



INTEGRAZIONE DI RETI E SERVIZI: IL “TRASPORTO DELLA VOCE” SULLE RETI IP

Nuove applicazioni e servizi di rete rendono difficile il compito di chi deve scegliere cosa implementare e soprattutto quali risorse sono necessarie alle sempre più diffuse applicazioni che si stanno facendo largo nell'ambito delle reti IP: tra le tante innovazioni una sfida interessante è quella di portare un servizio, come la fonia, sulla stessa infrastruttura utilizzata per il traffico dati. La migrazione verso il mondo “all IP” è sicuramente motivata dalla riduzione di tutti i costi associati al comparto fonia.

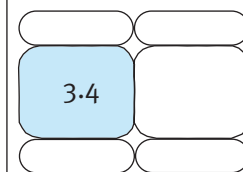
1. INTRODUZIONE

Da quando, nel febbraio 1995, Vocaltec rilasciò la prima versione dell'applicazione *Internet Phone* (versione β), la trasmissione di segnali vocali su reti IP ha fatto molta strada. Durante gli ultimi anni, si è assistito a una crescita eccezionalmente rapida della VoIP (*Voice over IP*) sia in termini di dimensioni di mercato, sia in termini di evoluzione tecnologica: la spinta principale che ha portato all'utilizzo della telefonia IP è, ad oggi, determinata dal minor costo consentito dalle reti IP rispetto a quelle tradizionali. Un utente che utilizza la VoIP per le sue comunicazioni interurbane o internazionali, a fronte di una qualità che, probabilmente, è leggermente inferiore rispetto a quella offerta dalla telefonia tradizionale, ma comunque superiore rispetto a quella, per esempio, della rete mobile GSM (*Global System for Mobile Communication*), può ottenere notevoli risparmi rispetto alle tariffe praticate dalle compagnie telefoniche. Innanzitutto, il chiamante paga una normale tariffa locale per collegarsi, con un circuito dedicato, a un *gateway* locale della

rete IP e utilizza, poi, la connettività IP su distanze nazionali o internazionali, a costi marginali, fino a un *gateway* terminale, ove si ricollega, a circuito e con tariffe ancora locali, al chiamato (*Toll Bypass*). Inoltre, le strutture di rete IP utilizzate sono più efficienti a causa della compressione del segnale vocale e della commutazione a pacchetti. Questo vale sia a livello *Intranet* che *Internet* [12].

È pensabile, tuttavia, che questa situazione è destinata a cambiare in tempi piuttosto rapidi: nel momento in cui la VoIP raggiungerà, infatti, una qualità del servizio paragonabile a quella tradizionale, interverranno nuove normative a regolamentare il settore, e questo porterà probabilmente a un aumento dei costi. D'altra parte senza una regolamentazione delle tariffe, né i *service provider*, né i costruttori di apparecchiature VoIP possono giustificare gli investimenti necessari a ottenere una VoIP qualitativamente accettabile: il 26 aprile 2001 l'“*US House of Representatives Telecommunications Subcommittee*” ha approvato un testo che regola l'offerta dei servizi voce su

Bruno Fadini
Antonio Pescapè
Giorgio Ventre



Internet. Le cause “scatenanti” che sono alla base del fenomeno “voce su IP” sono fondamentalmente le seguenti:

■ negli ultimi anni, il traffico dati, che tipicamente ha sempre viaggiato su reti a pacchetto, utilizzando in particolare il protocollo IP, ha superato in volume il traffico voce (Figura 1);

■ si è assistito, negli ultimi tempi, a una crescita esponenziale degli utilizzatori di reti IP. Un tale orientamento del mercato ha comportato una drastica riduzione dei prezzi di tutti gli apparati di rete concernenti il mondo IP. Pertanto, attualmente, un'impresa che decide di investire in una tecnologia di rete integrata basata su protocollo IP, può ipotizzare un risparmio sui costi globali di creazione dell'infrastruttura di rete che oscilla fra il 30 e il 50% (Figura 2).

FIGURA 1
Andamento del traffico dati e del traffico voce dal 1996 al 2001

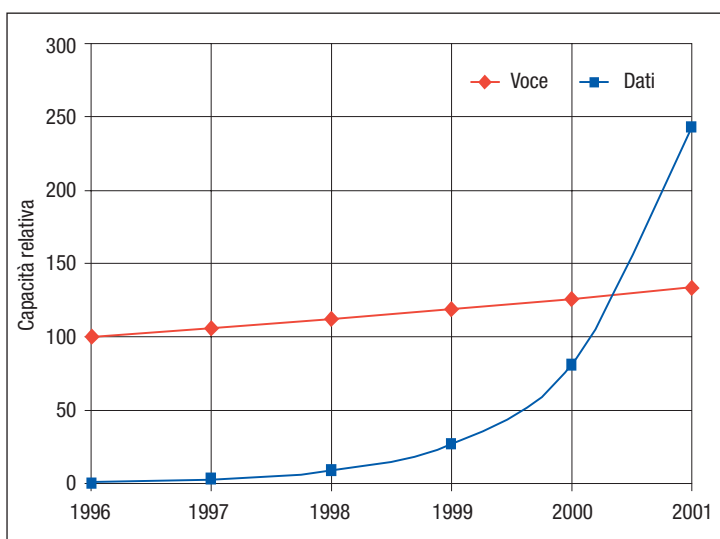
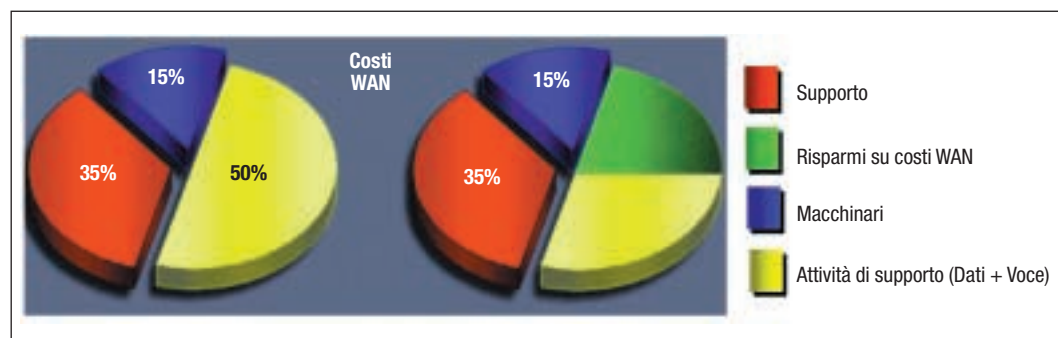


FIGURA 2
Riduzione dei costi associata all'integrazione di Dati/Video/Voce
Fonte: Cisco Systems



appare, quindi, come un elemento indispensabile per raggiungere questi obiettivi.

Le reti basate su protocollo IP sono state progettate e costruite per garantire la comunicazione ad applicazioni di tipo non *real time* (per esempio, *e-mail* e trasferimento di *file*) che sono caratterizzate da un traffico discontinuo che prevede picchi molto elevati di banda richiesta, ma anche tempi più o meno lunghi di inattività. Da questo punto di vista, quindi, voler utilizzare le reti a commutazione di pacchetto per la trasmissione di segnali vocali può apparire non conveniente, a meno di non inquadrare la cosa in un contesto più ampio che prevede la trasmissione attraverso un'unica rete di una moltitudine sempre crescente di servizi, che vanno dalla telefonia tradizionale alla telefonia con servizi avanzati, alla trasmissione dati, alla videoconferenza, al *broadcasting* televisivo ecc..

