca delle chiavi siano curate da una autorità garante che goda della piena fiducia degli utenti. E' importante notare a questo proposito come i sistemi asimmetrici possano essere organizzati in modo tale che i dati privati rimangano sconosciuti anche a chi amministra le chiavi pubbliche, fornendo un ulteriore livello di protezione non consentito dagli algoritmi simmetrici.

L'integrità impedisce la modifica o la distruzione delle informazioni (pur non impedendone la lettura). Anch'essa si basa su opportuni algoritmi che consentono di rivelare ogni modifica (intenzionale o accidentale) che sia avvenuta durante il trasferimento delle informazioni o la loro archiviazione.

L'utilizzo di semplici backup, infatti, può risolvere agevolmente il problema della perdita totale delle informazioni, ma la subdola modifica apportata a una parte di esse rischierebbe di passare inosservata se non si prendessero opportune contromisure.

Mentre confidenzialità e integrità impediscono a chi non è autorizzato di compiere determinate azioni (rispettivamente leggere e modificare), la disponibilità vuole garantire, a chi è effettivamente autorizzato, la possibilità di compierle. Infatti viene considerata una minaccia altrettanto pericolosa il fatto di bloccare delle risorse impedendo il loro utilizzo anche a chi ne è autorizzato (ad esempio guastandole o rallentandole al punto di non essere più efficienti).

Per motivi storici, la confidenzialità è senza dubbio l'aspetto della sicurezza che è stato indagato più in profondità soprattutto per le sue applicazioni in ambito militare. L'integrità è invece più recente; ha subito trovato possibili soluzioni attingendo in parte dalle tecniche di rilevamento degli errori di trasmissione dei dati (CRC check), in parte dalle conoscenze dei crittografi. Infatti i controlli tramite CRC da soli non bastano: contrastano efficacemente le modifiche accidentali, ma falliscono di fronte a quelle intenzionali (basta ricalcolare il CRC sul messaggio modificato e sostituirlo al precedente).

La disponibilità, infine, è un aspetto della sicurezza ancora in fase di studio. Il problema è complesso e confina con problematiche di "qualità" più generali; infatti, spesso è fin troppo semplice mandare in "tilt" un sistema o soffocarlo di richieste in modo da rallentarlo enormemente, soprattutto quando il sistema fornisce un servizio e quindi viene progettato proprio per rispondere alle richieste dell'utente.

In questi e tutti gli altri casi in cui vengono attuati attacchi alla sicurezza è importante individuare chi ha provocato l'inconveniente (tramite meccanismi di audit) per poterlo perseguire, e scoraggiare cosi` gli altri potenziali sabotatori. In questo modo, però, non si può più parlare di disponibilità (il sistema effettivamente si blocca), ma piuttosto di imputabilità (accountability).

Quest'ultima ricade in problematiche più ampie di autenticazione: l'identificazione sicura di tutte le entità che interagiscono in un servizio di telecomunicazione, e in particolare l'identificazione sicura dell'utente, risulta perciò fondamentale per garantire i requisiti minimi di sicurezza del servizio. Le tecniche maggiormente utilizzate per l'autenticazione dell'utente sono essenzialmente:

- le password (sotto forma di parole d'accesso, o codici identificativi segreti, ecc.)
- le tessere (magnetiche, ottiche o a chip)
- le misure biometriche (lettura della retina, della geometria della mano, del timbro vocale, ecc.)

Queste soluzioni sono via via più sicure, ma anche più costose. Nella pratica si tende ad abbinare due tecniche insieme (ad esempio la tessera magnetica abbinata al codice segreto di accesso) cercando di raggiungere un livello di sicurezza accettabile e costi contenuti. Tuttavia bisogna sempre tener presente il vantaggio economico derivante dalla frode e confrontarlo con i costi necessari per perpetrarla: duplicare una tessera magnetica può essere economicamente conveniente se l'entità della frode consente di recuperare i costi sostenuti (come può avvenire, ad esempio, con le carte di credito bancarie). Perciò se si pensa di estendere un meccanismo di autenticazione a diversi servizi bisogna anche considerare la maggiore appetibilità della frode e, dunque, prevenirla con meccanismi più robusti.

La robustezza delle varie tecniche dipende notevolmente da come esse sono implementate. Per esempio, le password possono essere fisse o dinamiche. Quelle fisse sono caratterizzate in base alla lunghezza, al periodo di validità, alla esclusione di parole d'uso comune. Le password dinamiche vengono invece generate ad ogni accesso da dispositivi software o hardware, comunemente indicati col nome di token. I token sono a loro volta dotati di un PIN (Personal Identification Number) che ne permette l'uso esclusivamente al legittimo proprietario. I processi di autenticazione che impiegano password dinamiche

sfruttano due meccanismi alternativi: il challenge-response e la sincronizzazione. Il primo prevede che in risposta ad una richiesta di accesso venga inviata all'utente una informazione random da immettere nel token come parametro di ingresso per la generazione della password. Il risultato di questa elaborazione viene parallelamente calcolato dal sistema di autenticazione che ha così modo di confrontarlo con quello fornito dall'utente. Se i due valori coincidono l'accesso viene consentito, altrimenti no.

La tecnica di sincronizzazione richiede che il token mantenga data e ora in quanto le password vengono cambiate automaticamente a intervalli di tempo regolari (dell'ordine del minuto). Affinché gli aggiornamenti del token siano coerenti con quelli del sistema di autenticazione, è necessario che gli orologi dei due dispositivi siano allineati. A differenza del metodo challenge-response, non sono necessari input per la produzione delle password, per cui la procedura d'uso risulta semplificata. In ambedue i casi il grado di sicurezza è direttamente legato alla difficoltà di ricostruire il funzionamento degli algoritmi che devono pertanto essere mantenuti segreti. Per quanto ciò possa sembrare ovvio, non è sempre indispensabile: esistono infatti meccanismi computazionali che pur essendo noti non possono essere invertiti e si prestano quindi ad applicazioni la cui robustezza non si poggia sulla riservatezza del metodo di codifica. Il calcolo dei codici di integrità è un tipico esempio in cui trovano impiego tali funzioni.

3. **I requisiti di sicurezza**

Storicamente i servizi pubblici di telecomunicazione sono nati con l'obiettivo di mettere in contatto la più vasta collettività di utilizzatori possibile, così che chiunque fosse in grado di parlare con qualsiasi altro interlocutore, ovunque nel mondo. Proprio in considerazione del carattere di universalità e apertura, la telefonia assolve una fondamentale funzione sociale e deve pertanto essere un servizio di facile fruibilità, economicamente accessibile a costi contenuti. Insieme alla diffusione del telefono si è andato consolidando nel tempo un codice comportamentale tacitamente adottato dalla maggior parte degli utenti. E', per esempio, accettata la possibilità che si possa digitare erroneamente il numero del chiamato raggiungendo la persona sbagliata, come è altrettanto condiviso il fatto di evitare di effettuare telefonate durante le ore notturne a meno di situazioni d'emergenza. Esiste in sostanza un "gentleman's agreement" non scritto, ma rispettato in buona misura, che regolamenta l'utilizzo del servizio. Ciò non esclude che taluni possano abusare della assenza di meccanismi specifici per forzare il rispetto delle regole suggerite dal comune buon senso. Esempi del genere sono le telefonate anonime, occasionali o ripetute. Si tratta tuttavia di fatti sporadici che come tali vanno gestiti. Si può, all'occorrenza, risalire all'utente disturbatore e predisporre il blocco delle chiamate da lui originate. A parte queste situazioni occasionali, i requisiti che caratterizzano la telefonia residenziale sono essenzialmente quelli di riservatezza, con un livello di sicurezza medio, capace di prevenire eventi accidentali o tentativi condotti da non professionisti. Tutto ciò è in accordo con la filosofia di base dei servizi rivolti alla collettività, per i quali non è economico, né per il gestore, né per l'utente, realizzare sofisticate procedure di accesso che richiederebbero investimenti sproporzionati rispetto alle reali necessità.

Esistono però settori in cui la segretezza e la disponibilità rivestono estrema rilevanza e per i quali occorre potenziare le contromisure idonee a contrastare gli attacchi condotti da personale competente e organizzato. Tradizionalmente le applicazioni militari sono quelle in cui le tecniche atte a proteggere la discrezionalità delle comunicazioni hanno avuto maggiore attenzione. Analoghe esigenze sono avvertite negli ambienti governativi e in tutti quelli che curano l'ordine pubblico (Polizia di Stato, Carabinieri). Costituiscono una categoria a parte i servizi che vengono richiesti prevalentemente in casi di emergenza (Pompieri, Croce Rossa, Pronto intervento, ecc.). In tutti questi casi il requisito che si deve soddisfare è di garantire una disponibilità praticamente assoluta, in quanto qualsiasi ritardo potrebbe avere pesanti conseguenze. I semplici esempi citati evidenziano la stretta relazione che intercorre tra gli accorgimenti da predisporre per migliorare la sicurezza dei sistemi di telecomunicazione e la criticità del servizio offerto in termini di valore delle informazioni da proteggere o di perdite derivanti da eventuali attacchi intenzionali o eventi accidentali. La prima regola di chi progetta la sicurezza delle reti telefoniche o telematiche deve, pertanto, essere quella di definire una politica di sicurezza, ovvero di individuare le risorse esposte a rischio, e valutare le potenziali perdite. In base a tali considerazioni si devono poi analizzare le minacce e le relative contromisure, prestando attenzione a che queste siano commisurate ai danni derivanti dalla loro violazione.

3.1 *Servizi di telefonia*

Per condurre una analisi sistematica dei requisiti di sicurezza è opportuno introdurre alcuni criteri di classificazione. E' innanzitutto utile distinguere i servizi dati da quelli in fonia. Sebbene le nuove tecnologie numeriche tendano decisamente ad uniformare le modalità trasmissive dei due diversi tipi di informazione (si pensi all'ISDN, o ancor di più all'ATM), permangono sostanziali differenze riguardo al genere di applicazioni e alla criticità dei dati che vengono trattati nei due casi. Di conseguenza, anche le problematiche di sicurezza sono solo in parte comuni. E' tuttavia interessante notare come, laddove si individuino esigenze condivise, il fatto di avere una tecnologia omogenea permette di definire soluzioni di validità generale, con una significativa

riduzione dei costi di analisi e realizzativi. Per quanto attiene alla telefonia, i principali aspetti di interesse sono i seguenti:

- addebito non autorizzato o evasione dell'addebito;
- intercettazione di comunicazioni;
- indisponibilità del servizio.

Ciascuno dei punti precedenti ha una duplice connotazione, può infatti essere ricondotto a eventi accidentali, o piuttosto, essere il risultato di attacchi fraudolenti. Un addebito non corretto è provocato sia da errori procedurali che dalla fruizione illegale del servizio a danno di altri utenti o del gestore. Le intercettazioni sono talvolta dovute ad interferenze casuali, ma spesso vengono realizzate intenzionalmente per ottenere informazioni di interesse commerciale. Anche la disponibilità dei sistemi di TLC è minata, oltre che dai guasti e dalle calamità naturali, anche da sabotaggi mirati a danneggiare particolari utenti (intrusioni nei PBX o nei sistemi di messaggistica vocale aziendali), o intere aree d'utenza. Altre caratteristiche, quali l'autenticazione o il non ripudio del chiamante e del chiamato, l'integrità e l'autenticità delle informazioni scambiate, che pure hanno notevole importanza nella trasmissione dati, possono ritenersi meno rilevanti per la telefonia. Infatti, è il colloquio stesso che fornisce il meccanismo di verifica dell'identità di entrambi gli interlocutori. Anzi, una delle tecniche usate nelle reti dati per validare l'accesso è proprio quella di imitare ciò che avviene tra le persone, tramite un meccanismo di domanda e risposta che riproduce una modalità di interazione pseudo-umana. Secondo tale schema le due macchine che si parlano devono condividere un insieme di nozioni. Facendo riferimento alla comune base di conoscenza, vengono formulate le interrogazioni che permettono di verificare se effettivamente le due parti in causa sono chi dichiarano di essere.

Il problema dell'integrità delle informazioni scambiate, e del non ripudio delle stesse, non è significativo nel caso dei servizi in fonia, in quanto, come avviene nei rapporti umani diretti, non si attribuisce valore legale a ciò che si dice e che non viene trascritto e sottoscritto su carta. Questi aspetti sono invece fondamentali per lo scambio di documenti per mezzo del fax. Il "fax sicuro" è il servizio a valore aggiunto che fornisce questo tipo di garanzie. Come la maggior parte delle transazioni di cui si vuole certificare la validità esso prevede il coinvolgimento di una terza parte fidata (un notaio elettronico) che svolge il ruolo di mediatore nel trasferimento dei documenti e si preoccupa di assicurare l'identità del mittente e del destinatario, conservando inoltre copia del fax allo scopo di dirimere eventuali contese.

Un'ultima osservazione va fatta in riferimento alla struttura della rete telefonica. L'attenzione del gestore verso le problematiche della sicurezza deve infatti essere concentrata più sulla rete di distribuzione che su quella di transito. La casualità dell'instradamento delle chiamate, e la multiplazione di esse su uno stesso portante fisico, complica enormemente il compito di chi volesse condurre attacchi mirati. E' invece molto più agevole cercare di intromettersi al livello del doppino d'utente sul quale sono direttamente disponibili le conversazioni di uno specifico utilizzatore. Le stesse considerazioni si possono estendere ai servizi dati.

3.2 *Servizi dati*

Nelle applicazioni telematiche tutti gli elementi che concorrono alla definizione di un ambiente sicuro, e che sono stati menzionati tra le caratteristiche funzionali del paragrafo precedente, hanno eguale dignità. Due sono le ragioni per le quali è necessario curare autenticazione, confidenzialità, controllo d'accesso, non ripudio, autenticità, integrità dei dati e accountability. La prima è l'esigenza di trasferire su processi automatici la stessa affidabilità che si esige nelle procedure manuali; la seconda è il crescente valore economico delle informazioni trasmesse. Entrambe le considerazioni sono il risultato dell'evoluzione avvenuta nei sistemi di elaborazione e di comunicazione.

Originariamente l'informatica e le reti per dati sono state introdotte nelle aziende per migliorare l'efficienza dei processi produttivi. In particolare le nuove tecnologie sono state concepite e impiegate come strumenti per accelerare le procedure di elaborazione ed il trasferimento dei risultati prodotti, sostituendosi all'uomo nell'esecuzione delle operazioni di routine in cui non fosse richiesto un intervento decisionale. Con il passare del tempo questo scenario è cambiato. Oggi i sistemi software vengono usati anche come "tools" di supporto alle decisioni strategiche, per cui mantengono e distribuiscono informazioni estremamente importanti. Sono quindi giustificati sia l'impegno che si deve porre nel proteggere il patrimonio dei dati aziendali che gli sforzi sempre più organizzati da parte della criminalità informatica.

E' tuttavia corretto puntualizzare che non sempre la trasmissione dati è il veicolo per applicazioni di tipo commerciale. Internet è l'esempio più eclatante di come i servizi di posta elettronica o di accesso a basi di dati siano utilizzabili in contesti aperti, con finalità promozionali, o come forum di discussione, senza voler introdurre meccanismi di protezione, ma cercando di promuovere la condivisione del patrimonio informativo. E' a questo punto essenziale mantenere ben distinti i due ambienti. Chiunque abbia una rete privata e voglia anche collegarsi con il mondo esterno deve porre particolare attenzione affinché gli ospiti che potranno accedere alla propria azienda non abbiano visibilità di ciò che non si vuole divulgare. I "firewall" sono elaboratori con particolari caratteristiche di sicurezza che vengono impiegati come controllori bidirezionali nell'interconnessione ad Internet. Essi verificano le autorizzazioni sia di chi accede che di chi

esce dalla rete privata, attuando il controllo sistematico sul tipo di servizio richiesto (posta elettronica, trasferimento di file, collegamento a computer remoti ecc.).

3.3 *Sicurezza del sistema di TLC*

Un secondo criterio di classificazione, oltre alla distinzione tra fonia e dati, consiste nel considerare, da un lato la sicurezza dei servizi erogati agli utenti e, dall'altro, la protezione dei sistemi di telecomunicazione. Ogni elemento di rete, sia esso un apparato di commutazione, di trasmissione o di elaborazione, ha una consol che permette di configurarlo e di riceverne le segnalazioni di allarme relative a guasti o a particolari condizioni di funzionamento. Tipicamente questi sistemi di gestione sono parte degli apparati e richiedono la presenza di un operatore per gli interventi di manutenzione. Il fatto che i dispositivi di rete siano ubicati in locali presidiati, e di non facile accesso, esclude che eventuali maleintenzionati possano sostituirsi all'operatore e manomettere il software.

Nel recente passato sono stati fatti notevoli investimenti nelle piattaforme di gestione, soddisfacendo una crescente domanda di centralizzare le funzioni di supervisione e controllo. Anche gli standard hanno ormai raggiunto un livello di maturità tale da rendere possibile la progettazione e l'implementazione di architetture distribuite per competenze funzionali o aree amministrative. I maggiori contributi in questo senso sono le raccomandazioni sulla TMN (Telecommunication Management Network) e i protocolli CMIP (Common Management Information Protocol) e SNMP (Simple Network Management Protocol). Oggi, gli elementi di rete sono dotati di agenti software che assolvono il duplice compito di notificare a distanza i cambiamenti di stato e gli allarmi, e di attuare sulle risorse gestite i comandi di configurazione impartiti da una consol remota.

I vantaggi che ne derivano dal punto di vista della riduzione dei costi di esercizio sono evidenti, è però altrettanto chiaro il rischio che si corre nel momento in cui l'accesso agli apparati non venga adeguatamente protetto. Gli accorgimenti adottati dagli operatori di rete sono diversi. L'uso di circuiti dedicati per il collegamento delle porte di gestione rientra tra i meccanismi di protezione fisica dell'accesso ed è una soluzione di gran lunga più sicura rispetto alle linee commutate. Ci sono però casi in cui i componenti che costituiscono la rete trasmissiva sono a loro volta dei computer con un indirizzo che li rende raggiungibili e quindi esposti a rischio. I router sono un tipico esempio di questo tipo, in quanto accettano il comando di login e permettono agli utenti autorizzati di eseguire le operazioni previste dal loro sistema operativo. In casi del genere, non essendo possibile predisporre delle contromisure di tipo fisico, è necessario avvalersi di meccanismi logici. Come osservato precedentemente, le password non possono considerarsi sufficienti ad assicurare un adeguato livello di sicurezza ed è quindi opportuno adottare tecniche più affidabili come le "smart card" (tessere magnetiche, ottiche o a chip).

Se da un lato appaiono preoccupanti le conseguenze di una intrusione sui singoli apparati, ben più gravi sono le minacce potenzialmente attuabili da chi riuscisse ad assumere il controllo dei sistemi di gestione. Le azioni criminose che un intruso potrebbe compiere vanno dal sabotaggio al furto intellettuale o monetario. Innanzitutto verrebbe ad essere compromessa la confidenzialità di tutti i dati amministrativi degli utenti. I dati di consumo e addebito sarebbero esposti al rischio di modifiche, con evidenti perdite economiche. Nella peggiore delle ipotesi, intere aree di servizio potrebbero essere mandate in tilt qualora la postazione di controllo avesse la supervisione di una molteplicità di risorse. E' questa una situazione che richiede misure preventive particolarmente efficaci e severe. Ogni politica di sicurezza deve in questi casi prevedere che l'accesso ai locali dove risiedono i sistemi di gestione sia rigidamente controllato e che siano programmati interventi di "disaster recovery". I centri di esercizio sono infatti duplicati e mantenuti allineati in modo che l'indisponibilità di uno dei due possa essere sopperita dall'altro.

4. **Valutazione della sicurezza**

Una peculiarità della sicurezza è che, a differenza di altre caratteristiche di un prodotto, come ad esempio la velocità o la facilità d'uso, non è direttamente visibile né misurabile e quindi l'utente non ha modo di giudicare la bontà del servizio; di conseguenza non può far altro che fidarsi di ciò che gli viene detto.

E' da sottolineare, a questo proposito, che non tutti i prodotti che vantano sicurezza sono poi davvero sicuri: esistono in commercio, ad esempio, pacchetti software per lo scambio di posta elettronica che vantano funzionalità di sicurezza e si basano su algoritmi crittografici facilmente forzabili.

Come può l'utente (ma anche il fornitore del servizio che voglia acquistare prodotti sicuri per il proprio servizio) verificare la bontà delle caratteristiche di sicurezza vantate?

Uno strumento di valutazione oggettiva è oggi disponibile: sono infatti attivi, in diverse nazioni europee, i laboratori di valutazione della sicurezza, che sulla base di criteri comuni a livello europeo (Information Technology Security Evaluation Criteria, abbreviato ITSEC) forniscono un giudizio oggettivo sulla sicurezza di prodotti o sistemi informatici che si esprime in un livello da E0 a E6 (E0 significa nessuna sicurezza, E6 massima sicurezza). Sottoporre un prodotto a valutazione in un laboratorio costa molto (si tratta di collaudi molto complessi), ma fornisce al produttore un marchio di qualità specifico per

la sicurezza riconosciuto in ambito internazionale, e all'utente un mezzo per distinguere la vera sicurezza dalle trovate pubblicitarie.

Un altro requisito fondamentale di cui tenere conto e la semplicità d'uso. Richiedere una password di 30 caratteri può rappresentare un'ottima sicurezza, ma verrà difficilmente accettato dall'utente. In questo caso, la probabilità che l'utente trascriva su un foglietto la parola d'ordine, non potendola ricordare, salgono vertiginosamente, rendendo l'effettiva sicurezza del meccanismo molto minore di una password numerica di sole 6 cifre, ma effettivamente ricordata a memoria.

Lo stesso si può dire per quel che riguarda l'utilizzo di particolari terminali o tastierini numerici: gli utenti, soprattutto i non giovanissimi, sono restii ad accettare tutto ciò che abbia più di 10 tasti e 3 "lucette"; trovano talvolta difficile riconoscere il significato dei led accesi o spenti e compiono spesso errori premendo i tasti. Un servizio, che dovesse disabilitare l'utente che compie due errori consecutivi nella password, è efficace per contrastare i tentativi di accesso fraudolenti, ma vedrebbe una schiera di legittimi utenti infuriati perché il servizio ha punito così severamente un loro errore.

5. **Aspetti legali**

Infine un cenno agli aspetti legali: tema estremamente delicato e importante. Soprattutto nel caso di servizi pan-europei, l'introduzione di funzionalità di sicurezza può avere interazioni di tipo legale molto complesse, sia a favore, sia contro.

Recentemente il governo italiano ha accolto le direttive CEE ("Green Paper on the Security of Information Systems", European Commission (DG XIII), Aprile 1994) con un disegno di legge contro i crimini informatici: sono previste pene per chi viola una banca dati, per chi diffonde virus, per chi aggira password e chiavi d'accesso o compie azioni volte a scoprirle. Leggi già precedentemente in vigore proteggono la privacy dell'individuo, intesa qui come il diritto di controllo sulla divulgabilità o meno di informazioni personali.

D'altro canto, le forze dell'ordine vogliono, in caso di particolare necessità, poter accedere a tali informazioni (si pensi alle intercettazioni telefoniche). L'utilizzo di robusti algoritmi crittografici rende questa materia estremamente difficile: scelte come quella francese, di vietare l'uso indiscriminato della crittografia rappresentano un drastico tentativo di soluzione.

6. **Evoluzione della sicurezza nei servizi d'utente**

Verranno ora analizzate alcune "famiglie" di servizi e applicazioni, sottolineando le esigenze di sicurezza di ciascuna di esse e la possibilità di una soluzione futura.

6.1 *Applicazioni telematiche*

L'introduzione di funzionalità di sicurezza all'interno delle applicazioni telematiche non è avvenuta in maniera uniforme. In particolare alcune applicazioni OSI presentano caratteristiche evolute di sicurezza, altre applicazioni (ad esempio nel mondo Internet) ne sono del tutto sprovviste, altre ancora sembrano considerare seriamente il problema, ma poi lo rimandano a future versioni dello standard.

Il gruppo di documenti standard OSI più avanzato nella specificazione di funzioni di sicurezza è costituito dall'MHS (CCITT X.400), dalla Directory (CCITT X.500) e dall'EDI (X.435).

Il Directory si propone come supporto per le altre applicazioni in veste di depositario delle chiavi di accesso. Rendere facilmente disponibili le chiavi pubbliche di ogni utente è il primo passo per una semplice implementazione di integrità, e di confidenzialità, e per una efficace autenticazione sia dell'utente nei confronti del servizio, sia dell'utente nei confronti di altri utenti.

Sulla base di questa struttura, MHS e EDI descrivono un ampio gruppo di funzioni di sicurezza che rappresentano a tutt'oggi lo stato dell'arte. EDI si differenzia dalla messaggistica elettronica tradizionale, orientata allo scambio di messaggi interpersonali, per il contenuto del messaggio che è di tipo commerciale, e orientato alla comunicazione tra applicazioni, e quindi bisognoso di maggiori protezioni; ma, d'altra parte, già l'MHS normale definisce una vasta serie di funzioni sicure che spaziano dalla possibilità di inviare un messaggio confidenziale, alla protezione di integrità, fino alla firma elettronica.

Per quel che riguarda le altre applicazioni OSI, la sicurezza è presente solo in misura minore.

La rete Internet, che costituisce oggi lo strumento telematico di gran lunga a maggior diffusione, soffre di notevoli carenze dal punto di vista della sicurezza. Sulle pagine di tutti i quotidiani e riviste (anche quelle non specializzate) appaiono sempre più frequentemente le "imprese" dei cosiddetti "hacker", cioè i "pirati" informatici, che, sfruttando le vulnerabilità di protocolli e sistemi operativi nati molti anni fa, hanno ormai automatizzato le loro procedure di attacco.

Il problema è più serio di quanto si possa immaginare: se un tempo per proteggere il proprio sistema era sufficiente conoscere un certo numero di norme generali di pubblico dominio, ora bisogna possedere competenze sempre più elevate per tener testa ad una classe di hacker sempre più esperta e sofisticata.

I casi più eclatanti di violazione di sistemi informatici sono segnalati su quotidiani e riviste, ma a questi si devono aggiungere molti altri episodi che riguardano l'attacco da parte di hacker che utilizzano il sistema violato come "ponte" per potersi collegare con altri sistemi informativi, in particolare all'estero, senza pagare la relativa bolletta, o che utilizzano furtivamente risorse informatiche, ad esempio gli hard disk, per scambiarsi programmi pirata.

Grande attenzione va inoltre rivolta ad una nuova tipologia di aggressione, il cosiddetto "phracking" (dalla contrazione dei termini phone-hacking): i pirati stanno infatti spostando le loro attenzioni dai sistemi informatici a quelli di telecomunicazione.

6.2 *Applicazioni di telemedicina*

I servizi di telemedicina rappresentano un'evoluzione molto particolare dei servizi di telecomunicazione; alcuni esempi sono:

- la gestione e l'archiviazione dei dati sui pazienti;
- il trasferimento di informazioni mediche (radiografie, analisi);
- il controllo remoto del paziente (elettrocardiogramma, teledialisi).

In questo settore i requisiti di sicurezza sono importantissimi. Se si pensa al primo esempio, si vede che due aspetti devono concorrere a mantenere sicuro il sistema: da un lato l'archivio contenente i dati sulla salute dei pazienti deve essere accessibile solo al personale autorizzato, dall'altro lato la trasmissione di ogni informazione dovrà avvenire in modalità confidenziale e l'identificazione dell'utente dovrà essere precisa, proprio per consentire un corretto controllo dell'accesso alla base dati.

Per questo tipo di applicazioni è anche fondamentale l'integrità dei dati: qualcuno potrebbe avere interesse a modificare delle informazioni o a mascherarne altre e, in ogni caso, anche l'accidentale modifica di qualche dato potrebbe avere conseguenze pericolose.

Assume, infine, un significato del tutto particolare un aspetto della sicurezza come la "disponibilità". Mentre per un servizio di telecomunicazione tradizionale si può accettare la temporanea indisponibilità dell'applicazione (a seguito di un blocco del sistema), nel caso della telemedicina l'arresto di una funzione del servizio può comportare serie conseguenze sul paziente. Si pensi, ad esempio, ai servizi che comportano un intervento attivo a distanza sulla persona, o anche quelli che consistono nella semplice consultazione di dati medici, ma in caso di bisogno urgente di quella informazione.

6.3 *Applicazioni multimediali*

Le applicazioni su cui si concentra attualmente lo studio in ambito internazionale sono, ad esempio, il reperimento di informazioni di tipo multimediale, la distribuzione di programmi televisivi a pagamento, la consultazione interattiva di banche dati contenenti sia documenti sia immagini, e cosi` via.

In tutti questi servizi gli aspetti di sicurezza sono importanti, sia per poter identificare gli utenti autorizzati all'accesso al servizio (o più in particolare, all'accesso alle singole informazioni da reperire), sia per impedire che altri si impossessino furtivamente delle informazioni mentre queste vengono trasferite all'utente finale.

Quindi entrambi gli aspetti di autenticazione dell'utente nei confronti del servizio, e di confidenzialità sono presi in considerazione per proteggere queste applicazioni.

Un aspetto nuovo e molto interessante legato a questo tipo di applicazioni è costituito dal problema del "copyright". Le informazioni che arrivano all'utente in formato digitale, saranno per lo più opere coperte da diritti d'autore: si pensi a un film trasmesso su Pay-TV, o ad un libro reperito da una banca dati. Come impedire che l'utente memorizzi l'informazione per poi produrne successivamente copie da distribuire a parenti e amici o addirittura da rivendere sottobanco? Non essendoci alcun deperimento nella copia di un'informazione digitale, bisogna impedire in qualche altro modo la duplicazione illegale e il problema è tutt'altro che di facile soluzione.

In ambito europeo si è costituito un consorzio denominato CITED (Copyright in Transmitted Electronic Documents) che vede la partecipazione di Francia, Germania, Olanda e Spagna, e che si sta occupando del problema in un apposito progetto ESPRIT II ("Electronic Copyright"). Un servizio dimostrativo è già stato approntato: è costituito da un sistema di reperimento di documenti in cui l'utente autorizzato, identificato tramite carte a chip, può ottenere copie dei documenti desiderati, ma sotto il controllo del sistema. Diverse tariffe sono previste a seconda che l'informazione sia solo visualizzata, stampata su carta o realmente duplicata. Una protezione impedisce a chi richiede, ad esempio, la sola visualizzazione, di ottenere un duplicato digitale del documento.

6.4 *Applicazioni per il servizio mobile*

Quando il terminale con cui l'utente accede al servizio è mobile, i requisiti di sicurezza non possono che aggravarsi. Soprattutto il problema dell'identificazione dell'utente diventa fondamentale per poter assegnare correttamente la bolletta del servizio.

Per questo motivo, sia il sistema GSM (Global System for Mobiles), sia quello DECT (Digital European Cordless Telecommunications) utilizzano tecniche di autenticazione dell'utente nei confronti del servizio.

Il metodo utilizzato dal GSM è il più sicuro: una smart card inserita nel telefono contiene tutti i dati e la chiave segreta necessari per l'identificazione dell'utente; è anche presente la protezione tramite PIN (modificabile a piacere dall'utente, fra 4 e 8 cifre). Il formato della smart card può essere quello standard (tipo tessera) o più piccolo (per i terminali di dimensioni ridotte) detto "plug-in".

Il DECT usa un algoritmo simile, ma più veloce: si può raggiungere un tempo di set-up della chiamata inferiore al secondo mentre, per contro, il GSM può richiedere qualche

secondo. La chiave segreta è attualmente presente all'interno dell'apparecchio DECT (e non su smart card come avviene nel GSM) e quindi presumibilmente non sarà difficile leggerla esaminando internamente il telefono. Ciò rappresenta una possibile vulnerabilità; c'è comunque una protezione tramite PIN (di 4 cifre) richiesto all'utente.

Dal punto di vista della confidenzialità entrambi i sistemi cifrano la conversazione con un algoritmo crittografico. Ciò è dovuto nuovamente alla natura mobile del terminale che renderebbe le intercettazioni via radio piuttosto semplici.

D'altra parte il fatto di rendere impossibili le intercettazioni ha lo svantaggio di non consentire alle Forze dell'Ordine di effettuare legittime indagini sulle conversazioni. Questo problema potrebbe essere superato introducendo appositi algoritmi sufficientemente sicuri per gli utenti, ma in ogni caso aggirabili da chi ne fosse autorizzato; studi sono in corso su questo tema e una simile proposta è emersa anche negli Stati Uniti per quel che riguarda la telefonia tradizionale. Si è suggerito di introdurre negli apparecchi telefonici un chip denominato "Clipper", in grado di rendere confidenziale la telefonata, ma contenente una "porta" d'accesso segreta (una cosiddetta "trap-door" nota solo alle agenzie di investigazione federali) cosi` da consentire comunque l'intercettazione.

Il problema sfocia nel campo politico, mantenendo comunque risvolti tecnici importanti: l'elaborazione di nuovi algoritmi e dispositivi sicuri, ma in qualche modo "controllabili", potrebbe essere nel futuro, un tema di grande importanza per la sicurezza dei servizi di telecomunicazione.

7. Conclusioni

La sicurezza è un argomento trasversale a tutti i servizi di telecomunicazione e presenta molteplici sfaccettature.

Molti aspetti della sicurezza di tali servizi sono però comuni e possono risolversi con soluzioni comuni.

Fra questi aspetti, la funzione che praticamente tutti i futuri servizi di telecomunicazione dovranno effettuare in maniera sicura è l'autenticazione dell'identità dell'utente.

Non ci sono differenze rilevanti fra i vari servizi per quel che riguarda l'esigenza di riconoscere con certezza l'identità dell'utente e quindi una soluzione comune potrebbe giovare sia in termini economici, che in termini di praticità per il pubblico. Nessuno vorrebbe portarsi in tasca, nei prossimi anni, molte schede (magnetiche e non) ciascuna con la sua password da tenere a mente. D'altra parte, se non si promuove una armonizzazione comune, si corre proprio questo rischio.

Al momento attuale la soluzione più sicura (e anche la più economica, quando si debba partire da zero nella definizione dei terminali d'utente) è costituita dalle smart card.

Le smart card giocheranno nei prossimi anni un ruolo sicuramente di primo piano: sono di facile accettazione per l'utente (tutto avviene come con il Bancomat odierno) e, grazie ai più recenti algoritmi crittografici (quelli asimmetrici), consentono una semplice gestione delle chiavi.

La realizzazione delle funzioni di confidenzialità e integrità resta, invece, su un piano puramente tecnologico: si tratta di applicare algoritmi crittografici in maniera relativamente invisibile all'utente e tutto ciò che si deve valutare è la velocità di trasmissione e la robustezza dell'algoritmo.

Bibliografia

[1] Degiovanni, B.: *Aspetti di sicurezza nei servizi e nei sistemi d'utente*. Luglio 1993.

[2] Lumello, N.; Cocino, R.: *Prodotti per la verifica dell'identità di utenti finali in ambiente rete*. Settembre 1994.

[3] *Green Paper on the Security of Information Systems*. European Commission (DG XIII), Aprile 1994.

[4] *OSI Security Architecture*. ISO 7498-2 (CCITT X.800).

[5] *Information Technology Security Evaluation Criteria (ITSEC)*. Version 1.2, Giugno 1992.

[6] Degiovanni, B.: *Il processo di valutazione della sicurezza dei sistemi informatici (ITSEC)*. CSELT RR 92.0405, Settembre 1992.

[7] Degiovanni, B.: *Sicurezza nei Servizi di Telecomunicazione; il problema delle frodi*. CSELT DTR 94.0749, Novembre 1994.

[8] *Authentication Services*. Progetto Eurescom P234, CSELT RR 93.0322, 1993.

[9] *Definition of a pan-european IC card for authentication; Deliverable Task 2*. Progetto Eurescom P401, CSELT DTR 94.0538, Agosto 1994.

[10] Zoreda, J.L.; Oton, J.M.: *Smart Cards*. Artech House, 1994.

[11] Pfleeger, C.P.: *Security in Computing*. Prentice-Hall, 1989.

[12] Amoroso, E.G.: *Fundamentals of Computer Security Technology*. Prentice-Hall, 1994.

Introduzione alla crittografia

C. Montechiarini (*)

L'articolo descrive i meccanismi fondamentali di crittografia, evidenziandone le applicazioni pratiche più significative. Sono trattate le problematiche relative all'autenticazione, la confidenzialità, l'integrità e l'autenticità dei dati, il non ripudio di documenti a valore legale. Vengono introdotti gli algoritmi simmetrici e quelli a chiave pubblica con cenni alle questioni relative alla loro gestione amministrativa. Alcuni esempi spiegano come l'uso di codici consenta di realizzare servizi che assolvono alle funzioni tipiche degli ambienti con requisiti di sicurezza. In appendice è riportata una trattazione dettagliata del sistema DES(1) allo scopo di dimostrare il grado di complessità dei sistemi reali e illustrare l'architettura interna di un prodotto commerciale.

1. Introduzione

L'esigenza di mantenere confidenziali le informazioni scambiate attraverso i mezzi di comunicazione è stata fin dall'antichità il motivo principale dell'interesse dell'uomo verso le tecniche crittografiche. Non a caso il codice di Cesare(2) è ancora oggi il più classico degli esempi che vengono menzionati quando si parla di algoritmi di codifica. Da allora si sono fatti importanti progressi ai quali hanno contribuito in primo luogo le teorie matematiche sulla complessità computazionale.

Se da un lato si è studiato come migliorare l'affidabilità e la riservatezza delle comunicazioni, dall'altro è stato altrettanto forte l'impegno nel cercare di violare le barriere predisposte per assicurare questi requisiti. La lotta contro i crittoanalisti(3) è dichiarata al punto che anche la valutazione della robustezza degli algoritmi viene espressa in termini di tempi o risorse necessarie a comprenderne il meccanismo di funzionamento ed essere in grado di riprodurlo. I metodi usati per produrre codici sicuri sono quindi basati su funzioni che realizzano trasformazioni difficilmente invertibili.