Il concetto fondamentale, dunque, è quello di integrazione: ovvero, un'integrazione di diversi servizi su di un'unica infrastruttura di rete, resa possibile da quella che in questi anni si è andata affermando come lingua franca, ossia il protocollo IP (*Internet Protocol*) (Figura 3). L'obiettivo di un'architettura di rete convergente, nel cui ambito si inquadra la VoIP, è, quindi, quello di realizzare una rete che da un lato, protegga gli investimenti economici sostenuti, supportando tutti gli elementi di rete e le interfacce analogiche e digitali presenti nella rete stessa e dall'altro, permetta di introdurre nuove tecnologie nella rete di accesso, di trasporto e negli elementi di commutazione, creando un'infrastruttura aperta e orientata ai servizi [1].

Agli evidenti benefici derivanti dall'adozione di una rete integrata si contrappongono, però, i problemi legati alla configurazione e alla gestione, che derivano dalla necessità

di supportare, garantendo gli opportuni requisiti di *Qualità del Servizio*, un traffico con caratteristiche assolutamente differenti (isocronia, *loss rate* ecc.) trasportato finora su reti differenti. Per fare ciò, è necessario capovolgere uno dei paradigmi fondamentali del mondo IP: non più tutto *Best Effort* ma classi di traffico a priorità differenti, introducendo opportune politiche di QoS (*Quality of Service*). Nella figura 4, è riportato un semplice grafico che mostra come in assenza di opportuni meccanismi di QoS la qualità di una conversazione telefonica su reti IP sia notevolmente compromessa. La migrazione del traffico voce sulle reti IP comporta, comunque, la necessità di poter

determinare l'ammontare di traffico generato dai canali vocali una volta noti il codec e il compressore di segnale.

Il modello di rete adottato sia dalle aziende che dalla Pubblica Amministrazione è tipicamente quello di una direzione generale (*headquarter* o centro stella) e di una serie di uffici distaccati (*branchoffice*) distribuiti sul territorio cittadino, nazionale o, addirittura, internazionale dipendente dalle dimensioni dell'azienda e quindi del business nella quale quest'ultima è impegnata. In un'architettura di questo tipo, i punti critici, in termini di banda, sono sicuramente i *link* di WAN (*Wide Area Network*) per i collegamenti tra le diverse sedi dell'azienda e, in situazioni di reti

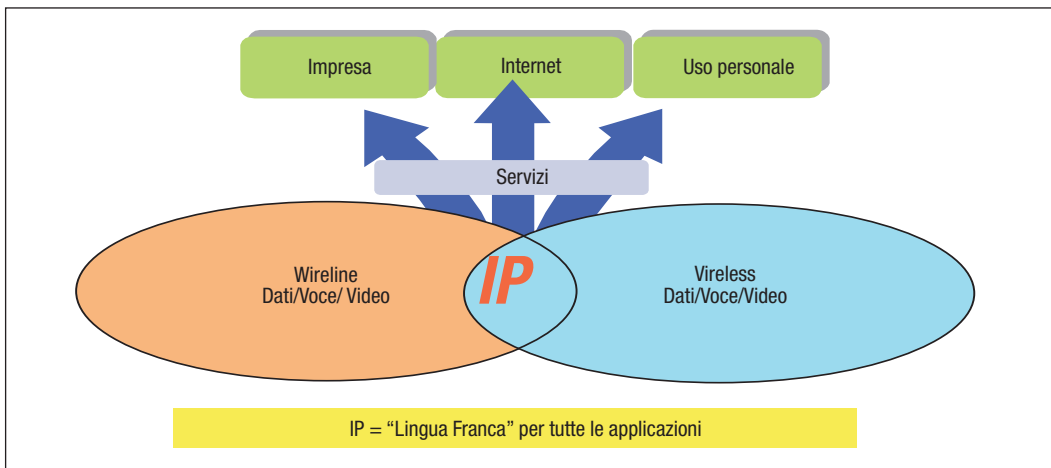


FIGURA 3
Il protocollo IP diviene il comune denominatore del trasporto dell'informazione

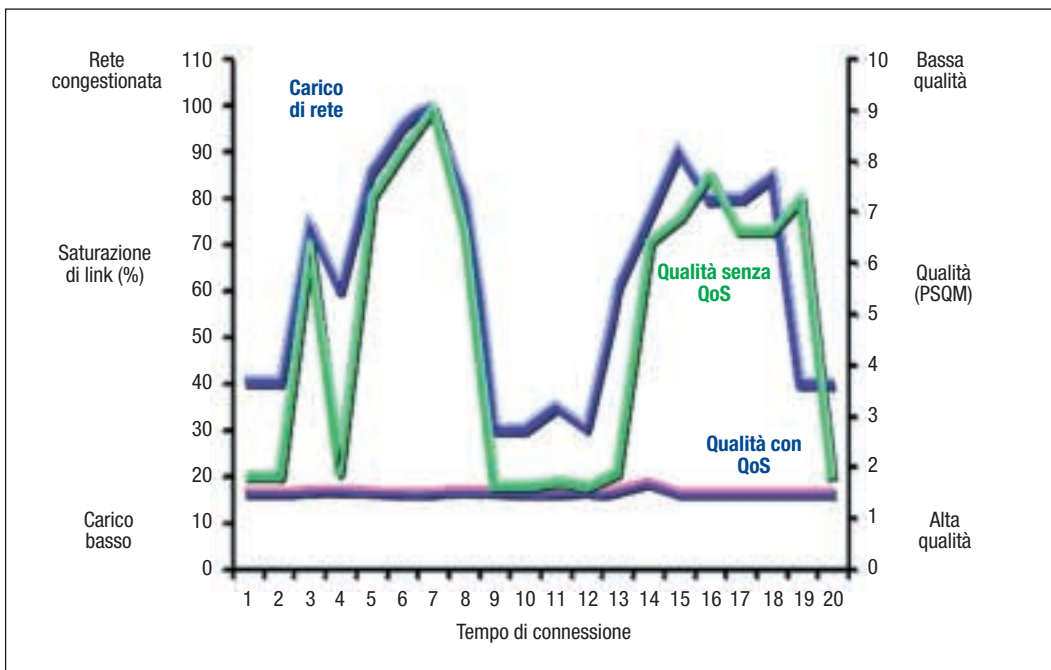


FIGURA 4
Traffico VoIP con e senza QoS
Fonte: Cisco Labs

particolarmente antiquate, il cablaggio interno di un edificio e, quindi, i collegamenti locali alla singola sede.

Una delle prime necessità di un progettista di architetture di rete, o semplicemente di un amministratore, propedeutica a un corretto funzionamento della rete stessa, è la corretta individuazione dei diversi applicativi che utilizzeranno tale rete, per poter poi individuare correttamente le risorse necessarie per ciascuno di essi.

Per i flussi dati tradizionali, che per molti anni sono stati i soli utilizzatori di reti IP, la progettazione è relativamente semplice. Normalmente, è già nota a priori la capacità di banda necessaria per ciascun servizio e, inoltre, poiché, in genere, si tratta di servizi sensibili alla perdita dei pacchetti (*loss sensitive*) ma non ai ritardi (*time sensitive*), anche in caso di momentanee congestioni della rete si verifica solo un rallentamento delle normali operazioni. Utilizzando reti IP per il trasporto di flussi di traffico real time, le condizioni che si presentano sono decisamente diverse, anche se il traffico voce, fra le varie caratteristiche peculiari, ha anche quella di avere un *bit rate* relativamente costante, rendendo possibile un calcolo più preciso di quella che è l'occupazione di banda. Il calcolo è reso meno agevole, anche, dalla necessità di considerare che per i segnali vocali sono stati sviluppati numerosi algoritmi di campionamento e compressione, ognuno con caratteristiche proprie e bit rate differenti [14, 17]. Inoltre, fino ad oggi, le reti dati e quelle telefoniche, anche in ambito locale, quale quello di un *Campus* o di una azienda, sono sempre state distinte, per cui non si hanno riferimenti sulle modalità di utilizzo della rete telefonica. Una soluzione, ampiamente diffusa, è basata sull'impiego di modelli più o meno raffinati per calcolare l'ammontare di traffico telefonico che bisogna supportare; inoltre, molti dei più recenti PBX (*Private Branch Exchange*) contengono una memoria storica, che consente sia di avere un *log* accurato delle chiamate effettuate, sia di conoscere verso quali direttrici principali sono state indirizzate (per esempio, se si tratta di chiamate interne alla rete locale oppure se si tratta, invece, di chiamate esterne).

2. STANDARD E ASPETTI CARATTERISTICI DELLE COMUNICAZIONI VOCALI SU RETI IP

Le prime applicazioni della tecnologia VoIP utilizzavano Internet quale mezzo trasmissivo per inviare traffico audio, in tempo reale, a due o più utenti collegati tramite *computer*. Il primo prodotto *software* per la telefonia su IP è stato presentato da Vocaltec all'inizio del 1995, come ricordato nell'introduzione: girando su un PC (*Personal Computer*) multimediale, questo software permetteva agli utenti di parlare nei microfoni e di ascoltare tramite gli altoparlanti o le cuffie. Da allora, la tecnologia è notevolmente migliorata, permettendo agli utenti di utilizzare anche i telefoni tradizionali per effettuare una chiamata su Internet e di gestire tipologie di servizi sempre più sofisticate, che integrano *media* differenti nelle loro comunicazioni. Sebbene all'inizio la VoIP sia stata utilizzata solo come un sistema relativamente semplice per fornire servizi di trasporto voce a basso costo fra due postazioni IP, da allora, l'interesse sempre crescente verso la possibilità di fornire servizi integrati di voce, video e dati ha provocato una notevole espansione degli scopi originari. La telefonia su IP abbraccia, oggi, un largo spettro di servizi, fra i quali, ad esempio, la conferenza tradizionale, la capacità di controllo della chiamata, il trasporto multimediale e l'integrazione fra la telefonia e il WEB, l'e-mail e l'*instant messaging*.

Inizialmente, il sistema di trasporto per i servizi VoIP era solamente la Internet pubblica, ma ora un numero sempre maggiore di società di trasporto (*carriers*) e di fornitori di servizi (*service provider*) stanno utilizzando (o, in alcuni casi, costruendo) le proprie dorsali (*backbone*) per fornire ai propri clienti una elevata qualità del servizio. Contrariamente a quello che si potrebbe pensare, il vero punto di forza della VoIP non riguarderà solo l'aver tariffe più ridotte rispetto alla telefonia tradizionale quanto, piuttosto, la capacità di fornire nuovi servizi a valore aggiunto.

L'introduzione della voce sulle reti a commutazione di pacchetto può essere realizzata secondo diversi modelli funzionali, ciascuno dei quali implica differenti soluzioni architetture. I modelli più diffusi sono: *PC-to-PC*, *PC-to-Pho-*

ne, Phone-to-Phone, fax-to-fax e, più in generale, tutti i modelli di *messaging service* (e-mail-to-fax, Web-to-fax, fax-to-e-mail, universal in-box, WWW/fax on demand server) [7].

Tale introduzione è, attualmente, in corso e, quindi, il processo di standardizzazione dei sistemi relativi all'*Internet Telephony* non si è ancora concluso. Al momento, sono coinvolti tre enti internazionali di standardizzazione: l'ITU (*International Telecommunications Union*), l'IETF (*Internet Engineering Task Force*) e l'ETSI (*European Telecommunication Standard Institute*) con alcuni consorzi (per esempio, Softswitch, H.323ORG, Vivida ecc.) [2]. La gestione delle chiamate voce sulla rete IP è, al momento, indirizzata da due differenti proposte, elaborate in ambito ITU e IETF, che sono rispettivamente **H.323** [11] e **SIP** (*Session Initiation Protocol*) [10].

Per quanto concerne l'interlavoro tra la rete telefonica a commutazione di circuito e la rete VoIP sembra, invece, essersi affermato il MGCP (*Media Gateway Control Protocol*), sviluppato, inizialmente, in ambito IETF e ora standardizzato, sotto l'acronimo H.248, in un *working group* congiunto ITU-IETF. Analogamente, le soluzioni per il trasporto

della segnalazione SS7 sulle reti IP sembrano, ormai, consolidate nell'ambito del gruppo di lavoro SIGTRAN del IETF.

3. IP TELEPHONY E VOIP: COSA SONO E QUANDO CONVENGO

I motivi principali che hanno giustificato l'introduzione prima e l'affermazione poi della voce su IP sono numerosi: la vasta diffusione delle reti dati e l'enorme diffusione di Internet (*la rete delle reti*); la larga diffusione dei PC; il basso costo per la realizzazione sia dei DSP (*Digital Signal Processing*), sia delle reti IP costituite da apparati passivi, schede di rete e cablaggio strutturato (Figura 5) e, infine, la possibilità di creare nuovi servizi.

L'introduzione di questa innovazione, in un'architettura di rete esistente, può seguire due approcci differenti che portano a scegliere due diverse architetture, diverse anche nel nome: *IP Telephony* e VoIP.

Con VoIP si intende il puro trasporto: l'utente continua a usare apparecchi telefonici comuni e non ha alcuna cognizione del fatto che la sua conversazione viene trasportata dalla rete dati. Ne vi è alcuna integrazione applicativa tra telefonia e sistema informativo.

Con IP Telephony si definisce un servizio telefonico completo, che considera come parte della rete IP anche il terminale d'utente. In pratica, l'utente dispone di un terminale telefonico interconnesso alla rete con un indirizzamento IP, oppure dispone di un programma software che realizza sul suo PC, con opportuno equipaggiamento multimediale, le analoghe funzioni.

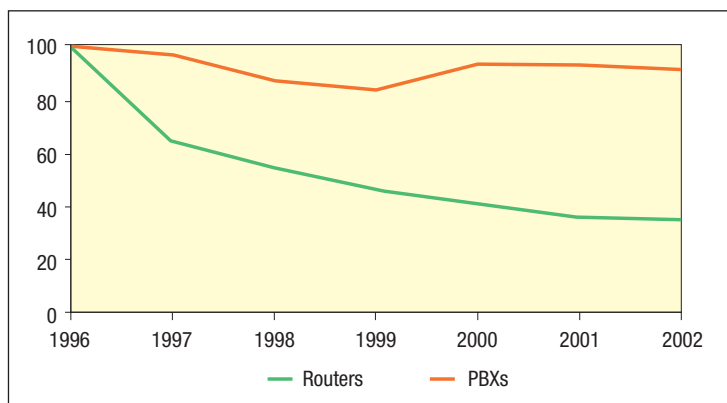
Con IP Telephony sono, quindi, possibili mag-

I sostenitori di ITU-T affermano che H.323, storicamente nato prima, abbia ormai ottenuto il supporto di tutti i fornitori di apparati VoIP, mentre i sostenitori di SIP dubitano dell'interoperabilità dei prodotti H.323 di differenti *vendor* e, allo stesso tempo, evidenziano i vantaggi di SIP in particolar modo per quanto riguarda la ridotta segnalazione durante l'attivazione della chiamata. A supporto di quest'ultima tesi, vi è il fatto che SIP sempre più viene introdotto tra i protocolli supportati dai *vendor*, che comunque mantengono la compatibilità "storica" con H.323 [18]. Volendo confrontare **H.323** e **SIP** bisogna comunque osservare che lo scopo dei due standard è piuttosto differente. Mentre SIP è un protocollo per la segnalazione e il controllo di sessioni multimediali, H.323 delinea un'architettura completa per lo svolgimento di conferenze multimediali, comprendente la definizione dei formati di codifica a livello applicativo, la definizione di protocolli per la segnalazione e il controllo, per il trasporto dei flussi audio, video e dati e per la gestione degli aspetti di sicurezza, tutto ciò con riferimento ad architetture di rete locali. SIP risulta una valida alternativa soprattutto per le sue caratteristiche di semplicità rispetto ad H.323 che deriva dal fatto di avere ereditato certe caratteristiche dall'ambiente delle reti locali e certe altre dallo standard H.320 (per ISDN, *Integrated Services Digital Network*).