Il principale criterio di classificazione delle tecniche di crittografia distingue tra sistemi simmetrici e asimmetrici. Entrambe sono utilizzabili per realizzare i più comuni servizi di sicurezza, ma implicano un differente impatto sugli aspetti prestazionali e amministrativi. La scelta tra le due dipende pertanto dal particolare contesto operativo. Sono comuni soluzioni ibride, in cui vengono sfruttate le migliori caratteristiche dell'una e dell'altra. Negli ambienti privati, dove gli interlocutori che devono interagire in condizioni protette sono pochi, i sistemi simmetrici sono maggiormente diffusi, ma quando si pensa a reti pubbliche di grandi dimensioni sembra sempre più realistica l'affermazione dei codici asimmetrici, che facilitano la gestione e sono più scalabili.

2. Algoritmi di codifica simmetrici

Oggetto della *crittografia* è lo studio della scrittura segreta, ovvero delle tecnologie attraverso le quali un testo, di cui si vuole proteggere il contenuto, viene elaborato così da non essere più intellegibile ad osservatori esterni, ed è poi ricondotto al suo formato originario una volta giunto in possesso del corretto destinatario. Tale trasformazioni, denominate rispettivamente *codifica* e *decodifica*, sono eseguite secondo regole ben precise che costituiscono nel loro insieme un *codice*.

La segretezza di un messaggio non può tuttavia essere affidata alla riservatezza del codice, in quanto l'uso ripetuto dello stesso algoritmo fornisce sempre maggiori informazioni a chi intende violare la sicurezza del sistema di crittografia. E' per questa ragione che si

(*) Ing. Claudio Montechiarini -Telecom Italia DG- Roma

(1) Data Encryption Standard.

(2) Consiste in un semplice algoritmo di sostituzione con K=3, vedi par. 2.

(3) La crittoanalisi è la scienza che si occupa dei metodi per rompere la segretezza di un codice crittografico.

sono sviluppati codici nei quali viene utilizzata una informazione ausiliaria, detta *chiave*, che rende parametrico il processo di codifica. Si consideri, a titolo di esempio, il caso elementare in cui, considerando l'alfabeto italiano come ciclico[4], ogni carattere di un testo venga sostituito con quello che lo segue di K posizioni. Se si assume $K = 3$, la frase "OGGI SPLENDE IL SOLE" diventa "RLLNVSOHQGHNOVROH". Se invece si suppone $K = 5$ si ottiene il risultato "TNNPAUQLSILPQATQL". Ebbene, l'algoritmo di *sostituzione* è chiaramente parametrico rispetto a K che ne è pertanto la chiave. In generale si può quindi rappresentare una operazione di codifica con una funzione E con due argomenti, il messaggio originario e la chiave, secondo la notazione:

$$C = E(M, K)$$

dove:
M è il messaggio nel formato originario,
C è il messaggio codificato,
K è la chiave di codifica.

Per ricostruire il testo iniziale si deve rimpiazzare ogni lettera con quella che la precede di K posizioni. Occorre quindi conoscere sia il valore di K che la regola di trasformazione inversa rispetto a quella applicata precedentemente. Generalizzando, si può rappresentare anche l'operazione di decodifica con una funzione D a due argomenti, il messaggio codificato e la chiave, secondo la notazione: $M = D(C, K)$ (fig. 1).

Nell'esempio i procedimenti di codifica e decodifica utilizzano la stessa chiave e presuppongono che essa sia nota esclusivamente al mittente e al destinatario dei messaggi. Gli algoritmi di questo tipo sono definiti *a chiave privata*, poiché si basano sull'assunzione che ciascuna coppia di interlocutori concordi e mantenga segreta la chiave con la quale proteggere i messaggi scambiati. Essi sono anche definiti *simmetrici*, in relazione al fatto che la chiave usata per la codifica è la stessa con la quale di esegue la decodifica. Sebbene ciò possa sembrare ovvio, esistono in realtà tecniche alternative, di notevole interesse applicativo, in cui le due chiavi sono diverse; di esse si parlerà in uno dei paragrafi successivi.

Oltre alla sostituzione, un altro meccanismo usato tipicamente nei codici è la *trasposizione*, ovvero la modifica dell'ordine dei caratteri componenti un testo. Per esemplificare questa tecnica prendiamo in esame, come è stato fatto in precedenza, la frase "OGGI SPLENDE IL SOLE" e riscriviamola su tre righe,

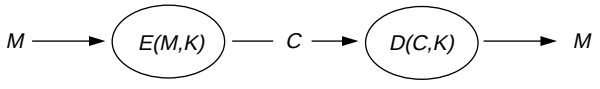


Figura 1 Sistema crittografico

seguendo un percorso a 'zig-zag', secondo quanto rappresentato in figura:

O S N L E
G I P E D I S L
G L E O

A questo punto rileggiamo la sequenza nell'ordine consueto, facendo seguire alla prima la seconda riga e a questa la terza. Otteniamo così il messaggio codificato OSNLEGIPEDISLGLEO. La chiave che caratterizza il codice così costruito coincide con il numero delle righe sulle quali si dispone il testo e, nel caso specifico, vale 3.

Entrambi i metodi di sostituzione e trasposizione sono di per sé piuttosto semplici, tuttavia il loro uso congiunto e iterato permette di costruire algoritmi di codifica sofisticati, il più noto dei quali è il DES. Il DES è stato sviluppato da IBM ed è adottato dal governo statunitense. Una sintetica esposizione del DES è fornita in Appendice.

3. **Requisiti funzionali**

Le funzioni di sicurezza sono classificabili in 10 categorie: identificazione, autenticazione, controllo d'accesso, confidenzialità, integrità e autenticità dei dati, affidabilità del servizio, non ripudio, audit, attribuzione di responsabilità (accountability).

Il requisito minimo di un sistema sicuro è che ciascuna delle azioni che in esso vengono compiute sia riconducibile ad un ben determinato esecutore. Affinché tale condizione venga realizzata è indispensabile che chiunque, prima di poter accedere al sistema, sia identificato e autenticato, ovvero fornisca la propria identità e una ulteriore prova del fatto che è effettivamente chi asserisce di essere. Esistono modalità più o meno affidabili per fornire questo genere di evidenza. Le *password* sono la soluzione più debole al problema. Esse si basano sul presupposto che esista una lista di tutti i legittimi utenti, ciascuno dei quali possiede una parola chiave (password), di sua esclusiva conoscenza. La segretezza della password impedisce a eventuali intrusi di spacciarsi per legittimi utilizzatori, in quanto il solo nominativo non è sufficiente per accedere al sistema. I malintenzionati possono tuttavia tentare di indovinare la password. Si può allora ottenere una maggiore sicurezza adottando ulteriori restrizioni sul vocabolario delle parole chiave o piuttosto usando informazioni di autenticazione più difficilmente riproducibili (per esempio le impronte digitali o il fondo della retina).

Supposto che sia stata superata con successo la fase di autenticazione, occorre limitare le risorse delle quali

(4) per cui dopo la lettera *'z'* si ricomincia con la *'a'*.

l'utente ha diritto di fruire. Tale funzione viene denominata *controllo d'accesso* e può essere realizzata secondo logiche diverse. Per esempio si può specificare per ogni utente la lista delle risorse a lui concesse, o viceversa, indicare, per ogni risorsa, la lista degli utenti che possono utilizzarla. Per semplificare la gestione di queste liste è talvolta consigliabile definire i criteri di accesso con riferimento ai *ruoli*, ovvero a gruppi di persone con specifiche abilitazioni, indicando a parte l'elenco di coloro che assolvono ai ruoli previsti.

La *confidenzialità* è invece la caratteristica per cui le informazioni riservate rimangono tali anche qualora vengano in possesso di qualcuno che non è autorizzato a conoscerle. Un altro requisito rilevante nello scambio di messaggi è la certezza che i dati ricevuti siano quelli effettivamente trasmessi. Si consideri ad esempio una transazione bancaria. E' chiaro che, insieme alla esigenza di proteggere la riservatezza del cliente è essenziale scongiurare la simulazione di operazioni critiche quali i depositi o i prelievi. Una frode del genere potrebbe essere realizzata ripetendo un messaggio intercettato durante una regolare transazione. Perché si possano escludere eventi simili deve essere assicurata l'*autenticità dei dati*. Analogamente, è importante preservare l'*integrità* delle informazioni. Alterare l'importo di un deposito, può infatti essere una tecnica persino più redditizia della replica di una precedente operazione.

Rimanendo nell'ambito commerciale, si comprende facilmente l'esigenza di strumenti mediante i quali dirimere eventuali contenziosi. Documenti come gli ordinativi o le fatture devono avere valore legale anche quando vengono scambiati in formato elettronico, e quindi non è accettabile che si possa negarne l'invio o la ricezione. Il *non ripudio*, sia da parte del mittente che del destinatario, è dunque una ulteriore funzionalità di sicurezza che deve essere fornita in ambienti informatici dove si vuole supportare le tipiche attività che intercorrono tra clienti e fornitori.

Una forma meno comune, ma tutt'altro che trascurabile, di minaccia alla sicurezza proviene dagli attacchi diretti alla *disponibilità del servizio*. Se un professionista utilizza la posta elettronica per mantenere i suoi contatti di lavoro, potrebbe essere fortemente danneggiato da chi, inondando la sua casella postale con un enorme numero di messaggi, lo costringesse a una affannosa ricerca di quelli realmente significativi. Analogamente, l'acquisizione e il mancato rilascio di risorse di elaborazione o informative danneggia chi in quel momento non ne può usufruire.

Infine, si comprende facilmente come, in aggiunta alle contromisure attuabili per garantire i requisiti appena menzionati, si debba considerare l'eventualità che esse vengano violate. Tutti i tentativi di aggirare o forzare i meccanismi di sicurezza dovranno pertanto essere registrati, così che si possano individuare i responsabili e le debolezze che hanno incoraggiato e, nella peggiore delle ipotesi, consentito le intrusioni fraudolente. L'attività di controllo degli eventi rilevanti per gli aspetti di sicurezza viene denominata *audit*, mentre gli strumenti che nel loro insieme concorrono ad attribuire la responsabilità delle azioni illegali realizzano la funzione di *accountability*.

4. Esempi d'impiego delle tecniche di crittografia

L'uso della crittografia permette di soddisfare alcune delle precedenti esigenze. In particolare è interessante vedere come si realizzano la confidenzialità, l'integrità, l'autenticità dei dati e il non ripudio. Gli algoritmi simmetrici forniscono una possibile soluzione sia per gli aspetti di *segretezza* che di *autenticità dei dati*. Entrambi sono garantiti dal fatto che la password è nota esclusivamente ai due interlocutori. Infatti, se viene ricevuto un messaggio decodificabile l'unico che può averlo inviato è l'altro che conosce la parola chiave e, d'altro canto, si è pure certi che nessuno sia nelle condizioni di violare la confidenzialità delle informazioni scambiate perché non possiede gli strumenti necessari per la decodifica.

E' importante notare che non è affatto garantito il *non ripudio*. Chi riceve un messaggio non ha mezzi per dimostrare di averlo avuto da un corrispondente, poiché, essendo anche lui in possesso della chiave di codifica, potrebbe generarlo autonomamente. Occorre quindi predisporre un meccanismo più articolato. Una delle soluzioni al problema prevede il coinvolgimento di una terza parte fidata. Sia il mittente (A) che il destinatario (B) condividono una password esclusivamente con la terza parte (C). Quando A trasmette un messaggio a B, B non è in grado di decifrarlo e deve pertanto richiedere l'intervento di C. C è l'unico che conosce la password di A e perciò può garantire l'autenticità del messaggio, decodificarlo e restituirlo a B dopo averlo codificato nuovamente con la password di quest'ultimo (fig. 2). In questo modo viene assicurato il non ripudio sia da parte del mittente che del destinatario. Quest'ultima osservazione

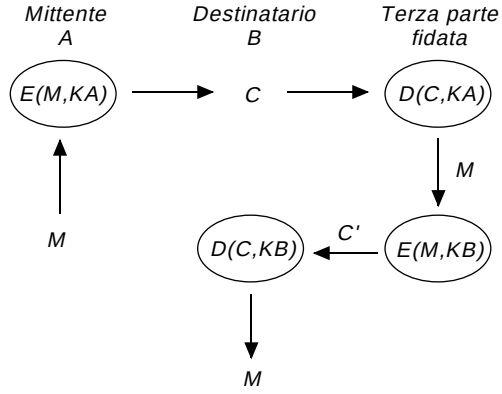


Figura 2 Protocollo a chiave privata con funzione di non ripudio

non è banale, esistono infatti tecniche di *firma elettronica*[5] (vedi paragrafo successivo) che consentono di verificare solo il non ripudio da parte del mittente.

La *integrità dei dati* viene invece verificata a posteriori. Le tecniche di crittografia permettono di rilevare l'avvenuta corruzione di un messaggio, ma non sono ovviamente in grado di prevenire questa eventualità. Le contromisure utilizzate si basano sulla generazione di un codice, indicato con la sigla MAC[6], dipendente dai dati e dalla sessione. L'uso di funzioni crittografiche previene la riproducibilità del MAC da parte di un attaccante privo della chiave. Uno dei metodi più diffusamente utilizzati per questo tipo di applicazioni sfrutta gli stessi meccanismi che permettono di riscontrare e correggere gli errori sulle memorie di massa o nella trasmissione dati ed è noto con l'acronimo CRC[7]. Il codice di controllo a ridondanza ciclica viene ottenuto mediante un'operazione di divisione il cui dividendo è costituito dalla stringa di bit da trasmettere ed il cui divisore è mantenuto segreto. In alternativa si usano pure funzioni hash[8] che generano i codici in modo tale che sia minima la possibilità di ottenere lo stesso risultato a partire da messaggi differenti. L'MCD[9], questo è il nome con il quale si indica il codice ottenuto, viene aggiunto in coda e cifrato insieme ai dati utili.

5. Algoritmi di codifica asimmetrici

I sistemi di crittografia asimmetrici utilizzano, per ogni utente, due chiavi: una privata e l'altra pubblica. Esse godono di una importante proprietà biunivoca: ciascuna delle due permette di decodificare un messaggio che è stato crittografato facendo uso dell'altra. Inoltre sono generate in modo tale che, una volta definita la chiave privata, la corrispondente pubblica è calcolata con una trasformazione non invertibile. In altre parole ciò comporta che la conoscenza della chiave privata implica quella della chiave pubblica, ma non il viceversa.

$$C = E(M, Kpb) \rightarrow D(C, Kpv) = M$$
$$C = E(M, Kpv) \rightarrow D(C, Kpb) = M$$

L'utilità di queste caratteristiche si comprende alla luce dei requisiti di confidenzialità e autenticità formulati precedentemente. La nomenclatura usata per denominare le due chiavi, pubblica e privata, è autoesplicativa. Infatti la chiave pubblica è nota a tutti, mentre la chiave privata è riservata. Quando si vuole assicurare la segretezza di un messaggio lo si codifica con la chiave pubblica del destinatario (B), che sarà il solo a poterlo leggere in quanto possessore della corrispondente chiave privata (fig. 3). Quando invece si vuole assicurare l'autenticità occorre che il mittente (A) codifichi il messaggio con la sua chiave privata. Il fatto che unicamente la sua chiave pubblica permetta di risalire al testo originario dimostra che nessun altro, eccetto lui, può averlo trasmesso (fig. 4).

Le due esigenze possono presentarsi congiuntamente. In questa ipotesi il mittente provvederà a codificare il messaggio con la sua chiave privata per garantirne l'autenticità, quindi, eseguirà una ulteriore codifica, con la chiave pubblica del destinatario, per tutelare gli aspetti di segretezza. In fase di ricezione, il destinatario ricostruirà il testo originario con due decodifiche successive, per le quali userà in sequenza la sua chiave privata e la chiave pubblica del mittente (fig. 5).

A differenza di quanto detto per gli algoritmi simmetrici, insieme all'autenticità, viene garantito anche il non ripudio da parte del mittente (firma elettronica).

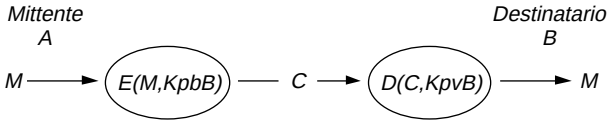


Figura 3 Protocollo a chiave pubblica con caratteristiche di confidenzialità

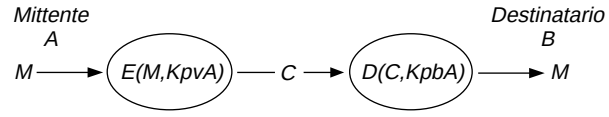


Figura 4 Protocollo a chiave pubblica con caratteristiche di autenticità

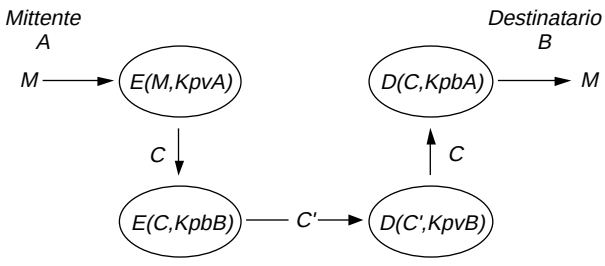


Figura 5 Protocollo a chiave pubblica con caratteristiche di confidenzialità e autenticità

(5) La firma elettronica deve garantire anche l'autenticità dei dati.

(6) Message Authentication Code.

(7) Cyclic Redundancy Checks.

(8) Una one-way hash function, $H(\cdot)$ è una funzione che trasforma un messaggio di lunghezza arbitraria, M, in una stringa di lunghezza fissa, $h=H(M)$, chiamato hash value o message digest se riferito ad un documento. Le funzioni di hashing a senso unico hanno la particolarità di poter essere facilmente calcolabili, ma estremamente difficili da invertire. Dato il valore h, è quindi computazionalmente impossibile trovare il messaggio da cui è stato originato; questa caratteristica, unita alla bassissima probabilità che da due argomenti distinti, $M1$ ed $M2$, si ottenga lo stesso risultato, fa sì che h sia utilizzabile come prova della autenticità e dell'integrità del documento che lo ha generato.

(9) Manipulation Detection Code.

Non esiste la possibilità che venga simulata la ricezione di un messaggio in quanto la chiave di codifica è quella privata e non è nota la destinatario. Per escludere il ripudio da parte del ricevente è comunque indispensabile l'intervento di una terza parte fidata, secondo uno schema del tipo descritto precedentemente.

6. **Servizio di Certificazione**

Quanto detto finora a proposito dei sistemi di crittografia asimmetrici prescinde da come gli utenti siano venuti in possesso delle loro chiavi e di quelle dei loro interlocutori. Chi fornisca queste informazioni è un problema tutt'altro che marginale che trova la sua più naturale soluzione nella istituzione di una infrastruttura pubblica per la gestione delle chiavi. Quando il mittente codifica un messaggio deve essere certo che la chiave pubblica che sta utilizzando sia quella del corretto destinatario. Tale consapevolezza viene garantita mediante il rilascio di *certificati* emessi da una terza parte fidata.

Un certificato associa l'identificativo di un utente alla sua chiave pubblica ed è inoltre protetto dalla firma elettronica di chi lo emette. Gli enti di certificazione sono organizzati secondo una struttura gerarchica che stabilisce una relazione di fiducia basata sul rapporto di parentela. Tale relazione impone che ciascuna delle parti coinvolte nel processo di autenticazione riconosca i certificati emessi dai suoi antenati e sia quindi in grado di leggerli utilizzando la chiave pubblica con la quale sono stati codificati.

Per chiarire come si attua il processo di acquisizione delle chiavi pubbliche, si consideri la situazione (fig. 6) in cui A sia il mittente di un messaggio e non conosca la chiave del destinatario B. Si supponga inoltre ch'egli possa consultare un servizio di directory dal quale desumere il percorso (D-C-B) che, a partire dal primo antenato comune, D, gli permette di raggiungere B. A è pertanto nelle condizioni di richiedere a B il certificato emesso da D a favore di C. Conoscendo la chiave pubblica di D, A può decifrare il certificato e individuare così la chiave pubblica di C. Analogamente, A richiede a B il certificato emesso da C a favore dello stesso B e ne verifica la validità mediante le informazioni ricavate al passo precedente.

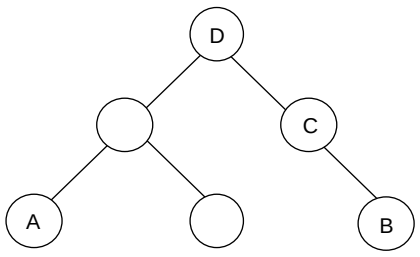


Figura 6 Albero di certificazione

Questo schema ha il vantaggio di riflettere da un lato la struttura aziendale e dall'altro il principio di località delle comunicazioni. Se si pensa di far corrispondere ai nodi dell'albero gerarchico le unità organizzative, dalle funzioni aziendali ai singoli dipartimenti, si comprende come in questo modo si tenga automaticamente conto del fatto che gli scambi informativi siano, in buona misura, circoscritti nell'ambito dello stessa unità, all'interno della quale è perfettamente logico collocare anche la gestione degli aspetti di sicurezza. Solo per relazioni che coinvolgono dipartimenti distinti vengono interessati i livelli intermedi della gerarchia.

E' importante notare che, laddove sussistano interessi significativi tra utenze strutturalmente 'lontane' (ovvero che richiedono l'attraversamento di livelli multipli per individuare il primo antenato comune), può essere prevista la *certificazione incrociata*, mediante la quale viene limitata la propagazione delle richieste. Se, per esempio, A fosse stato certificato anche da C non ci sarebbe stato bisogno di coinvolgere D.

7. **Aspetti gestionali**

Chi amministra un sistema di crittografia deve prendere in considerazione l'eventualità che una chiave venga rubata o smarrita. La probabilità che ciò avvenga aumenta col passare del tempo per cui è opportuno che le chiavi siano rinnovate periodicamente, mantenendo traccia delle versioni precedenti e dei relativi periodi di validità. Conseguentemente anche i messaggi dovranno riportare l'orario e la data di emissione, in modo che si possa controllare se la chiave usata per codificarlo era valida al momento in cui è stato trasmesso. In proposito è da sottolineare come non può essere il mittente ad apporre queste informazioni, in quanto sarebbe fin troppo facile falsificarle per riutilizzare una chiave scaduta. E' perciò necessario che tale funzione venga assolta da una terza parte fidata che, a dimostrazione del fatto di essere intervenuta nel processo di validazione, provvederà a sua volta a firmare il messaggio, come illustrato nella fig. 7 (quando B riceve *C'*, e ricava *C*, è in grado di verificare con T se *KpbA* è effettivamente la chiave pubblica usata da A all'istante *t*).

Un altro esempio in cui è evidente il ruolo di primaria rilevanza che gli aspetti di gestione assumono nelle applicazioni basate su algoritmi di crittografia è la *distribuzione delle chiavi di sessione*. Nella scelta tra meccanismi di codifica si deve tener conto anche degli aspetti prestazionali. Dal punto di vista computazionale i codici a chiave pubblica sono più onerosi di quelli a chiave privata. Per questo motivo si usa spesso una soluzione di compromesso, impiegando gli algoritmi asimmetrici per lo scambio delle chiavi private che vengono poi utilizzate nella codifica dei messaggi mediante algoritmi simmetrici.

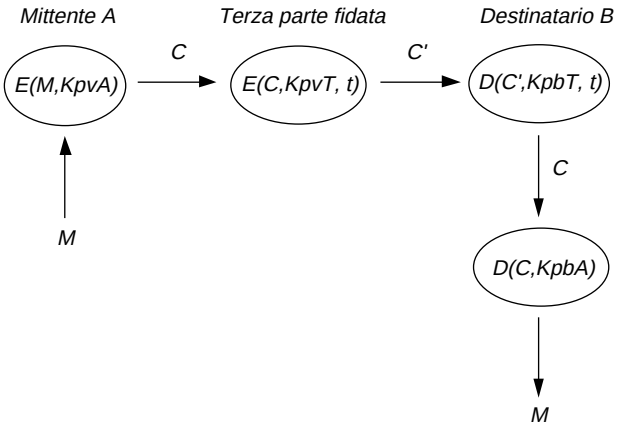


Figura 7 Protocollo a chiave pubblica con time stamp

Un semplice protocollo di distribuzione (fig. 8) prevede che il mittente A invii alla terza parte fidata il suo identificativo e quello del destinatario B, ricevendo in risposta i rispettivi certificati *Ca* e *Cb*. Da questi vengono ricavate le chiavi pubbliche dei due interlocutori. Il passo successivo consiste nel comunicare a B la chiave di sessione *K*, generata da A, e codificata prima con la chiave privata di A e poi con la chiave pubblica di B, allo scopo di assicurarne l'autenticità e la riservatezza. Onde evitare una ulteriore richiesta delle stesse informazioni alla terza parte fidata, insieme alla chiave di sessione vengono trasmessi a B anche i due certificati. B dispone a questo punto di tutti gli strumenti per ricostruire *K*. Per una maggiore sicurezza B può generare un numero random, mandarlo ad A, dopo averlo codificato con *K*, e ricevere indietro lo stesso numero modificato secondo una regola predefinita (per esempio, incrementato di uno) e nuovamente codificato con *K*.

Esistono anche protocolli per la distribuzione delle chiavi di sessione che non richiedono l'uso di algoritmi asimmetrici. Essi presuppongono il coinvolgimento di una terza parte fidata che comunica con gli utenti mediante canali sicuri protetti da chiave privata (fig. 9). La richiesta di una chiave di sessione è perfettamente identica a quella descritta in precedenza e prevede l'invio alla terza parte fidata degli identificativi dei due interlocutori A e B. La terza parte risponde ad A con un messaggio codificato in cui sono contenuti due elementi: la chiave di sessione *K* e un messaggio addizionale da trasmettere a B. Le informazioni per B sono codificate con la sua chiave privata e comprendono *K* e l'identità di A. Anche questa soluzione può prevedere un handshake aggiuntivo analogo a quello descritto sopra.

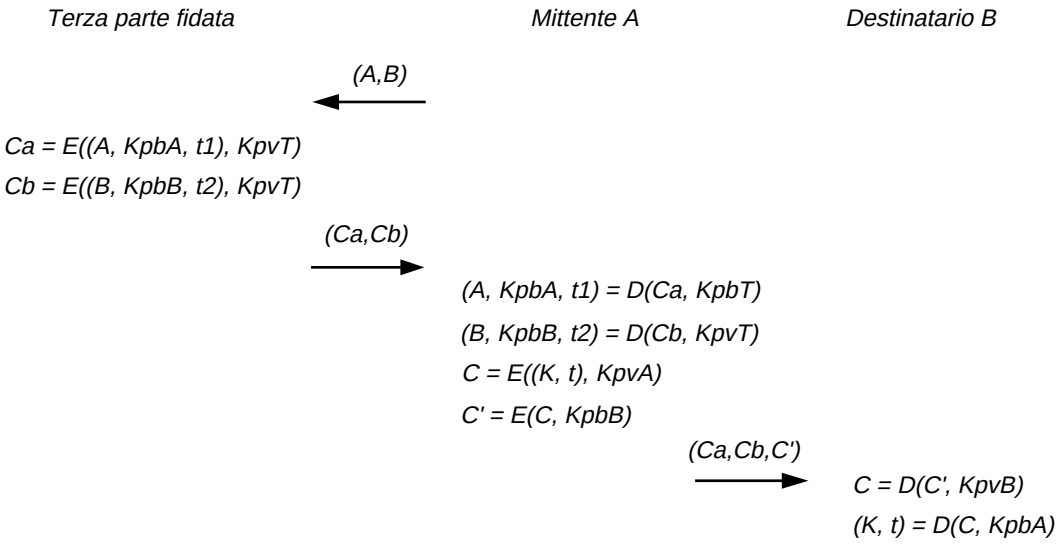


Figura 8 Protocollo asimmetrico per la distribuzione delle chiavi di sessione

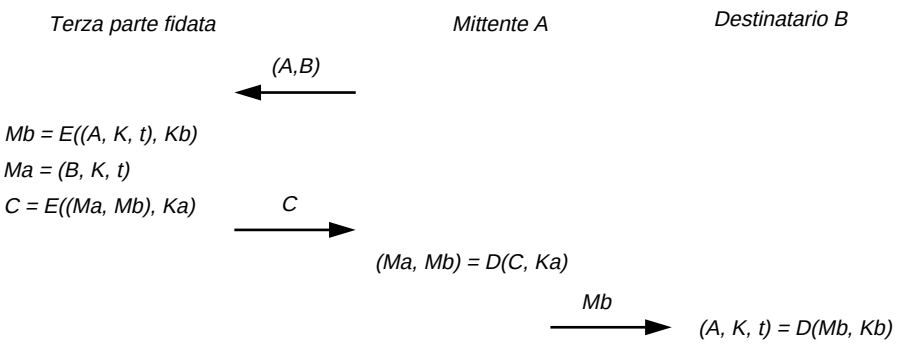


Figura 9 Protocollo simmetrico per la distribuzione delle chiavi di sessione

8. **Conclusioni**

Gli strumenti oggi disponibili per lo sviluppo di funzionalità di sicurezza hanno raggiunto un adeguato livello di maturità. Esiste altresì una crescente domanda verso questo tipo di servizi, che finora è stata soddisfatta con sistemi privati e spesso proprietari. In considerazione della forte tendenza verso reti di estensione mondiale e della apertura dei mercati, è sempre più importante l'adozione di soluzioni standard e pubbliche.

Sebbene le tecnologie siano in grado di supportare questa evoluzione esistono ancora forti barriere regolamentari. Come si è visto, in molte applicazioni è essenziale il ruolo svolto da una terza parte fidata che deve assicurare la regolarità delle transazioni elettroniche. Affinché si possa fare un uso commerciale delle reti telematiche è opportuno che tale funzione notarile abbia un riconoscimento legale. In attesa che siano istituite le organizzazioni che potranno svolgere questo compito, si stanno diffondendo offerte basate sui meccanismi di crittografia asimmetrici. Il credito che essi saranno in grado di guadagnare presso gli utenti sarà certamente indicativo delle loro prospettive di sviluppo.

Appendice

Data Encryption Standard (DES)

Il DES è un algoritmo di codifica a chiave simmetrica sviluppato da IBM e adottato dal governo statunitense nelle applicazioni non classificate. Si tratta di un codice a blocchi in cui il flusso dei dati binari è suddiviso in gruppi di 64 bit, ciascuno dei quali viene elaborato usando una stessa chiave di 56 bit.

Il DES si basa su una serie di operazioni di sostituzione e trasposizione opportunamente combinate secondo lo schema di fig. 10 i cui componenti essenziali sono le funzioni P_1 ed f. P_1 è una semplice permutazione e, come tale, è descritta dalla tabella T_1 (tab. 1) che, letta da sinistra a destra e dall'alto in basso, definisce la sequenza dei bit di output in funzione di quelli di input. Per esempio, il primo elemento della tabella è 58; ciò significa che il 58-esimo bit del blocco di ingresso diventerà il primo in uscita.

Analogamente il 50-esimo passerà in seconda posizione, il 42-esimo in terza e così via.

Dopo la permutazione i 64 bit sono divisi in due parti di 32 elementi, L_0 e R_0, che vengono modificati secondo le relazioni iterative:

$$\begin{cases} R_i = L_{i-1} \oplus f\left(R_{i-1,}\, K_i\right) \\ L_i = R_{i-1} \end{cases} \quad i = 1,\ldots,16.$$

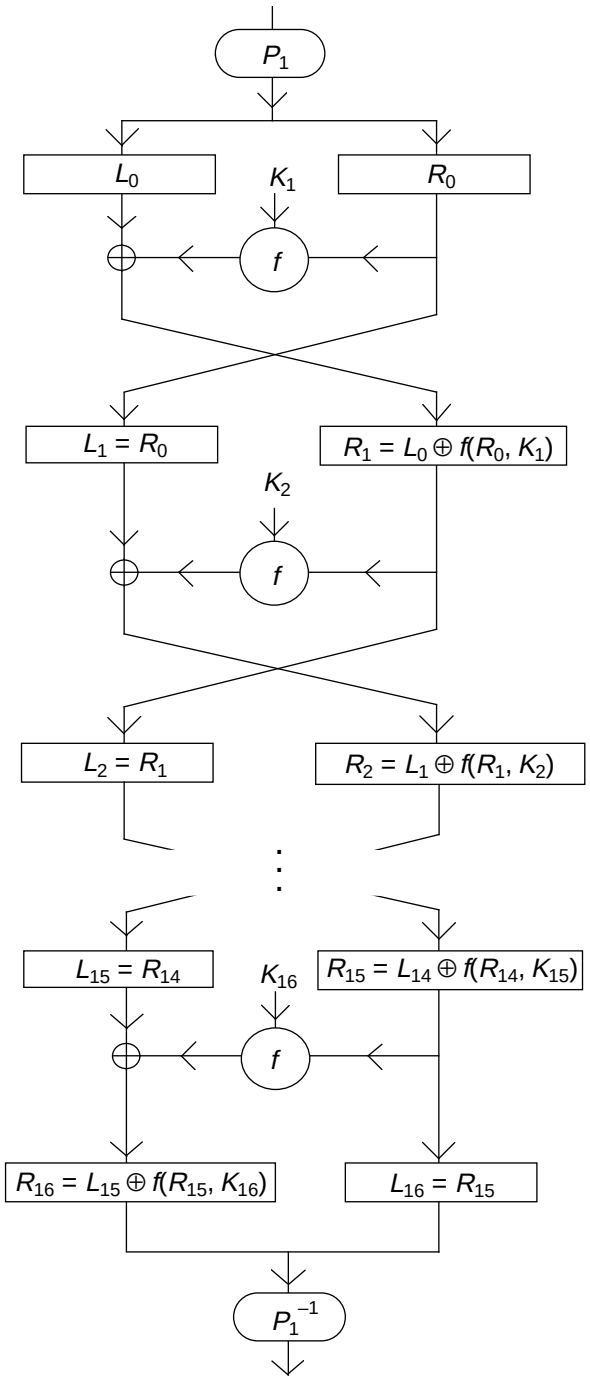


Figura 10 Algoritmo di codifica DES

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

Tabella 1 Tabella T_1 dello standard DES: permutazione P_1

La metà di destra R_{i-1} è portata a sinistra e diventa L_i; quella di sinistra, combinata in or esclusivo (operazione indicata con $\oplus$) con il risultato della funzione f, genera R_i. Questa elaborazione viene ripetuta 16 volte. E' da notare che nell'ultimo passo del ciclo, diversamente dai precedenti, le due metà non vengono scambiate. Tale accorgimento rende l'algoritmo perfettamente simmetrico e ne consente l'uso sia per la codifica che per la decodifica. Non è quindi necessario sviluppare un ulteriore chip per la ricezione dei messaggi criptati, con conseguenti economie di scala nella realizzazione dei prodotti commerciali che implementano il DES.

Al termine del ciclo i 64 bit $R_{16}L_{16}$ sono nuovamente permutati utilizzando una trasformazione (P_1^{-1}) che inverte quella iniziale (tabella T_2 - vd. tab. 2). Quest'ultima operazione fornisce il risultato definitivo.

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

Tabella 2 Tabella T_2 dello standard DES: permutazione P_1^{-1}

La funzione f ha due argomenti R_{i-1} e K_i che per il momento supporremo entrambi noti. Vedremo poi come si calcola K_i. Come illustrato nel diagramma in fig. 11, R_{i-1} viene esteso da 32 a 48 bit usando una tavola (tabella T_3 - vd. tab. 3) analoga a quella di P_1 e P_1^{-1}, ma con un numero di elementi (48) pari ai bit che si vogliono ottenere in uscita. L'uso della tabella è lo stesso descritto prima, per cui leggendola da sinistra a destra e dall'alto in basso si individua la sequenza di output in funzione di quella di input. E' evidente che alcuni riferimenti saranno duplicati in quanto si desidera avere 16 bit in più rispetto a quelli di partenza.

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

Tabella 3 Tabella T_3 dello standard DES: permutazione P_2

$P_2(R_{i-1})$ e K_i vengono combinati dall'operatore di or esclusivo ed il risultato viene suddiviso in 8 blocchi di 6 bit ciascuno:

$$P_2(R_{i-1}) \oplus K_i = B_1 \ldots B_8.$$

Ogni blocco B_i ($b_{i1}...b_{i6}$) è oggetto di una sostituzione anch'essa definita in forma tabellare (tabelle S_1,...,S_8 - vd. tab. 4). Le sostituzioni S_1,...,S_8 differiscono l'una dall'altra. Ognuna delle S_i ha 4 righe e 16 colonne e consiste di una quaterna di bit (interi da 0 a 15). Il meccanismo di sostituzione prevede che la coppia di bit $b_{i1}b_{i6}$ sia utilizzata per selezionare una delle 4 righe, mentre $b_{i2}...b_{i5}$ individuano una delle 16 colonne. Se per esempio $B_1 = 100001_2$ sarà presa la quaterna di bit posta sull'ultima riga (b_{11}=1, b_{16}=1 e quindi l'indice di riga è $11_2=3_{10}$) e sulla prima colonna ($b_{12}=b_{13}=b_{14}=b_{15}=0$ e quindi l'indice di colonna è $0_2=0_{10}$). Nel caso specifico $S(3,0)=14$ per cui $B_1=100001_2$ diventa $1110_2=14_{10}$.