FIGURA 5

Andamento dei prezzi relativi di router e PBX in Europa

Fonte: Gartner Group



giori interazioni tra il mondo della telefonia e il mondo delle applicazioni dati. È, per esempio, possibile la gestione di una rubrica in linea, con chiamata da telefono su segnalazione iniziata da un PC, oppure l'automazione delle funzioni di ricerca da parte di un *call-center* tramite interazione diretta tra il telefono dell'operatore e il sistema informativo.

Ma quanto è realmente conveniente l'integrazione del servizio di fonìa alla preesistente infrastruttura di rete dati? Tra le diverse osservazioni, quella che forse merita maggior interesse si riferisce al peso relativo che hanno la voce e i dati nei confronti della infrastruttura generale. Se si è, infatti, nella situazione in cui il numero di conversazioni contemporanee tra due sedi è "relativamente" limitato e la quantità di dati scambiati è molto grande, è allora probabile che la rete dati abbia un proprio naturale "sovradimensionamento". In queste condizioni, trasportare la voce significa semplicemente aggiungere qualche kbit/s al traffico dati senza che ciò debba comportare degli oneri aggiuntivi.

Una decina di canali fonici su VoIP richiedono, per esempio, un centinaio di kbit/s. Se la rete dati utilizza linee a 2 Mbit/s o più, la voce può transitare sopra senza che sia necessario alcun ampliamento di capacità particolare. Il costo del VoIP è, pertanto, il solo costo di ammortamento degli apparati, o meglio, delle parti di apparati necessarie a introdurre la funzione di *voice gateway*. In merito all'IP Telephony, occorre fare anche una considerazione di opportunità: è, ormai, prassi comune considerare come apparati di distribuzione di LAN (*Local Area Network*) degli *switch* (o *hub*) a 10/100 Mbit/s. I 100 Mbit/s, inoltre, per alcune stazioni *end-user* sono un prerequisito funzionale (*workstation* con sistemi CAD, sistemi per le elaborazioni e le analisi geofisiche). Esistono, tuttavia, edifici con cablaggi aventi un'età superiore ai 10/12 anni e che, probabilmente, presentano infrastrutture di rete antiquate e cablaggi "non strutturati". In queste situazioni, quindi, può nascere l'esigenza di ristrutturare i cablaggi.

In quest'ottica, una postazione di lavoro tradizionale richiede un impianto a tre cavi/PDL (uno per i dati, uno per la fonìa e un terzo per un eventuale seconda uscita dati *network printer* o per un apparato telefonico/fax),

mentre l'uso di telefoni IP con funzione di hub e di switch integrato, porta all'impiego di un solo cavo per la PDL. Questo significa poter dividere per tre il costo di installazione del cablaggio. In più, si avrebbe una maggiore flessibilità conseguente al fattore IMAC (*Installation, Move, And Change*): se tutte le porte LAN in distribuzione agiscono da *trunk* e l'indirizzamento in VLAN (*Virtual LAN*) di quanto è presente sulla scrivania viene fatto dal "telefono/switch", spostare una stazione significa, semplicemente, disconnetterla dal cavo di distribuzione, trasferirla (telefono compreso, che diventa personale come il PC o la workstation) e riconnetterla con la stessa configurazione nella nuova postazione senza che sia necessario riconfigurare gli apparati di accesso e di distribuzione, o le permutazioni o le predisposizioni di utente in centrale: si conseguono, così, una serie di risparmi "indotti" che hanno, comunque, un peso di rilievo [9]. Per terminare questa semplice analisi e completare le considerazioni di ordine pratico, un'ultima variabile da tenere presente è il costo, attualmente critico, dei collegamenti geografici (WAN). Nel corso degli ultimi mesi, il costo delle CDN (*Collegamenti Diretti Numerici*) è decisamente diminuito. Nel 1996, un circuito CDN 64 kbit Milano-Roma costava quasi 18 milioni di lire, oggi, il costo è sceso di un ordine di grandezza ed è grossomodo di circa € 775. Inoltre, attualmente, esistono diverse soluzioni architetture che prevedono la possibilità di acquistare da un operatore di rete il servizio di VPN (*Virtual Private Network*) grazie al quale è possibile ottenere la garanzia di determinati requisiti di QoS a costi contenuti. Pertanto, in situazioni in cui il costo dei *link* di WAN non è critico le ragioni dell'integrazione di dati/voce risiedono, perciò, altrove e cioè nella capacità di integrare facilmente, e a un costo marginale, applicazioni CTI (*Computer Telephony Integration*) e CRM (*Customer Relationship Management*), nella manutenzione effettuabile da persone non esperte di telefonia standard, nel poter disporre di sistemi *unified messaging* integrati, nel poter attivare *voicemail* sofisticate e di dimensioni fuori dall'ordinario e, infine, nel gestire una sola rete su cui convergono dati e voce.

4. LA "QUALITÀ" DEL SEGNALE VOCALE

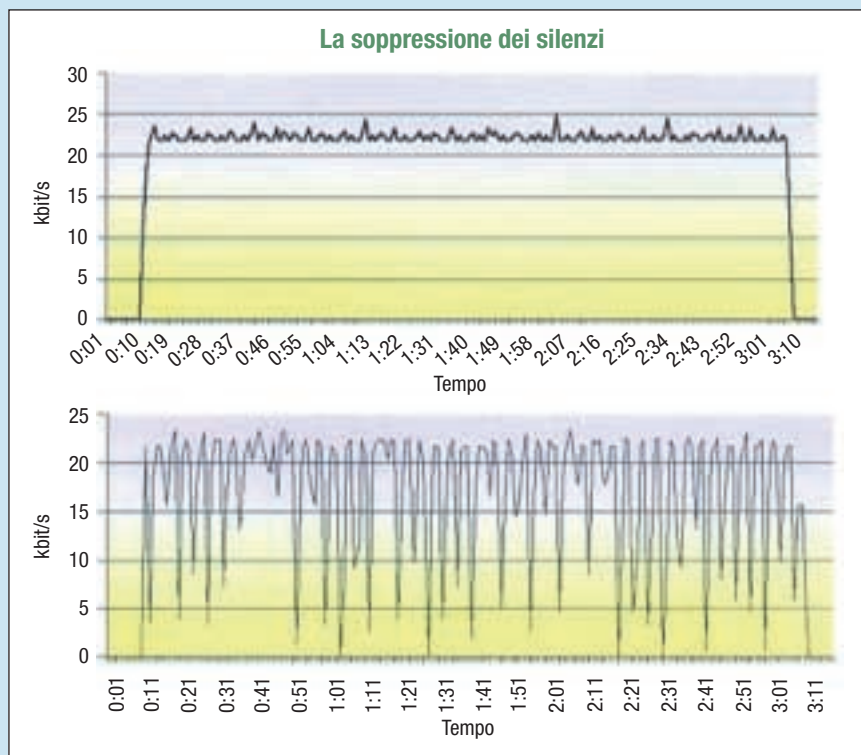
Per comprendere gli aspetti legati al trasporto della voce su reti IP è necessario soffermarsi prima sulle caratteristiche del segnale vocale, e poi sui problemi del trasporto su reti a commutazione di pacchetto. Per garantire determinati livelli di servizio sulla rete PSTN (*Public Switched Telephone Network*) la qualità della voce è legata al livello sonoro, al ritardo, alla eco, alla comprensione, all'intelligibilità, al rumore, al *fading* e alla diafonia. In aggiunta, sulle reti IP, bisogna risolvere problemi legati al ritardo, al *jitter*, alla comprensione, alla perdita di pacchetti, alla disponibilità di banda e alla compressione. È importante sottolineare come i parametri relativi alla qualità del servizio offerto siano soggettivi; il valore che ogni utente assegnerà a ciascuno dei servizi elencati è una questione individuale: per esempio, i viaggiatori abituali daranno la preferenza a servizi quali le carte di credito o la portabilità del numero, mentre una famiglia con bambini può preferire piuttosto avere un numero maggiore di caselle di posta. Viceversa, per quanto riguarda, invece,

la qualità della voce, è possibile definire in maniera oggettiva alcuni vincoli. Un approccio abbastanza frequente, anche se non sempre utile, per verificare la qualità della voce, consiste nell'utilizzare le tecniche della rete telefonica tradizionale, ovvero di confrontare a schermo le forme d'onda e misurare il rapporto segnale/rumore e la distorsione armonica totale. Quest'analisi può avere senso, però, solo nel caso in cui una qualunque differenza nella forma d'onda sia indice di una distorsione non voluta introdotta dalla rete: utilizzando alcuni *codec* a bit rate molto basso il segnale ricevuto, in realtà, non ha la stessa forma d'onda del segnale vocale trasmesso, bensì ne è una riproduzione soggettiva dovuta al tipo particolare di codec utilizzato. Pertanto, appare evidente che per misurare la qualità del traffico voce che attraversa reti IP sono necessari diversi metodi di prova.

Un metodo utilizzato è quello basato sulla misura della qualità vocale percepita (PSQM, *Perceptual Speech Quality Measure*, definito in *ITU-T Recommendation P.861*). Tale metodo analizza in modo oggettivo segnali vocali con una larghezza di banda di 300-3400 Hz.

La soppressione delle pause o dei silenzi viene effettuata mediante un rilevatore di attività vocale o **VAD (Voice Activity Detector)**. Quando si ha l'attività vocale, il VAD consente ai pacchetti di essere trasmessi. Quando, invece, si ha lo stato di silenzio, il VAD "chiude" momentaneamente la trasmissione. Poiché, normalmente, la conversazione umana è fatta per circa la metà da silenzi, o per meglio dire è *half-duplex* (in genere, si parla uno per volta), l'utilizzo di tale dispositivo permette un risparmio medio di banda, in termini di flusso aggregato, di circa il 30-40%. Nella figura, si può apprezzare la variazione di capacità di canale necessaria alla trasmissione di un flusso voce codificato con la raccomandazione ITU-T, G.729, nel caso in cui il VAD sia disattivato o meno: il risparmio di banda nel secondo caso è sensibile.

Confronto tra l'occupazione di banda in presenza e in assenza di VAD



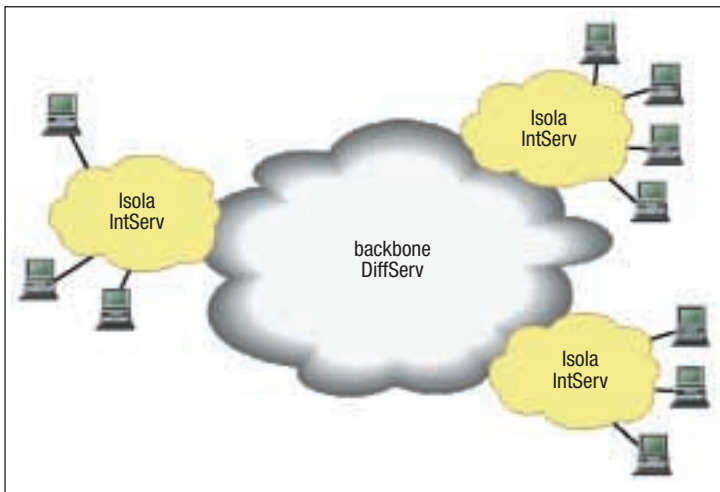


FIGURA 6
Architettura
integrata Int-Serv-
Diff-Serv

In modo molto semplice, si può affermare che il PSQM è una sorta di ascoltatore automatico, che confronta segnali vocali di riferimento con i campioni registrati dopo il passaggio attraverso la rete; applicando vari modelli uditivi e cognitivi, il PSQM è in grado di fornire una stima della degradazione percepita, che corrisponde abbastanza fedelmente ai risultati forniti da ascoltatori umani. Per quanto riguarda i cancellatori d'eco, anche in questo caso l'ITU-T ha definito alcuni metodi di misura (ITU-T *Recommendation* G.165 e G.168), che utilizzano metriche basate sulla risposta al rumore bianco oppure a segnali pseudo-vocali. Sono, tuttavia, ancora in fase di elaborazione soluzioni complete per la misura della qualità della voce su reti VoIP in maniera completa e affidabile, che sono la premessa indispensabile per utilizzare con successo tale tecnologia.

5. LA "QUALITY OF SERVICE" IN UNA RETE VOIP

Il servizio offerto dalle reti basate sull'utilizzo del paradigma Internet è, tradizionalmente, caratterizzato dall'assenza di qualsiasi tipo di garanzia sulle prestazioni della comunicazione (best effort). Tuttavia, lo sviluppo e la rapida diffusione di nuove applicazioni real-time, quali la telefonia su IP, l'audioconferenza e la videoconferenza, hanno portato a un radicale mutamento nelle richieste dell'utenza, tale da rendere più che mai urgente la realizzazione di architetture di rete che consentano di offrire presta-

zioni adeguate. Per realizzare una differenziazione del servizio si è reso necessario introdurre il concetto di Classi di Servizio (CoS): ogni classe raggruppa tipi di traffico con requisiti analoghi, ai quali la rete deve essere in grado di offrire prestazioni adeguate alle singole richieste. Con l'espressione "Qualità del Servizio" si fa, invece, riferimento a una serie di parametri che misurano la prestazione della comunicazione che è offerta dalla rete: caratteristiche di *throughput* e di banda, tempi di latenza, probabilità di perdita di pacchetti.

Un servizio può garantire un limite alla probabilità di scarto di pacchetti appartenenti a un certo flusso, oppure può garantirne l'assenza. Per soddisfare i requisiti sempre più stringenti relativamente ai parametri di QoS, l'IETF ha provveduto a definire due architetture avanzate per la rete Internet, denominate rispettivamente *Internet Integrated Services* e *Differentiated Services* [5,13]. Senza scendere in confronti dettagliati tra le due architetture, si può sicuramente affermare che il modello di utilizzo universalmente riconosciuto è quello mostrato in figura 6. Tecniche di ingegneria del traffico vengono poi utilizzate per l'allocazione ottima del carico e delle risorse sulla rete.

La qualità di servizio *end-to-end* relativa a una sessione VoIP è caratterizzata sia dalla qualità relativa alla fase di instaurazione della chiamata (*call set-up quality*), sia dalla qualità relativa alla fase attiva della chiamata (*call quality*). Complessivamente, la qualità di servizio end-to-end è influenzata da tutti i componenti intermedi attraversati, e cioè dal terminale IP, dalla rete d'accesso IP, dal backbone IP, dal VoIP gateway, dal VoIP *gatekeeper*, dalla rete telefonica, dal terminale telefonico.

La *call set-up quality* dipende dal ritardo nella instaurazione della chiamata (*call set-up delay*), percepito in termini di rapidità di risposta del servizio.

I fattori che contribuiscono al *call set-up delay* sono, sostanzialmente, il ritardo di set-up nella rete telefonica e nella rete di accesso IP, il ritardo di segnalazione nel backbone IP e di set-up negli elementi di rete (*media gateway*, *signalling gateway*, *media gateway controller*, *gatekeeper* ecc.), e, infine, il tempo di ac-



cesso e di elaborazione nei *server* dei servizi di *directory* o di autenticazione.