Concatenando i risultati parziali ricavati da $B_1...B_8$ si

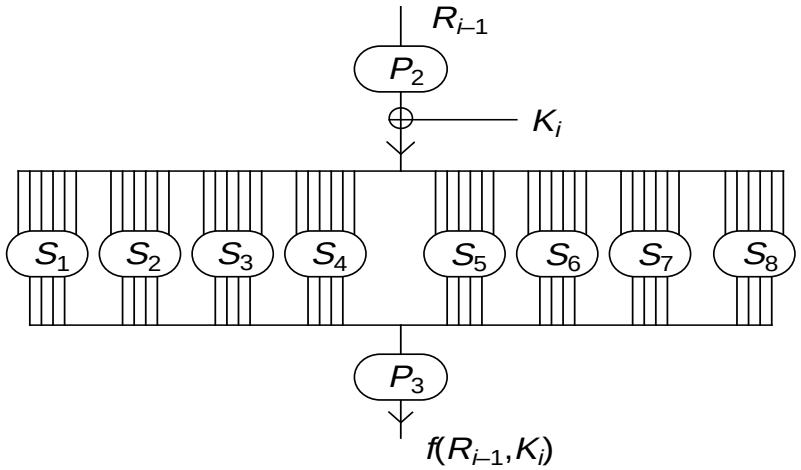


Figura 11 Calcolo della funzione f

S_1	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
	0	15	7	4	14	2	31	1	10	6	12	11	9	5	3	8
	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13
S_2	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9
S_3	10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8
	13	7	0	9	3	4	6	10	2	8	5	14	12	11	15	1
	13	6	4	9	8	15	3	0	11	1	2	12	5	10	14	7
	1	10	13	0	6	9	8	7	4	15	14	3	11	5	2	12
S_4	7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15
	13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9
	10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
	3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14
S_5	2	12	4	1	7	10	11	6	8	5	3	15	13	0	14	9
	14	11	2	12	4	7	13	1	5	0	15	10	3	9	8	6
	4	2	1	11	10	13	7	8	15	9	12	5	6	3	0	14
	11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3
S_6	12	1	10	15	9	2	6	8	0	13	3	4	14	7	5	11
	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
	9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6
	4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13
S_7	4	11	2	14	15	0	8	13	3	12	9	7	5	10	6	1
	13	0	11	7	4	9	1	10	14	3	5	12	2	15	8	6
	1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
	6	11	13	8	1	4	10	7	9	5	0	15	14	2	3	12
S_8	13	2	8	4	6	15	11	1	10	9	3	14	5	0	12	7
	1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2
	7	11	4	1	9	12	14	2	0	6	10	13	15	3	5	8
	2	1	14	7	4	10	8	13	15	12	9	0	3	5	6	11

Tabella 4 Tabelle di sostituzione S_i dello standard DES

ottengono 8·4=32 bit che subiscono una ulteriore permutazione P_3 in accordo alla tavola T_4 (tab. 5).

Riassumendo, la funzione f può essere rappresentata formalmente con la notazione:

16	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

Tabella 5 Tabella T_4 dello standard DES: permutazione P_3

$$f(R_{i-1},K_i) = P_3\big(S_1(B_1)\ldots S_8(B_8)\big),$$

dove $B_1\ldots B_8 = P_2(R_{i-1}) \oplus K_i$.

Ultimo aspetto da chiarire è come si calcola K_i. Ogni iterazione usa una chiave diversa ottenuta da K che, come si è detto all'inizio, consta di 56 bit. Il procedimento per la generazione di K_i è schematizzato in fig. 12. In generale:

$$K_i = P_4(C_i, D_i), \quad i=1,...,16,$$

dove P_4 è una permutazione descritta dalla tabella T_5 (tab. 6) e

$$\begin{cases} C_i = LS_i(C_{i-1}) \\ D_i = LS_i(D_{i-1}) \end{cases} \quad i = 1,\ldots,16.$$

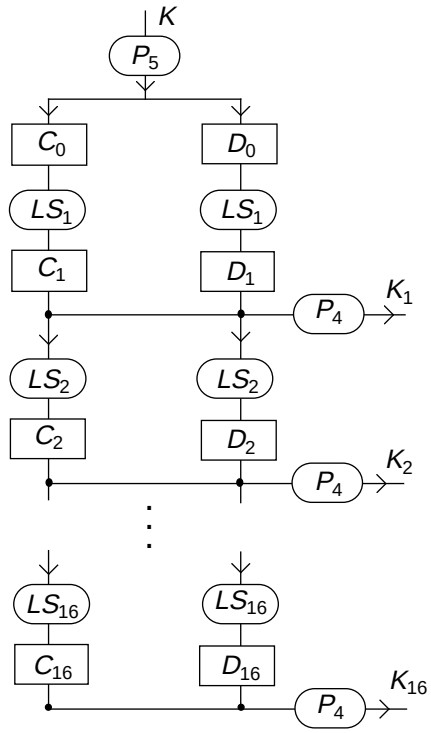


Figura 12 Calcolo delle chiavi K_i

14	17	11	24	1	5
3	28	15	6	21	10
23	19	12	4	26	8
16	7	27	20	13	2
41	52	31	37	47	55
30	40	51	45	33	48
44	49	39	56	34	53
46	42	50	36	29	32

Tabella 6 Tabella T_5 dello standard DES: permutazione P_4

LS_i sta ad indicare una operazione di rotazione a sinistra. L'indice i non è superfluo in quanto da esso dipende il numero di posizioni di cui vengono ruotate le sequenze di bit C_{i-1} e D_{i-1} (tabella T_6 - vd. tab. 7).

Le precedenti relazioni sono ricorsive per cui il processo è completamente definito quando siano noti C_0 e D_0. Essi si ottengono tramite la permutazione P_5 (tab. 8 - T_7) che interpreta K come un blocco di 64 bit supponendo che le posizioni multiple di 8 (8, 16, ...64) siano occupate da bit di parità che non entrano nel successivo calcolo (8+56=64). Il risultato $P_5(K)$ viene infine suddiviso a metà (28 bit per parte) e fornisce la coppia C_0 e D_0.

57	49	41	33	25	17	9
1	58	50	42	34	26	18
10	2	59	51	43	35	27
19	11	3	60	52	44	36
63	55	47	39	31	23	15
7	62	54	46	38	30	22
14	6	61	53	45	37	29
21	13	5	28	20	12	4

Tabella 8 Tabella T_7 dello standard DES: permutazione P_5

La decodifica è eseguita usando le chiavi K_i nell'ordine inverso a quello della codifica, ma mantenendo lo stesso algoritmo. E' infatti immediato verificare le relazioni inverse

$$\begin{cases} R_{i-1} = L_i \\ L_{i-1} = R_i \oplus f(L_i, K_i) \end{cases} \quad i = 1, \ldots, 16,$$

che discendono direttamente da quelle per il calcolo di R_i e L_i.

Bibliografia

[1] Denning, D.: *Cryptography and Data Security*. 1983.

[2] Ellman, M.E.: *La crittografia a chiave pubblica*. «Le Scienze», dicembre 1979.

[3] Muftic, S.; Sloman, M.: *Security architecture for distributed systems*. «Computer Communications», luglio 1994.

[4] Chokhani, S.: *Toward a national public key infrastructure*. «IEEE Communications Magazine», settembre 1994.

[5] Sandhu, R.; Samarati, P.: *Access control: principles and practice*. «IEEE Communications Magazine», settembre 1994.

[6] Martin, D.: *Trust in the new information age*. «AT&T Technical Journal», ottobre 1994.

[7] Valocchi, P.: *Introduzione alla crittografia*. 1993.

[8] DiFonso, M.; Sconci, M.; Valocchi, P.: *La sicurezza nei sistemi distribuiti*. 1993.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
#LS	1	1	2	2	2	2	2	2	1	2	2	2	2	2	2	1

Tabella 7 Tabella T_6 dello standard DES: funzione LS_i: numero di rotazioni in funzione dell'indice i

I protocolli TCP ed IP

F. Antonelli, M. Carissimi, F. Iuso, F. Pugliese (*)

La pila di protocolli TCP/IP, introdotta alla fine degli anni '60 per permettere l'interlavoro tra elaboratori aventi sistemi operativi differenti, deve la sua diffusione al trend di sviluppo molto accentuato avuto dalla rete Internet negli ultimi anni ed anche alla sua semplicità implementativa che ne permette l'utilizzo su diverse tecnologie di trasporto (linea dedicata, X.25, Frame Relay, SMDS, ATM). Nel presente articolo si introducono i fondamenti dell'architettura di rete riportandone caratteristiche e vincoli, e vengono descritti la pila protocollare e i meccanismi di trasporto. Si discute, inoltre, riguardo le modalità funzionali dei protocolli TCP ed IP nello scambio di messaggi, riportando l'interazione con gli altri protocolli della pila che ne completano le funzionalità. Si riporta, infine, l'analisi tecnica di dettaglio dei protocolli della pila TCP/IP.

1. Introduzione

Intorno agli anni settanta, il Ministero della Difesa degli Stati Uniti (DoD), attraverso l'Advanced Research Project Agency (ARPA), finanziò un progetto ed una sperimentazione su larga scala di protocolli di comunicazione a pacchetto.

Questo progetto nasceva alla fine degli anni '60, all'inizio del periodo della guerra fredda tra USA e U.R.S.S., ed aveva l'obiettivo di garantire con la massima affidabilità e sicurezza lo scambio di messaggi tra sedi militari americane e centri di ricerca anche in caso di un'azione bellica improvvisa ed inaspettata. Inoltre, i requisiti della sperimentazione imponevano da un lato l'interlavoro e la comunicazione tra sistemi di elaborazione basati su tecnologie differenti e dall'altro che l'architettura di rete e per i relativi protocolli fossero in grado di garantire la connettività anche in caso di malfunzionamenti e guasti. D'altro canto la possibilità di poter utilizzare collegamenti di elevata capacità e affidabilità quali quelli basati su fibra ottica, ha portato, fin dalle prime ipotesi, a definire un protocollo di rete che consentiva la realizzazione di una piattaforma di rete a commutazione di pacchetto senza connessione e senza garanzie di qualità del servizio, essendo queste ultime rimandate ai livelli applicativi.

La rete di calcolatori progettata e realizzata in campo, a cui fu dato il nome ARPAnet, è stata il primo esempio di rete geografica a commutazione di pacchetto. Essa fu il primo nucleo della rete Internet da si sviluppò in seguito .

Nel seguito sono presi in esame l'architettura di rete, lo stack protocollare e le modalità di istradamento dinamico e riconfigurazione della topologia della rete.

2. Architettura di rete e protocollare

L'architettura (fig. 1) prevede una rete non gerarchica costituita dall'interconnessione di sottoreti omogenee, cioè che utilizzano il protocollo di rete IP, mediante dispositivi denominati *IP router* e di sottoreti eterogenee mediante dispositivi chiamati gateway. Il router IP è un apparato che permette l'interconnessione di sistemi a livello di rete , mentre il gateway realizza l'interconnessione di sistemi con protocolli di comunicazioni diversi interessando i livelli OSI superiori a quello di rete. Gli utenti, infine, possono accedere alla rete con diverse tipologie elaboratori, dal semplice PC alla complessa Workstation fino ai grossi elaboratori: essi vengono identificati comunemente con il nome di Host.

(*) Ing. Ferruccio Antonelli, sig. Mauro Carissimi, ing. Francesco Iuso, ing. Francesco Pugliese -Telecom Italia DG- Roma

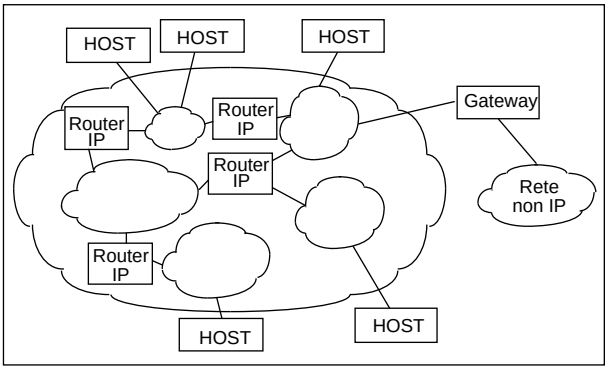


Figura 1 La struttura di riferimento dell'architettura di rete

I protocolli progettati per espletare le funzionalità richieste sulla rete ARPAnet furono chiamati TCP/IP (*Transport Control Protocol*/*Internet Protocol*).

IP è il protocollo di livello rete del modello OSI (cfr. fig. 2) che offre un servizio di tipo *datagram*, non affidabile e non orientato alla connessione (*connectionless*).

Le funzionalità svolte dal protocollo IP sono quelle tipiche dei protocolli di rete (livello 3 OSI): indirizzamento, instradamento, frammentazione, riassemblaggio dei pacchetti e loro inoltro in rete. Il protocollo gestisce sia pacchetti destinati a entità di livello superiore sia pacchetti in transito verso altre entità di pari livello.

Lo scopo del protocollo IP è quello di trasferire i pacchetti attraverso un insieme di reti interconnesse tra di loro. Questo trasferimento si ottiene facendo transitare i pacchetti da un nodo IP all'altro fino alla destinazione mediante opportuna interpretazione degli indirizzi IP. Ciò mette in luce una delle caratteristiche fondamentali del protocollo: l'indirizzamento universale. Esso consiste nella definizione di indirizzi univoci per gli host connessi alle reti.

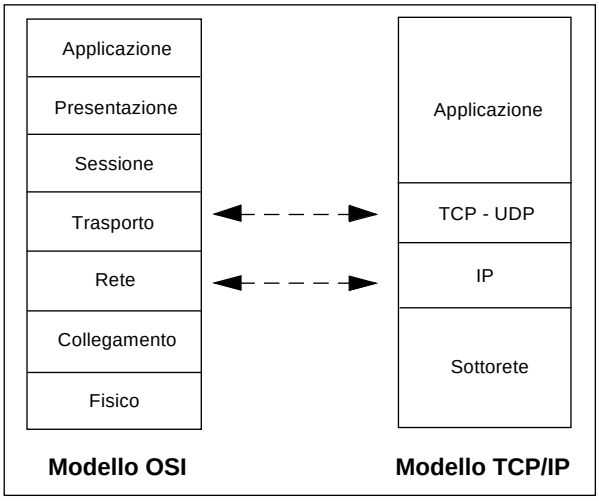


Figura 2 Confronto tra il modello OSI e la pila dei protocolli TCP/IP

Il protocollo IP non impone vincoli di nessun tipo sulla tecnologia delle reti di trasporto (X.25, Frame Relay, SMDS[(1)], ATM, CDN) utilizzabili per il trasferimento dei pacchetti dati da un nodo all'altro. Come ovvia conseguenza la velocità di trasmissione delle informazione è fortemente legata alla tecnologia di trasporto su cui IP si poggia.

Il trasferimento dei pacchetti può richiedere una segmentazione degli stessi laddove le dimensioni dei pacchetti gestiti dalle reti non coincidano con le dimensioni massime consentite. A tale scopo il protocollo fornisce un meccanismo specifico per la segmentazione e il riassemblaggio dei pacchetti.

Il protocollo IP tratta ciascun pacchetto come un messaggio indipendente da tutti gli altri pacchetti; non esistono pertanto i concetti di connessione o di circuiti logici (il protocollo IP appartiene pertanto alla classe connectionless).

Il livello di trasporto viene realizzato dai protocolli UDP (User Datagram Protocol) e TCP; TCP offre ai livelli applicativi un servizio affidabile di tipo orientato alla connessione (connection oriented), in grado quindi di supportare un numero elevato di applicazioni, da quelle interattive (ad es. Telnet [1]) a quelle caratterizzate da solo trasferimento di dati (File Transfer Protocol, Simple Mail Transfer Protocol).

TCP fornisce comunicazioni affidabili tra processi che risiedono su host diversi, collegati mediante reti di comunicazione interconnesse. Questa affidabilità viene raggiunta senza che lo stesso TCP imponga molti requisiti a quelli che sono i protocolli sui quali si poggia. Difatti, il protocollo TCP è stato sviluppato nel presupposto di funzionare anche sopra un servizio datagram potenzialmente inaffidabile; TCP può operare in un ampio spettro di piattaforme comunicative, dalla semplice connessione mediante cavo fino alla commutazione di pacchetto o di circuito.

TCP è in grado di trasferire un flusso continuo di dati tra i suoi due utenti in entrambe le direzioni nello stesso istante (full duplex), creando dei segmenti di dati da trasferire attraverso la rete che sta utilizzando.

I problemi tipici che si riscontrano nel trasferimento dei dati, presupponendo che il mezzo di trasporto non sia affidabile, sono il danneggiamento, la perdita, la duplicazione e la consegna fuori sequenza dei dati stessi. Il TCP cerca di sopperire a questi problemi assegnando un numero di sequenza a ciascun segmento dati trasmesso e richiedendo all'utente remoto un riscontro positivo (servizio confermato) su quello che effettivamente ha ricevuto.

Lo scambio dei dati, tra due entità TCP, è gestito mediante un semplice meccanismo che consente di adeguare il volume di dati trasmesso alle reali capacità

(1) Switched Multimegabit Data Service.

di ricezione ed invio dei processi TCP coinvolti. Ciò viene realizzato utilizzando un meccanismo a finestra variabile.

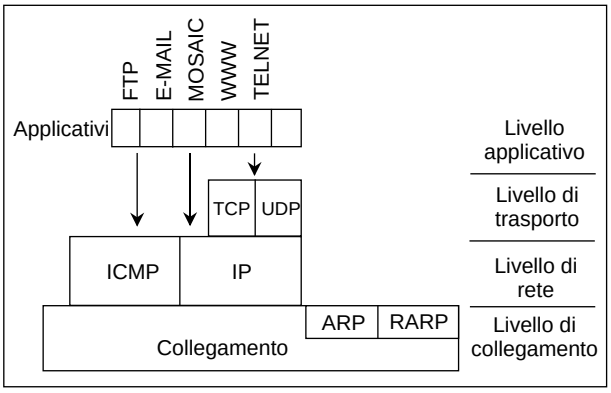
Il protocollo UDP, invece, offre ai livelli applicativi un servizio non confermato, e quindi non affidabile, di tipo senza connessione (c*onnectionless*), quindi orientato all'impiego in ambienti con requisiti di qualità di servizio meno stringenti.

Nella fig. 3 sono mostrate le relazioni tra gli applicativi e i protocolli della pila TCP/IP; in particolare ICMP (Internet Control Message Protocol) viene utilizzato per notificare situazioni di malfunzionamento ed i protocolli ARP e RARP supportano il meccanismo di indirizzamento fornendo l'indirizzo fisico di un host dato il suo indirizzo IP (ARP) e viceversa (RARP). Tali protocolli sono descritti in dettaglio nel seguito.

3. **I protocolli TCP ed IP**

3.1 *Protocollo IP*

Le funzionalità svolte dal protocollo IP possono riassumersi nell'indirizzamento, nell'instradamento, nella frammentazione e nel riassemblaggio dei pacchetti. L'instradamento di pacchetti presuppone la conoscenza sia dell'indirizzo IP che dell'indirizzo fisico del destinatario da parte dell'host sorgente. Tale associazione è resa possibile dai protocolli ARP, che consente di conoscere l'indirizzo fisico di un host a partire dal suo indirizzo IP, e RARP che realizza la funzionalità opposta dell'ARP.



TCP: Transport Control Protocol
UDP: User Datagram Protocol
ICMP: Internet Control Message Protocol
IP: Internet Protocol
ARP: Address Resolution Protocol
RARP: Reverse Address Resolution Protocol

Figura 3 Ambiente TCP/IP e relazione con gli applicativi

3.1.1 Formato del pacchetto IP

Il datagramma IP è formato da una parte intestazione (header), in cui è contenuto anche l'indirizzo, e da una parte dati (fig. 4).

I sottocampi costituenti il pacchetto IP hanno i seguenti significati ed utilizzi:

- *Version*: specifica la versione del protocollo IP impiegata.
- *Internet Header Length (IHL)*: specifica la lunghezza dell'header in parole di 32 bit.
- *Type of Service*: specifica parametri che riguardano la qualità del servizio, come il grado di precedenza del pacchetto (una sorta di priorità), quale tipo di servizio deve essere privilegiato nel routing tra basso ritardo, elevato throughput ed elevata affidabilità.
- *Total Length*: specifica la lunghezza del pacchetto, misurata in ottetti, includendo l'header e i dati. Il numero standard di ottetti trasmessi è 576.
- *Identification*: è un valore identificativo assegnato dal processo sorgente al pacchetto o ai suoi frammenti. Questo valore è fornito da un contatore dell'host sorgente che viene incrementato ogni qualvolta l'host genera un pacchetto. I router di transito ricopiano questo campo nell'header di ogni frammento in cui viene suddiviso un pacchetto.
- *Flags*: è un campo di 3 bit. Il primo bit è riservato ed è posto a 0; il secondo (DF: Don't Fragment) se 0 indica che il pacchetto può essere frammentato, se 1 no; il terzo bit (MF: More Fragment) se 0 indica che quello è l'ultimo frammento, se 1 che ci sono altri frammenti.
- *Fragment Offset*: indica l'offset del frammento all'interno del pacchetto di partenza misurato in numeri di ottetti a partire da zero. I frammenti non arrivano necessariamente in ordine e, se uno o più frammenti non viene ricevuto, verrà scartato l'intero pacchetto.
- *Time to Live*: indica quanto tempo il pacchetto può rimanere all'interno della rete. Quando un host invia un pacchetto, posiziona questo campo al massimo

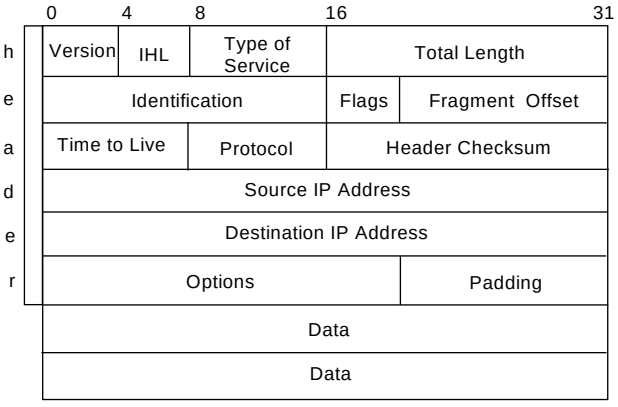


Figura 4 Formato di un pacchetto IP

tempo di vita del pacchetto all'interno della rete. Questo valore verrà decrementato dai router di transito; quando arriverà a zero il pacchetto verrà scartato. Questa tecnica impedisce ad un pacchetto di girare indefinitamente nella rete, se, a causa di errori, il suo instradamento risultasse un cammino chiuso.

- *Protocol*: specifica il tipo di protocollo di livello superiore utilizzato nella parte dati del pacchetto (ad es. TCP, UDP).
- *Header Checksum*: realizza un meccanismo di controllo di errore del solo header del pacchetto. Il contenuto di questo campo è ottenuto considerando i bit dell'header a gruppi di 16 alla volta, se ne effettua la somma e si memorizza il complemento a 1 del risultato.
- *Source IP Address e Destination IP Address*: campi riservati agli indirizzi IP sorgente e destinazione.
- *Options*: campo di lunghezza variabile (multipli di 8 bit) opzionale per la richiesta di prestazioni aggiuntive.
- *Padding*: rende la lunghezza dell'header multiplo intero di 32 bit mediante introduzione di zeri.

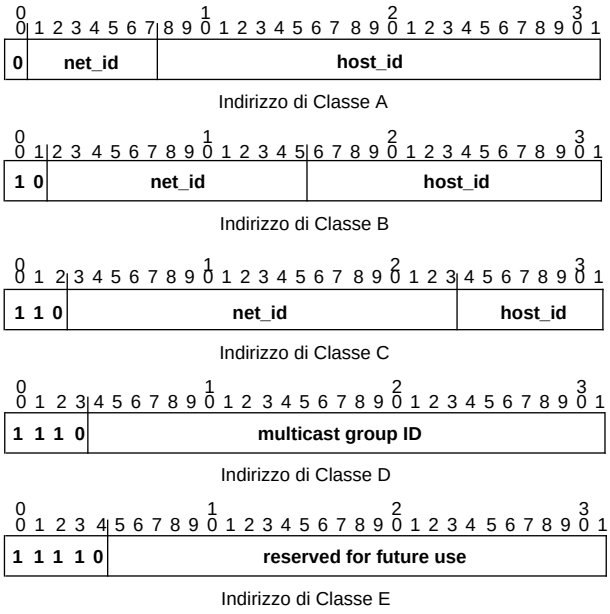


Figura 5 Formati degli indirizzi IP

3.1.2 Modalità di indirizzamento

Il sistema di indirizzamento IP è di tipo universale e permette di identificare un qualsiasi host collegato ad una qualsiasi rete interconnessa.

Una notazione comunemente utilizzata per gli indirizzi IP divide i 32 bit che compongono l'indirizzo in quattro campi di 8 bit e specifica il valore di ciascun campo come numero decimale, utilizzando come separatore dei campi un punto ("."). Questo tipo di rappresentazione dell'indirizzo è chiamata *notazione dot* (es. 151.99.250.4). A livello concettuale i 32 bit dell'indirizzo IP vengono suddivisi in due campi <net_id, host_id> in cui il primo identifica la rete mentre il secondo un singolo host di questa rete.

Esistono delle limitazioni ai valori che questi campi possono assumere, poiché il protocollo associa un particolare significato ad alcuni indirizzi riservati (come sarà descritto nel prossimo paragrafo). Le dimensioni di questi due valori possono variare dando origine ai seguenti formati (riportati nella fig. 5).

- Classe A**:** il bit di peso più elevato è posto a 0. Tale formato riserva 7 bit per l'identificazione della rete e 24 bit per l'identificazione dell'host all'interno della rete; l'indirizzo di Classe A si adatta bene al caso di reti con numerosi nodi al loro interno poiché con 24 bit si identificano potenzialmente fino a 2^{24} = 16.777.216 nodi distinti; in notazione "dot", le reti che possono essere definite in questo indirizzamento sono nell'intervallo da 1 a 128.
- Classe B**:** i due bit di peso più elevato sono posti rispettivamente a 1 e 0. tale formato riserva 14 bit per l'identificazione della rete e 16 bit per l'identificazione dell'host all'interno della rete; l'indirizzo di Classe B si adatta al caso di reti con numero di nodi interni (host) tali da consentire l'indirizzamento di un maggior numero di reti rispetto all'indirizzamento di Classe A. Infatti, con 16 bit si identificano potenzialmente fino a 2^{16} = 65.536 nodi distinti; in notazione "dot", le reti che possono essere definite in questo indirizzamento sono nell'intervallo da 128.1 a 191.254.
- Classe C: i tre bit di peso più elevato sono posti rispettivamente a 1,1 e 0. Tale formato riserva 21 bit per l'identificazione della rete e 8 bit per l'identificazione dell'host all'interno della rete; l'indirizzo di Classe C si adatta al caso di reti con numero di nodi interni basso così da consentire l'indirizzamento di un numero di reti ancora maggiore rispetto all'indirizzamento di Classe B. Infatti, con 8 bit si identificano potenzialmente fino a 2^8 = 256 nodi distinti; in notazione "dot", le reti che possono essere definite in questo indirizzamento sono nell'intervallo da 192.1.1 a 223.254.254.
- Classe D: i quattro bit di peso più elevato sono posti rispettivamente a 1,1,1 e 0. Tale formato riserva 28 bit per l'identificativo di gruppo multicast[(2)]; l'indirizzo di Classe D è utilizzato per indirizzare un insieme di host che appartengono ad un gruppo multicast. Il numero di gruppi multicast distinti è

(2) Un gruppo multicast è un insieme di host caratterizzati da un unico indirizzo detto “indirizzo multicast”. L'indirizzo di multicast può essere utilizzato come indirizzo destinazione per inviare un messaggio a tutti i componenti del gruppo. Qualora, invece, si vuol comunicare simultaneamente un messaggio a tutti gli host connessi ad una stessa rete si utilizza un indirizzo di destinazione riservato detto “indirizzo di broadcast”. Questi messaggi sono chiamati “messaggi broadcast”.

potenzialmente fino a 2^{28} = 268435456; in notazione "dot", i gruppi multicast che possono essere definiti in questo indirizzamento sono nell'intervallo da 224.0.0.0 a 239.255.255.255.

- Classe E**:** i cinque bit di peso più elevato sono posti rispettivamente a 1,1,1,1 e 0. L'uso dei rimanenti bit di tale formato non sono stati definiti; in notazione "dot", il numero di valori che possono essere definiti in questo indirizzamento sono nell'intervallo da 240.0.0.0 a 247.255.255.255.

Alcuni indirizzi IP assumono una codifica particolare allorquando ad essi si voglia associare un significato specifico; la casistica possibile è riportata nella tab. 1.

3.1.3 L'instradamento in IP

La funzionalità di instradamento (*routing*) implementata nel protocollo IP prevede l'individuazione della entità di rete successiva, sia essa l'host di destinazione o un router di transito, nel percorso che il pacchetto IP deve seguire per raggiungere la destinazione finale.

Sia gli host che i router partecipano dunque all'instradamento dei pacchetti. I router si interfacciano a due o più reti e provvedono a ricevere pacchetti da una interfaccia e a rilanciarli attraverso un'altra interfaccia opportunamente individuata; gli host invece sono (generalmente) connessi ad una sola rete, per cui non trasferiscono pacchetti da una rete all'altra. E' possibile fare una prima distinzione sulla base delle reti attraversate dal pacchetto IP per raggiungere la destinazione:

- si parla di r*outing diretto* nello scambio di pacchetti tra host attestati ad una medesima rete;
- si parla di *routing indiretto* se nel trasferimento dei pacchetti a destinazione vengono interessati router IP.

In base alla distinzione ora fatta emergono due questioni. In primo luogo la modalità di riconoscimento di appartenenza alla medesima rete e in secondo luogo la determinazione della router successivo.

Ricordando il significato del **net_id** all'interno dell'indirizzo IP, per determinare l'appartenenza ad una data rete sarà sufficiente confrontare il net_id con l'indirizzo di tale rete.

Nel caso di routing indiretto, il protocollo IP ricava l'indirizzo del router successivo a cui inviare il pacchetto consultando una apposita tabella (*routing table*) contenente, per un data rete di destinazione, l'indirizzo del router successivo a cui deve essere inviato il pacchetto.

In tale tabella oltre alle corrispondenze <rete-destinazione, router successivo> viene anche indicata la distanza della destinazione espressa in una metrica specificata. Tale metrica può essere ad esempio il numero di router da attraversare, il costo dei link, l'affidabilità dei link, il tempo di attraversamento, la massima lunghezza dei pacchetti.

E' possibile inoltre specificare all'interno della routing table degli instradamenti verso specifici host, e cioè corrispondenze del tipo <host destinazione, router successivo, distanza>. Questa opportunità può essere utilizzata dal gestore a scopo di controllo e sicurezza.

Indirizzo IP			**Può apparire come**		**Descrizione**
net_id	**subnet_id** (*)	**host_id**	**Source ?**	**Dest ?**	
tutti i bit a 0	non presente	tutti i bit a 0	si	mai	Questo host in questa rete.
tutti i bit a 0	non presente	host_id	si	mai	Specifica un host in questa rete.
127	non presente	qualsiasi	si	si	Indirizzo di Loopback (**)
255	non presente	255	mai	si	Broadcast Limitato alla propria rete.
net_id	non presente	255	mai	si	Broadcast diretto alla rete net_id.
net_id	subnet_id	255	mai	si	Broadcast diretto alla rete net_id e sottorete subnet_id.
net_id	255	255	mai	si	Broadcast diretto a tutte le sottoreti della rete net_id.

(*) E' pratica comune applicare una ulteriore segmentazione della parte <host _i d> in <subnet_id, host_id>. L'introduzione di questo ulteriore livello gerarchico (subnet_id) rende ancora più flessibile lo schema di indirizzamento delle classi A, B, C, poiché consente di ottimizzare il dimensionamento di una sottorete al numero effettivo di host che la compongono. Ciò viene realizzato dall'utente introducendo un meccanismo di maschera che permette di individuare, in termini di bit, i campi <subnet_id> ed <host_id>.

(**) L'indirizzo di loopback permette di effettuare test funzionali sul proprio host.

Tabella 1 Casistica delle codifiche particolari per gli indirizzi IP

Può essere previsto l'instradamento verso un r*outer di default* per tutti quei pacchetti IP destinati a reti non previste esplicitamente nella tabella; ciò consente di ridurne le dimensioni.

Per poter consentire l'effettiva trasmissione del pacchetto IP, questo viene incapsulato in un *frame* e indirizzato sulla base dell'*indirizzo fisico* della destinazione successiva, sia essa il router di transito o la destinazione finale. Si rende pertanto necessario far corrispondere all'indirizzo logico di rete (indirizzo IP) l'indirizzo fisico.

IP ricava questa corrispondenza invocando l'ARP, che gestisce una apposita tabella denominata *ARP cache*, ossia una zona di memoria dell'host contenente le corrispondenze <indirizzo IP-indirizzo fisico> già risolte. Qualora tale corrispondenza non sia risolta, provvede a risolverla con un opportuno scambio di messaggi. Viene prevista un cancellazione periodica delle informazioni contenute nell'*ARP cache* per garantire la consistenza delle informazioni con la condizione reale della rete.

Nel processo di routing del pacchetto IP, quest'ultimo rimane inalterato nell'attraversamento della rete, mentre nel trasferimento da un router all'altro cambia l'indirizzo fisico da inserire nell'*header* del frame.

Infine, al processo di routing prende parte anche l'ICMP. Infatti oltre a notificare alla sorgente l'eventuale mancato recapito del pacchetto, il protocollo ICMP prevede un messaggio per la modifica delle informazioni contenute nella routing table. Qualora un pacchetto venga instradato erroneamente verso un router, quest'ultimo provvede ad inviare alla sorgente un opportuno messaggio ICMP di redirect, che modifica le informazioni della routing table.

Da quanto finora descritto emerge che protocollo IP si avvale di due tabelle: la routing table e l'ARP cache.

Vi sono due modi per la loro gestione:

- la *gestione statica*: prevede che le tabelle vengano costruite e gestite dal system manager mediante operazioni di management, effettuate manualmente alla console di gestione;
- la *gestione dinamica*: prevede che appositi protocolli, mediante opportuno scambio di informazioni, effettuino l'aggiornamento dinamico delle tabelle.

Di seguito viene descritto il protocollo ARP, che risolve dinamicamente le corrispondenze tra indirizzo logico di rete e indirizzo fisico.

3.1.4 Address Resolution Protocol (ARP)

Ogni host collegato ad una rete ha un numero identificativo biunivocamente legato all'hardware, chiamato indirizzo fisico (es. indirizzo Ethernet nel caso si utilizzi un supporto trasmissivo del tipo IEEE 802.3). Il protocollo ARP permette di associare dinamicamente l'indirizzo fisico all'indirizzo IP e aggiorna contemporaneamente l'ARP cache. La funzionalità del protocollo è legata unicamente alla rete a cui è collegato l'host e alla possibilità di trasmettere trame in broadcast o in multicast. La tabella svolge la funzione di cache, cioè permette di ottenere rapidamente quelle informazioni che vengano richieste ripetutamente in brevi intervalli di tempo. La cancellazione periodica delle informazioni garantisce la consistenza delle informazioni.

Si supponga, come mostrato in fig. 6, che due nodi distinti A ed B appartengano alla stessa rete e che i loro indirizzi fisici siano rispettivamente HA(A) e HA(B) e che entrambi siano dotati di un indirizzo IP: IP(A) e IP(B). Si supponga, inoltre, che il nodo A è un nodo "nuovo" della rete, cioè non conosciuto prima dagli altri nodi nella rete, e che si trovi di fronte alla necessità di inviare datagram al nodo B.

Il livello IP del nodo A conosce soltanto l'indirizzo IP del nodo B e lo comunica al driver hardware per l'invio del pacchetto. Il driver consulta le tabelle ARP in suo possesso per convertire l'indirizzo IP(B) nell'indirizzo hardware HA(B). Poiché il nodo A è nuovo non avrà questa informazione nelle sue tabelle. Il protocollo ARP, su richiesta del driver, crea al suo posto uno specifico pacchetto di richiesta che contiene informazioni per l'identificazione univoca a livello IP del nodo di destinazione.

Sfruttando la funzione di broadcasting, il driver invierà la richiesta ARP a tutti i nodi situati in rete, i quali sono in ascolto per verificare se il pacchetto è indirizzato a loro oppure no.

Il nodo B, verificato che il pacchetto ARP contiene il suo indirizzo IP, estrarrà da tale pacchetto la coppia di indirizzi del nodo A (IP(A) ed HA(A)), con i quali aggiorna le sue tabelle ARP, e costruisce un pacchetto ARP di risposta che invia direttamente al nodo A.

Il nodo B è, a questo punto, in grado di indirizzare correttamente il nodo A ma non è vero il viceversa, ossia A non conosce ancora come indirizzare il nodo B.

Tale lacuna viene colmata alla ricezione da parte del nodo A del pacchetto inviato da B, il quale trasporta le informazioni necessarie per mappare l'indirizzo IP di B in quello hardware (IP(B) ed HA(B)).

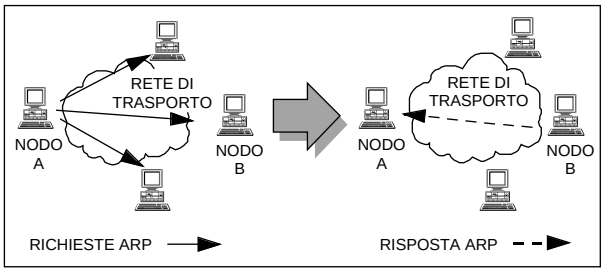


Figura 6 Funzionamento dell'ARP

I successivi tentativi da parte del nodo A di inviare datagram al nodo B, e viceversa, andranno a buon fine grazie all'aggiornamento delle tabelle ARP che il processo descritto sopra ha generato.

3.1.5 Formato del pacchetto ARP

Di seguito vengono descritti i campi costituenti il messaggio utilizzato dall'ARP, rappresentato in fig. 7:

- **Hardware Type (16 bit)**: specifica il tipo di interfaccia hardware di cui il mittente sta richiedendo l'indirizzo (ad es. Ethernet, Token Ring);
- **Protocol Type**: specifica il tipo dell'indirizzo di protocollo di livello superiore che il mittente ha fornito;
- **HLEN (8 bit)**: indica la lunghezza dell'indirizzo fisico in byte;
- **PLEN (8 bit)**: indica la lunghezza dell'indirizzo Internet in byte;
- **Operation (16 bit)**: permette di distinguere tra richiesta ARP, risposta ARP, richiesta RARP, risposta RARP;
- **Sender HA**: contiene l'indirizzo fisico del mittente;
- **Sender IP**: contiene l'indirizzo Internet del mittente;
- **Target HA**: contiene l'indirizzo fisico della stazione ricercata;
- **Target IP**: contiene l'indirizzo Internet della stazione ricercata.

3.1.6 Reverse Address Resolution Protocol (RARP)

Il protocollo RARP è utilizzato da host connessi in rete e privi di sistemi di memorizzazione di massa (hard disk) per determinare, durante la fase di inizializzazione, il proprio indirizzo IP a partire dal indirizzo fisico MAC (Medium Access Control) della sua interfaccia di rete. Il protocollo assume che in rete siano presenti uno o più *RARP server* a cui inviare la richiesta RARP per conoscere l'indirizzo IP associato al suo indirizzo fisico.

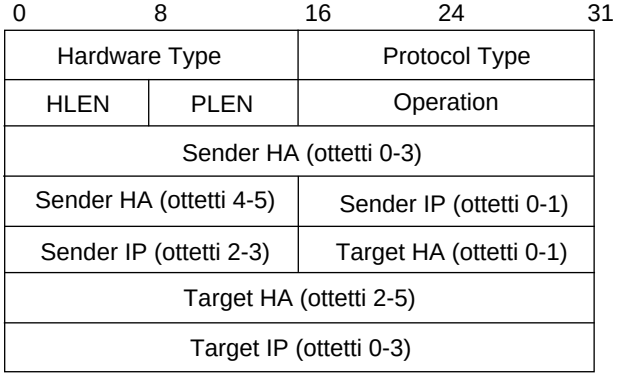


Figura 7 Formato del messaggio ARP

Tale procedura è realizzata inviando il pacchetto RARP, identico a quello del protocollo ARP, in broadcast nella propria sottorete. In tale pacchetto RARP l'indirizzo sorgente è posto uguale all'indirizzo fisico dell'host. Il RARP server, una volta ricevuto il pacchetto, risponde inviando indietro un pacchetto in cui è presente l'indirizzo IP cercato.

3.1.7 Modalità di segmentazione

La segmentazione di un datagramma IP si rende necessaria quando anche solo una delle reti attraversate ha una MTU (Message Transfer Unit) inferiore alla dimensione del datagramma; in questo caso i pacchetti IP eventualmente contrassegnati come non segmentabili verranno persi poiché incompatibili con la capacità di trasporto della rete e verrà generato un messaggio ICMP.

Le procedure di segmentazione e riassemblaggio devono essere in grado di frammentare il pacchetto originario in un numero arbitrario di unità che, giunte a destinazione, devono poter essere ricomposte nella forma originaria. Il destinatario utilizzerà il campo *"Identification"* del frammento per garantire che pacchetti originatisi da processi diversi non siano confusi tra loro. Tale campo identificativo, che sarà univoco per tutti i processi operanti in quel momento tra unità sorgenti e unità remote, sarà assegnato al datagramma trasmesso dalla sorgente.

Ad ogni frammento è inoltre assegnato un campo *offset* (Fragment Offset) che permette al destinatario di risalire alla posizione occupata nel datagram originario. Il frammento dotato del Flag MF posizionato a zero indica che è l'ultimo segmento del pacchetto.

Per riassemblare i frammenti di un datagramma, la destinazione combinerà insieme i frammenti che hanno le stesse informazioni relative all'identificazione, alla sorgente, al destinatario e al tipo di protocollo. Ciò viene fatto ponendo la parte dati del frammento nella posizione indicata dal campo offset che si trova nell'intestazione dello stesso frammento.

3.1.8 Funzionalità di controllo: ICMP

La funzionalità di comunicazione di situazioni anomale alla sorgente è realizzato tramite il protocollo ICMP (Internet Control Message Protocol). Esso è parte integrante del protocollo IP, anche se è collocato logicamente, nella pila protocollare, in posizione superiore.

Fa parte integrante dell'IP, anche se eseguito da un apposito protocollo collocato logicamente in posizione superiore, la funzionalità di comunicazione di situazioni anomale alla sorgente. Di ciò si occupa il protocollo ICMP.

In caso di malfunzionamento della rete, il protocollo di controllo provvede a uno scambio di messaggi fra le macchine in rete per notificare l'errore o indicare le circostanze inaspettate che causano il comportamento anomalo del sistema.

Non vengono previste ulteriori azioni.

Il meccanismo consiste in uno scambio di messaggi utilizzando il supporto fornito da IP come un protocollo di livello inferiore. Viene di seguito riportato l'elenco dei tipi di messaggio:

- messaggio "Echo Reply": verifica il buon funzionamento di una certa destinazione, sia essa un host o un router;
- messaggio "Destination Unreachable": viene emesso da un router per notificare alla sorgente del datagramma che non è in grado di instradarlo;
- messaggio "Source Quench": l'host destinazione informa l'host sorgente che il traffico generato è superiore alle sue capacità ricettive con conseguenti perdite. La sorgente provvede a ridurre il numero di datagrammi inoltrati in rete;
- messaggio "Redirect": un router può informare l'host sorgente che l'instradamento prescelto non è il migliore e ne notifica uno nuovo;
- messaggio "Time Exceeded for a Datagram": viene in questo modo informato l'host sorgente che un datagramma IP è stato eliminato dalla rete per aver superato il limite temporale di esistenza nella rete;
- messaggio "Parameter Problem on Datagram": viene generato quando vengono rivelati errori nell'header di un datagramma per informare la relativa sorgente;
- messaggi "Timestamp Request" e "Timestamp Replay": il primo viene inviato da una sorgente per richiedere alla destinazione l'ora, il secondo viene utilizzato per la risposta;
- messaggi "Information Request" e "Information Reply": vengono utilizzati dalle macchine per ottenere l'indirizzo Internet delle reti a cui sono collegate, in alternativa al RARP;
- messaggi "Address Mask Request" e "Address Mask Reply": vengono utilizzati dalle macchine per ottenere la parte dell'indirizzo Internet che riguarda la rete e la rete di appartenenza.

Il messaggio ICMP viene incapsulato all'interno del campo dati del datagramma IP, così come mostrato in fig. 8.

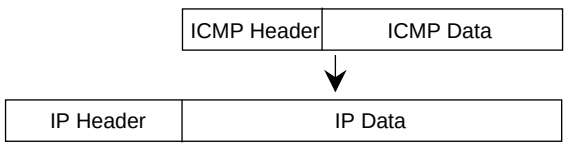


Figura 8 Incapsulamento di un datagramma ICMP in un datagramma IP

3.2 *Il protocollo TCP*

Il TCP è stato progettato al fine di offrire un servizio end-to-end affidabile alle applicazioni che lo utilizzano, tenendo conto che la sottorete potrebbe non garantire la consegna dei pacchetti dati. E' un protocollo di trasporto di tipo connection oriented e quindi realizza una prima fase di instaurazione della connessione prima di inviare i messaggi all'entità destinataria di stesso livello.

TCP accetta dal livello superiore messaggi senza vincoli sulla loro lunghezza, li frammenta in pacchetti di piccole dimensioni e li invia in datagrammi distinti. Ogni trasmissione di dati deve essere preceduta da una fase di attivazione della connessione e seguita da una fase di rilascio. Poiché IP offre un servizio di consegna non garantito, TCP deve verificare la corretta ricezione dei datagrammi ed, eventualmente, attuare le procedure necessarie alla ritrasmissione. Inoltre è compito del TCP verificare che i datagrammi giungano a destinazione nella stessa sequenza con cui sono stati trasmessi, che non vi siano duplicati o datagrammi mancanti. Queste funzionalità vengono garantite mediante la numerazione dei datagrammi e l'invio di messaggi di "acknowledgment" da parte della destinazione ogni qual volta viene ricevuto correttamente il giusto datagramma della sequenza.

TCP offre anche meccanismi di controllo di flusso in grado impedire il sovraccarico della rete che può causare situazioni di congestione.

Evidentemente tutte le funzionalità aggiunte dal TCP hanno un costo in termini di ritardo di trasmissione e di overhead per la rete, che ne impediscono l'applicazione in reti ad alta velocità. Nel seguito vengono descritti i meccanismi sin qui elencati e vengono illustrate alcune possibili estensioni per l'applicazione a reti ad alta velocità.

3.2.1 Il formato del pacchetto TCP

Il pacchetto può essere scisso in due parti: i dati e l'header (vedi fig. 9). Nello spazio dedicato ai dati viene posta una porzione dell'insieme delle informazioni che, di volta in volta, il livello applicativo offre al protocollo TCP per la trasmissione. La suddetta porzione è comunemente chiamata s*egmento*.

L'header di TCP, invece, è costituito da un certo numero di campi:

- *Source Port (16 bit)*: definisce l'indirizzo logico del processo sorgente dei dati.
- *Destination Port (16 bit)*: definisce l'indirizzo logico del processo destinatario dei dati.
- *Sequence Number (32 bit)*: contiene il numero di sequenza del primo byte di dati contenuti nel segmento.
- *Acknowledgement Number (32 bit)*: nei pacchetti in cui il bit ACK, presentato più avanti, è settato a uno,

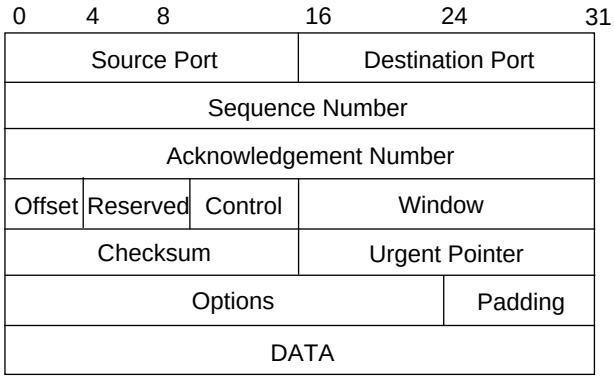


Figura 9 Formato del pacchetto TCP

contiene il numero di sequenza del prossimo byte che il trasmettitore del segmento si aspetta di ricevere. Come si può intuire nel caso di connessioni interattive bidirezionali avviene il *piggybacking* degli acknowledgement (nel senso che si utilizzano i segmenti di risposta per inviare gli Ack al trasmettitore senza dover inviare dei segmenti appositi).

- *Offset (4 bit)*: contiene il numero di parole di 32 bit contenute nell'header di TCP. L'header di TCP non supera quindi i 60 byte ed inoltre è sempre costituito da un numero di bit multiplo di 32.
- *Reserved (6 bit)*: riservato per usi futuri per ora contiene degli zeri.
- *Control bit (6 bit)*: i bit di controllo sono:
  - URG: viene settato a uno quando il campo urgent pointer contiene un valore significativo;
  - ACK: viene settato a uno quando il campo Acknowledgement Number contiene un valore significativo;
  - PSH: viene settato a uno quando l'applicazione esige che i dati forniti vengano trasmessi e consegnati all'applicazione ricevente prescindendo dal riempimento dei buffer allocati fra applicazione e TCP e viceversa (solitamente infatti è il riempimento dei suddetti buffer che scandisce la trasmissione e la consegna dei dati);
  - RST: viene settato a uno quando un malfunzionamento impone il reset della connessione;
  - SYN: viene settato a uno solo nel primo pacchetto inviato durante il 3-way handshaking (fase di sincronizzazione fra le entità TCP);
  - FIN: viene settato a uno quando la sorgente ha esaurito i dati da trasmettere.
- *Window (16 bit)*: contiene il numero di byte che, a cominciare dal numero contenuto nel campo Acknowledgement Number, il trasmettitore del segmento è in grado di ricevere. É utile notare fin d'ora che il controllo di flusso è orientato al byte.
- *Checksum (16 bit)*: contiene la sequenza che permette al TCP ricevente di verificare la correttezza del pacchetto.
- *Urgent Pointer (16 bit)*: contiene il numero di sequenza del byte che delimita superiormente i dati che devono essere consegnati urgentemente al processo ricevente. Tipicamente sono messaggi di controllo che esulano dalla comunicazione in senso stretto. A tale traffico ci si riferisce di solito con il nome di *out-of-band*.
- *Options (di lunghezza variabile)*: sono presenti solo raramente: le più note sono End of option List, No-operation e Maximum Segment Size (più brevemente MSS). Ci si soffermerà, in seguito, solo sull'ultima opzione citata.
- *Padding (di lunghezza variabile)*: contiene sempre degli zeri. Serve come riempitivo aggiunto per far sì che l'header abbia una lunghezza multipla di 32 bit.

3.2.2 Il meccanismo 3-way handshaking

Il protocollo TCP è un protocollo di tipo connection-oriented. Questo significa che prima di intraprendere un qualsiasi trasferimento di informazioni, verso un utente remoto, esso deve instaurare una connessione con l'interlocutore in questione. Le due entità TCP interagenti si sincronizzano scambiandosi il proprio numero di sequenza iniziale, che rappresenta il numero a partire dal quale tutti i byte trasmessi, una volta instaurata la connessione, saranno sequenzialmente numerati.

In particolare quando deve essere instaurata una connessione (vedi fig. 10) fra un processo applicativo ULP[3] A, residente nella stazione A, ed un ULP B, residente nella stazione remota B, il primo passo che si deve compiere è l'invio di una *active open*, primitiva di Service-Request, da parte dell'ULP A al TCP A, con la quale quest'ultimo viene messo al corrente di tale desiderio. Il TCP A risponde ad ULP A tramite la primitiva *open id*, primitiva di Service-Response, ed avvia il meccanismo *3-way handshaking* inviando, al TCP ricevente, un SYN-segment con il bit SYN settato a uno e con l'*Initial Sequence Number* (ISN), all'interno del campo Sequence Number, definito da un generatore a 32 bit, il quale si incrementa ogni 4 microsecondi.

Alla ricezione di tale segmento, il TCP B risponderà, nel caso si trovi nelle condizioni di poter accettare la connessione (cioè nel caso abbia ricevuto precedentemente una *passive open* dall'ULP B), con un segmento in cui sono settati a uno il bit SYN ed il bit ACK ed in cui sono definiti l'ISN per il ricevitore e, nel campo Ack number, il numero di sequenza del primo byte di informazioni atteso in ricezione (fig. 9). La procedura si conclude con l'invio da parte del TCP A, una volta ricevuto il SYN-segment, di un ACK-segment.

(3) Upper Layer Protocol.

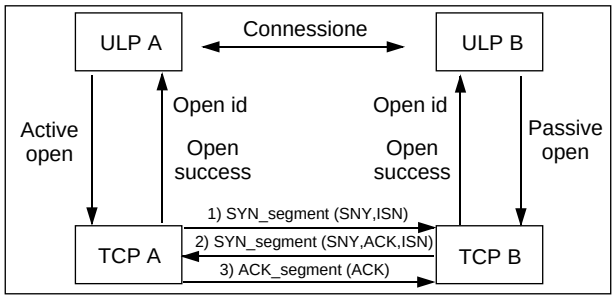


Figura 10 Fasi del meccanismo 3-way handshaking

La ricezione del SYN-segment permetterà al TCP A di informare il proprio ULP A, mediante una primitiva op*en success*, che la connessione è attivata. D'altro canto, il TCP B completerà la procedura informando il proprio ULP della ricezione dell'ACK-segment conclusivo.

A proposito del generatore di numeri di sequenza osserviamo che, aggiornandosi ogni 4 microsecondi, esso presenterà lo stesso numero ogni 4,7 ore circa. Per evitare la possibilità di confondere i numeri di sequenza dei segmenti è sufficiente fare in modo che il pacchetto rimanga in rete per un tempo limitato. A questo scopo, come abbiamo già osservato, esiste il campo Time-To-Live (TTL) del protocollo IP che assicura un'adeguata protezione: ogni router IP dopo aver "processato" un datagramma ne aggiorna il campo TTL e se il valore contenuto si annulla il pacchetto viene scartato.

3.2.3 **M**aximum Segment Size

Nel momento in cui il TCP trasmittente invia il primo segmento (SYN) per instaurare la connessione con un TCP remoto, esso può inserire in tale segmento un'informazione particolare che rappresenta la massima lunghezza di segmento che è in grado di trattare.

La scelta della massima lunghezza dipende da due fattori: dalla lunghezza del buffer a disposizione e dalla MTU, resa nota al TCP dal driver che lo interfaccia alla rete.

Il calcolo della Maximum Segment Size (MSS) viene effettuato sottraendo alla MTU la lunghezza degli header introdotti dal formato IP e TCP. Nel calcolo un problema potrebbe essere rappresentato dalla variabilità della lunghezza degli header di IP e TCP ma in realtà, fatta eccezione per i rari casi in cui si usano le opzioni, si avrà sempre a che fare con entrambi gli header aventi lunghezza pari a 20 byte.

La scelta della MSS usata nel trasferimento dei dati viene fatta, di solito, scegliendo il minimo fra la MSS offerta dal TCP remoto e quella a disposizione. Esistono però dei casi in cui questo non succede per cui il protocollo IP opera una frammentazione alla sorgente. Nel caso in cui l'opzione non venga utilizzata si impone l'uso, per default, di una MSS pari a 536 byte.

Oggi, con i servizi dati a disposizione, si va delineando la possibilità di usare valori molto più grandi della MSS. In passato si raccomandava, nel caso di stazioni interfacciate su reti differenti interconnesse da router o bridge, di usare una MSS di 536 byte; attualmente si usano valori anche molto più grandi e sono state fatte proposte che nel contesto di una rete locale come FDDI porterebbero alla possibilità di usare MSS di 4096 byte.

3.2.4 Stima del Round Trip Time medio e calcolo del Retransmission Timeout

Il Round Trip Time è il tempo che intercorre fra l'istante in cui si inizia la trasmissione di un segmento e l'istante in cui se ne riceve l'Acknowledgement.

Il Retransmission Timeout è il timeout alla cui scadenza si inizia una ritrasmissione. Ogni qualvolta il TCP trasmette un segmento ne memorizza il suo istante di partenza in un buffer dedicato a contenere le informazioni di gestione della connessione. Alla ricezione di un Acknowledgement, che informi della corretta ricezione del medesimo, si ha quindi a disposizione un campione di intervallo di tempo necessario a trasferire i dati al ricevitore. Il trasmettitore, man mano che raccoglie i campioni, può stimare il Round Trip Time medio operando una media, chiamata "running average".

Due fattori hanno un certo peso in questa procedura: il numero di campioni coinvolti nel meccanismo a finestra per il controllo di flusso e l'algoritmo di calcolo (di entrambi si parlerà nei prossimi paragrafi). Le prime implementazioni utilizzavano un solo campione per finestra, ma ovviamente questo modo di procedere aveva dei grossi limiti. Le implementazioni più recenti usano le stime legate a ciascun segmento, trascurando però i campioni relativi ai segmenti che abbiano subito ritrasmissioni perché, relativamente ad essi, non si è in grado di stabilire a quale delle trasmissioni l'Ack si riferisca. Infatti dopo aver trasmesso e ritrasmesso lo stesso segmento, nel momento in cui se ne riceve l'Ack non si è in grado di capire a quale dei due pacchetti tale Ack si riferisca e quindi in tale caso la misura risulta inaffidabile.

3.2.5 Le ritrasmissioni e gli acknowledgement

Quando al ricevitore pervengono dei dati corretti ed in sequenza esso deve prontamente inviare un riscontro dell'avvenuta ricezione o "acknowledgment" (Ack), al fine di evitare ritrasmissioni inutili e di sollecitare l'invio di nuovi dati.

Se la comunicazione è bidirezionale, è possibile effettuare il piggybacking(4) degli Ack settando a uno il bit ACK del segmento usato per rispondere e ponendo nel campo Ack number dello stesso il numero del prossimo pacchetto atteso. Nel caso di bulk transfer unidirezionali, si rende necessaria invece la trasmissione di Ack-segment appositi. Il limite di questa tecnica sta nel fatto che il ricevitore, per ora, non ha alcuna possibilità di informare il trasmettitore delle eventuali ricezioni corrette di segmenti avvenute fuori sequenza.

Se il trasmettitore non riceve l'Ack di un segmento prima che sia scaduto il relativo timeout si rende necessaria la ritrasmissione. La peculiarità di questa procedura sta nel fatto che il Retransmission Timeout del segmento viene raddoppiato ad ogni ritrasmissione del medesimo fino al raggiungimento di un fattore moltiplicativo pari a 64 (ottenuto alla settima trasmissione). Si prosegue, quindi, con il fattore moltiplicativo suddetto fino al numero limite di ritrasmissioni permesso dall'implementazione, oltre il quale la connessione abortisce. Tipicamente viene adottato un numero massimo di ritrasmissioni pari a dodici o tredici.

3.2.6 Il controllo di flusso e il controllo della congestione

Il controllo di flusso, in questo contesto, è una procedura tra la sorgente ed il destinatario delle informazioni inteso a limitare, in funzione delle risorse a disposizione, il flusso dei dati, prescindendo dal traffico presente nella rete. Lo scopo del controllo di flusso è assicurare la ricezione del destinatario di uno o più pacchetti trasmessi dalla sorgente.

Il controllo della congestione ha invece lo scopo di recuperare situazioni di sovraccarico nella rete.

Il controllo di flusso nel protocollo TCP è implementato mediante un meccanismo a finestra di tipo *sliding window* (vedi fig. 11) orientato al byte, nel senso che la finestra rappresenta istante per istante il numero massimo di byte che possono essere trasmessi verso il destinatario. La finestra offerta dal ricevitore, Ad*vertised window*, rappresenta la disponibilità di buffer in ricezione. Il ricevitore la può variare, informando il trasmettitore, avendo presente però che quest'ultimo la modificherà solo dopo aver ricevuto dati, corretti ed in sequenza, che abbiano "riempito" le finestre precedentemente offerte.

(4) Con PiggyBack si identifica quella tecnica per l'invio di una informazione di controllo (ad es. ACK) ad un destinatario congiuntamente all'invio del prossimo messaggio diretto a questo così da inglobare l'informazione in esso. Ciò evita di sovraccaricare la rete con messaggi contenenti solo informazioni di controllo.

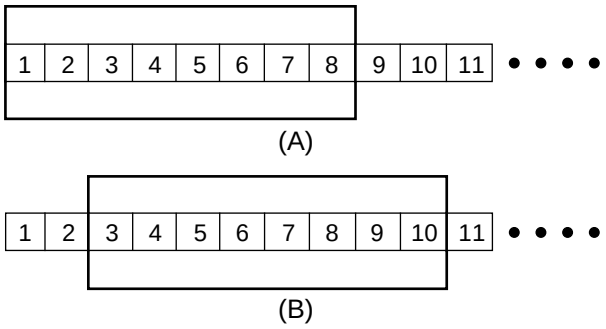


Figura 11 Il meccanismo Sliding window a 8 byte

Le prime implementazioni di TCP utilizzavano la Advertised window ed una politica di tipo *go-back-n* per quanto concerneva il recupero degli errori, ossia permettevano al ricevitore di rifiutare i segmenti non compresi nella sequenza definita dalla finestra. Ad esempio, nel caso (A) della fig. 11, i pacchetti che possono essere ricevuti dal destinatario sono quelli etichettati da 1 a 8 e quelli da 9 in poi sono considerati fuori sequenza; nel caso (B) invece i pacchetti etichettati con 1 e 2 sono stati riconosciuti e la finestra si è spostata in avanti potendo accettare ora i pacchetti con etichetta 9 e 10 ma non quelli da 11 in poi. Parallelamente a questo esse utilizzavano un meccanismo di ritrasmissione a intervalli crescenti ed il protocollo ICMP per il controllo della congestione. Il modulo ICMP, tuttora utilizzato, è implementato nei nodi terminali e nei nodi di interconnessione. Il suo scopo è quello di rallentare il ritmo di trasmissione della stazione, mediante l'invio di messaggi opportuni (Source Quence etc.), nel momento in cui il nodo si trovi a dover rifiutare pacchetti a causa della mancanza di risorse di memoria di ricezione.

Questi meccanismi, nel caso di rapide variazioni del traffico, sembrarono del tutto insufficienti nel contesto di reti ad alta velocità (LAN). Diversi studi presentano così, fin dalla seconda metà degli anni '80, l'intento di implementare un controllo della congestione basata sui Timeout e che prescinda da ICMP. In sostanza si considera lo scadere di un Timeout come un sintomo di congestione delle risorse di interconnessione. Ad esempio, l'algoritmo CUTE (Congestion control Using Timeouts of the End-to-end layer) prende in considerazione il valore della finestra a disposizione del trasmettitore non pari al valore della Advertised window ma bensì variabile tra un minimo e un massimo. Tale algoritmo utilizza i seguenti parametri:

- Massimo: è il valore massimo della finestra in trasmissione: in generale è pari al valore della Advertised window offerta dal ricevitore;
- Minimo: è il valore minimo della finestra: tipicamente pari alla lunghezza di una MSS;
- Inizializzazione: rappresenta il valore iniziale della finestra: su reti molto cariche è preferibile partire dal valore minimo;

- Incremento: si incrementa la finestra in trasmissione della lunghezza pari ad un segmento (senza però mai superare il valore massimo) ogni N segmenti ricevuti corretti dal TCP remoto;
- Decremento: il decremento può essere di tipo
  – *sudden,* se allo scadere di un Timeout la finestra è posta uguale al valore minimo;
  – *gradual,* se la si riduce di un segmento;
  – *binary,* se la si divide a metà.

3.2.7 La Sindrome Silly Window

Questo fenomeno si manifesta nel momento in cui il ricevitore rende disponibile al trasmettitore una finestra di dimensioni ridotte cosicché possono essere inviati dalla sorgente al destinatario solo una limitata quantità di dati oppure è il trasmettitore che invia piccole quantità di dati invece di attendere segmenti di dimensioni maggiori. Questo comportamento può causare una diminuzione dell'"Average Segment Size" ed una diminuzione della efficienza di trasmissione, in quanto il trasmettitore invia segmenti brevi.

Nelle implementazioni più recenti di TCP sono stati adottati dei semplici algoritmi che permettono di proteggersi da questa sindrome. Un metodo, per esempio, prevede che il ricevitore ritardi l'aggiornamento della finestra finché questa non abbia raggiunto una dimensione almeno pari ad una MSS, mentre un altro prevede che il trasmettitore eviti di trasmettere nuovi segmenti se la finestra disponibile non raggiunge almeno la dimensione di una MSS.

3.2.8 Le estensioni proposte ed i problemi di compatibilità

Le estensioni che sono state proposte per TCP sono funzionali ad un suo migliore comportamento in reti ad alta velocità. Il protocollo TCP, infatti, così come si presenta nella sua implementazione tradizionale, non sembra poter offrire prestazioni soddisfacenti. In quest'ottica sono state studiate alcune modifiche orientate a migliorare l'efficienza del protocollo in questo nuovo scenario.

Innanzitutto viene osservato che, in reti a fibra ottica, aumenta il prodotto definito da banda per ritardo di propagazione. Quindi un primo passo per migliorare le prestazioni è senz'altro quello di offrire al protocollo la possibilità di operare con una finestra massima più ampia di quella che per ora ha a disposizione: i 16 bit a disposizione attualmente limitano la ampiezza della finestra a 65 Kbyte mentre, per aumentare il "pipeline" dei dati, sarebbe opportuno avere la possibilità di lavorare almeno con 32 bit. A questo scopo è stata introdotta la "Window Scale Option" che permette alle due stazioni TCP di accordarsi, durante il 3-way handshake, sulla massima dimensione.

Il più grosso limite delle implementazioni tradizionali di TCP risiede, però, nella scarsa accuratezza con cui viene stimato il Round Trip Time dei segmenti trasmessi. Si pensi che in certi casi si basa l'aggiornamento della stima solo su un pacchetto per finestra, creando quindi notevoli problemi di "aliasing" dovuti alla stima ed alla incapacità di adattarsi a cambiamenti nell'andamento del traffico. L'inadeguatezza di questo metodo cresce inoltre al crescere delle dimensioni della finestra disponibile. D'altra parte anche se si aumenta il campionamento rimane sempre il problema della incapacità di stimare il Round Trip Time nel caso in cui avvengano delle ritrasmissioni di un segmento: questo perché quando si riceve l'Ack di un segmento ritrasmesso non si è in grado di capire a quale dei segmenti si riferisca.

Una soluzione a questi problemi viene offerta dalla "Timestamp Option". Usando tale opzione il trasmettitore scrive l'istante di partenza su ogni segmento trasmesso in modo tale che il ricevitore, quando ne invia l'Ack, possa scrivere sul segmento di risposta l'istante di partenza del pacchetto a cui si riferisce. In trasmissione basterà operare una semplice differenza per avere una accurata misurazione del Round Trip Time.

Un'ulteriore limitazione alle prestazioni del TCP risiede nel fatto che il ricevitore non ha la possibilità di informare il trasmettitore della eventuale corretta ricezione di pacchetti che siano giunti a destinazione fuori sequenza. Per superare tale problema è stata definita la opzione SACK (Selective Acknowledgement).

Il problema principale che le nuove implementazioni di TCP incontreranno sarà quello relativo alla compatibilità nelle interazioni con le versioni tradizionali. Tutte le implementazioni dovrebbero (almeno ci si aspetta che la maggior parte si comporti così) ignorare le opzioni sconosciute che vengono inserite nei SYN-segment utilizzati durante il set-up della connessione. D'altro canto, però, è possibile che talune implementazioni di TCP diano luogo a malfunzionamenti nel momento in cui ricevono un segmento, diverso dal SYN, contenente opzioni sconosciute. Quindi la soluzione prospettata è quella di usufruire delle estensioni al protocollo, in segmenti diversi dal SYN, solo se lo scambio di opzioni, durante il 3-way handshake, ha indicato che entrambe le implementazioni sono in grado di comprendere le estensioni.

3.3 *Il protocollo UDP*

User Datagram Protocol

UDP è un protocollo di livello *trasporto* che utilizza i servizi di offerti da IP per il trasferimento di messaggi di lunghezza variabile. Tali messaggi vengono denominati *datagrammi utente* e il loro formato è mostrato in fig. 12.

Source port	Destination port
Lunghezza	Checksum
Dati	
....	

Figura 12 Formato dei messaggi scambiati da UDP

Il protocollo UDP offre ai livelli superiori un servizio datagramma non affidabile e il suo impiego consente lo scambio di informazioni tra più sorgenti/destinazioni all'interno di uno stesso host attraverso un unico canale IP di comunicazione. Infatti i processi all'interno di un host sono identificati da un indirizzo di porta e pertanto se l'indirizzo IP identifica un dato host, l'indirizzo di porta identifica un dato processo su quell'host.

I campi sourc*e port* e *destination port* contengono l'indirizzo di porta in base al quale multiplare i messaggi tra i processi che, all'interno dell'host, sono in attesa per lo scambio di dati. L'impiego del campo source port è opzionale.

Il campo *lunghezza* esprime la lunghezza del datagramma UDP espressa in ottetti, incluso l'header.

Il checksum del messaggio è opzionale al fine di ridurre, se necessario, il carico elaborativo per la trasmissione e la ricezione di un messaggio UDP. Infatti se si prevede l'impiego di reti altamente affidabili, la verifica di errori non è necessaria.

E' necessario però osservare che IP non effettua alcun controllo di errore sul suo payload, per cui il checksum di UDP costituisce l'unico strumento per verificare che i dati siano giunti a destinazione correttamente. Qualora poi venga impiegato, tale controllo riguarda non solo il datagramma UDP, ma anche uno p*seudo header*, come mostrato in fig. 13, che viene considerato al solo fine del calcolo del checksum, che risulta costituito dagli indirizzi IP della sorgente e della destinazione (contenuti nell'header del pacchetto IP) , dal numero di protocollo corrispondente all'UDP, dalla lunghezza del messaggio UDP e da un byte di padding per fare in modo che la lunghezza complessiva sia multipla di 16 bit.

Il motivo per cui viene considerato un tale pseudo header è quello di verificare che il messaggio UDP abbia raggiunto la destinazione corretta, ossia il protocollo corretto sull'host corretto.

Indirizzo IP sorgente		
Indirizzo IP destinazione		
Padding	Protocollo	Lunghezza messaggio UDP

Figura 13 Il formato dello *pseudo header*

4. Conclusioni

La sempre più crescente diffusione dei sistemi di elaborazione personale e la riorganizzazione dei processi aziendali verso lo sviluppo di metodologie di lavoro cooperativo delineano uno scenario in cui l'interazione a distanza assume un ruolo centrale nello sviluppo della comunicazione, coinvolgendo aspetti non solo legati alla comunicazione vocale o visiva, ma anche multimediale e interattiva, assumendo infine il carattere di comunicazione globale.

In tale prospettiva, soluzioni tecnologiche quali l'architettura TCP/IP, nate in contesti particolari quali quelli della ricerca e militari, assumono un carattere completamente nuovo e rivoluzionario, ponendosi come fattori abilitanti alla comunicazione universale multimediale.

L'enfasi posta in questo articolo su TCP/IP cerca di definire e chiarire gli elementi tecnici che sono alla base di tale impostazione, illustrando quali meccanismi possono essere utilizzati per creare una piattaforma di comunicazione che sia in grado di integrare tecnologie e metodologie differenti. In seguito potranno essere delineate le possibili soluzioni tecnologiche per la realizzazione su rete pubblica, in particolare basata ATM, di tale piattaforma.

Bibliografia

[1] Carissimi, M.; Guadagni, F.; Pugliese, F.: *La rete Internet: ambiente di comunicazione e servizi*. «Notiziario Tecnico Telecom Italia», Vol. 3, n. 2, Agosto 1994, pp. 40-53.

[2] Comer, D.: *Internetworking with TCP/IP, Volume I - Principles, Protocols Architecture*. Prentice Hall, 1991.

[3] Stevens, W.R.: *TCP/IP Illustrated, Volume 1, The Protocols*. Addison Wesley, 1994.

[4] Comer, D.; Stevens, D.: *Internetworking with TCP/IP, Volume II - Design, Implementetion and Internals*. Prentice Hall, 1994.

[5] Comer, D.; Stevens, D.: *Internetworking with TCP/IP, Volume III - Client-Server Programming and Applications for the AT&T TLI Version*. Prentice Hall, 1994.

[6] Comer, D.; Stevens, D.: *Internetworking with TCP/IP, Volume III - Client-Server Programming and Applications for the BSD Socker Version*. Prentice Hall, 1993.

[7] Perretti, E.; Antonelli, F.; Iuso, F.; Montechiarini, C.; Pugliese, F.; Carissimi, M. et altri: *TELECOM ITALIA, "Progetto IP" Studio di fattibilità del servizio IP su rete pubblica e sperimentazione tecnologica*. Settembre 1994.

La qualità della trasmissione telefonica - Parte seconda

L. Bonavoglia (*)

Questa è la seconda parte di un lavoro del quale la prima parte è stata pubblicata nel n. 2, Anno 3 del Notiziario Tecnico Telecom Italia.

Nella prima parte si è cercato di dare un quadro dei metodi in uso per determinare la qualità trasmissiva, e di mettere in evidenza i pericoli che si possono correre utilizzando in futuro alcuni criteri di valutazione, come quelli esclusivamente basati sulla comprensione del parlato.

In questa seconda parte si cerca di determinare l'influenza che le code del circuito numerico, fra utenti e centrali terminali, hanno sulla qualità globale di un collegamento telefonico; si analizzano a tal fine le diverse situazioni (e sono in numero non piccolo), causate dalla diversità di realizzazione delle code di utente.

Il lavoro si chiude con un veloce esame di quanto si sta facendo in questo settore della qualità della voce da parte di ITU-T (ex CCITT) ed ETSI. L'esame è veloce perché non si fa molto: purtuttavia, qualcosa comincia a muoversi, e occorre spingere perché il moto si acceleri.

1. **Premessa**

Nel numero 2 dell'Agosto dello scorso anno del "Notiziario" è apparsa la prima Parte di questo lavoro, che illustrava le ragioni per cui oggi sembra opportuno parlare di qualità della trasmissione telefonica e dei metodi per la sua valutazione usati nel passato recente e quelli ormai dominanti oggi.

Nella descrizione generale dell'evoluzione dei metodi passati e odierni si è dato risalto a due dei parametri trasmissivi, e cioè alla attenuazione del segnale sonoro (dalla bocca all'orecchio dei due utilizzatori del circuito), ed al rumore presente in ricezione, da qualunque fonte esso derivi.

Si sono soltanto menzionati altri parametri trasmissivi come l'effetto locale, i suoi effetti sull'intensità sonora del parlatore e sul rumore captato dall'ambiente, il ritardo di trasmissione e i suoi effetti dannosi come l'eco e il disturbo psicologico che nasce perché chi smette di parlare deve attendere un tempo (doppio del ritardo) per percepire la risposta dell'interlocutore e così via; diversi altri parametri si sono trascurati perché col tempo sono divenuti oggi meno importanti.

Per evitare a chi legge il fastidio di andare a cercare, sui libri specialistici, gli elementi necessari a una buona conoscenza degli effetti che i più importanti parametri hanno sulla qualità, si è pensato di aggiungere a questa parte qualche Appendice; la prima su uno dei metodi per il trattamento del segnale vocale (il "Vocoder"), una seconda sugli effetti del ritardo trasmissivo, ed una sul come si determina il rumore globale in ricezione.

Un altro punto, su cui ci si è soffermati poco, è invece così importante per la sua incidenza sui metodi di valutazione della qualità, che abbiamo pensato di premettere un certo approfondimento a questa seconda parte del lavoro.

Il lettore, si sarà chiesto perché si sia accennato in diversi punti della prima parte al fatto che il metodo oggi più usato è basato su un modello matematico, derivato da una larga messe di dati sperimentali sulla facilità di una conversazione; questi dati sono in pratica i giudizi ottenuti da soggetti sulla *comprensibilità* del parlato, al termine di un collegamento telefonico, in relazione alle caratteristiche trasmissive del circuito stesso. Il modello matematico, cioè, fornisce una misura della qualità trasmissiva, quasi esclusivamente dal punto di vista della *comprensibilità*, in pratica della *facilità* di conversazione.