La call quality dipende, invece, dal ritardo end-to-end (*end-to-end delay*), che ha effetti sulle situazioni interattive della chiamata, ovvero in tutte quelle situazioni in cui il dialogo fra i due interlocutori necessita di risposte frequenti da parte dell'uno o dell'altro e dalla qualità del parlato end-to-end (*end-to-end speech quality*), che coinvolge anche situazioni non interattive della chiamata. In questo caso, quella che viene a mancare è la possibilità di comprendere le parole pronunciate dall'interlocutore, indipendentemente dal fatto che queste prevedano o meno replica. Anche l'eco contribuisce alla qualità della comunicazione, divenendo tanto più fastidiosa quanto maggiore è il ritardo end-to-end. Per eliminarla è necessario utilizzare appositi cancellatori d'eco.

Per quanto concerne l'end-to-end delay, il ritardo percepito dagli interlocutori deve essere costante e contenuto, al fine di mantenere la comunicazione interattiva. Ritardi inferiori a 15 ms non creano alcun fastidio, mentre l'interattività diventa difficoltosa per ritardi superiori a 400 ms. Valori intermedi di ritardo consentono un'interattività accettabile, ma richiedono la presenza di cancellatori d'eco ed è comunque consigliabile che il ritardo non superi i 150 ms.

Un problema tipico delle reti dati è la variabilità del ritardo, nota come jitter, che tende a corrompere l'intelligibilità della comunicazione. È, quindi, necessario rimuovere il jitter, compensandolo con l'ausilio di memorie tampone (*buffer*). L'end-to-end delay è determinato da molteplici contributi quali il ritardo di *bufferizzazione* nel terminale IP (scheda audio o scheda telefonica, *modem* o adattatore di rete), il ritardo di codifica e decodifica (formazione della trama, elaborazione, *look-ahead*), il ritardo per costituire i pacchetti (in trasmissione) e quello dei buffer (in ricezione) e il ritardo di rete (propagazione, trasmissione, accodamento ed elaborazione nei *router* e nei gateway e ritardi dovuti ai meccanismi protocollari), trascurabili nelle reti telefoniche standard.

Con riferimento alla qualità del servizio, è interessante valutare l'efficienza delle reti a

commutazione di pacchetto nel caso del trasporto di segnali vocali. Il confronto con la commutazione tradizionale di circuito può avvenire sia da un punto di vista percettivo dell'utente finale (MOS, *Mean Opinion Score*), sia prendendo in esame la caratterizzazione di parametri end-to-end; ma deve tenere in considerazione anche l'efficienza di utilizzo delle risorse di rete [3].

Per quanto concerne quest'ultimo aspetto, sono state identificati due differenti tipi di efficienza: quella real-time, intesa come la capacità della rete di trasportare traffico real-time, senza influenzare il servizio offerto ad altri tipi di traffico e l'efficienza di trasporto che identifica, invece, la capacità di sfruttare le risorse di rete combinando al meglio traffico real time con quello best-effort, rispettando le richieste in termini di QoS.

Gli studi effettuati in proposito hanno evidenziato che garanzie deterministiche sul ritardo di trasmissione in una rete a pacchetto richiedono una sovra-allocazione di risorse; in altre parole, per ogni chiamata devono essere allocate più risorse di quante ne sarebbero state impiegate calcolando il bit rate di codifica del flusso dati e l'instanziazione (l'*overhead*) introdotta dagli *header*. Sono di conseguenza definiti *rate reale* per chiamata e *rate apparente*. Il primo considera unicamente l'occupazione di banda dovuta alla codifica e agli header, mentre il secondo considera quanta banda deve essere allocata a ogni chiamata in modo da rispettare i requisiti sui parametri end-to-end. Inoltre, è stato altresì evidenziato che l'efficienza real-time dipende direttamente dalla dimensione dei pacchetti trasportati sulla rete. A sua volta, la dimensione ottima dipende da parametri relativi alla singola chiamata, per esempio devono essere tenuti in considerazione il numero di salti (*hop*) attraversati dal flusso dati. Al crescere della dimensione dei pacchetti diminuisce l'efficienza real time in quanto si verificano ritardi di pacchettizzazione maggiori e, quindi, un maggiore *rate* apparente.

Analogamente, al crescere del numero di nodi attraversati dalla chiamata sono richieste maggiori risorse per poter rispettare i requisiti di QoS. Diminuisce di conseguenza la capacità dei link in termini di *Erlang*

(unità di misura del traffico telefonico: rappresenta il tempo di utilizzo continuativo, valutato in ore, di un canale voce). Se si vogliono, infatti, rispettare i vincoli temporali imposti in fase di progetto, nel caso aumenti il numero dei nodi da attraversare, si deve, ovviamente, diminuire il tempo che il pacchetto impiega per attraversare il singolo nodo e, quindi, anche in questo caso, aumenta il rate apparente del singolo flusso.

6. IMPATTO DELLA QOS SUL DIMENSIONAMENTO DELLA RETE

L'impatto dei requisiti in termini di ritardo end-to-end della chiamata sull'attività di dimensionamento delle capacità dei link della rete richiede un'analisi di tutti i fattori che influenzano il ritardo di trasmissione dei dati da sorgente a destinazione. In prima approssimazione, è possibile considerare come termini significativi i ritardi dovuti alle seguenti operazioni, effettuate dai terminali o dai nodi della rete IP: ritardo di Codifica/Decodifica, reti voce locali di origine e destinazione e *Core Network*.

Per quanto riguarda i ritardi introdotti in fase di codifica e di decodifica è conveniente utilizzare la stima proposta nell'ambito del progetto TIPHON [10], nella quale si evidenziano tre fattori: *ritardo di bufferizzazione del terminale IP* che riguarda, generalmente, ritardi introdotti dall'*hardware* del terminale utilizzato (per esempio, i buffer in ingresso ai convertitori A/D); *ritardo del codificatore* che comprende i tempi di elaborazione, di *look ahead* e di bufferizzazione di un frame di campioni su cui si esegue la codifica e *ritardo di pacchettizzazione* dovuto alla necessità di dover considerare la possibilità di inserire più di un frame generato dal codec in un solo pacchetto IP.

Ipotizzando trascurabile il primo, il progetto TIPHON suggerisce come stima del ritardo del codificatore la seguente formula:

$$\tau_c = 2f_s + \Delta_a + \Delta_{la}$$

dove f_s rappresenta la durata del frame, Δ_a il ritardo di elaborazione dell'algoritmo (DSP) e Δ_{la} la componente look ahead (il fattore

"look ahead" tiene conto del fatto che i codificatori, al fine di migliorare le prestazioni in termini di bit rate, considerano durante la codifica di un voice frame anche alcuni campioni del frame successivo).

Nel *decoder* si assume che vi sia ulteriore ritardo dovuto al tempo di elaborazione e all'uso di buffer di uscita. Il ritardo di elaborazione totale attraverso *encoder* e *decoder* ($2\Delta_a$) si assume sia sensibilmente inferiore alla lunghezza del buffer di uscita, solitamente scelto in modo da contenere un *voice frame*.

Questo conduce alla valutazione approssimativa del ritardo coder/decoder, nella quale si trascura il tempo di *processing* del DSP:

$$\tau_c = 3f_s + \Delta_{la}$$

Il ritardo di pacchettizzazione incide in modo direttamente proporzionale al numero di voice frame inclusi in ogni pacchetto IP: per ogni frame aggiuntivo si deve considerare un tempo di pacchettizzazione pari alla durata del frame.