Allora fu appena adombrato il fatto che un giudizio sulla sola comprensibilità non è completo, perché è già in atto per alcuni tipi di servizio telefonico un trattamento del segnale vocale (a volte pesante) che riduce l'ingombro di banda, ma che porta con sé diverse conseguenze, fra

(*) prof. ing. Luigi Bonavoglia

le quali emergono come più importanti: la buona comprensibilità ottenuta sacrificando a volte, oltre la naturalezza della voce, anche la piena riconoscibilità del parlatore; inoltre il ritardo piuttosto rilevante introdotto se i processi di trattamento in partenza e all'arrivo richiedono forti elaborazioni.

Per chiarire bene le conseguenze che nascono facciamo un esempio, sia pur paradossale, basato sull'impiego di apparati ancora lontani dalla produzione industriale, ma già esistenti in diversi laboratori. E' noto che oggi è a un discreto stato di sviluppo la dattilografa automatica, cioè una macchina che riceve la voce di chi detta e fornisce lo scritto corrispondente oppure i relativi codici elettrici. E' anche disponibile, con maggiore perfezione, quello che potremmo chiamare il lettore vocale automatico, che da un dattiloscritto o dai segnali elettrici equivalenti, sintetizza la voce che corrisponde a quello scritto.

Ed ecco allora l'esempio di come si può trasmettere un messaggio (vocale all'inizio e alla fine) con perfetta comprensibilità e nessuna relazione fra la voce ricevuta e quella di partenza; basta supporre, ad esempio, che la persona che parla alla dattilografa automatica sia una donna, e che i caratteri dello scritto vengano inviati all'altro estremo del collegamento con un canale di tipo telegrafico, quindi occupante una banda di frequenza molto ridotta, e che, all'estremo ricevente, dai segnali telegrafici si sintetizzi parola per parola quanto viene dettato; la sintesi può dare tranquillamente una voce molto diversa da quella di partenza, addirittura quella di un uomo, fornendo un parlato perfettamente comprensibile. Tutto il processo introduce, inoltre, un ritardo non indifferente.

Che tipo di processo si è fatto? In definitiva, una analisi in partenza della voce di chi parla volta a ottenere la sua rappresentazione fonetica con simboli e, in arrivo, la sintesi di una voce a partire da quella rappresentazione. Un processo cioè basato sulla volontà di ottenere una buona comprensione e non sulla volontà di riprodurre più o meno fedelmente quella voce. E' vero che i trattamenti in uso oggi non arrivano a questo eccesso: ma la tendenza a sacrificare la riconoscibilità del parlatore pur di diminuire la banda di frequenza occupata, mantenendo buona la sola comprensibilità, è piuttosto marcata[1].

Ovviamente per il futuro si aprono nuove possibilità di comunicazione: da uno scritto, si può inviare il segnale vocale, oppure da un segnale vocale far arrivare al corrispondente lo scritto relativo, e sbrigliando la fantasia altre combinazioni, compresa in futuro la traduzione della lingua. Ma non è più telefonia: per rendersi conto di quanto stia avvenendo in questo servizio dal punto di vista del trattamento del segnale si guardi la fig. 1, tratta da un articolo di Jayant [1], che riassume la situazione riportando la qualità ottenibile -misurata con il metodo della valutazione media (MOS = Mean Opinion Score) [2]- in funzione del ritmo di bit che trasporta il segnale vocale: questo esame è fatto per vari casi e cioè per il PCM a 8 bit, per l'ADPCM che è un trattamento della forma d'onda con previsione del codice futuro in funzione dei codici (2 o più) passati, codec ibridi (basati sulla forma d'onda e sullo spettro), e del vocoder, che è una tipica macchina ad analisi e sintesi (vedi app. 1). Ancora più interessante è la fig. 2, ripresa da un lavoro piuttosto vecchio, del 1983, ma molto chiaro ed anche preveggente, di Crochiere e Flanagan [3].

Qui la qualità (in unità arbitrarie) è riportata in funzione del ritmo di bit (fino a 64 kbit/s), ed è da notare che andando da 64 kbit/s in giù si percorre, secondo la previsione degli autori, una strada (in discesa come qualità) verso un maggior risparmio di flusso di bit, ma verso una maggiore complicazione: si passa dal trattamento che all'arrivo cerca di ricostruire la "forma d'onda"[2] (denominato "Codifica della forma d'onda") a un trattamento basato su parametri fisici come lo spettro del segnale, il valore della fondamentale, ecc.; si arriva finalmente ai metodi di analisi e sintesi, chiamati "modellamento della sorgente", esemplificati nel vocoder. L'acutezza di visione dei due autori è stata evidenziata da quanto è successo nei fatti: infatti la strada descritta è anche quella temporale degli sviluppi

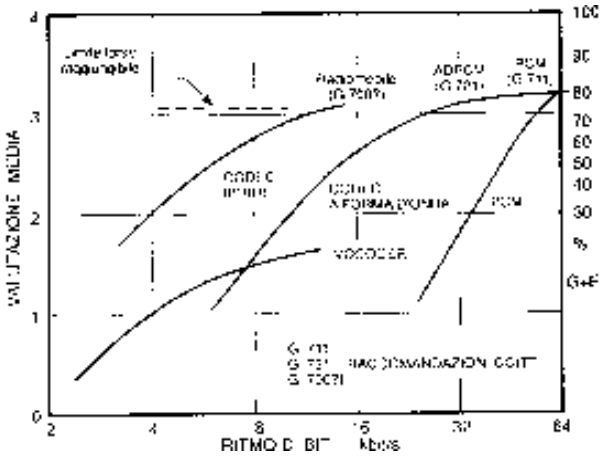


Figura 1 Valutazione media in funzione del ritmo di bit e del metodo di trattamento. Disegnato in base ai dati di Jayant: "High Quality Coding of Telephone Speech and Wideband Audio", IEEE Comm., January 1990, pag. 10

(1) L'autore ritiene che i gestori di un servizio telefonico dovrebbero preoccuparsi di questo fatto e introdurre nel metodo di valutazione della qualità trasmissiva quel tanto di dipendenza dalla naturalezza della voce percepita e dal riconoscimento del parlatore tale che impedisca ai cultori del trattamento della voce di scivolare verso eccessi del tipo di quello illustrato poco fa; certo che se ci sono utenti che sono interessati solo alla comprensione si può fare molto nel senso detto; resta però il dubbio se non sia meglio usare per costoro la trasmissione della parola scritta per farla comparire così su un visore, partendo dalla analisi della voce lontana.

(2) In questo caso il giudizio sulla bontà del trattamento era basato sull'errore quadratico medio tra forma d'onda iniziale e forma riprodotta.

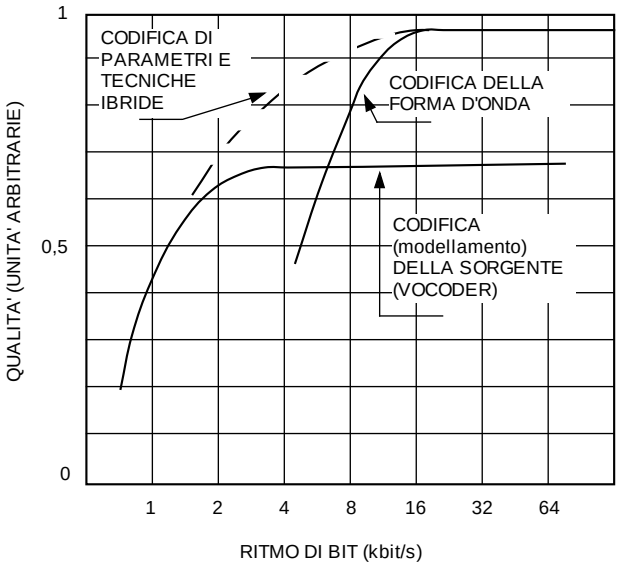


Figura 2 La qualità ottenibile con vari ritmi di bit e con vari tipi di trattamento

industriali nel trattamento della voce, che ci hanno portato alla situazione di oggi (per esempio nel GSM), in cui si comincia a intravedere quanto si è detto all'inizio: la poca connessione fra comprensibilità e naturalezza unita a riconoscibilità della voce, insieme all'introduzione di un notevole ritardo globale.

La fig. 3 riporta dati più moderni da una memoria di C. Mossotto presentata all'EUROSPEECH nel settembre 1991 a Genova [4].

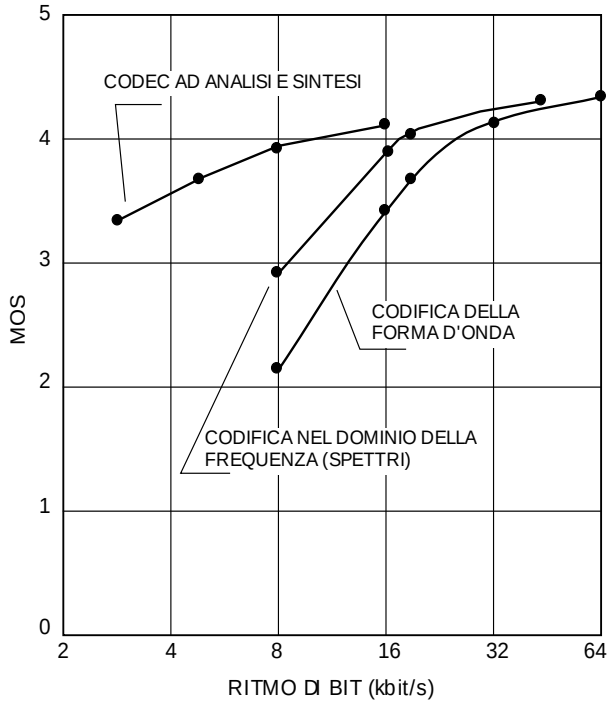


Figura 3 Una recente valutazione (C. Mossotto) della qualità ottenibile con i vari metodi di trattamento

Tutte le figure mostrate danno, come misura della qualità, la valutazione media (MOS) che giudica la *facilità* (cioè la comprensione) di una conversazione.

Dopo questa premessa passiamo all'esame della situazione odierna, nei vari tipi di collegamento che abbiamo elencato nella prima parte di questo lavoro.

2. **Esame dello scenario attuale**

Riferiamoci allo scenario illustrato nella I parte, e precisamente alle figg. 2 e 3 di pag. 28 e 29 del n. 2, anno terzo; dal loro esame scaturiscono come casi di interconnessione possibile i seguenti:

- Telefonia fissa
  1) Telefono fisso analogico (TFA) con TFA
  2) Telefono fisso numerico (TFN) con TFN
  3) Telefono fisso numerico (TFN) con TFA
- Telefonia fra mobile e fisso
  4) Radiotelefono analogico (RTA) con TFA
  5) Radiotelefono analogico (RTA) con TFN
  6) Radiotelefono numerico (RTN) con TFA
  7) Radiotelefono numerico (RTN) con TFN
- Radiotelefonia fra mobili
  8) RTN con RTN
  9) RTN con RTA
  10) RTA con RTA

  Esaminiamo la situazione nei vari casi.

2.1 *Caso TFA-TFA*

Questo caso è, ancor oggi, quello più comune; infatti, anche se ci avviamo velocemente a realizzare collegamenti completamente numerici a 4 fili con equivalente costante fra le centrali che li alimentano, la sostituzione di milioni di telefoni e relative coppie in rame non può avvenire di colpo ma solo con una ragionata gradualità.

Il caso è trattabile con semplicità, pur con l'incertezza circa la conoscenza delle caratteristiche dei telefoni costruiti ed installati molti anni fa, che esistono ancora in buon numero sulla rete; quanto ai telefoni TFA più moderni, essi devono rispondere a norme che tengono conto dell'IIS (LR), la curva di risposta, l'effetto locale, ecc.

a) *Caso dei telefoni "vecchi"*

Cominciamo dai telefoni vecchi: per il complesso telefono + rilegamento + ponte di alimentazione, esistevano norme che fissavano l'ER (equivalente di riferimento) in valori che sono cambiati, nel tempo, seguendo in genere le raccomandazioni del CCITT e suoi predecessori. Cercando fra gli antichi documenti e tenendo presenti due lavori degli ingg. R.Casale e G.Tortia [5] [6], entrambi della fine degli anni '80, la situazione per i vecchi telefoni si può riassumere nel modo esposto qui di seguito.

Per prima cosa va ricordato che fino agli anni settanta non si costruivano telefoni (ovviamente analogici) con elementi attivi incorporati, e che il microfono usato era nella stragrande maggioranza dei casi quello a carbone, con i suoi pregi e i suoi difetti. Nel corso degli anni settanta è iniziato lo sviluppo di telefoni con elementi attivi e capsule, quindi, liberate dal vincolo della massima efficienza e più rispettose del principio di ottenere una buona curva di risposta. Il progresso mondiale è riassunto nella fig. 4, che riporta l'equivalente di riferimento globale dei telefoni (con ponte di alimentazione ma senza rilegamento) più rappresentativi (di massima quelli del vecchio gruppo Bell) fino al 1970 circa. Insieme all'ER (Equivalente di riferimento allo SFERT o NOSFER - curva continua) è riportato anche l'IIS ottenuto per calcolo (curva a trattini), usando cioè la formula ricordata nella parte I di questo lavoro.

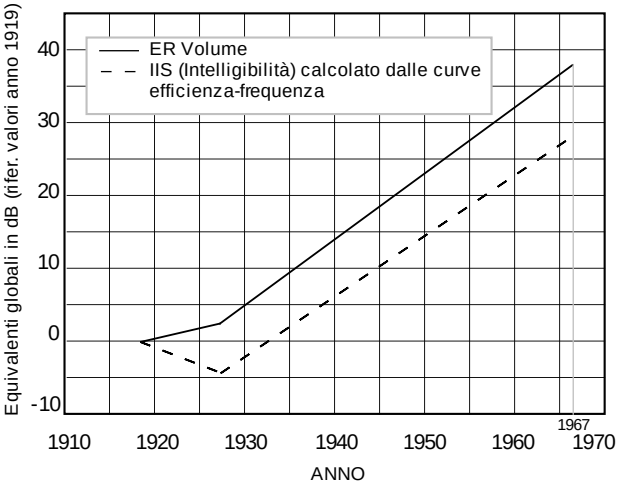


Figura 4 Miglioramento nel tempo di microfoni, ricevitori e complessivo dell'efficienza degli apparecchi telefonici del gruppo Bell negli Stati Uniti

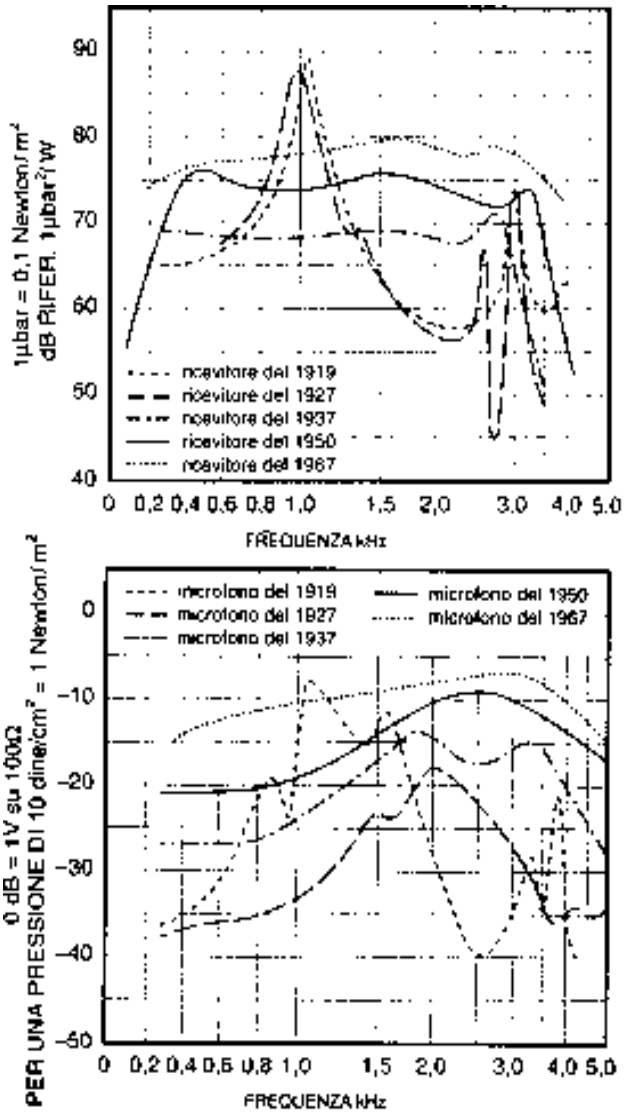


Figure 5 e 6 Evoluzione nella efficienza e nella curva di risposta di ricevitori e microfoni telefonici; dopo il 1970 si ha nel telefono una evoluzione non più basata sul miglioramento di questi elementi. Le figure sono dedotte da descrizioni dei telefoni usati nel Gruppo BELL

Si nota il buon progresso (circa 20 dB) dagli anni '20 agli anni '70, dovuto sia all'aumento della efficienza, che al miglioramento nella curva di risposta. Le figg. 5 e 6 mostrano questo secondo aspetto sempre per i telefoni Bell. La situazione italiana, più variegata perché risente dell'uso di telefoni di diverse manifatturiere, non si è mai discostata molto da quella americana.

Vediamo allora cosa accade agli utenti che usano telefoni "vecchi", fermandoci però a quelli considerati da Casale-Tortia nel lavoro del 1985. Dalla loro indagine statistica risulta una distribuzione di ER (equivalente di riferimento) globale dei sistemi locali (trasmissione + ricezione) definita da un valore medio di 2 dB circa e uno scarto quadratico medio (s.q.m.) dell'ordine di 4 dB (nel seguito lo s.q.m. sarà indicato con σ).

Questa valutazione è stata fatta considerando i valori forniti da Casale-Tortia per i telefoni più diffusi allora, supponendo poi uguali le loro diffusioni e le possibilità di combinazioni incrociate: è un metodo molto sbrigativo, ma ci dà una sufficiente idea della situazione di allora; non dimentichiamo che gran parte di quei telefoni sono ancora in servizio, ma non tutti; le nostre considerazioni vanno perciò considerate come di larga massima. Dal valore medio e dallo s.q.m. su riportato deriva un valore di ER globale del sistema locale emittente + ricevente non superato per il 97,7% dei casi pari a $2+2\sigma = 10$ dB (cioè *quasi* massimo), e un *quasi* minimo pari a $2-2\sigma = -6$ dB.

Il circuito numerico fra le centrali terminali ha un ER = 8 dB (ER e IIS sono quasi coincidenti su un buon circuito numerico regolato con equivalente a 1020 Hz pari a 7 dB) e quindi il valore quasi massimo di ER fra

i due interlocutori è 8+10 = 18 dB; il valore quasi minimo vale 8–6 = 2 dB.

Per passare da questi valori alla qualità, cioè ai giudizi favorevoli espressi dagli utenti, dobbiamo far riferimento al diagramma di fig. 8 basato sul rumore e sull'ER e non sull'IIS perché i valori che conosciamo sono di questo tipo (vedi fig. 7).

Quanto al rumore da considerare al telefono in ricezione, nell'appendice 3 è mostrato come lo si può determinare; assumiamo un campo di variabilità minore che nelle reti analogiche, dato che il circuito numerico viaggiante sulla rete di giunzione porta, dal punto di vista telefonico, un contributo veramente basso; un campo di rumore fra –65 e –60 dBmp sembra ragionevole.

Disegnamo sul diagramma un rettangolo fra i rumori –60 e –65 dBmp e i valori +18 e +2 dB per l'ER. Si vede che i giudizi favorevoli vanno da un massimo di circa 95 a un minimo dell'ordine dell'80÷75%. (Il diagramma riportato nella I parte conteneva un errore, in quanto la curva di migliore giudizio era quotata con il 100% di giudizi favorevoli anziché 95% come invece è giusto. Il diagramma riportato qui è invece esatto). La collocazione del rettangolo attuale, riferita alla collocazione del rettangolo ottenibile sulla vecchia rete analogica, mostra chiaramente il notevole progresso ottenibile con la introduzione della tecnica numerica anche solo nella rete di giunzione fra centrale terminale e terminale.

Resta però, come si vede, un certo divario fra utenti meglio trattati e quelli peggio trattati: per farlo sparire, o quanto meno ridurlo, non sembra necessario aspettare che anche le code divengano numeriche, ma basterebbe che i telefoni analogici moderni fossero progettati con un controllo automatico di efficienza in funzione della resistenza di linea e con tolleranze più strette, ma questa considerazione, si vedrà subito, coinvolge anche i telefoni "nuovi".

Passiamo ora a:

b) *Caso dei telefoni "nuovi"*

Passiamo a cercare le notizie riguardanti i telefoni analogici che sono venduti e si installano oggi, e consideriamo, data la liberalizzazione avvenuta, non già la produzione esistente, ma le norme CEI del 1993, condensate nelle tabb. 1.a e 1.b.

Linea	0 ohm	300 ohm	700 ohm	1000 ohm	1400 ohm
ISE	+4 dB	+4 dB	+5 dB	+6,5 dB	+9 dB
Tolleranza	±3 dB	±3 dB	±3 dB	±3 dB	±3 dB

ISE_{min} = +1 dB ISE_{max} = +12 dB

Tabella 1.a Indice d'intensità soggettiva in *emissione* in funzione della lunghezza (elettrica) della linea (Norma CEI-1993)

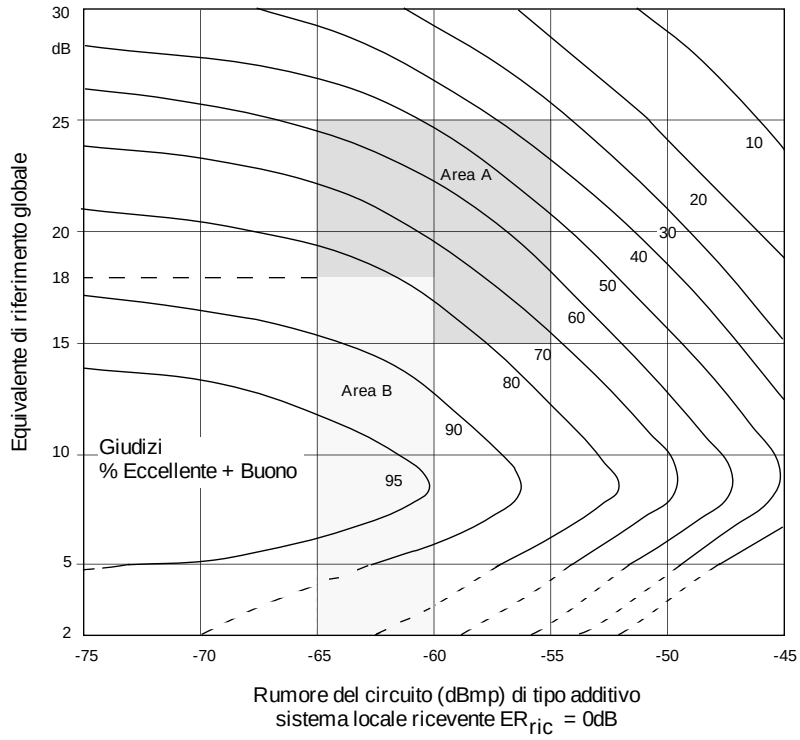


Figura 7 Percentuale di utenti soddisfatti (E+G) al variare dell'equivalente di riferimento e del rumore di tipo additivo (Libro Giallo CCITT, 1981, Vol. V, pag. 190). L'Area A è quella relativa alla vecchia rete tutta analogica, quella B è relativa alla rete di giunzione numerica con telefoni analogici "vecchi"

Linea	0 ohm	300 ohm	700 ohm	1000 ohm	1400 ohm
ISR	–6,5 dB	–6,5 dB	–5,5 dB	–4,5 dB	–3,5 dB
Tolleranza	±2,5 dB	±2,5 dB	±2,5 dB	±2,5 dB	±2,5 dB

$ISR_{min} = -9$ dB $ISR_{max} = -1$ dB

Tabella 1.b Indice d'intensità soggettiva in *ricezione*, in funzione della lunghezza (elettrica) della linea (Norma CEI-1993)

Queste norme stabiliscono i limiti per il sistema d'utente, in funzione della lunghezza del rilegamento (espressa in Ω); viene preso in considerazione il valore dell'IIS da ottenere in emissione ISE e in ricezione ISR.

In base ai limiti riportati in tabella, e alla distribuzione statistica delle resistenze dei rilegamenti della nostra rete, si può instaurare un calcolo probabilistico non difficile in base ad alcune assunzioni che qui riferiamo.

La prima assunzione è che la distribuzione delle resistenze dei rilegamenti in rame della nostra rete, derivata da misure alquanto vecchie e riportata in fig. 8, sia ancora valida oggi; questa assunzione è certamente molto plausibile data l'enorme quantità di rilegamenti già esistenti al tempo delle misure[3].

La seconda assunzione è che la variabilità delle caratteristiche per ISE e ISR ammessa come tolleranza in tabella, segua una legge di probabilità gaussiana troncata ai limiti. Si è assunto che il troncamento avvenga alla probabilità di 1,3 per mille, cioè nel punto a 3 σ dal valore mediano. Questo significa assumere un σ per ISE pari a 1 dB e per ISR pari a 0,833 dB.

Con i soliti calcoli si trova allora che il valore dell'IIS totale delle due code ha, con buona approssimazione, un andamento gaussiano (ovviamente troncato) con un valore mediano pari a –3,6 dB e un σ nella parte gaussiana pari a 1,3 dB circa. La variabilità fra i punti a 0,13% e 99,87% è di $\pm 3,9 \text{ dB} = \left(3^2 + 2,5^2\right)^{1/2}$.

Il troncamento avviene per il valore ottimo (pari a $ISE_{min} + ISR_{min} = -8$ dB) con probabilità dello 0,05% (cioè 5 su 10.000) e quello per il valore peggiore (pari a $ISE_{max} + ISR_{max} = +11$ dB) avviene con probabilità evanescente[4].

La forte dissimetria dei due troncamenti è dovuta al fatto che gran parte dei rilegamenti ha resistenza sotto i 700 ohm (come si rileva dalla fig. 8) e quindi questa gran parte (~80%) presenta un ISE medio ≤5 dB, un ISR medio ≤–5,5 dB con un IIS totale ≤–0,5 dB, mentre i rilegamenti con R>1000Ω sono nettamente meno dell'1%.

A questo punto dobbiamo decidere quali valori assumere come *quasi* massimo e *quasi* minimo; sembra ragionevole assumere i valori con valore medio ±3σ, cioè rispettivamente $IIS_{max} = -3,6+3,9 = +0,3$ dB e $IIS_{min} = -3,6-3,9 = -7,5$.

Aggiungiamo a questi valori l'IIS del circuito sulla rete di giunzione e otteniamo per IIS_t (cioè IIS totale)

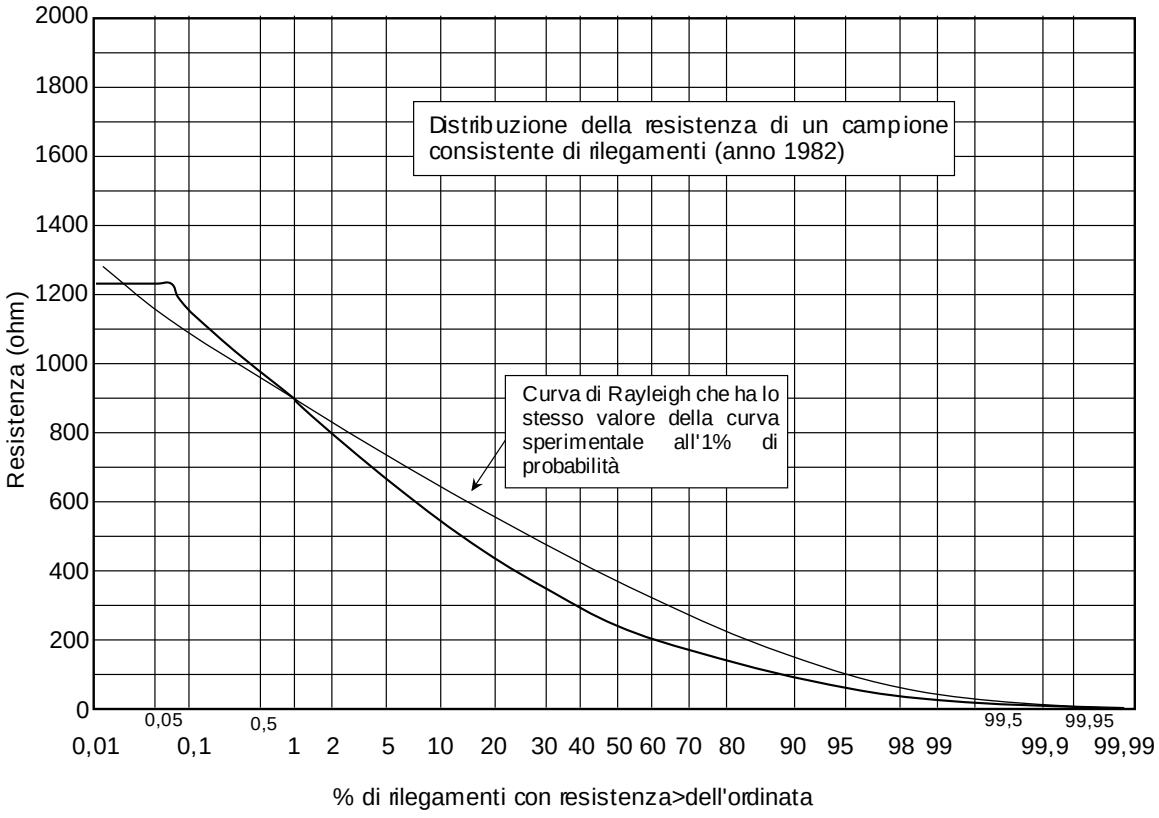


Figura 8 Distribuzione statistica della resistenza del rilegamento (ohm) nella rete italiana

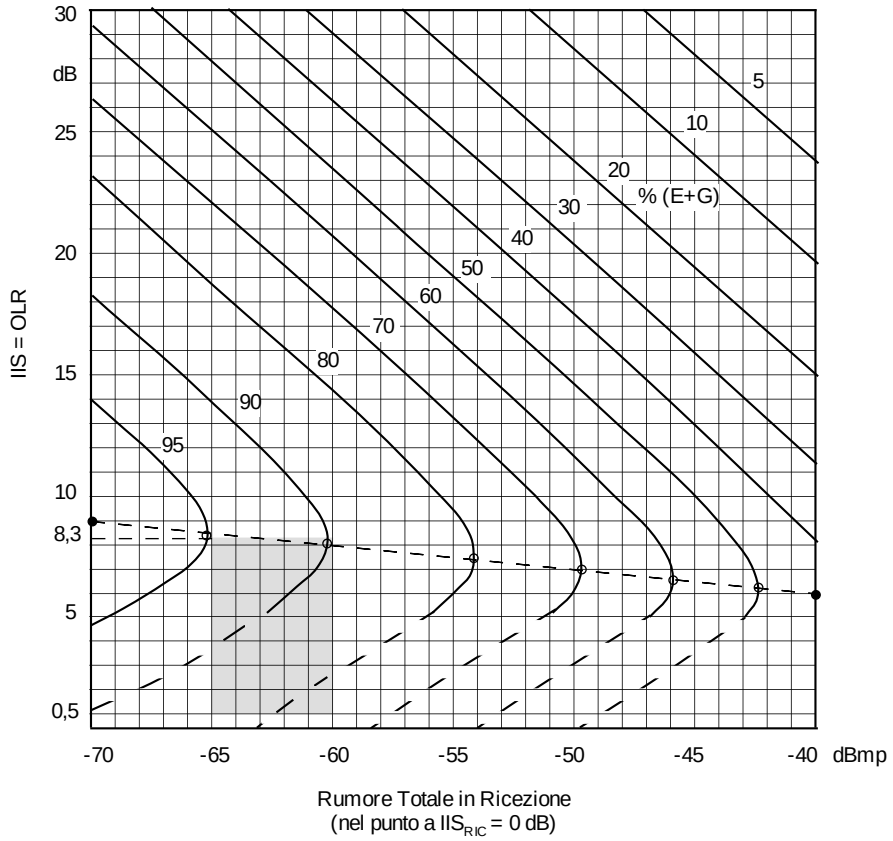


Figura 9 Percentuale di utenti soddisfatti in presenza delle sole degradazioni dovute ad attenuazione, misurata dell'IIS totale, e al rumore. L'Area campita rappresenta la situazione che si ha con rete di giunzione numerica a cui sono connessi utenti analogici provvisti di telefoni "nuovi"

$$IIS_{tmax} = +0,3+8 = 8,3 \text{ dB}$$

$$IIS_{tmin} = -7,5+8 = +0,5 \text{ dB}$$

Consideriamo ora il diagramma di fig. 9 (riportato dalla I parte del lavoro) e disegnamo il rettangolo con

ascisse –65 e –60 dBmp per il rumore

+0,5 e +8,3 dB per IIS_t

Vediamo che i giudizi favorevoli degli utenti (E+G) vanno dal 95% a circa l'80% (questo valore ottenuto per estrapolazione).

Si nota che il campo di variabilità (il rettangolo) è sfasato rispetto alla posizione ottima: occorrerebbe alzare tutti i valori di IIS totale e stringere le tolleranze sui telefoni. Tutto sommato, la situazione è però buona perché l'utente stesso si adatterà a segnali in arrivo troppo forti, allontanando un po' (basta poco) il ricevitore dall'orecchio. Ricordiamo che per l'IIS totale il Piano Regolatore Nazionale per le Telecomunicazioni chiede, come suggerito dalle Raccomandazioni del vecchio CCITT (1990, epoca dell'entrata in vigore del Piano), che esso resti nel campo da 8 a 12 dB.

D'altronde, la situazione che deriva dalla norma CEI del '93 favorisce quei pochi utenti che, serviti da linee lunghe (resistenza fra 700 e 1400 Ω), si avvicinano al massimo dei massimi di IIS totale = 11+8 = 19 dB; essi ottengono una qualità passabile (E+G dell'ordine 70÷80%).

E' anche evidente che se si usasse un unico tipo di telefono la dispersione di valori di ISE e ISR sarebbe inferiore: infatti le tolleranze di ±3 dB per ISE e ±2,5 dB per ISR tengono conto non solo delle tolleranze di fabbricazione dei telefoni, ma anche delle differenze del prodotto da costruttore a costruttore; la norma CEI,

(3) La fig. 8 riporta anche una distribuzione di Rayleigh, calcolata in modo che alla probabilità dell'1% essa abbia lo stesso valore della curva sperimentale. Oggi si cominciano a introdurre multiplex d'utente, cosa che provocherà un aumento relativo dei rilegamenti con resistenze di poche centinaia di ohm (intorno a 200÷300) e una diminuzione relativa delle forti resistenze: la distribuzione generale delle resistenze sarà quindi più favorevole a un trattamento uniforme degli utenti. (Di passaggio va notato che la curva si avvicinerà a quella di Rayleigh).

(4) Il valore ottimo si ottiene per linea a 0 ohm che dà IIS = ISE+ISR–5,5 = 4–6,5–5,5 = –8 dB, quello peggiore (per linea a 1400 ohm) dà un IIS = ISE+ISR+5,5 = 9–3,5+5,5 = 11 dB; essendo 5,5 dB la tolleranza in più o in meno ammessa in totale dalla CEI.

cioè, prevede una più larga tolleranza per permettere a più costruttori di entrare sul mercato. Se, per esempio, si usasse ovunque un solo tipo di telefono (per esempio il SIRIO), il valore di IIS globale (telefono emittente + telefono ricevente + rilegamenti + ponti) sarebbe meno variabile; il 90% dei collegamenti avrebbe un IIS globale compreso fra –4 e –5 dB, e il *quasi* massimo (0,13% di probabilità di essere sorpassato) sarebbe di –1 dB circa. Aggiungendo gli 8 dB del circuito numerico si avrebbe un minimo IIS totale di +3 dB e un quasi massimo di circa +7 dB, con una situazione alquanto migliore di quella ottenibile con l'uso di telefoni disparati. In ogni caso, sembra che la norma generale dovrebbe essere ritoccata per centrare meglio, rispetto al campo raccomandato dagli organismi internazionali (HS totale fra +8 e +12), il campo di variabilità effettivo.

Da quanto si è visto, si può trarre la conclusione che la numerizzazione della rete di giunzione, anche con telefoni e rilegamenti di tipo analogico, ha comportato un grosso miglioramento sul passato, in cui tutte le tecniche trasmissive erano analogiche. Va detto però che se ci fosse stato un miglioramento nelle norme e nella produzione dei telefoni si sarebbero ottenuti risultati ancora migliori: infatti, la dispersione dell'IIS totale che deriva dalla norma CEI, di circa ±4 dB intorno ai valori medi richiesti, è la massima causa della dispersione della qualità. Se, per esempio, si restringesse il limite a ±2 dB, il campo di variazione si restringerebbe molto con uno spostamento dei giudizi favorevoli (E+G) verso il meglio.

Non trattiamo il caso di un telefono analogico "vecchio" con uno "nuovo", perché si otterrebbero risultati medi fra quelli estremi che abbiamo ora determinato.

Si può constatare, in definitiva, che con la riduzione dell'equivalente della rete delle giunzioni numeriche (da centrale terminale a terminale) al valore di 7÷8 dB, si è migliorata molto la situazione quando si hanno combinazioni delle code peggiori (con TFA), ma si incorre nel pericolo di livelli troppo alti in ricezione per le combinazioni di code migliori, e questo fatto incita a procedere verso la diminuzione della dispersione dell'IIS dei telefoni. Una accurata indagine andrebbe fatta circa la ripartizione dell'IIS totale fra emissione e ricezione, tenendo conto dei volumi sonori attuali dei parlatori telefonici. Si ricorda che questi volumi sonori si sono notevolmente abbassati nel tempo per il miglioramento generale della qualità.

2.2 *Caso TFN-TFN*

Questo caso è presto risolto: l'IIS dei telefoni numerici è fissato già dal vecchio CCITT per i due obiettivi riportati nella tab. 2.

E' certo che quando introdurremo la numerizzazione su vasta scala nella rete d'utente saremo al "lungo termine"; e quindi non si vede ragione per non rispettare i valori relativi. Tenuto conto che il circuito numerico andrà quindi da utente a utente, e che il segnale numerico sarà codificato e decodificato nei telefoni, l'IIS totale dei telefoni sarà quello dell'intero collegamento: il rumore sarà solo quello di quantizzazione, in pratica non influente. Resterà il rumore di fondo dei codec che si potrà tenere <–65 dBmp; anzi sarà bene non eliminarlo del tutto per non dare la sensazione di telefono muto. Basta dare un'occhiata alla fig. 9 per vedere che un IIS totale pari a 10 dB con rumore <–65 dBmp ci pone al meglio della qualità. E tutti i collegamenti avranno la stessa qualità dal punto di vista dell'IIS e del Rumore in ricezione.

	IISE	IISR	IIS globale
Obiettivo a breve termine	5÷11 dB	–1÷5 dB	4÷16 dB
Obiettivo a lungo termine	8 dB	2 dB	10 dB

Tabella 2

Ma non dimentichiamo altri elementi influenti sulla qualità: la limitazione di banda fonica, l'effetto locale, il ritardo di trasmissione, eventuali errori sui bit trasmessi e la distorsione.

A pag. 36 della I parte è calcolato l'effetto della limitazione di banda: ma con i codec situati nel telefono non si vede perché, con capsule adeguate, non si possa ottenere la banda da 150 (oppure 200) Hz fino a 3400, in modo da non portare praticamente nessuna conseguenza negativa sul giudizio degli utenti.

Quanto all'effetto locale, esso sarà dominabile in pieno entro lo stesso telefono, e portato al valore tale da non causare degradazioni della qualità (vedi pag. 35, parte I, formula 4 e definizioni).

L'appendice 2 dà una sintetica illustrazione dei problemi che nascono a causa del ritardo di trasmissione, e qui, per collegamenti fra due telefoni numerici, va subito tenuto presente che l'eco deriva solo da eventuali accoppiamenti acustici nel microtelefono, ed è quindi pressoché impercettibile.

Infatti, il circuito da telefono a telefono è a due vie (cioè senza forchette). Se però il ritardo è forte si ha un altro tipo di inconveniente, cioè un ritardo alla risposta notevole, fra il momento in cui il parlatore smette (pari al doppio del ritardo) e la risposta, cosa che può dare un fastidio psicologico. Escludendo però ogni trattamento (non è così per i telefoni mobili), cioè con circuito e telefoni tutti a 64 kbit/s, il ritardo nell'intera Europa è tale da non dare fastidi del genere.

Le distorsioni presenti dipenderanno, praticamente, solo da quelle presenti nelle capsule e nei codec

terminali, e non si vede perché non possano essere tenute a un livello molto basso in apparecchi telefonici moderni correttamente concepiti. Quanto a quelle dovute a eventuali trattamenti, si veda quanto detto per i telefoni mobili. Lo stesso vale per gli errori di trasmissione sui bit.

2.3 *Caso TFA-TFN*

Si tratta di applicare a questo caso gli stessi criteri finora usati e si trova che i giudizi positivi sono abbastanza soddisfacenti in percentuale anche con telefoni analogici vecchi. Va ricordato che l'eco scompare solo per l'interlocutore presente al telefono analogico, mentre l'altro, pur avendo un telefono più progredito, si trova in presenza dell'eco (infatti la forchetta è presso il TFA) e questo problema va risolto a seconda dei casi.

2.4 *Casi RTA-TFA, RTA-TFN, RTN-TFA, RTN-TFN*

Tratteremo questi casi prima in generale e poi scenderemo ai casi di telefoni mobili analogici e numerici.

Sia il radiotelefono su auto che il cosiddetto telefonino (personale) sono del tipo a 4 fili, e quindi i problemi di eco sono gli stessi del caso precedente, ma con conseguenze più evidenti nel caso del telefono mobile numerico.

Occorre ovviamente dare una idea, sia pur sommaria, di come è costituita la coda radio, dal punto di vista trasmissivo, senza entrare nei problemi di segnalazione, reperimento nell'area di copertura del radiomobile chiamato ecc.; tratteremo ovviamente solo i moderni sistemi cellulari.

La copertura radio della zona da servire è assicurata suddividendo la zona in aree più o meno grandi, dette celle, ciascuna delle quali è servita da una propria stazione radio. Ad ogni cella è assegnato un certo numero di canali radio, canali che possono essere riutilizzati in altre celle con opportune precauzioni. Ovviamente celle adiacenti non utilizzano gli stessi canali, e si sono sviluppate regole da seguire per l'assegnazione dei canali disponibili in modo da coprire l'intera zona desiderata. E' anche ovvio che il numero totale dei canali radio, e il traffico da smaltire, insieme all'allocazione in frequenza dei canali radio, incidono sulle dimensioni delle celle; non ci occuperemo di questi problemi.

Per quanto ci riguarda, dobbiamo vedere dove si attesta il circuito di giunzione numerico (a 64 kbit/s) e come viene trattato per entrare nella stazione radio, e come di qui giunge all'utente. E' noto che la rete numerica (Piano Regolatore del 1990) considera un piano cosidetto dei transiti con centrali denominate SGT, alle quali si attestano le centrali SGU, che a loro volta servono le centrali terminali (vedi I parte del lavoro).

Un utente fisso che voglia connettersi a un mobile, raggiunto il suo SGT fisso, viene da questo instradato su una centrale della rete mobile denominata MSC (Mobile Switching Center), anch'essa fissa nonostante il suo nome.

Questa è connessa (sempre a 2 vie, cioè 4 fili) con le stazioni radio di un certo numero di celle: in una di queste si trova l'utente mobile desiderato (il metodo per determinare questa cella, l'abbiamo detto, non sarà qui descritto).

Nel caso del radiotelefono analogico (in Italia denominato ETACS) o nella centrale MSC o nella stazione radio, si passa via codec alla frequenza vocale, e con questa si modula (angolarmente) una radio-frequenza. Ogni radio frequenza (per un collegamento ne servono una in andata e una in ritorno) porta un segnale vocale (cioè un utente). Il sistema viene chiamato SCPC-FDMA (Single Channel Per Carrier-Frequency Division Multiple Access).

La stazione radio emette e riceve quindi tante coppie di frequenze quanti sono gli utenti presenti nell'area che chiamano o sono chiamati. Le radiofrequenze nella banda dei 900 MHz (in passato si è usata una banda più bassa) sono spaziate di 25 kHz, e la modulazione (d'angolo) è analogica.

Il telefono radio-mobile è dotato di un modem per la modulazione d'angolo come la stazione radio, e quindi il circuito giunge a 2 vie (4 fili) fino ad esso; generalmente su ogni circuito viene equipaggiato, per migliorare la situazione nei confronti del rumore, un compandor sillabico.

La tratta radio completa è progettata (apparati, antenne, area coperta) per dare un circuito con banda standard (300-3400 Hz) e una rumorosità entro determinati limiti. L'IIS è in genere buono, ma non si dispone di dati che descrivano in modo significativo la casistica attuale; e ciò perché i radiotelefoni e i telefonini sono di molti e svariati produttori. La qualità ottenibile è dominata dal rumore che nasce sulla tratta radio, in definitiva dalla posizione e dalle caratteristiche di propagazione nel tratto fra il mobile e la radiostazione.

Concludendo, i telefoni mobili analogici, tutti ormai di recente costruzione, sono certamente buoni come risposta in frequenza e come IIS (taluni possono variare l'IIS in ricezione con un comando manuale), l'eco è inesistente al telefono fisso, e al telefono mobile è poco ritardato e tollerabile, dato il fatto che praticamente non si introducono ritardi oltre a quelli delle vie di trasmissione. Diviene quindi dominante nei riguardi della qualità il rumore della tratta radio, maggiore di quello presente sulla rete delle giunzioni e sul rilegamento dell'utente fisso.

Esiste inoltre la possibilità di interferenze da canali sulle radio-frequenze eguali riutilizzate in celle più o meno vicine.

Questi rumori sono molto variabili con le condizioni di propagazione (dipende molto dalla posizione e velocità del mobile) e quindi la qualità non è costante, e in qualche istante il collegamento può interrompersi.

Passiamo ora al caso del telefono radiomobile numerico.

Da noi e in tutta Europa si è data la preferenza a un sistema denominato GSM (Groupe Special Mobiles), con accesso multiplo alla stazione radio a divisione di tempo. Questo sistema è stato sviluppato, dalla fine del 1985, in unione da molti enti europei, e ha portato all'inizio del servizio in varie Nazioni europee, compresa l'Italia, in questi ultimi due o tre anni.

Il sistema è molto progredito, rispetto a quelli passati, dal punto di vista dell'utilizzazione dello spettro, della sicurezza contro le intercettazioni e inserimenti fraudolenti, della interconnessione internazionale (un utente GSM italiano può usare, per esempio in Francia, il suo radiotelefono) e altre interessanti caratteristiche per la trasmissione dati, l'identificazione del chiamante, ecc.

Dal punto di vista della qualità interessa conoscere il metodo di codifica, di protezione degli errori di linea, e del metodo usato per la definizione della qualità stessa.

Cominciamo dall'ultimo punto: per quanto abbia cercato fra i vari documenti esistenti, ho trovato che il solo metodo seguito ufficialmente per le scelte del codec è stato quello del MOS (Mean Opinion Score) che, come abbiamo ricordato nella premessa, giudica la *facilità* della conversazione, in pratica giudica la *comprensione* del parlato.

L'obiettivo richiesto al codec è stato quello di ridurre il ritmo di bit (64 kbit/s per il PCM classico a 8 bit) intorno a 13 kbit/s, con un basso ritardo nell'elaborazione necessaria, una buona robustezza agli errori di trasmissione, insieme alla facilità di realizzazione con un basso consumo di potenza (quest'ultima caratteristica è molto importante per il radio-telefono).

Furono confrontati diversi codec, alcuni basati sulla divisione in sottobande (8 oppure 16) e codifica a bit variabili nelle varie sottobande, altri codec basati su predizione lineare e a lungo termine, con eccitazione da parte di impulsi regolari o variabili.

Dalle prove (sempre in termini di MOS) [7] risultò:

- in genere la qualità dei codec in esame dipende dal parlatore più che nei sistemi FM (analogici con modulazione d'angolo);
- tutti i codec superano la qualità del sistema FM a pari tasso d'errore quando questo è medio o alto (fino a circa 10^{-3});
- per il tasso d'errore di 10^{-2} tutti i codec degradano notevolmente.

Le prove effettuate portarono a proporre un nuovo codec, ottenuto da un misto dei due migliori codec chiamato RPE-LPC (Regular Pulse Exitation-Long Term Prediction).

Nel sistema è compreso un rivelatore di attività vocale, che permette l'uso di una emissione in antenna discontinua, con riduzione quindi del consumo d'energia. Anche questo sistema RPE-LPC è piuttosto sensibile agli errori di linea, che fra l'altro, durante il movimento del radiomobile, arrivano a raffiche. Perciò, l'affasciamento di 8 canali (a 13 kbit/s) sulla portante radio in TDM (Time Division Multiplex), avviene con un metodo che tende a diminuire l'effetto negativo sulla qualità di queste raffiche di errori.

Si procede così (i valori indicati sono quelli del libro di Grimaldi e Zingarelli citato in bibliografia [7]; possono cambiare ma il procedimento resta): in ognuna delle portanti radio di cui una stazione è dotata (sono disponibili 124 canali spaziati di 200 kHz in tutto, da suddividere fra le varie celle) si multiplano nel tempo 8 canali telefonici (o equivalenti servizi) ciascuno a 13 kbit/s lordi. Per proteggersi dagli errori a raffiche, ogni flusso a 13 kbit/s viene suddiviso in spezzoni di 260 bit, e ad ogni spezzone si aggiungono 196 bit per la protezione, e 8 bit di preambolo e identificazione: ogni spezzone contiene quindi 260+196+8=464 bit; questo blocco viene diviso in otto sottoblocchi di 58 bit ciascuno, e questi sottoblocchi vengono intervallati con quelli degli altri 7 flussi di bit (provenienti dagli altri codec) secondo lo schema di fig. 10. Al tutto vengono accomunati segnali di controllo e servizio.

In definitiva, su una portante radio vengono immessi 270,83 kbit/s; ogni canale telefonico contribuisce quindi, compresi i segnali di servizio, per 270,83/8 ≅ 33,85 kbit/s, flusso che non è molto diverso da quello che richiede un ADPCM (Adaptive Differential PCM) o un ADM (Adaptive Delta Modulation) che a 32 kbit/s danno una buona qualità. Qui però si è ottenuta anche una buona protezione dagli errori, in particolare quelli a raffica; in

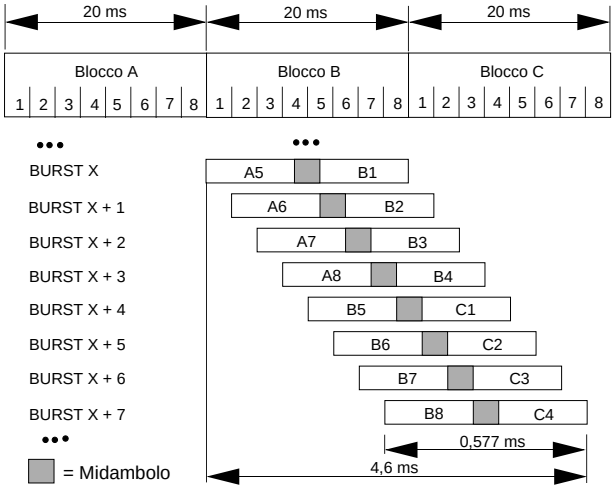


Figura 10 Schema di formazione dei burst di segnale partendo dagli spezzoni di segnale vocale (canali telefonici A, B, C) e del loro "interleaving diagonale" (dal volume di Grimaldi e Zingarelli [7])

sintesi, è come se ogni segnale di un canale a 13 kbit/s venisse ripetuto 33,85/13 ≅ due volte e mezza. Tutto il processo del GSM qui descritto determina in emissione e ricezione un ritardo dell'ordine di 80÷90 ms e richiede perciò la cancellazione dell'eco, sulla tratta radio. L'altro utente non riceve eco. Va infine detto che:

- la tecnica di modulazione della portante radio è del tipo GMSK (Gaussian Minimun Shift Coding);
- si usa in ricezione un equalizzatore adattativo;
- nella stazione radio di cella o nel MSC si torna al PCM classico a 64 kbit/s che costituisce il veicolo verso la rete fra MSC e SGT.

Questa lunga descrizione mi è sembrata necessaria per ben capire la situazione che si genera fra un radiomobile e un utente telefonico fisso.

Di tutto quanto abbiamo detto vanno ricordati i punti salienti del sistema GSM:

- buona protezione dagli effetti dei buchi di trasmissione che provocano le raffiche di errori;
- qualità (definita in termini di MOS) decente anche in presenza di affievolimenti profondi e interferenze da iso-canale;
- buona utilizzazione della banda;
- ritardo di trasmissione, introdotto dai codec e dal processo di protezione, dell'ordine di 80÷90 ms (molto alto);
- necessità di cancellatori d'eco nella tratta radio;
- ritardo alla risposta notevole (anche con eco cancellata).

Gli ultimi tre punti mettono in evidenza il prezzo pagato per ottenere le buone prestazioni dei primi tre punti.

Va aggiunto anche un punto di debolezza del sistema, che le prove di MOS non possono mettere in evidenza: e cioè la scarsa riconoscibilità della voce del parlatore. Molte persone, specialmente al primo approccio col radiotelefono GSM, si lamentano infatti di non riconoscere l'interlocutore anche se molto familiare.

In recenti riunioni (ad Helsinki e Parigi), un gruppo di esperti ha esaminato, fra gli altri, questo problema; l'obiettivo del gruppo è di proporre un nuovo codec GSM, migliorato. Al gruppo sono pervenuti contributi, fra i quali uno mette in evidenza la troppa dipendenza dal tipo di parlatore, nel codec attuale, della qualità ottenibile.

Si sta discutendo se mantenere fissi alcuni punti come, per esempio, il flusso di 13 kbit/s, e il tipo di codifica, o passare addirittura ad altri tipi. In USA si stanno sviluppando forti preferenze, anche ad alto livello accademico, verso la codifica CDMA (Code Division Multiple Access) che è ben nota e presenta un grosso pregio: la degradazione della qualità in presenza di fading e interferenze è graduale e non brusca come nel sistema usato per il GSM. Attendiamo, quindi, le proposte di questo gruppo, che però ha già dato un nome alla versione migliorata del codec attuale: GSM-EFR (in cui EFR = Enhanced Full Rate, cioè Migliorato 13 kbit/s).

Concludendo: il telefono radiomobile numerico in collegamento con un TFA dà buone prestazioni:

a) ma provoca nell'interlocutore un po' di difficoltà nel riconoscere chi parla;
b) sopprime l'eco, ma resta un notevole ritardo alla risposta (cioè un effetto psicologico di incertezza); questo effetto vale per entrambi gli interlocutori. Il ritardo è circa 2·90=180 ms.

Certamente questi effetti possono essere tollerati, ma occorre farci un po' l'abitudine: sembra peraltro che sia opportuno prendere in considerazione questi problemi per cercare di migliorare i punti deboli, cosa che comunque è già allo studio del gruppo prima menzionato; questo gruppo ha anche preso in considerazione la possibilità di tandem (TFN con TFN per esempio), condizione che, ovviamente, esalta i problemi.

2.5 *Casi RTA-RTA, RTN-RTA, RTN-RTN*

Il primo di questi casi, radiomobile analogico con radiomobile analogico, non presenta difficoltà di esame. Sembra utile ricordare che il collegamento risulta completamente a 4 fili con la scomparsa pressoché totale di ogni eco. Le due code dal MSC, al radiotelefono e viceversa, non comportano ritardi di trasmissione apprezzabili, e quindi non si ha peggioramento della buona situazione generale presente sulla rete delle giunzioni (che ha un basso ritardo): per esempio, su tutta la rete italiana il circuito numerico MSC-SGU-SGU-MSC ha un ritardo veramente basso: su fibra ottica (con velocità di trasmissione di 200.000 km/s) su 2.000 km, che è una distanza massima, il ritardo è 10 ms.

Ovviamente, il pericolo di incontrare buchi di propagazione, o interferenze, cresce data la presenza di due tratte di radiomobile anziché di una. Però, questo non sembra un grosso problema, e il collegamento RTA con RTA sulla nostra rete non presenta difficoltà e può avvenire con buona qualità.

Il caso RTA con RTN presenta sicuramente, in linea di principio, una buona qualità; c'è anche, rispetto al caso di telefoni radiomobili con telefoni fissi analogici, il miglioramento dovuto alla sparizione dell'eco. Resta però il ritardo alla risposta, dovuto in gran parte al telefono RTN a causa del trattamento del segnale.

Anche il caso RTN con RTN, essendo a 2 vie, non dà fenomeni di eco. Purtroppo, il ritardo alla risposta diviene (come sarebbe stato per l'eco) quello che compete alla presenza di due tratte radio equipaggiate con codec GSM; questo comporta circa 160÷180 ms di ritardo in un senso più il ritardo di trasmissione della rete e quindi ritardo alla risposta di 0,32÷0,36 secondi fino anche a 0,4 secondi, valore che rende certamente molto fastidioso l'effetto psicologico di cui abbiamo già parlato; questo valore è da tutte le organizzazioni internazionali considerato al limite della tollerabilità.

La qualità della conversazione rimane buona dal punto di vista della comprensibilità. Resta il fatto che il doppio trattamento sul segnale audio, basato su analisi e sintesi sulle due tratte radio, non può che dare un risultato meno buono, dal punto di vista della dipendenza della qualità dall'identità del parlatore. In poche parole, la comprensibilità resterà abbastanza buona, ma la riconoscibilità del parlatore e la naturalezza della voce non potranno che soffrire per causa di questo abbinamento RTN con RTN.

3. **Conclusioni**

Dopo il diffuso esame dell'effetto delle code sulla qualità trasmissiva per il servizio telefonico, possiamo riassumere i risultati in questo modo:

- la numerizzazione della rete da centrale terminale a centrale terminale, che è quasi completa oggi e lo sarà del tutto tra breve, comporta un notevole effetto migliorativo sulla qualità telefonica in ogni caso, sia per i collegamenti fra telefoni fissi analogici o numerici, sia per i collegamenti con o fra radiomobili. Questo avviene a causa delle prestazioni del circuito numerico (giunzione) e cioè : a) equivalente fisso fra due centrali estreme tenute a un valore basso (7 dB di equivalente a 1020 Hz, il che significa IIS ≈ 8 dB); b) ottima curva di risposta di questa giunzione dovuta alla presenza di due soli codec a 64 kbit/s e transiti tutti in fase numerica; c) basso rumore della giunzione dovuto quasi completamente alla quantizzazione; gli eventuali rumori dovuti ad errori di linea isolati o raffiche sono tollerati dal segnale telefonico fino a valori molto alti (per esempio un tasso di errore di 10^{-3} non porta conseguenze percepibili se non come piccoli disturbi momentanei);
- la buona qualità fornita dalla giunzione numerica fa sì che anche con telefoni fissi analogici "vecchi", connessi alla centrale (SL) [10] con rilegamenti in rame e dimensionati con ER (equivalenti di riferimento) in funzione della vecchia rete di giunzioni analogiche, la qualità complessiva migliori notevolmente rispetto a quella che era presente sulla vecchia rete. Il fatto che in molti casi l'IIS totale è piuttosto basso (voce troppo forte all'orecchio), non è un grosso inconveniente; certo è meglio sia così piuttosto che una voce troppo debole all'orecchio; infatti basta scostare il microtelefono dall'orecchio anche di poco e si rimedia;
- per i telefoni analogici "nuovi", cioè installati in questi anni, si riproduce la situazione descritta per i "vecchi", però un po' migliorata; purtroppo la norma CEI, sui sistemi d'utente, del 1993 non ha tenuto molto conto dell'obiettivo proposto per l'IIS globale dal Piano Regolatore Nazionale delle Telecomunicazioni del 1990. Si ha una distribuzione dei valori IIS dei collegamenti completi un po' meno dispersa che per i telefoni "vecchi", con tendenza a IIS troppo bassi. Se si usasse un solo tipo di telefono (per esempio), la situazione generale migliorerebbe. In ogni caso sembrerebbe utile ritoccare la norma CEI, dato che fra non molto tutte le giunzioni saranno numeriche. E' ovvio che va ripetuta la considerazione di prima: è meglio un IIS troppo basso che il contrario (l'IIS troppo alto non è rimediabile dall'utente se non con telefoni con regolazione manuale del volume ricevuto);
- i collegamenti fra telefoni radiomobili analogici e quelli fissi sono senz'altro di buona qualità, fino a che guai di propagazione e interferenze non influiscono;
- i collegamenti fra telefoni radiomobili numerici (GSM) e quelli fissi, sono anch'essi buoni e più resistenti ai fading e interferenze che possono avvenire nella tratta radio. Si ha però, dato il ritardo dovuto al trattamento della voce e alla protezione dagli errori, la necessità di cancellazione dell'eco. Esiste un effetto psicologico negativo dovuto al ritardo della risposta (piuttosto pesante). Il trattamento della voce è tale da provocare una scarsa riconoscibilità della voce del parlatore, e incide anche sulla naturalezza;
- infine i collegamenti tra telefoni radiomobili, senz'altro possibili, aumentano gli effetti presenti nei collegamenti fra radiomobili e fissi. Più importanti sono gli effetti per il collegamento RTN con RTN, in cui si arriva a ritardi della risposta troppo alti.

In conclusione, in questo periodo di transizione verso una rete completamente numerica fino ai telefoni, il progettista di rete si troverà di fronte a questa situazione:

- i problemi posti dai telefoni analogici non sembrano molto pesanti: non sarebbe però male por mano alle necessarie modifiche sulle norme CEI. In fin dei conti telefoni nuovi si introdurranno in rete ancora (credo) per un decennio e forse più; sarebbe bene prendere l'occasione per chiedere l'abbassamento[5] della frequenza inferiore della banda trasmessa a 150÷200 Hz;
- i problemi maggiori sono quelli presentati dai ritardi di trasmissione, e dai trattamenti sulla voce che incidono sulla naturalezza e riconoscibilità del parlatore. Qui occorre una riflessione, come sembra stia per fare, ad esempio, il gruppo costituito in seno al GSM.

4. **Evoluzione delle norme internazionali**

4.1 *CCITT e ITU-T*

La raccomandazione P80 del CCITT, che riguarda il modo per eseguire misure soggettive di qualità, è stata revisionata e una delle varianti apportata sembra degna di essere ricordata: essa riguarda la scala della valutazione immediata[6] (MOS), essendo stato variato il punteggio

da dare ai giudizi. Anche altre raccomandazioni sulla qualità (P81, P83, P85) hanno subito ritocchi che non sembra qui il caso di riportare, perché non affrontano il problema "naturalezza" e "riconoscibilità".

	Nuovo	Vecchio
E (excellent)	5	4
G (good)	4	3
F (fair)	3	2
P (poor)	2	1
B (bad)	1	0

Tabella 3

Non è chiaro perché sia stata fatta la variante sulla scala dei MOS, che, di fatto, può causare confusione nel confronto fra vecchie e nuove misure; va quindi sempre fatta molta attenzione alla data delle misure.

Sono state poi ridefinite le scale per la valutazione nelle prove di

"Listening-Quality" (da 1 a 5)
"Listening-Effort" (da 1 a 5)
"Loudness-Preference" (da 1 a 5)

Va anche ricordata la contribuzione n. 6 allo Study Group XII [11] dell'ITU-T circa una proposta di valutazione della voce ottenuta a mezzo di dispositivi automatici: in questa (doc. citato, pag. 9) esiste una scala per la gradevolezza (pleasantness) della voce. Per la voce umana sembra non vi sia ancora nessuna proposta per la "naturalezza" e "riconoscimento del parlatore".

4.2 *ETSI*

L'ETSI (European Telecommunications Standards Institute) ha prodotto nell'ottobre del '94 un documento, per ora in bozza, molto pregevole. Esso riguarda gli obiettivi di progetto circa il ritardo di trasmissione sulle reti numeriche che si stanno sviluppando. E' ovvio che questi obiettivi possono valere anche per le reti analogiche o le parti di esse ancora rimanenti.

Il documento comincia con l'osservare che oggi c'è poca esperienza sugli effetti dell'eco e del ritardo, e quindi i loro effetti sono visti spesso come "nuove" degradazioni della qualità. (Ovviamente la "novità" non riguarda chi ricorda come si parlava con eco molto forte sui vecchi e lunghi circuiti pupinizzati dotati per questo di soppressori d'eco).

In un punto del documento l'ETSI riporta la raccomandazione G114 dell'ITU-T (ex CCITT) e classifica i ritardi di trasmissione così:

campo 0-25 ms — non occorrono speciali precauzioni;
campo 25-150 ms — occorrono apparati per la riduzione dell'eco;
campo 150-400 ms — come sopra, ma esiste una difficoltà (crescente col ritardo) per l'effetto psicologico dovuto al ritardo della risposta. La conversazione diviene frammentaria (disruption of conversational flow);
campo >400 ms — collegamenti siffatti vanno evitati.

Sempre l'ETSI ha prodotto un altro documento (12 gennaio 1995) dal titolo "Transmission and Multiplexing" sulla qualità della voce trasmessa con banda larga 3,1 kHz.

E' un bel documento riassuntivo che dovrebbe essere letto dai programmatori della rete: dopo una rassegna generale dei parametri trasmissivi che riguardano la voce, a partire dalla attenuazione della pressione da bocca a orecchio, si arriva a suggerire i limiti entro cui i vari parametri trasmissivi sono accettabili. Della rassegna ricordiamo alcuni punti importanti:

- ritardo ed eco nei sistemi mobili;
- degradazione (valutata in QDU) per i vari codec nei sistemi mobili;
- modello ETSI per la valutazione della qualità vocale su un collegamento. Questo modello usa parametri già noti, ma il tutto è presentato molto coerentemente;
- esame dei metodi di misura in servizio non intrusivi.

Il documento è corredato da una serie di Annessi (da A a K), alcuni semplici da leggere, altri richiedenti una buona preparazione sull'argomento.

In definitiva, sembra che qualcosa si stia muovendo negli organismi internazionali: ritengo che per ogni gestore esista la convenienza a spingere perché la qualità telefonica non venga trattata come merce: cioè una qualità mediocre si fa pagare poco, e l'inverso. Un livello di qualità tale da essere accettabile per tutti gli utenti va forse discusso ed eventualmente determinato dall'autorità di controllo.

Appendice 1

Analisi e sintesi della voce: il vocoder

Molti anni fa (anni '30) si osservò che la voce, o meglio il segnale elettrico corrispondente uscente da un microtelefono, avrebbe potuto essere trasmesso in una banda di frequenza inferiore a quella occupata naturalmente.

(5) Questo abbassamento coinvolge anche i ponti di alimentazione in centrale, ed eventuali rumori sui rilegamenti. Non v'è dubbio che va richiesto anche e soprattutto per i telefoni numerici sia fissi che mobili.

(6) E' stata in pratica variata la scala dei giudizi 0÷4 in 1÷5 aggiungendo 1 a tutti i punteggi.

Questa osservazione partiva dall'esame spettrale della voce che veniva fatto analizzando lo spettro vocale all'uscita di un buon microfono con stretti filtri (molto usati quelli di un terzo di ottava); ci si accorse che l'ampiezza del segnale in uscita da ogni filtro non fluttuava molto, anzi le sue variazioni temporali più veloci avvenivano con un gradiente che corrispondeva a frequenze non maggiori di una ventina di Hz.

Questa considerazione, insieme al fatto che si capì che la sorgente fonica era o periodica (come per i suoni vocalici) o costituita da un rumore a spettro distribuito (come per gli altri suoni), diede modo di effettuare la valutazione che faremo qui di seguito; anche la frequenza e l'ampiezza della fondamentale e quella del rumore generato dalla sorgente variavano lentamente.

Supponiamo di usare da 100 Hz a 3200 Hz (5 ottave) 15 filtri di un terzo di ottava, di misurare con continuità la frequenza e l'ampiezza della fondamentale (oppure l'ampiezza necessaria per il generatore di rumore), e parimenti l'ampiezza (efficace) del segnale in uscita di ognuno dei quindici filtri; ebbene, siamo in presenza di:

- 15 segnali (ampiezze in uscita dei filtri);
- 1 segnale per la frequenza della fondamentale;
- 1 segnale per l'ampiezza della fondamentale;
- 1 segnale per l'ampiezza della sorgente di rumore.

Tutti questi segnali sono compresi in una banda di circa 20 Hz data la loro poca variabilità e contengono, si pensava, tutta la informazione necessaria a ricostruire la voce.

Consegue, da tutto quanto detto poc'anzi, che l'informazione essenziale circa lo spettro vocale è contenuta entro 360 Hz = (15+3)·20 Hz.

Sembrò quindi, in base a questa considerazione, che il contenuto spettrale della voce potesse essere trasmesso entro poche centinaia di Hz. Basandosi su questi ragionamenti, verso la fine degli anni '30 furono realizzati un "Voder" [8] (sintetizzatore della voce) e un "Vocoder" [9] (analizzatore e successivo sintetizzatore della voce). Le figg. A1-1 e A1-2 mostrano il principio di funzionamento di questi apparati.

Il Voder era uno strumento manuale: aveva la banda vocale suddivisa fra 10 filtri contigui attivati con tasti sotto le dita delle mani dell'operatore; col polso si commutava la sorgente dall'oscillatore (non sinusoidale ma di rilassamento, cioè ricco di armoniche) al generatore di rumore, e con un piede si regolava un pedale che comandava la frequenza dell'oscillatore. Altri tre tasti (sotto le dita) permettevano di generare le consonanti occlusive. A parte il fatto che occorreva un operatore molto abile, i risultati furono molto incoraggianti. Ma risultati ancora più interessanti si ottennero dal vocoder(7); in questo i filtri di analisi e poi di sintesi erano sempre dieci, con larghezza di banda di 300 Hz (fissa) a partire da 0 fino a 3000 Hz; l'intelligibilità ottenuta tramite questo apparato era buona, e era anche possibile riconoscere dalla voce chi fosse la persona che parlava. Si otteneva un forte risparmio di banda. A Dudley seguirono molti altri, studiando e realizzando altri vocoder: M.R. Schroeder sviluppò il vocoder a correlazione, J.L. Flanagan il vocoder a formanti, e poi lo stesso Flanagan altri tipi basati su principi diversi dal puro andamento spettrale, come per esempio la divisione per due delle frequenze in uscita dai filtri (che erano soltanto 3), la trasmissione di una banda ridotta a 1/2, e il recupero mediante moltiplicazione per 2 delle frequenze delle 3 bande all'arrivo.

Questi tentativi iniziali, durati però a lungo, cioè fino agli anni '60, anche se davano una qualità decente, richiedevano apparati complessi; se ci si limitava al guadagno di banda ottenibile pur assicurando una buona qualità (in genere da 3 a 1) il costo della loro presenza non ripagava il risparmio di kHz-km di linea.

Il problema dell'alto costo dell'apparato di analisi e sintesi era intrinsecamente agganciato alla impossibilità di operare sul segnale in maniera semplice seguendo le tecniche analogiche; si ottenne invece una grande semplificazione degli apparati quando si fu in grado di procedere alla numerizzazione del segnale: infatti se il segnale è numerizzato è possibile memorizzare per il tempo necessario all'analisi una breve serie di numeri, e le operazioni fattibili si ampliano enormemente; per esempio effettuare una FFT (Fast Fourier Transform) diviene semplice.

I principi finora esposti sono solo la base su cui si è innestato lo straordinario sviluppo di questo ramo del trattamento del segnale, per la conoscenza del quale non si può che rimandare ai lavori specialistici.

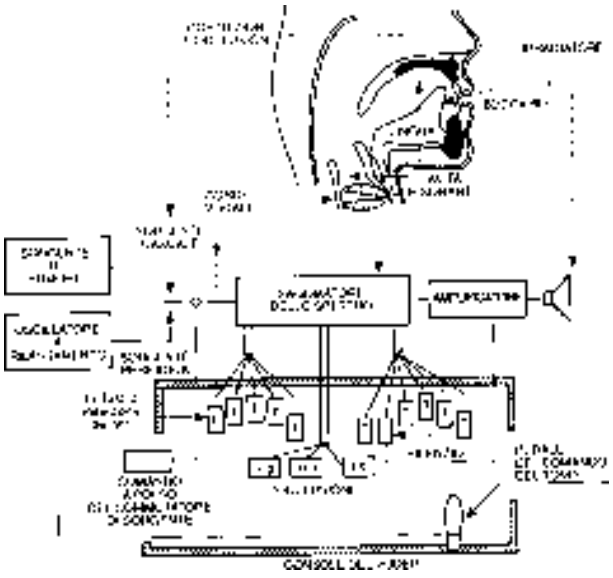


Figura A1-1 Schema di principio del "Voder", sintetizzatore della voce

(7) Il Vocoder fu usato molto a scopi militari perché consentiva di ottenere un linguaggio segreto, alterando l'ordine dei canali informativi in partenza e ripristinando l'informazione corretta in arrivo.

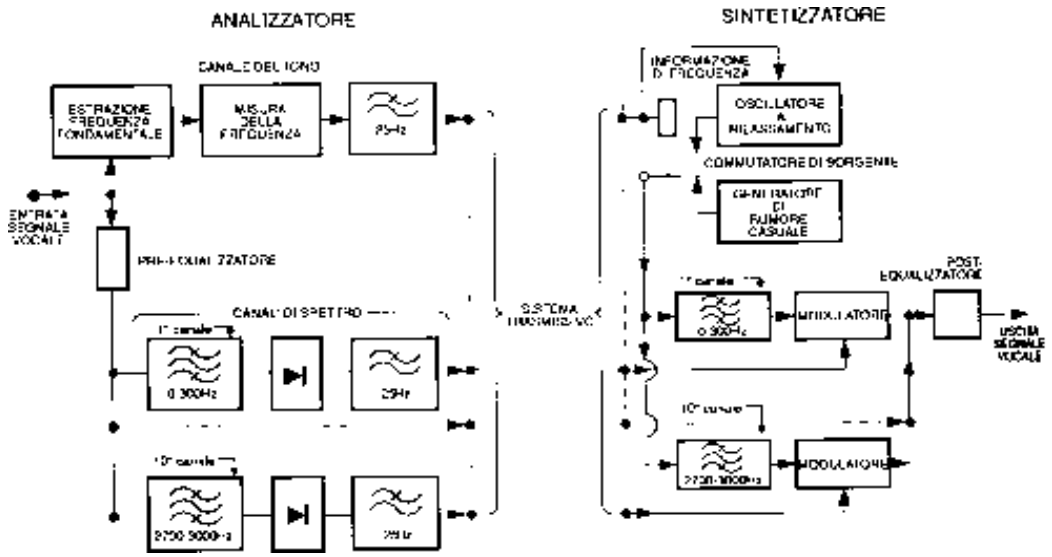


Figura A1-2 Schema di principio del "Vocoder" di Dudley. I canali che trasportano l'informazione possono essere realizzati su un sistema trasmissivo in qualunque modo; si vede subito che la banda necessaria (utile) è di 11·25 = 275 Hz. Undici canali di questa larghezza occuperebbero, se realizzati in FDM, una banda lorda fra 350 e 400 Hz; in PCM occorrerebbero almeno (con 4 bit per campione mediamente) 4·11·50 = 2200 bit/s

Appendice 2

Ritardo di trasmissione e suoi effetti sulla telefonia

Il ritardo di trasmissione fra l'istante in cui la parola è pronunciata e quello in cui è ascoltata nasce in vari punti e per varie ragioni in una rete e cioè, ricordando solo le fonti più importanti:

a) velocità di propagazione finita sui mezzi trasmissivi (all'incirca 300.000 km/s via etere e coassiali normalizzati terrestri, 200.000 km/s su fibra ottica, e così via);
b) trattamento del segnale nei codec; questo ritardo è molto basso per il codec PCM a 8 bit/campione, può diventare notevole se si fa trattamento del segnale, per portare ad esempio la parola a qualche kbit/s;
c) trattamento del segnale dopo codifica sulla linea per ottenere particolari prestazioni trasmissive: rientra in questo caso il trattamento per la riduzione degli errori di linea su segnali numerici;
d) ogni nodo di commutazione a divisione di tempo introduce un ritardo: minimo nelle centrali che operano sul PCM classico, ma a volte notevole in centrali che operano, per esempio, con commutazione di blocchi di bit con indirizzamento (pacchetto, ecc.);
e) anche i centri in cui si opera uno smistamento di gruppi di canali su segnali numerizzati introducono un ritardo, in genere molto basso;
f) anche una eventuale cifratura e decifratura (con lo scopo di assicurare la riservatezza) introduce un ritardo sul segnale.