La trattazione proseguirà ipotizzando una tipologia di rete dove sono indicati anche i vari tipi di ritardo sperimentati nelle diverse parti in cui viene suddivisa la rete. Per stimare il ritardo di accodamento dei pacchetti IP a ogni hop attraversato non è possibile utilizzare l'approccio tradizionalmente utilizzato nella telefonia, basato sui modelli markoviani o reti di code M/M/1 o più in generale sulla formula di Erlang, non essendo distribuito secondo un andamento poissoniano il tempo di interarrivo dei pacchetti al nodo IP. Al contrario, il traffico IP in ingresso a ogni nodo può essere considerato deterministico, in quanto i buffer di uscita dei codificatori tendono a generare un bit rate costante. Inoltre, assumendo di avere pacchetti a lunghezza costante, anche il traffico in uscita dal nodo in questione risulta essere di tipo deterministico, essendo uguale il tempo di smaltimento impiegato dal nodo per ogni pacchetto. Analizzato il ritardo dovuto alla codifica, si passa direttamente a definire i ritardi dovuti alla rete locale di partenza o di arrivo (*Originating/Terminating Network*): i due contributi possono es-



sere visti come simmetrici, e quindi il computo sul ritardo viene effettuato considerando una sola delle due tratte. Per avere il ritardo totale basta raddoppiare il risultato. Sostanzialmente, si possono individuare i seguenti contributi [10]:

■ **Fixed Switching Delay:** è quello dovuto ai meccanismi di forwarding interni allo switch locale; può diventare significativo nel caso di switch lenti.

■ **Variable Voice Contention Delay:** è il ritardo dovuto alla contesa per l'accesso alle risorse del link da parte dei pacchetti voce; poiché si tratta di pacchetti aventi stesse caratteristiche e priorità, può essere applicato il modello della teoria delle code relativo a un traffico a bit rate costante che condivide un'unica coda senza priorità;

■ **Variable Data/Voice Contention Delay:** è il ritardo dovuto alla contesa fra i pacchetti voce e i pacchetti dati; poiché si ipotizza l'utilizzo di un meccanismo di code a priorità, questo caso si può verificare solo nel momento in cui un pacchetto dati abbia già impegnato il canale al momento dell'arrivo del pacchetto voce.

■ **Fixed Serialization Wan Delay:** è il ritardo di trasmissione dei pacchetti sulla WAN che porta alla Core Network. Per quanto concerne l'architettura di Core Network, in bibliografia esistono diversi modelli a diversa complessità: per una stima di primo livello è possibile scegliere un modello molto diffuso in cui si considerano significativi nei confronti del ritardo di trasmissione end-to-end solamente i nodi aventi N interfacce in ingresso concorrenti verso una o più interfacce di uscita. Nei casi in cui il nodo IP abbia una sola interfaccia in ingresso, si è supposto che la/e interfaccia/e di uscita e la capacità di processing siano state opportunamente dimensionate in modo da non generare ritardi di accodamento: ogni pacchetto in ingresso viene

immediatamente servito e inoltrato sull'opportuna interfaccia di uscita. Considerato, quindi, un nodo IP isolato e analizzandone il comportamento in termini di ritardo di bufferizzazione, l'adozione di tali modelli porta a concludere che poiché solitamente le velocità dei link appartenenti alla Core Network spaziano solitamente nella cosiddetta "larghissima banda", a meno di non trovarsi a coprire distanze tali da rendere significativo il ritardo dovuto alla propagazione dell'onda luminosa nelle fibre (un valore tipico è di circa $5 \mu\text{s}$ per km) oppure a effettuare un numero molto elevato di passi (hop), tale ritardo sarà trascurabile rispetto a quello dovuto alla periferia della rete e ai Codec.

7. CONCLUSIONI

Questa rapida carrellata sui benefici e, in parte, sulle problematiche legate all'introduzione della fonia sulle reti dati, con particolare enfasi sugli aspetti legati alle **QoS**, ha mostrato come la VoIP, intesa come tecnologia, permette riduzioni di costo in quanto utilizza l'infrastruttura di rete IP per la realizzazione delle tradizionali applicazioni di fonia: l'utilizzo di Internet per chiamate a lunga distanza e l'utilizzo di normali PC come PBX aziendali (*Computer Telephony Integration*). Utilizzando le nuove tecnologie disponibili per la codifica e la compressione della voce, in aggiunta a tecniche quali la compressione dei silenzi è possibile il trasporto di un numero di canali vocali maggiore rispetto al caso della telefonia classica. La convergenza di voce e dati permette la realizzazione di nuovi servizi realizzati secondo modelli funzionali tra cui PC-to-PC, PC-to-Phone, Phone-to-Phone, fax-to-fax, PC-to-fax, Web-to-fax, E-mail-to-fax. Infine, servizi quali IP-PBX, Web-call center, IP-call

La **QoS** end-to-end dipende da tutti i componenti coinvolti nell'attivazione e gestione della chiamata. L'interesse maggiore è rivolto agli aspetti di QoS associati ai componenti di tipo IP, essendo ormai consolidati tutti gli aspetti relativi alla QoS nelle reti telefoniche a commutazione di circuito. Tra i principali problemi che caratterizzano le reti telefoniche a circuito vi è l'eco. La criticità di tale fenomeno risulta accentuata in un contesto VoIP nel quale il ritardo introdotto dalla trasmissione della voce a pacchetto contribuisce a diminuire sensibilmente la tolleranza umana all'effetto eco. Il fenomeno è affrontato e risolto tramite l'introduzione di opportuni soppressori d'eco a livello di media gateway e, infatti, la maggior parte dei gateway disponibili sul mercato offrono questa importante funzione.

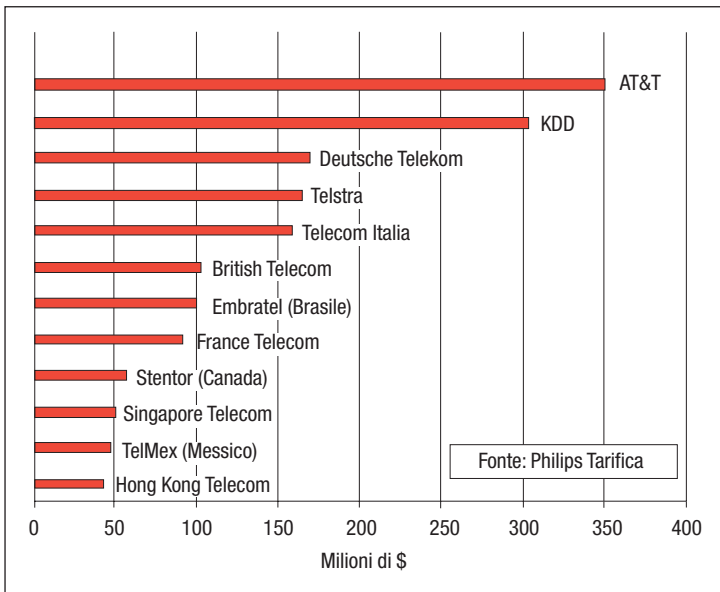


FIGURA 7 Perdite sulle telefonate internazionali nel 2001

center, *Internet Call Waiting*, fanno di VoIP una tecnologia che crea nuove fonti di fatturato per le imprese.

Per avere un'idea delle proporzioni del fenomeno in gioco, in termini di perdite economiche della telefonia classica si osservi la figura 7 che mostra la previsione fatta da Philips Tarifica riguardo le perdite sulle telefonate internazionali nel 2001 per alcuni tra i più importanti operatori telefonici mondiali. La previsione è stata ampiamente rispettata.

Se è vero che una rete IP senza il supporto della QoS non è in grado di fornire gli opportuni requisiti alle applicazioni real time, è anche vero che la QoS, e la sua gestione, possono avere un costo molto elevato in termini di risorse richieste alla rete, sia come capacità di banda che come capacità di elaborazione. È necessaria, pertanto, un'attenta pianificazione dell'utilizzo della QoS, che deve tenere conto della parte della architettura di riferimento che si sta trattando, della struttura fisica della rete, della velocità dei link e così via. È molto importante, ad esempio, valutare la capacità dei link su cui si vogliono configurare politiche di QoS. Il traffico voce ha una banda estremamente limitata, in genere si parla di poche decine di kbit/s per flusso voce, per contro, le attuali reti locali (almeno quelle basate su protocollo *Ethernet*) hanno velocità tipiche che partono dai 100 Mbit/s.