Per citare un caso, il ritardo globale "trasmissione più ricezione" sui sistemi cellulari per telefonia GSM, dovuto alla codifica sull'ordine dei 13 kbit/s, e al trattamento per la riduzione degli errori e alla cifra e decifra è dell'ordine di parecchie decine di millisecondi.

Il ritardo di trasmissione ha diversi effetti sulla voce: il primo più noto e più studiato è l'eco; dai suoi effetti dannosi sulla conversazione ci si può difendere con diversi mezzi, dei quali il più usato oggi è la compensazione del segnale d'eco grazie all'introduzione del segnale in avanti con fase, ampiezza e ritardo opportuni sul segnale di ritorno, in modo da portare la somma dei due a un valore molto basso. Ovviamente, se le sorgenti di eco sono più d'una, il problema si complica e conviene studiare la situazione e collocare il sistema o i sistemi di neutralizzazione nella posizione più adatta.

Il sistema di neutralizzazione oggi viene comunemente chiamato "cancellatore d'eco", per quanto non si tratti di vera cancellazione, ma compensazione fino a una buona riduzione del livello d'eco. Le figg. A2-1, A2-2 e A2-3 mostrano gli effetti dell'eco sulla qualità della conversazione.

Il secondo effetto notevole del ritardo di trasmissione è un effetto psicologico sulla persona che parla e al termine attende la risposta: questa giunge con un ritardo un po' maggiore del ritardo dell'eco (cioè 2 volte il tempo di trasmissione + il tempo dovuto ai riflessi dell'interlocutore). Si è già notato che quando questo ritardo totale giunge a qualche frazione di secondo, il parlatore iniziale tende a riprendere a parlare troppo

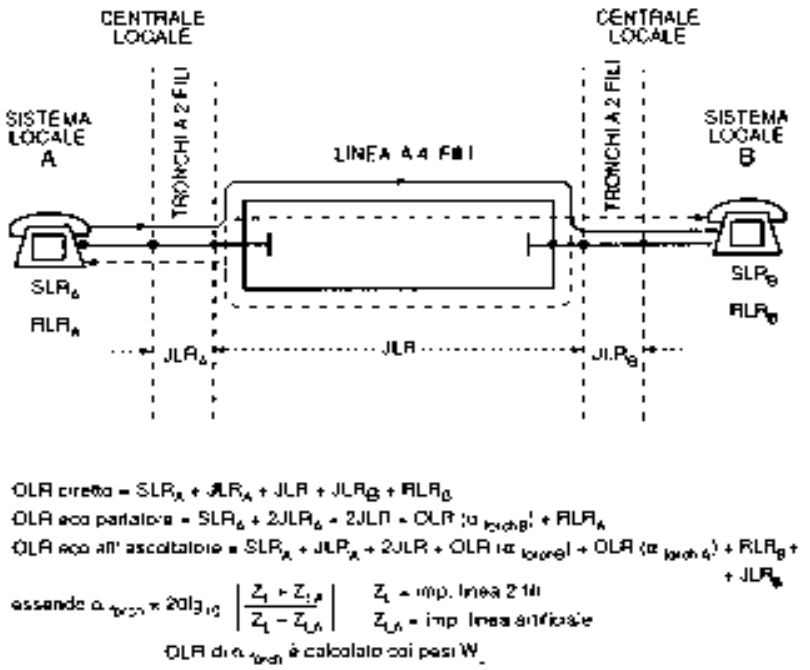
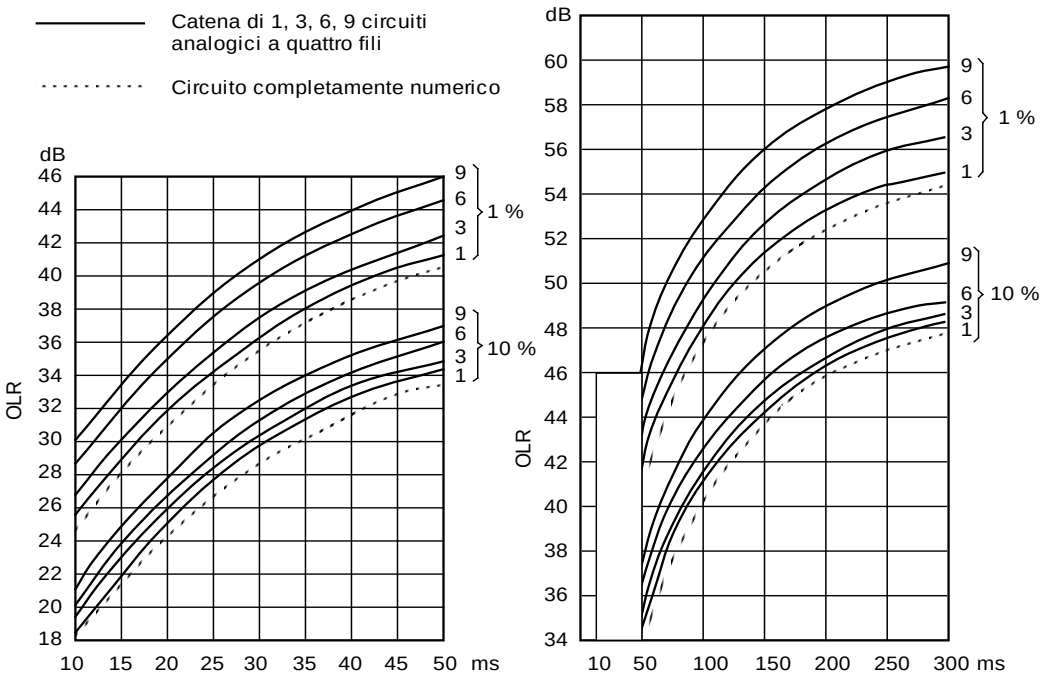


Figura A2-1 Definizione delle grandezze interessanti la valutazione dell'attenuazione d'eco



Nota 1 - La figura di sinistra presenta in dettaglio l'OLR per i tempi di propagazione compresi tra 10 e 50 ms

Nota 2 - Le percentuali si riferiscono alla probabilità di incontrare un'eco fastidiosa per il parlatore

Nota 3 - Il valore di OLR dell'eco si ottiene, qui, sommando:
- gli LR nelle due direzioni del sistema locale di chi parla;
- gli LR nelle due direzioni della catena di circuiti fra la terminazione a 2 fili verso il sistema locale dalla parte del parlatore e la terminazione 4/2 fili del collegamento a 4 fili dalla parte dell'ascoltatore;
- il valore medio dell'attenuazione di equilibrio della stessa terminazione 4/2 fili

Figura A2-2 Valori di OLR dell'eco, raccomandati dal CCITT, sotto i quali si può incontrare un'eco fastidiosa per il parlatore (con probabilità dell'1% o del 10%)

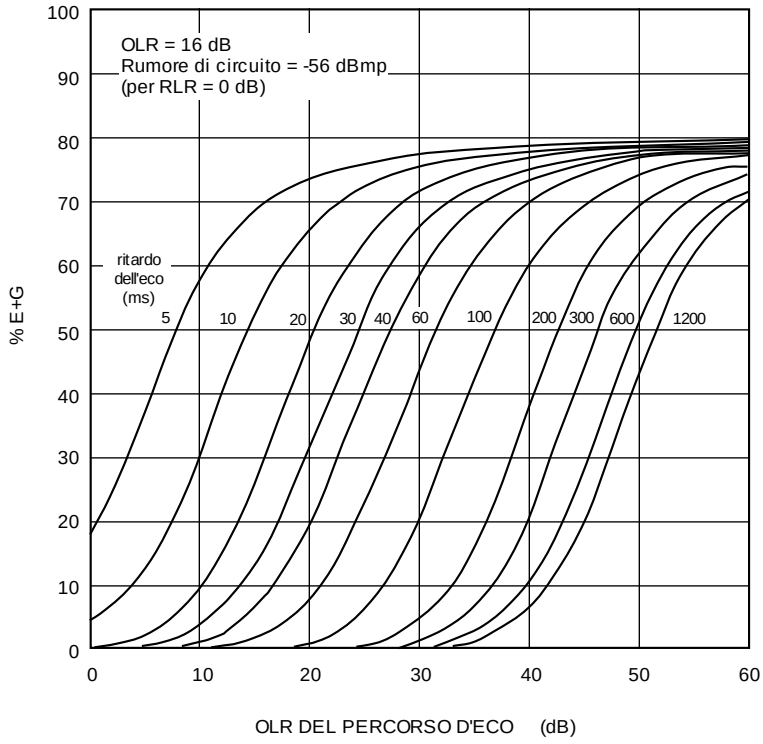


Figura A2-3 Percentuali di giudizi favorevoli in funzione dell'OLR e del ritardo d'eco

presto perché crede che l'altro non abbia capito. Su questo inconveniente non c'è nessun rimedio tecnico; sarà solo l'abitudine degli utenti a parlare su collegamenti con questo difetto che consentirà una agevole conversazione.

Purtroppo, può darsi che in futuro anche su brevi collegamenti si avranno notevoli ritardi, come sulle code dei telefoni cellulari GSM. La tab. A2-1 mostra i ritardi alla risposta in varie situazioni, per collegamenti in cui domina il ritardo di propagazione nel mezzo trasmissivo.

Distanza →	100 km	1000 km	10.000 km	20-30.000 km (antipodi)
fibra	1	10	100	∼ 300
1 salto di satellite geostazionario	500	500	500	—
2 salti di satelliti geostazionari	—	—	—	1000
misto fibra, satellite	—	—	—	∼ 600
Velocità sulle fibre ∼ 200.000 km/s				

Tabella A2-1 Ritardo in ms alla risposta (doppio del tempo di propagazione) in assenza di ritardi per eventuali elaborazioni

Appendice 3

Rumore totale in ricezione

Ci riferiamo solo al caso che stiamo esaminando, e cioè rete completamente numerica (a 64 kbit/s per ogni canale telefonico) fra le centrali che servono l'utente e code dei vari tipi elencate nel testo.

Alla capsula ricevente del telefono terminale arrivano questi rumori dominanti:

a) rumore di quantizzazione del circuito utilizzato sulla rete di giunzione, attenuato dalla linea di rilegamento (coda). Dobbiamo tradurre questo rumore in rumore di tipo additivo (R_Q) secondo la formula

$$R_Q = -3 - \mathrm{IIS} - 2{,}2\,Q \quad \mathrm{dBmp}$$

in cui R_Q è il rumore di tipo additivo al terminale del collegamento PCM, e quindi in entrata al rilegamento, IIS è l'indice di intensità soggettiva del circuito (numerico) di giunzione che il Piano Regolatore Nazionale indica di regolare a 7 dB di equivalente a 1020 Hz, Q è il rapporto segnale rumore nel circuito stesso. Tenuto conto che Q vale circa 37 dB per un vasto campo di livelli sonori, che IIS è circa 8 dB, si ha che R_Q può essere tranquillamente preso eguale a

$$R_Q = -3 - 8 - (2,2 \cdot 37) = -92,4 \quad \text{dBmp}$$

Un rumore del genere può essere tranquillamente ignorato. Non così per il rumore di fondo del codec terminale che è normalmente di –65 ÷ –68 dBmp in uscita.

Questo rumore va all'orecchio, attenuato dall'ISR del sistema d'utente, nel caso di telefoni analogici. Se questo IIS in ricezione vale *x* dB il rumore proveniente dalla centrale fino al telefono ($R_{tel\,L}$)

$$R_{tel\,L} \cong -67 - x \quad \text{dBmp}$$

Ma al ricevitore giunge anche il rumore d'ambiente attraverso l'effetto locale: questo rumore si può considerare come un rumore di linea pari a R_{eqA} deducibile dalla fig. A3-1.

La figura è parametrata in STMR, che è l'attenuazione fra la potenza del rumore di ambiente R_A e quella che giunge all'orecchio (riferita come al solito all'attenuazione dell'IRS). La potenza R_A del rumore di ambiente è quella che si misura con un particolare strumento (specifiche IEC - Rec. Pubblication 179 del 1965). Si ha una indicazione di 50 dBA, quando la banda di 1/3 di ottava intorno a 800 Hz presenta una densità spaziale di potenza pari a 40,4 dB sul livello di 10^{-16} W/cm^2.

Determinato R_{eqA}, esso va sommato in potenza a R_{telL} e questo è il rumore totale con cui entrare sulle ascisse dei diagrammi che in funzione di *R* e IIS danno le % di giudizi favorevoli.

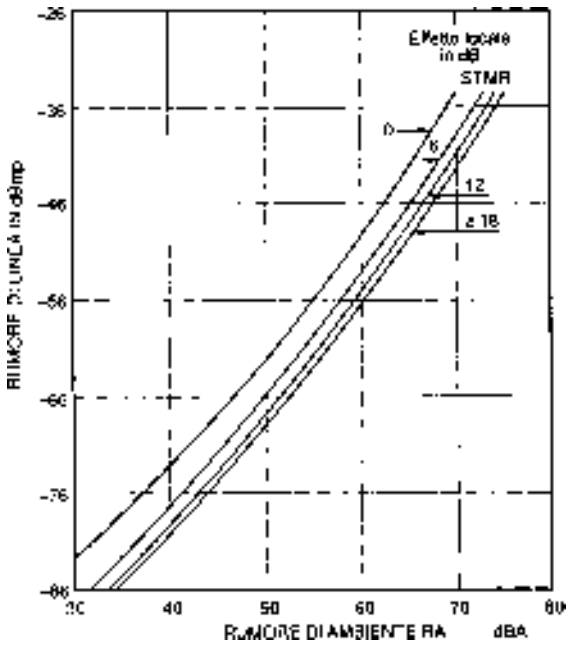


Figura A3-1 Equivalenza del rumore di ambiente RA captato tramite l'effetto locale e il rumore di linea R_{eqA}. L'effetto locale viene misurato come STMR e varia normalmente da pochi dB a circa 20 dB

Ovviamente, se il rilegamento pesca rumori per conto suo, come quelli di induzione da linee di potenza, o diafonie da coppie contigue, questo rumore va aggiunto ancora per ottenere il rumore complessivo.

In ambienti riparati, come uffici, abitazioni abbastanza protette dai rumori stradali, e su rilegamenti in coppie in rame non antichissimi (escluse linee in fili nudi aerei) si può contare su valori di R_A dell'ordine di 45÷55 dBA e su rumori captati dal rilegamento R_L molto bassi (–70 ÷ –60 dBmp).

In definitiva, componendo i contributi

del circuito di giunzione $R = -67 - x$ dBmp

del rilegamento $R_L = -70 \div -60$ dBmp

d'ambiente $R_{eqA} = -70 \div -60$ dBmp

si arriva a rumori globali per la rete che stiamo considerando, e sistemi di utente analogici compresi fra

$$-67 < \text{Rumore totale} < -57 \text{ dBmp}$$

Una buona parte degli utenti sta verso l'estremo inferiore.

Infine, resta un altro tipo di rumore che su reti ben costruite ed esercite è generalmente trascurabile per la telefonia; si tratta del rumore dovuto agli errori di trasmissione sui circuiti numerici. E' chiaro che, nel caso della telefonia, dato che i vari bit dell'ottetto significano valori molto diversi fra loro, a seconda del bit colpito dall'errore, si ha un impulso di rumore di entità diversa (impulso che poi viene allungato e spianato dal filtro passa basso finale del codec). Pur tuttavia si sono scritte formule che, in base a considerazioni statistiche, danno il rumore equivalente sul circuito in funzione del tasso di errore che chiameremo *p*. Se $p<10^{-4}$, non vale la pena di effettuare calcoli. Ad ogni modo, il rapporto Segnale su Rumore per questo tipo di rumore vale

$$(S/R)_{dB} = 10 \lg\left(\frac{1-4p}{4p}\right) \approx -10 \lg 4p = -6 - 10 \lg p \quad \text{dB}$$

Per $p=10^{-4}$, che è un tasso di errore oltre il quale non si va nelle reti fisse, si ha *S/R* = 34 dB, cioè all'incirca eguale al rumore di quantizzazione (*S/R* cioè = 37 dB) e quindi possiamo trascurarlo. Anche un tasso $p=10^{-3}$ per telefonia a 64 kbit/s non porta notevole danno. Notevole è invece il danno, come si è specificato nel testo, che *p* elevati portano al segnale codificato a pochi kbit/s, cioè dopo trattamento del segnale.

Questo concetto, ovviamente, non vale nel caso di trasmissione di dati.

Bibliografia

[1] Jayant: *High Quality Coding of Telephone Speech and Wideband Audio*. «IEEE Comm.», January 1990, pag. 10.

[2] Bonavoglia, L.: *La qualità della trasmissione telefonica - Parte prima*. «Notiziario Tecnico Telecom Italia», Vol. 3, n. 2, Agosto 1994, pp. 27-39.

[3] Crochiere, R.E.; Flanagan, J.L.: *Current Perspective in digital Speech.* «IEEE Comm. Magazine», 1983, p. 32.

[4] Mossotto, C.: *Speech Technology and Telecommunications.* «CSELT Technical Reports», March 1992, p. 5, fig. 5.

[5] Casale, R.; Tortia, G.: *Evoluzione delle caratteristiche trasmissive dalla rete analogica a quella numerica*. «Elettronica e Telecomunicazioni», 1985.

[6] Casale, R.; Tortia, G.: *Digital goes to subscriber: discussion on trasmission aspects.*

[7] Grimaldi, F.; Zingarelli, V.: *Sistemi radiomobili cellulari*. SSGRR, L'Aquila, 1991.

[8] Dudley, H.W.; Riesz, R.R.; Watkins, S.A.: *A Synthetic Speaker*. «J. of the Franklin Institute», Vol. 227, n. 6, p. 748, Giugno 1939.

[9] Dudley, H.W.: *The Vocoder*. Bell Laboratories Record, Vol. 17, p. 123, 1939.

[10] Piano Regolatore Nazionale delle Telecomunicazioni, 1990.

[11] ITU-T, Study Group XII, Period 1993-1996, Contribution 6.

ALFA: un parametro per la valutazione dell'efficienza degli autocommutatori numerici UT

F. Lentini, A. Renna, M. Venuto (*)

Un impianto di commutazione numerica fornisce una quantità rilevante di segnalazioni di allarme che, pur dettagliando le cause delle anomalie, non consentono di avere un valore di sintesi dell'efficienza dell'impianto e quindi dello stato dell'autocommutatore. La mancanza di un indice quantitativo rende quindi difficile valutare il grado di efficacia degli interventi di manutenzione preventiva effettuati dal centro di lavoro.

Per cercare di soddisfare queste esigenze è stato sviluppato dalla Direzione Territoriale Rete Roma un prodotto informatico denominato ALFA (Automatic fauLt and eFficiency Analysis) che sintetizza, mediante tabelle facilmente consultabili ed interpretabili, gli allarmi spontanei emessi dagli autocommutatori numerici UT in tecnica ITALTEL e fornisce il valore di efficienza "α" dell'impianto legato al numero di allarmi rilevati in opportuni intervalli temporali ed all'importanza delle parti degradate.

1. Introduzione

L'inserimento in rete delle centrali di commutazione numerica, insieme alla numerizzazione della rete trasmissiva, consente di raggiungere valori sempre più elevati di trasparenza della rete di telecomunicazioni, manifestati dal miglioramento dell'indicatore *TER* (*Tasso di Efficacia di Rete*) il quale misura la percentuale delle connessioni che la rete riesce ad instaurare sul totale delle richieste dell'utenza. A tutto ciò le centrali numeriche uniscono la possibilità di offrire al cliente una varietà crescente di servizi aggiuntivi arricchendo di fatto le prestazioni ottenibili dalla rete.

Oltre a questi aspetti, l'attenzione del gestore di una rete di TLC è anche rivolta al controllo ed al mantenimento degli standard qualitativi dell'impianto e quindi all'analisi degli indicatori di controllo dell'impianto ed alla pianificazione degli interventi di manutenzione preventiva.

L'esperienza maturata con l'esercizio delle centrali di tecnica analogica aveva portato a sviluppare tutta una serie di attività di manutenzione preventiva (controllo a vista dei selettori, revisione periodica, chiamate di prova manuali ed automatiche, ecc.) che consentivano al responsabile dell'impianto di avere sempre degli indicatori di sintesi che rendessero conto dello stato dell'impianto e sui quali basarsi per indirizzare le attività di manutenzione.

Visto nell'ottica della manutenzione e non delle prestazioni fornite, un impianto di commutazione numerica fornisce una quantità rilevante di segnalazioni che, pur rendendo conto di singoli eventi anomali, non consentono di avere un valore di sintesi dell'efficienza dell'impianto e quindi dello stato dell'autocommutatore. La mancanza di un indice globale inoltre rende difficile valutare il grado di efficacia degli interventi di manutenzione preventiva effettuati dai tecnici.

Per cercare di venire incontro a queste esigenze è stato sviluppato della DTRE-RM (ex DR/RM) un prodotto denominato *ALFA* (Automatic fauLt and eFficiency Analysis) che, attraverso l'analisi degli allarmi spontanei emessi dagli autocommutatori numerici UT100 (tecnica ITALTEL), fornisce un valore di efficienza dell'impianto legato al numero di allarmi rilevati in opportuni intervalli temporali ed all'importanza delle parti degradate.

Prima di presentare le caratteristiche del prodotto ed i risultati della sperimentazione sul territorio della DTRE ROMA, è opportuno richiamare alcuni concetti sull'architettura dell'autocommutatore numerico UT100.

(*) Ing. Fabrizio Lentini, sig. Alberto Renna, sig. Matteo Venuto -Telecom Italia DTRE Roma

2. Richiami sull'architettura del sistema UT

2.1 *Generalità*

Il sistema UT, prodotto dalla società ITALTEL, svolge funzioni di autocommutatore numerico di media/grande capacità in configurazione UT100 e di piccola capacità in configurazione UT20.

Può essere inserito nella rete telefonica come Stadio di Gruppo Urbano (SGU in configurazione UT100) con i suoi relativi Stadi di Linea (SL) con una capacità massima totale (tra SGU e SL) di 100000 utenti (in funzione del traffico richiesto).

Può inoltre svolgere funzioni di Stadio di Gruppo Teleselettivo (SGT in configurazione UT100) con capacità massima di 60000 giunzioni.

Nella Direzione Territoriale Rete Roma sono presenti 43 impianti UT (64% del totale degli impianti numerici) che gestiscono circa 1.270.000 utenti (65% del totale degli utenti numerici) e 435.000 giunzioni (53% del totale delle giunzioni su impianti numerici).

Nell'area urbana di Roma sono presenti 36 impianti UT di cui 33 svolgono funzioni di SGU (e corrispondenti SL) e 3 svolgono funzioni di SGT.

Nei distretti di Civitavecchia e di Tivoli, la centrale UT effettua inoltre funzioni di SGU di decade 1.

2.2 *Tipologie ed interconnessione dei moduli*

La peculiarità principale dell'autocommutatore UT è la sua struttura. Infatti il sistema UT è composto da moduli indipendenti con propria capacità elaborativa e specializzati a svolgere le diverse funzioni telefoniche.

Le tipologie di moduli sono (fig. 1):

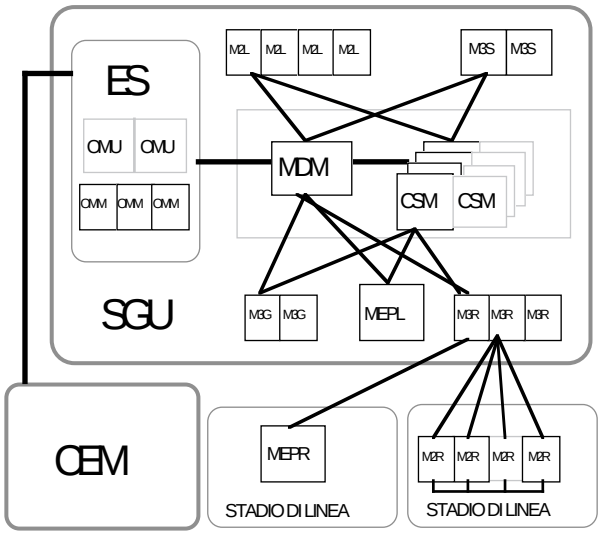


Figura 1 Struttura dell'impianto UT

Moduli utente: M2L, M2R, MEPL, MEPR
Moduli giunzione: M3G, M3S, M3R
Moduli centralizzati: MDM, CSM, OMM, OMU.

La struttura dell'impianto UT in base al tipo di interconnessione tra i moduli può essere di due tipi:

1. stella (configurazione UT100)
2. maglia completa (configurazione UT20 o stadio di linea)

La prima struttura, (utilizzata per configurare centrali di grande/media capacità) utilizza i moduli specializzati denominati CSM e MDM per connettere tra loro, rispettivamente in fonia e in segnalazione, i moduli periferici M2L (utenti) e M3 (giunzioni).

Il sistema ES (Elaboratore di Supporto) svolge funzione di memoria di massa per il data base dei programmi applicativi telefonici (release), di configurazione e di tassazione. Esegue inoltre funzione di interfacciamento con l'operatore di centrale per la gestione locale o remota (tramite CEM Centro di Esercizio e Manutenzione) delle procedure di esercizio e manutenzione.

Il limite di connessione locale della struttura a stella è di 102 moduli locali (tra M2 e M3).

La struttura a maglia completa (utilizzata per configurare le centrali di piccola capacità UT20 e Stadi di Linea) esegue l'interconnessione dei moduli periferici (solo di tipo M2) in maniera diretta (un modulo verso tutti). Il limite di connessione è di 16 moduli.

2.3 *Diagnostica e report di manutenzione*

Ogni parte centralizzata del modulo, quale: rete di commutazione, comando, temporizzazioni, distribuzione dei messaggi intramodulo, ecc, è duplicata (lato 0 - lato 1 in modalità riserva calda) per aumentare l'affidabilità del sistema e ridurre al minimo la probabilità del fuori servizio completo verso l'utenza attestata.

Gli applicativi di diagnostica, residenti nel comando di modulo, effettuano tramite la rete di controllo, un continuo monitoraggio sullo stato qualitativo delle parti hardware: piastre, cavi, ecc.

L'identificazione del guasto è effettuata dagli applicativi di diagnostica che, associando ad ogni piastra hardware uno o più blocchi logici software, denominati *blocchi funzionali* (BLCFUN), individua automaticamente la piastra o il gruppo di piastre sospette guaste.

Più BLCFUN che eseguono funzioni omologhe sono raggruppati in un *blocco di sicurezza* (BLCSIC).

Si esamini ad esempio la fig. 2. La rete di commutazione interna del modulo M2, che esegue la connessione della fonia tra due utenti in conversazione, viene identificata con il blocco di sicurezza "pcm". Tale blocco di sicurezza è l'insieme di diversi blocchi funzionali denominati FBKTSC (matrice temporale che effettua lo stadio di compressione), FBKTSE (matrice temporale che effettua

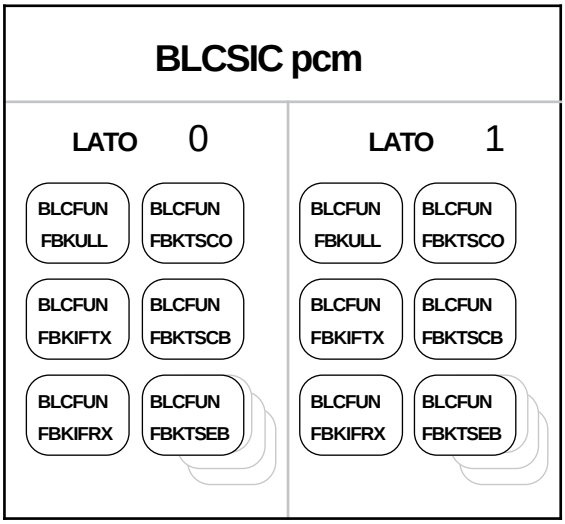


Figura 2 L'insieme dei BLCFUN formano un BLCSIC

lo stadio di espansione), FBKIFTX (interfaccia di trasmissione), FBKULL (interfaccia di trasmissione/ ricezione verso il CSM) e così via.

In caso di una rilevazione di un allarme sulla piastra che svolge la funzione di interfaccia verso il CSM, i programmi di autodiagnosi residenti nel comando di modulo emettono una segnalazione di guasto e lanciano la procedura di individuazione, riconfigurazione e scambio del blocco di sicurezza guasto con il gemello sano.

Parallelamente i programmi di diagnosi richiedono all'elaboratore di supporto (ES) di emettere sulla stampante di sistema (si consulti la fig. 3) la segnalazione codificata SINTOMO (riga 1). Tale segnalazione è accompagnata da una serie di codici alfanumerici che hanno il compito di descrivere il tipo di guasto individuando le coordinate della piastra.

Se la procedura di individuazione del guasto conferma la situazione anomala, vengono messe fuori servizio le piastre associate a quel blocco funzionale (riga 3). Il BLCSIC "pcm" associato (nell'esempio, quello del lato 1) viene posto (riga 2) in OUT (fuori servizio). Immediatamente i programmi di autodiagnostica pongono (riga 4) il lato 0 che svolgeva funzioni di "*riserva calda*" (situazione di BAK), nello stato ONL (in linea).

Queste quattro segnalazioni ed il conseguente allarme codificato saranno poi rilevate sotto forma di allarmi luminosi sulle piastre guaste e come righe di stampa sulla stampante di sistema dell'elaboratore di supporto (ES).

Se successivamente i programmi di autodiagnosi non rilevano più alcuna anomalia su tale piastra, automaticamente vengono rimesse in servizio le piastre associate al BLCFUN ed il BLCSIC viene riportato in situazione di BAK (*caso di autoripristino*).

L'insieme delle righe di stampa delle segnalazioni di allarme costituiscono il "*report di manutenzione*" dell'impianto come nell'esempio di fig. 3.

I report di manutenzione sono stampati on-line sulla stampante di sistema e contemporaneamente registrati sulle memorie di massa dell'ES, essendo a disposizione degli operatori locali o delle postazioni remote per essere consultati nelle operazioni di manutenzione della centrale.

Il report di manutenzione degli impianti UT, oltre a contenere le segnalazioni di allarme di tipo SINTOMO, BLCFUN e BLCSIC, presenta anche segnalazioni di allarme di tipo ERRORE (transizione precedente al SINTOMO e visualizzabile su richiesta) e di tipo CTRSOF e EVENTSW (incongruenze sui programmi software, visualizzabili a richiesta ed interpretabili solo dalla casa costruttrice).

Sono presenti anche segnalazioni di controllo di esecuzione delle principali procedure automatiche quali il salvataggio dei contatori di utenti su ES, esecuzione degli utenti morosi, ecc..

Mediamente per un autocommutatore di grande capacità (40000 numeri corrispondenti a circa 90 moduli) con un basso tasso di guasto, il report di manutenzione è formato da circa 15 pagine giornaliere (circa 900 righe di segnalazione).

Da quanto esposto si evince che l'autocommutatore UT è un impianto che fornisce una notevole quantità di informazioni di manutenzione.

Ne derivano:

- aspetti positivi:
1. L'allarme è identificato con un elevato grado di dettaglio;
2. Tutta la catena delle dipendenze del guasto è sotto controllo.

- aspetti negativi:
1. la mole eccessiva delle segnalazioni può indurre confusione ed errori durante le fasi di analisi del guasto;
2. la ricerca di guasti ripetitivi diventa laboriosa e abbastanza difficoltosa.

riga 1 M2 052* SINTOMO <> ULLINK0 * 100000 000000 000000 000000
riga 2 M2 052* BLCSIC <> pcm * IM DG100 (DG000) OUT ONL 000001
riga 3 M2 052* BLCFUN <> FBKULL * DG100 (DG000) 000000
riga 4 M2 052* BLCSIC <> pcm * IM ONL BAK 000000

Figura 3 Esempio di report di manutenzione

3. **Descrizione del progetto ALFA**

3.1 *Caratteristiche tecniche*

Il Progetto ALFA è un supporto informativo che permette da un lato di ricavare un indice di efficienza e dall'altro di sintetizzare gli allarmi emessi dall'impianto in poche tabelle facilmente consultabili ed interpretabili anche dai non specialisti. Il software è stato sviluppato in ambiente turbo-basic e si presenta come un applicativo di tipo interattivo, strutturato a menu, in cui i dati sono richiesti on-line dal programma.

La fonte di elaborazione è il report di manutenzione dell'impianto che deve essere stato precedentemente acquisito su file di tipo ASCII dall'Elaboratore di Supporto (ES) mediante un software di emulazione.

Tale emulatore permette al Personal Computer (PC) di comportarsi come un terminale di centrale al quale è collegato in maniera diretta tramite linea seriale verso ES o in modo remoto tramite linea dati verso il CEM.

La fig. 4 descrive le modalità di collegamento.

Il file di manutenzione acquisito dall'impianto può essere elaborato da ALFA e fornire le tabelle di efficienza e di statistica sotto forma di report su schermo, su stampante o su file dati registrato su disco fisso del PC o su floppy

3.2 *Definizione e calcolo dell'indice di efficienza*

L'indice di efficienza α che esprime il buon funzionamento dell'impianto UT può essere sia globale (per tutta la centrale) che parziale per tipologia di modulo M2, M3 e PC (CSM e MDM).

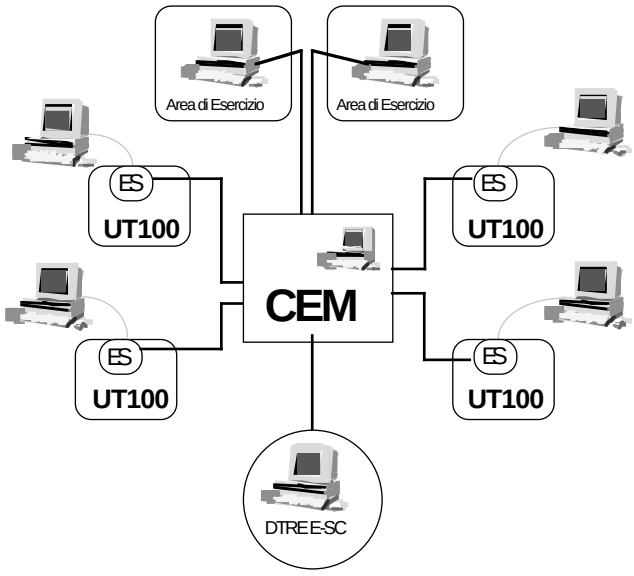


Figura 4 Modalità di collegamento dei PC

Il parametro α (in percentuale) è ottenuto confrontando l'efficienza del modulo di riferimento con l'efficienza dei moduli di tipo M2, M3, PC (Parti Centralizzate) sotto esame moltiplicato 100.

Si è definito come modulo di riferimento quello che emette 1 allarme ogni *KP* ore. *KP* è funzione della tipologia di modulo:

*KP*2 = 15 h (moduli M2)
*KP*3 = 24 h (moduli M3)
KPC = 48 h (moduli centralizzati).

Quindi ad esempio, il modulo di riferimento M3 emette 1 allarme ogni 24 ore.

I coefficienti *KP* tengono conto sia del grado di affidabilità del modulo, dipendente dal livello di tecnologia impiegata e dall'epoca di progettazione, che dall'importanza funzionale dello stesso all'interno della struttura UT100.

I valori di KP sono stati assegnati in maniera empirica e sono il frutto dell'esperienza maturata dopo numerosi anni di esercizio del sistema UT a Roma.

L'efficienza della tipologia di modulo sotto esame è ricavata dalla relazione:

$$\frac{Na}{M \cdot NH}$$

dove:

- *Na* = Numero degli allarmi (BLCFUN+BLCSIC) emessi dai moduli in esame durante l'intervallo di osservazione
- *M* = numero dei moduli
- *NH* = numero di ore di osservazione

Se ne ricava:

$$\alpha\% = 100\frac{M \cdot NH}{Na \cdot KP}$$

In base alle tre tipologie di modulo si considereranno:

$$\alpha M2\% = 100\frac{M2 \cdot NH}{NaM2 \cdot KP2}$$

$$\alpha M3\% = 100\frac{M3 \cdot NH}{NaM3 \cdot KP3}$$

$$\alpha PC\% = 100\frac{PC \cdot NH}{NaPC \cdot KPC}$$

L'efficienza totale di impianto risulterà la media pesata delle tre precedenti ed in particolare:

$$\alpha\% = 100\frac{\alpha M2 \cdot KP2 + \alpha M3 \cdot KP3 + \alpha PC \cdot KPC}{KP2 + KP3 + KPC}$$

Riassumendo, l'efficienza risulta inversamente proporzionale alla quantità degli allarmi emessi e direttamente proporzionale al numero di moduli ed alle ore di osservazione. L'efficienza totale dell'impianto è quindi maggiormente influenzata dalle parti centralizzate (*KP* maggiore) che dai moduli M3 e M2.

Tale incidenza è pienamente confermata dalla struttura dell'impianto UT100 come già descritto nel paragrafo 2.2.

Per descrivere come l'indice di efficienza α possa variare in base ai diversi parametri, si analizzi il grafico di fig. 5 derivato dalle definizioni analitiche descritte precedentemente. Il grafico riassume l'andamento degli indici di efficienza α*M*2, α*M*3, α*PC* al variare del numero degli allarmi *Na* (per valori da 1 a 200) su un impianto di medie dimensioni (40 M2, 25 M3 e 9 PC) con intervallo di osservazione di 24 ore.

Si può osservare come le tre curve di efficienza abbiano andamenti sostanzialmente diversi, variando questi in funzione del valore di *KP* assunto. In particolare:

- la curva di efficienza per gli M2 (α*M*2) manterrà il valore massimo del 100% fino ad un numero di 65 allarmi, oltre il quale scende gradatamente fino al 40% per valori superiori a 200;
- la curva di efficienza per gli M3 (α*M*3) manterrà il valore massimo fino ad un numero di 26 allarmi (circa 50% inferiore a quelli di M2) oltre il quale scende di valore in maniera più ripida rispetto a α*M*2 attestandosi al 20% per la quantità di allarmi uguale a 120;
- α*PC* sarà al 100% solo per 4 allarmi, dopodiché assumerà il valore del 20% con 25 allarmi.

Da tale analisi si evince che la ripidità della curva è *proporzionale al valore di KP*, mentre il valore soglia degli allarmi, oltre il quale l'efficienza scende sotto il 100%, è proporzionale alla quantità dei moduli e delle ore di osservazione. Questa ultima considerazione è avvalorata dalla fig. 6 che illustra l'andamento della curva di α*M*2 in funzione degli allarmi in tre centrali di diverse dimensioni (56 M2, 33 M2 e 17 M2) con stesso intervallo di osservazione di 24 ore e naturalmente con stesso *KP*2=15.

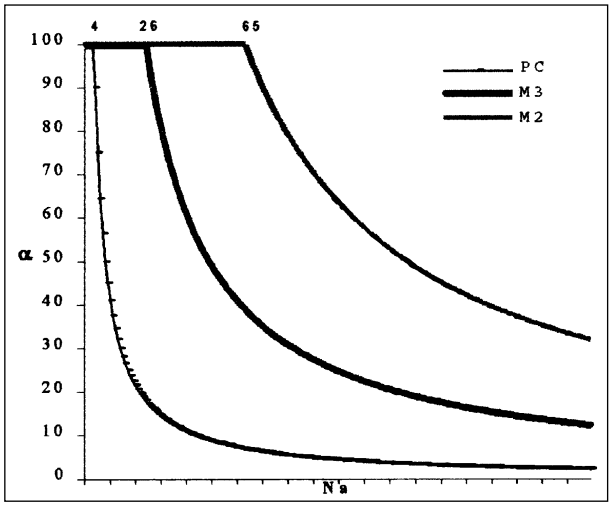


Figura 5 Andamento di α*M*2, α*M*3, α*PC* in funzione di *Na*

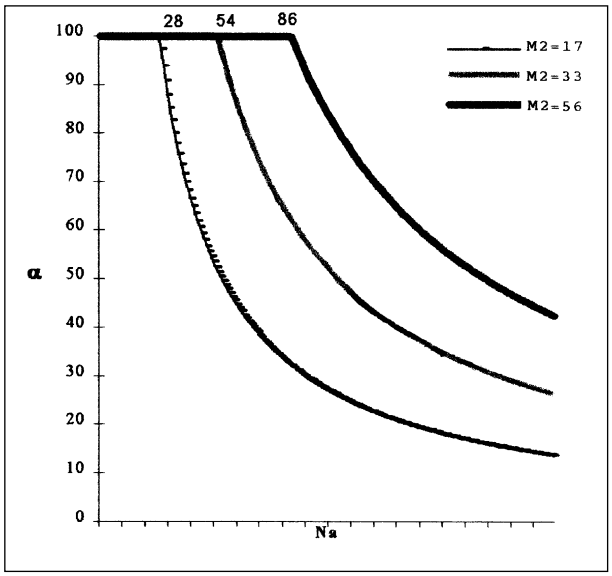


Figura 6 Andamento di α*M*2 in funzione di *Na* su impianti con diverse consistenze di *M*2

3.3 *Gli output di ALFA*

L'indice di efficienza α è ricavato elaborando i dati dell'impianto contenuti nel file di tipo ASCII acquisito da UT con le modalità descritte nel paragrafo 3.1 e produce l'output di fig. 7.

In particolare:

A = quantità utenti (richiesto da operatore)
B = quantità giunzioni (richiesto da operatore)
Tali dati non influenzano il calcolo dell'indice α
C = quantità dei moduli M2 (automatico)
D = quantità dei moduli M3 (automatico)
E = quantità delle PC (automatico)
F = quantità delle ore di osservazione (automatico)
G = quantità allarmi dei moduli M2 (automatico)
H = quantità allarmi dei moduli M3 (automatico)
I = quantità allarmi delle PC (automatico)
L = efficienza moduli M2 (calcolato)
M = efficienza moduli M3 (calcolato)
N = efficienza PC (calcolato)
O = efficienza impianto (calcolato)
P = legenda dei calcoli e simboli

I punti L, M, N ed O sono calcolati secondo le modalità descritte precedentemente nel paragrafo 3.2.

Gli output di sintesi (statistici e di ricerca) sono ottenuti dall'elaborazione di ALFA mediante l'aggregazione delle righe di allarme rilevate dal report di manutenzione secondo opportune modalità.

3.3.1 Statistiche

I programmi statistici forniscono:

- *quantità* dei SINTOMI, BLCFUN e BLCSIC suddivisi per modulo
- *ripetitività* dei SINTOMI, BLCFUN e BLCSIC selezionati da operatore e, se richiesto, suddivisi per modulo.

Nella parte A della fig. 8 vengono riassunte in maniera

tabellare le quantità delle segnalazioni per caratteristica di allarme e per le varie tipologie di moduli. Nella parte B della fig. 8, la statistica si spinge nel dettaglio, riassumendo le caratteristiche degli allarmi (SINTOMI, BLCSIC, BLCFUN, ecc) per ogni modulo. Tale output è particolarmente utile per evidenziare le criticità e quindi dare la giusta priorità sugli interventi di manutenzione correttiva. Per eseguire l'analisi approfondita di tali criticità

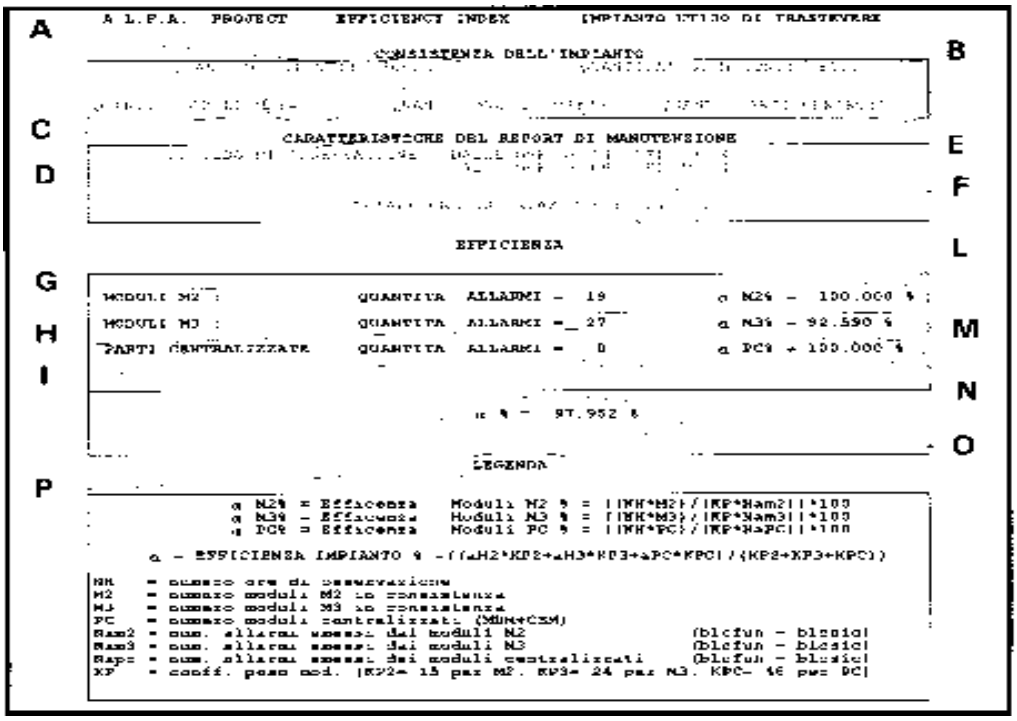


Figura 7 Indice di efficienza α

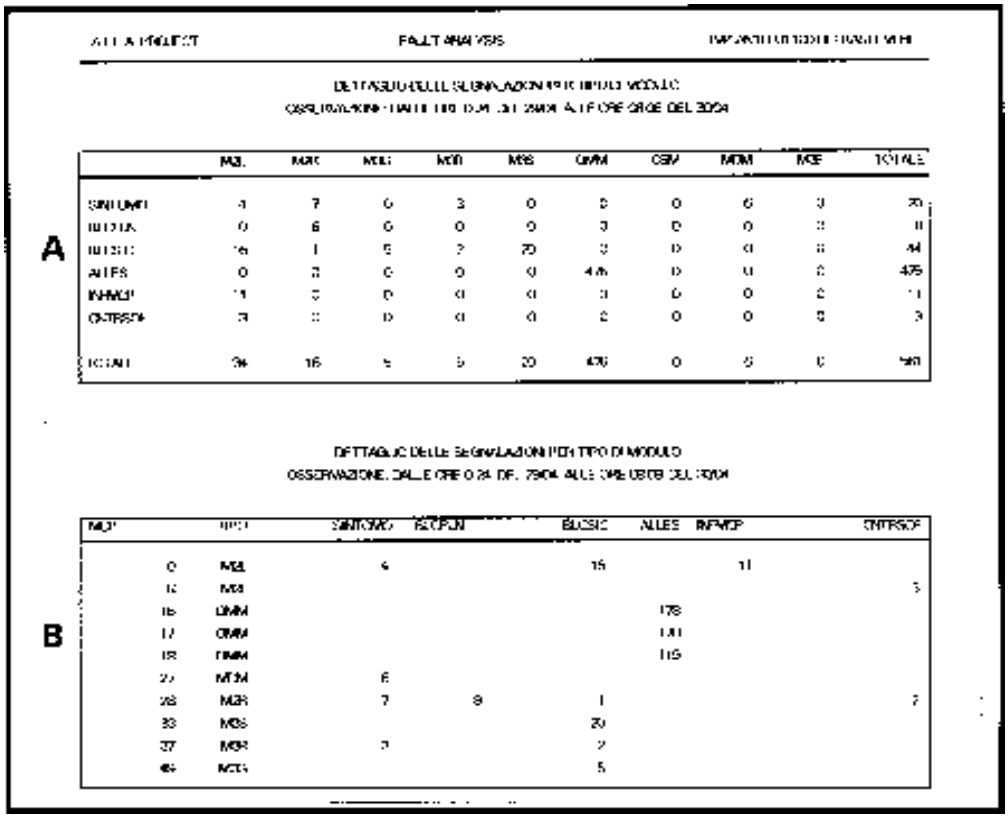


Figura 8 Statistiche per modulo

saranno di aiuto le routines di ricerca per modulo.

Nella parte A della fig. 9 vengono riportate le ripetitività degli allarmi che erano stati precedentemente selezionati nella apposita routine prevista da ALFA con una eventuale possibilità di dettaglio.

Nella parte B della fig. 9 vengono riportate le quantità delle segnalazioni precedentemente definite come dettaglio suddivise per modulo. Tali output sono necessari per monitorare segnalazioni di allarme gravi oppure ripetitive. Per approfondire tale analisi saranno di aiuto le routines di *ricerca per allarme*.

3.3.2 Ricerche

Le routine di ricerca sono indispensabili per identificare situazioni di criticità sui moduli dando come chiave di ricerca un particolare allarme, modulo o orario.

Esse forniscono (vedi fig. 10) la stampa della riga completa di allarme del report di manutenzione selezionata per:

1. modulo
2. allarme
3. intervallo di orario
4. uno o più parametri definiti precedentemente.

4. Applicazioni sul territorio

4.1 *Sviluppo del progetto ALFA*

Il software ALFA per le centrali UT, è stato sviluppato dall'ex DR ROMA R/ETR-STN (Supporto Tecniche Numeriche) nel corso del 1^ trimestre 1993.

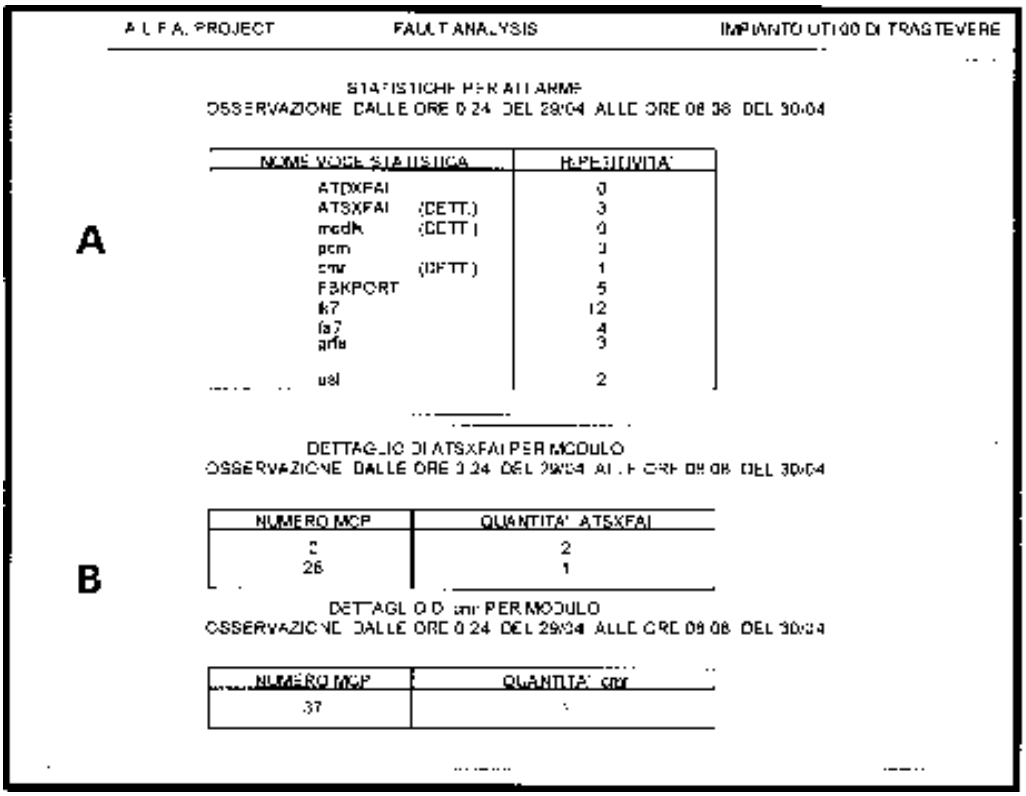
Dopo il rilascio della prima versione, sono state eseguite circa 100 sessioni di prova sugli output dei report di manutenzione ricavati da CEM di 10 impianti UT con diverse caratteristiche strutturali.

Nell'Aprile del 1993 un gruppo di lavoro costituito da personale di DR (STN), ed UTR ha utilizzato il prodotto ALFA su un campione significativo di impianti in una sorta di pre-esercizio.

Implementate le nuove prestazioni, richieste dagli utilizzatori finali, e risolti alcuni inconvenienti software, il 30 giugno 1993 è stata rilasciata la versione software 2.1 considerata operativa a tutti gli effetti.

4.2 *L'utilizzo di ALFA da parte delle AIC*

A partire dal 30 giugno 1993 ogni AIC ha utilizzato l'applicativo ALFA come strumento di monitoraggio e


A L F A. PROJECT FAULT ANALYSIS IMPIANTO UT100 DI TRASTEVERE

STATISTICHE PER ALLARME
OSSERVAZIONE DALLE ORE 0.24 DEL 29/04 ALLE ORE 08.38 DEL 30-04

A

NOME VOCE STATISTICA		RIPETITIVITA'
ATDXFAI		0
ATSXFAI	(CETT.)	3
modfk	(CETT.)	0
pcm		0
cmr	(CETT.)	1
FBKPORT		5
fk7		12
fa7		4
grfa		3
usl		2

DETTAGLIO DI ATSXFAI PER MODULO
OSSERVAZIONE DALLE ORE 0.24 DEL 29/04 ALLE ORE 08.38 DEL 30/04

B

NUMERO MCP	QUANTITA' ATSXFAI
2	2
26	1

DETTAGLIO DI cmr PER MODULO
OSSERVAZIONE DALLE ORE 0.24 DEL 29/04 ALLE ORE 08.38 DEL 30/04

NUMERO MCP	QUANTITA' cmr
37	1

Figura 9 Statistiche per allarme

RICERCA PER MODULO

```
A.L.F.A.     PROJECT 2.1    FAULT ANALYSIS
RICERCA PER MODULO   Modulo n.28

-MNTZ- 1993-04-29 14:35:40 TRASTEVERE M2L 028* BLCSIC  <>
alc       * MN              OFF       ONL        000125
-MNTZ- 1993-04-29 14:38:23 TRASTEVERE M2L 028* CNTRSOF <>
MPDSP     *   A-o  000004 117070 101041 000016 000061
-MNTZ- 1993-04-29 14:38:41 TRASTEVERE M2L 028* BLCSIC  <>
alc       * MN              ONL       OFF        000125
-MNTZ- 1993-04-29 19:05:27 TRASTEVERE M2L 028* SINTOMO <>
ATSXFAI   *   000521  000000  000220 000000
-MNTZ- 1993-04-29 23:52:43 TRASTEVERE M2L 028* BLCFUN  <>
FBKPORT   *    DG100  (DG000)                    001741
```

RICERCA PER ALLARME

```
A.L.F.A.     PROJECT 2.1    FAULT ANALYSIS
RICERCA PER ALLARME Allarme: ATSXFAI

-MNTZ- 1993-04-29 19:05:27 TRASTEVERE M2R 028* SINTOMO <>
ATSXFAI   *    000521  000000  000220  000000
-MNTZ- 1993-04-29 19:26:03 TRASTEVERE M2R 028* SINTOMO <>
ATSXFAI   *    000565  000010  000000  000000
-MNTZ- 1993-04-30 00:55:07 TRASTEVERE M2R 028* SINTOMO <>
ATSXFAI   *    000521  010000  000220  000000
```

RICERCA PER ORARIO

```
A.L.F.A.     PROJECT 2.1    FAULT ANALYSIS
RICERCA PER ORARIO Dalle 0400 alle 0500

-MNTZ- 1993-04-30 04:34:35 TRASTEVERE OMM 038* ALLES  <>
procmessi *   mess inizio procedura morosi da cem.
-MNTZ- 1993-04-30 04:35:15 TRASTEVERE M2R 028* BLCFUN <>
FBKPORT   *    DG100  (DG000)                    000361
-MNTZ- 1993-04-30 04:34:35 TRASTEVERE OMM 038* ALLES  <>
procmessi *   mess fine procedura morosi da cem.
```

RICERCA COMBINATA

```
A.L.F.A.     PROJECT 2.1    FAULT ANALYSIS
RICERCA COMBINATA
Modulo: tutti  allarme: FBKPORT
dalle ore: 0000 alle ore:0200

-MNTZ- 1993-04-30 00:46:25 TRASTEVERE M2L 000* BLCFUN <>
FBKPORT   *    DG100  (DG000)                    000367
-MNTZ- 1993-04-30 00:56:07 TRASTEVERE M2R 028* BLCFUN <>
FBKPORT   *    DG100  (DG000)                    000521
-MNTZ- 1993-04-30 01:13:25 TRASTEVERE M2R 028* BLCFUN <>
FBKPORT   *    DG000  (DG100)                    000565
```

Figura 10 Output di ricerca

di ausilio per le operazioni di manutenzione.

Giornalmente all'inizio del turno di lavoro il responsabile ed il suo staff hanno a disposizione l'indice di efficienza ALFA relativo allo stato di manutenzione del giorno precedente con le statistiche degli allarmi sintetizzate in più tabelle.

Con tale metodo si può immediatamente stabilire una linea di condotta per affrontare le problematiche legate alla manutenzione e tesa alla prevenzione dei guasti delle parti centralizzate e degli attacchi di utente.

Con le statistiche dettagliate ed i programmi di ricerca forniti da ALFA è possibile poi scendere nel dettaglio del guasto ed attuare le eventuali azioni correttive.

Inoltre, effettuando una analisi mirata e continuativa, è possibile identificare guasti saltuari e/o ripetitivi che, se non monitorati, potrebbero sfociare in situazioni di grave disservizio.

Le reazioni al nuovo tipo di gestione sono state decisamente positive da parte dei centri di lavoro. I responsabili hanno avuto finalmente la possibilità di "tastare il polso" al proprio impianto con dati semplici ed univoci, prendendo coscienza in prima persona della situazione manutentiva della centrale. ALFA ha contribuito a coinvolgere maggiormente i gestori degli impianti creando inoltre una sana "competizione" per il raggiungimento di valori di elevata Qualità.

L'analisi dei dati è diventato un momento di discussione tra il personale AIC che tra l'altro ha la possibilità di avere riscontro di quanto viene effettuato in termini di azioni di manutenzione preventiva.

4.3 *L'utilizzo di ALFA da parte delle AE e della DTRE*

ALFA ha introdotto un nuovo modo di effettuare il controllo della qualità degli impianti di commutazione anche alle linee di AE (ex UTR) e di DTRE (ex DR).

Infatti si è sfruttata la peculiarità dell'indice di efficienza come parametro di confronto fra impianti di tecniche omologhe (di qualunque grandezza e struttura), nonché per la stessa centrale in funzione del tempo.

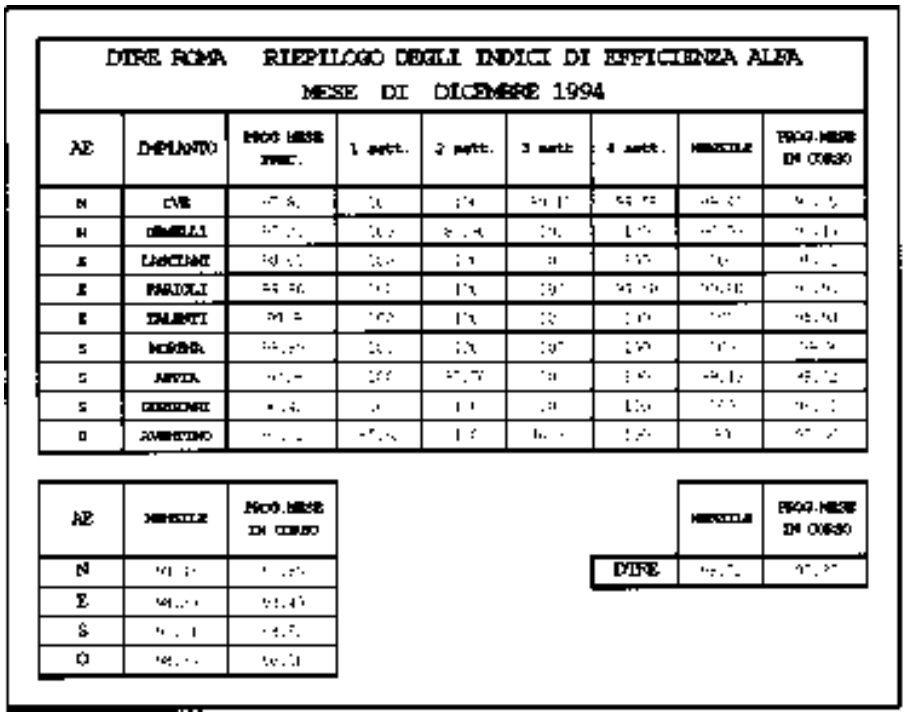
In particolare i CSE producono un report settimanale sugli indici di efficienza di tutti gli impianti di loro competenza, elaborando i report di manutenzione ricavati direttamente dalle centrali UT mediante collegamento con CEM.

L'indice di efficienza in esame è calcolato considerando l'intervallo di tempo intercorrente tra le ore 18 del venerdì e le ore 8 del lunedì successivo. In tal modo l'indice sarà scevro da qualsiasi allarme derivato da interventi di manutenzione programmata o da lavorazioni varie, ottenendo l'efficienza reale dell'impianto.

I risultati dei report settimanali vengono analizzati dagli specialisti dei CSE che possono consultare le varie statistiche a loro disposizione, offrendo il loro contributo verso le AIC per il miglioramento della manutenzione.

Il report settimanale, viene poi inviato a DTRE E-SC (ex STN) che ne esegue ulteriore interpretazione, se necessario approfondisce eventuali problematiche, e produce un report globale settimanale, mensile ed annuale suddiviso per impianto, per AE e per l'intera DTRE.

In tali report, di cui un esempio parziale è riportato nella fig. 11, vengono riportati l'andamento mensile e progressivo dall'inizio dell'anno e sono riportati nel *fascicolo mensile della qualità edito da DTRE Roma*.



DTRE ROMA RIEPILOGO DEGLI INDICI DI EFFICIENZA ALFA

MESE DI DICEMBRE 1994

AE	IMPIANTO	PROG MESE PREC.	1 sett.	2 sett.	3 sett.	4 sett.	MENSILE	PROG.MESE DI CORSO
N	CVE							
N	GRADOLI							
E	LANCIANI							
E	PARIOLI							
E	TALENTI							
S	MONZA							
S	APRILIA							
S	CERVERONI							
O	AVENTINO							

AE	MENSILE	PROG.MESE DI CORSO
N		
E		
S		
O		

	MENSILE	PROG.MESE DI CORSO
DTRE		

Figura 11 Esempio di report mensile

4.4 *Risultati dell'esperienza in campo*

Il controllo dell'indice di efficienza è stato perseguito operativamente dalla DTRE RM secondo due obiettivi:

- monitorare l'andamento nel tempo di α per i vari impianti al fine di intervenire prontamente su problematiche di manutenzione per garantire elevati standard qualitativi nei confronti del cliente;
- fissare una soglia minima di α sotto la quale l'impianto è definito non efficiente. Tale soglia è stata fissata nel 1993 al valore del 90%.

Le strategie di trend analisys e le modalità di utilizzo di ALFA sono state illustrate al territorio nel giugno 1993 con una convention in Direzione Regionale nella quale hanno partecipato STN, CSE e responsabili di esercizio delle UTR.

In tale incontro, oltre ad illustrare gli obiettivi finali, si è sottolineato che ALFA è soprattutto uno strumento gestionale nato per supportare i responsabili AIC, molto spesso non specialisti della tecnica in quanto provenienti da esperienze lavorative differenti dalla commutazione numerica, nella gestione dei propri impianti.

In particolare ALFA ha eliminato:

a) difficoltà dei responsabili a:
   - interpretare correttamente le segnalazioni degli impianti;
   - valutare e organizzare gli interventi di manutenzione;

b) disuniformità di comportamento sul territorio riguardo alle azioni di manutenzione

c) mancanza di procedure di sintesi per il controllo degli interventi del sistema numerico.

Successivamente sono stati effettuati nelle singole UTR ulteriori incontri tra i responsabili di esercizio, CSE ed AIC.

Sono stati eseguiti, inoltre, ulteriori interventi nelle sedi AIC da parte di STN e CSE per approfondire alcune operatività su ALFA e delineare i piani di azione per il miglioramento dello stato manutentivo della centrale, nei casi di rilevamento di bassi valori di α.

L'indice di efficienza medio degli impianti presi a campione nel 1^ semestre 1993 era di circa il 73%. A Settembre, la media progressiva di D.R. Roma si era attestata al 93%.

Un impianto con gravi problematiche di manutenzione, dovute a fornitura critica di piastre e cavi, è passato dal 68% nei primi mesi di sperimentazione, al 78% di Luglio, al 90% di Settembre, chiudendo l'anno 1993 a 95,27%.

In questo caso la numerosa sintomatologia riguardava in massima parte blocchi duplicati con allarmi ripetitivi che talvolta si autoripristinavano. Con la segnalazione del basso indice α e con il monitoraggio della sintomatologia effettuato con le statistiche, si sono evidenziate le cause dei guasti che altrimenti sarebbero difficilmente venute alla luce nel notevole quantitativo di carta che la stampante di sistema emetteva. In tale impianto sono state sostituite 65 piastre centralizzate e 12 cavetti di collegamento interni al modulo.

Le opere di bonifica non sono state chiaramente così pesanti in tutti gli impianti, ma ci sono state importanti iniziative a livello UTR e DR per eliminare situazioni di scarsa efficienza.

Per l'anno 1994 si è posto l'obiettivo al 96%, superato dalla media di DTRE con 97,3%.

Il monitoraggio dell'indice di efficienza α, e quindi l'utilizzo in termini di manutenzione preventiva, ha contribuito in modo fondamentale al miglioramento di alcuni indicatori di qualità degli impianti di commutazione:

- *indisponibilità di utente*: definita come la quantità media di minuti/anno per i quali un utente risulta privo del servizio telefonico a causa del fuori servizio dell'autocommutatore;
- *tasso di allarme*: definito come il rapporto tra la quantità di allarmi fuori orario di presidio e le linee pesate;
- *tasso di intervento*: definito come il rapporto tra la quantità di interventi del reperibile fuori orario di presidio e linee pesate.

La fig. 12 illustra chiaramente come il costante aumento dell'indice α corrisponda ad un decremento degli altri tre parametri.

Il grafico sintetizza i risultati ottenuti con il trend analysis di α e con gli interventi di manutenzione mirati sugli impianti UT.

Tali operatività hanno da un lato comportato la diminuzione di guasti gravi sulle parti centralizzate e dall'altro hanno influito positivamente sull'indicatore di indisponibilità utente degli impianti UT.

In particolare a fronte di 4.86 minuti/anno di indisponibilità (competenza TELECOM) e di α di 96.73 nell'Aprile 1994 il primo parametro è sceso gradatamente fino a 2.63 minuti/anno a seguito dell'aumento di α a 97.35 di Dicembre 1994.

Parallelamente anche il tasso di allarme (e di riflesso anche quello di intervento) degli impianti UT è sceso da 2.61 allarmi per 10000 linee pesate di Aprile 1994 a 2.06 allarmi di Dicembre.

Tale situazione è facilmente riconducibile al fatto che vengono emessi un minor numero di allarmi dall'impianto e, quindi, risultano più efficaci anche gli interventi da parte del personale AIC (in orario base) e dei CEM (fuori orario base).

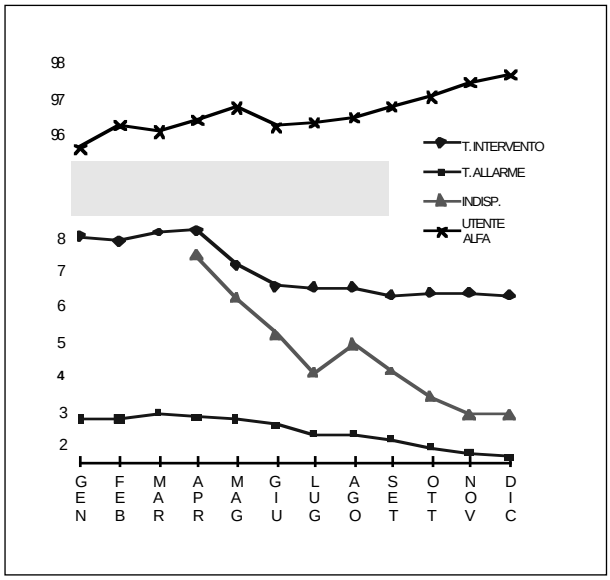


Figura 12 Andamento degli indici di qualità per l'anno 1994 di DTRE Roma

4.5 *Sviluppi futuri di ALFA*

Nel corso del 1994 si è deciso di estendere il progetto ALFA anche agli impianti AXE, che rappresentano circa il 30% degli impianti numerici della DTRE Roma.

Il programma è stato sviluppato, anche se con parametri di calcolo diversi, in base alla struttura centralizzata che caratterizza il sistema AXE.

Infatti la caratteristica che AXE non emette gli allarmi di tipo SINTOMO, ma direttamente l'allarme del blocco guasto, provoca una diversa tipologia e quantità di allarmi rispetto ad UT.

Sono state eseguite varie sperimentazioni sugli impianti a campione. Attualmente il software è stato consegnato ad un gruppo di lavoro che sta valutando l'impatto in campo per ciò che riguarda la pesatura dei vari allarmi.

A seguito di queste considerazioni, è stato coinvolto lo "CSELT" al fine di migliorare l'algoritmo di calcolo dell'indice di efficienza α per UT, per omogeneizzarlo con quello per AXE e per renderlo più scientifico e meno empirico.

Inoltre è stata interessata la DRE per:

1. cercare di ottenere su AXE un comportamento analogo ad UT (emissione dei SINTOMI) per ciò che riguarda l'allarmistica per i motivi precedentemente accennati;
2. ingegnerizzare ALFA nell'ambito del progetto DAMA.

Parallelamente alle attività svolte da CSELT e DRE, sono in corso da parte di DTRE RM E-SC gli aggiornamenti del software ALFA UT per includere nel conteggio dell'indice di efficienza, anche i nuovi moduli ME e OMU, gli allarmi di tipo SINTOMO e quindi rendere il monitoraggio ancora più spinto.

Ciò rende necessario, però, eseguire la fase di pesatura (coefficiente *KP*) non soltanto a livello di modulo, ma anche a livello di allarme.

In particolare è necessario effettuare il controllo sul nome dell'allarme e darne un valore in base ad una banca dati. Tali valori sono stati già definiti da un gruppo di lavoro, formato da STN e CSE, costituitosi allo scopo nel corso del 1994 e che ha discusso tutti i vari aspetti legati agli effetti degli allarmi che ogni singola irregolarità provoca.

Attualmente il nuovo software ALFA è in fase di approntamento e si prevede che le prime sperimentazioni negli impianti campione possano avvenire già ad Aprile 1995. Durante queste fasi sperimentali si metteranno a

punto le banche dati dei pesi degli allarmi, per avviare la fase operativa nel 2^ semestre 1995.

5. Conclusioni

Ci sentiamo di poter affermare che l'esperienza ottenuta con ALFA è sicuramente positiva sia in termini di miglioramento degli indicatori di *efficienza/efficacia* (indisponibilità utente, tasso d'allarme, tasso d'intervento) sia perché riteniamo che i responsabili AIC, CSE e STN siano diventati più consapevoli dello stato di efficienza degli impianti.

L'indice α si è rivelato un valido mezzo con cui incrementare la responsabilizzazione ai vari livelli e fornire un approccio più tecnico alle problematiche delle centrali UT non solamente basato su sensazioni e considerazioni personali.

Infine l'utilizzo di ALFA per fissare *obiettivi quantitativi* che, anche se hanno un fondamento empirico, e non scientifico (che si sta migliorando mediante l'aiuto di CSELT) ha dimostrato *grande valenza in termini di confronto* tra impianti.

Bibliografia

[1] Italtel : *Descrizione di prodotto UT*. Edizione 94.

[2] Italtel: *Manuale di manutenzione-Allarmi UT*. Edizione 05/94.

[3] Telecom Italia DR Roma R/ETR-STN: *Corso UT100-Principi ed inserimento in rete*. Edizione Aprile 1994.

[4] Telecom Italia DR Roma R/ETR-STN: *PROGETTO ALFA-Manuale operativo*. Edizione Marzo 1993.

Acronimi

AE	Area Esercizio
ALFA	Automatic fauLt and eFficiency Analysis
AIC	Area Impianti Commutazione
CEM	Centro di Esercizio e Manutenzione
CSE	Centro di Supervisione Esercizio
DAMA	Diagnosi ed Analisi MAnutenzione (Progetto di Telecom Italia DG DRE/S-CSS che ha l'obiettivo di realizzare un sistema di supporto all'operatore nelle attività di manutenzione degli elementi di rete, offrendo strumenti di analisi statistica dei malfunzionamenti della rete)
SC	Sistemi di Commutazione
SGT	Stadio di Gruppo Teleselettivo
SGU	Stadio di Gruppo Urbano
SL	Stadio di Linea
TER	Tasso di Efficacia di Rete
UTR	Unità Territoriale di Rete

Glossario dei moduli degli autocommutatori Italtel UT100

CSM	Circuit Switching Module (Modulo a commutazione di circuito)
ES-E	Elaboratore di Supporto Evoluto per l'esercizio e la manutenzione (in sostituzione degli OMM)
MDM	Message Distributor Module (Modulo di distribuzione messaggi)
MEPL	Modulo Evoluto Periferico Locale (modulo per utenti con capacità massima di 4000 terminazioni, utilizzato sia per utenza POTS (telefonia di base) che per utenza ISDN)
MEPR	Modulo Evoluto Periferico Remoto (modulo per utenti con capacità massima di 4000 terminazioni, utilizzato come isola remota)
M2	Modulo di tipo 2 (modulo con capacità massima di 2000 terminazioni miste utenti-giunzioni)
M2L	Modulo di tipo 2 Locale (modulo con capacità massima di 2000 terminazioni miste utenti-giunzioni, sito nella sede dello Stadio di Gruppo)
M2R	Modulo di tipo 2 Remoto (modulo con capacità massima di 2000 terminazioni miste utenti-giunzioni, utilizzato nelle isole remote)
M3	Modulo di tipo 3 (modulo per il trattamento delle giunzioni)
M3G	Modulo di tipo 3 Giunzioni (modulo con capacità massima di 1000 giunzioni, sito nella sede dello Stadio di Gruppo)
M3R	Modulo di tipo 3 Remoto (modulo per la connessione tra l'impianto e i moduli remotizzati (M2R, MEPR), sito nella sede dello Stadio di Gruppo)
M3S	Modulo di tipo 3 Segnalazione (modulo che gestisce le segnalazioni delle giunzioni in Canale Comune CCS#7, sito nella sede dello Stadio di Gruppo)
OMM	Operation and Maintenance Module (Elaboratore di supporto per l'esercizio e la manutenzione)
OMU	Operation and Maintenance Unit (parte dell'elaboratore di supporto evoluto ES-E)

Ringraziamenti

Gli autori ringraziano il sig. Emilio Patrizi ed il sig. Daniele Rossi di Telecom Italia DTRE Roma E-SC per la collaborazione offerta nella preparazione delle illustrazioni.