Anche considerando una rete locale che debba gestire 100 flussi voce contemporanei, considerando per ogni flusso una capacità di banda necessaria a livello *Data Link* di 40 kbit/s, che è un valore molto conservativo, si arriva a un valore di circa 4 Mbit/s, ovvero il 4% della capacità di banda disponibile, che è un valore del tutto marginale. Si nota, pertanto, che è molto più importante la corretta progettazione e realizzazione della rete, intesa come pianificazione dei domini broadcast, limitazione dei domini di collisione tramite un massiccio utilizzo di switch, ove possibile, e così via, piuttosto che cercare di trasmettere traffico real-time su un'architettura di rete mal configurata, imponendo pesanti sovrastrutture di QoS che potrebbero portare facilmente al collasso. Tipicamente, in ambito locale quello che si fa è utilizzare VLAN [15, 17] per suddividere la rete voce da quella dati, combinato con la creazione di code a stretta priorità che gestiscono il traffico real-time. Si utilizzano, poi, meccanismi quali WRED (*Weighted Random Early Detection*) per evitare per quanto possibile congestioni.

Il discorso cambia decisamente quando si passa dalla LAN alla WAN, dove al traffico locale si somma il traffico generato dalle altre LAN. Inoltre, normalmente, i collegamenti LAN-WAN sono a velocità molto più bassa, per cui l'eventuale traffico voce comincia a incidere in modo tutt'altro che marginale.

Nel caso di collegamenti a velocità inferiore ai 2 Mbit/s, è senz'altro utile prendere in considerazione strutture di QoS più stringenti, quali IP precedence o RSVP (*Resource Reservation Protocol*), l'utilizzo di code CB-WFQ (*Class Based Weighted Fair Queueing*) per una gestione ottimale della capacità di banda disponibile e, inoltre, tutti quegli accorgimenti utili ad aumentare l'efficienza della trasmissione, quali per esempio LFI (*Link Fragmentation and Interleaving*) e CRTP (*Compressed Real Time Protocol*). La qualità del servizio offerta al flusso voce è fortemente dipendente dal livello di saturazione della rete e, pertanto, l'utilizzo di politiche di QoS adeguate permette di avere una qualità costante, e rispondente alle specifiche desiderate, anche in condizioni di forte saturazione.

Bibliografia

- [1] AA.VV.: *Towards an Integrated Architecture for the Harmonization for PSTN and Internet Services*. Proceedings Actes – 6th International Conference on Intelligence in Networks, Gennaio 2000.
- [2] Arora R: *Voice over IP: Protocols and standards*. 23 novembre 1999.
- [3] Baldi M, Bergamasco D, Risso F: *On the Efficiency of Packet Telephony*. 7-th IFIP International Conference, on Telecommunication Systems, Nashville, Tennessee, USA, Marzo 1999.
- [4] Baldi M, Bergamasco D, Guarene E: *Architectural choices for packet switched telephone networks*. International Switching Symposium (ISS '97), Settembre 1997.
- [5] Braden R, Zhang L, Berson S, Herzog S, Jamin S: *Resource ReSerVation Protocol*. RFC 2205, IETF, Settembre 1997.
- [6] Casner S, Jacobson V: *Compressing RTP/UDP/IP Headers for Low-Speed Serial Links*. RFC 2508, IETF, Febbraio 1999.
- [7] Decina M: *Stato dell'Arte: Voice over IP Architetture e servizi*. Politecnico di Milano/Cefriel, Seminario Teach, Maggio 2000.
- [8] ETSI Technical Report: *Design Guide for Elements of a TIPHON connection from an end-to-end speech transmission performance point of view*. Draft RTR 101 329-7, ETSI TIPHON Working Group 5, Febbraio 2001.
- [9] Garavaglia E: *Mailing List Teach Innovazione*. www.teach.it
- [10] Handley M, Schulzrinne H, Schooler E, Rosenberg J: *SIP: Session Initiation Protocol*. RFC 2543, IETF, Marzo 1999.
- [11] International Telecommunication Union (ITU-T): *Packet Based multimedia communication systems*. ITU-T Recommendation H.323 Version 3, Maggio 1999.
- [12] McPhillips E: *The Factors Affecting the Growth of VoIP*. HP Labs Technical Reports, Maggio 1999.
- [13] Nichols K, Blake S, Baker F, Black D: *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*. RFC 2474, IETF, Dicembre 1998.
- [14] Pescapè A, De Petris R, Ventre G, Fadini B: *IP Telephony Bandwidth Calculator: a web-based application for network planning support*. AICA 2001, XXXIX Annual Congress p. 363-374.
- [15] Pescapè A, Ventre G, Barone GB: *Vlan Manager: An Architecture for Automatic VLAN design, configuration and management*. AICA 2000, XXXVIII Annual Congress, p. 327-342.
- [16] Pescapè A, Esposito M, Romano SP, Ventre G: *A simulation model for bandwidth allocation of voice channels over IP networks*. ISCS 2001, p. 32-37 – ISBN 88 7146611-X. Cuen Editore.
- [17] Pescapè A, D'Antonio S, Ventre G: *Un'applicazione per il Supporto del traffico real time su reti IP*. AICA 2002, XL Annual Congress p. 645-659.
- [18] Pescapè A, Barone GB, Ventre G: *Traffico voce su reti IP: analisi e misure di protocollo*. AICA 2002, XL Annual Congress - p. 661-676.
- [19] Pracht S: *Factors in the success of Voice Quality in Converging Telephony and IP Networks*. *International Engineering Consortium – Annual Review in Communications*, Vol. 53, section IV, Settembre 2000.
- [20] Schulzrinne H, Casner S, Frederick R, Jacobson V: *RTP: A Transport Protocol for Real Time Applications*. RFC 1889, IETF, Gennaio 1996.

BRUNO FADINI è Professore Ordinario di Calcolatori Elettronici presso la Facoltà di Ingegneria dell'Università di Napoli "Federico II". Tiene sin dal 1963 un corso dove tratta argomenti di architettura dei sistemi per l'elaborazione dell'informazione. È stato direttore del Dipartimento di Informatica e Sistemistica e Presidente del corso di Laurea in Ingegneria Elettronica della Facoltà di Ingegneria dell'Università di Napoli "Federico II" e presidente dell'AICA. Attualmente è direttore del CINI.
E-mail: fadini@unina.it

ANTONIO PESCAPÈ è stato docente di Calcolatori Elettronici I, Elementi di Informatica e Reti Logiche presso la Facoltà di Ingegneria dell'Università di Napoli "Federico II". È membro del gruppo COMICS del GRID e attualmente frequenta il Dottorato di Ricerca in Ingegneria Informatica e Automatica presso il Dipartimento di Informatica e Sistemistica della Federico II di Napoli; svolge inoltre attività di ricerca presso il laboratorio ITEM del CINI di Napoli.
E-mail: pescapè@unina.it

GIORGIO VENTRE è Associato di Sistemi per l'Elaborazione dell'Informazione presso il Dipartimento di Informatica e Sistemistica dell'Università di Napoli Federico II. Dal 1990 al 1991 è stato ricercatore a contratto presso il Centro di Ricerca sul Calcolo Parallelo e i Supercalcolatori del CNR. Dal 1991 al 1993 è stato Post-doctoral fellow presso l'International Computer Science Institute di Berkeley. Dal 2000 è inoltre Direttore del Laboratorio Nazionale per l'Informatica e la Telematica Multimediali (ITEM) del Consorzio Interuniversitario Nazionale per l'Informatica.
E-mail: giorgio@unina.it