

La tecnologia VoIP: architettura e protocolli

Mariano Severi

Sono ormai passati oltre cento anni da quando, nel 1876, Alexander Graham Bell realizzò la prima trasmissione vocale su cavo mediante un semplice circuito 'ring-down': due persone ai capi di una connessione punto-punto riuscivano a comunicare in maniera unidirezionale. Il

Nel 1867 Alexander Graham Bell fece comunicare con un circuito ring-down due persone, oggi le trasmissioni vocali su rete IP avvengono con tecnologia VoIP

secolo scorso ha visto una rapida evoluzione della tecnologia con l'affermazione delle reti Pstn (Public Switched Telephone Network) e la diffusione dei servizi telefonici in ambito consumer. Mediante una struttura a switch su più livelli, le reti Pstn sono in grado di connettere due utenze; la voce viene codificata in formato PCM e trasmessa in maniera sincrona su un canale TDM (Time Division Multiplexing). Signaling System 7 (SS7) è quindi lo standard adottato per la gestione e il 'routing' delle chiamate, lo scambio delle informazioni, il computo e l'addebitamento dei costi presso l'utente, il supporto per le funzionalità di 'intelligent network'. Negli anni, diversi ulteriori servizi sono stati introdotti nelle reti Pstn per rispondere alle esigenze di diverse applicazioni; tra i più diffusi, ad esempio, l'avviso e il trasferimento di chiamata, le chiamate in teleconferenza o, più recentemente, il blocco in ingresso o in uscita di specifici numeri telefonici o la capacità di richiamare in automatico in caso di linea occupata o ricevente assente. La tecnologia Pstn presenta tuttavia delle limitazioni intrinseche che sono diventate sempre più pesanti con la crescente convergenza delle applicazioni nel settore multimediale. L'architettura di base, con la maggior parte dei servizi strettamente legati tra loro, non è sufficientemente flessibile per consentire prestazioni elevate; non è possibile, ad esempio, utilizzare una singola linea analogica per ricevere contemporaneamente dati, telefonate e video. Inoltre non consente un utilizzo efficiente della banda di trasmissione dal momento che, come noto, lo stabilirsi della connessione rende il canale occupato durante l'intera comunicazione. Paradossali, in questo

senso, sono le situazioni in cui entrambi gli utenti ai capi della linea telefonica restano in silenzio; il canale è occupato ma non sono inviati dati. Anche in situazioni più realistiche, dal momento che il canale è bidirezionale ma le voci normalmente non si sovrappongono, e considerate le inevitabili pause della conversazione, l'occupazione di banda è significativamente inferiore al 50 per cento.

Dall'esercito all'utente finale

Il VoIP (Voice Over IP) è la nuova tecnologia per le trasmissioni vocali su rete IP. Rispetto alle reti Pstn, la comunicazione avviene mediante trasmissione asincrona sulla rete in modo da ottimizzare l'occupazione di banda; il segnale audio viene rilevato, convertito in digitale e codificato in pacchetti da inviare. Tecniche di compressione appropriate sono applicate per ridurre la quantità di informazioni trasmesse.

Il primo esempio di trasmissione vocale su rete risale al 1973 con l'implementazione del Network Voice Protocol su Arpanet; l'obiettivo era dimostrare la capacità di trasferire voce a basso 'bit rate' e a elevata qualità in maniera sicura nell'ambito di comunicazioni militari su scala mondiale. La tecnologia è stata resa disponibile a utenti finali dagli inizi degli anni '80, ma le difficoltà tecnologiche e la scarsa capillarità della rete non hanno consentito una significativa diffusione. È soltanto nel 1995 che viene commercializzato il primo software per applicazione VoIP su Internet da parte di VocalTec Communication, la compagnia israeliana produttrice di apparecchiature per telefonica fondata da Alan Cohen e Lior Haramatry, inventori del primo transceiver audio; le comunicazioni sono tuttavia ancora limitate a utenze connesse alla rete Internet. Per collegare i tradizionali terminali bisogna attendere il 1997, con la realizzazione di Softswitch di Level3, prototipo dei moderni dispositivi per l'interoperabilità di reti a circuito e a pacchetto che ha spianato la strada all'integrazione delle classiche reti Pstn con le piattaforme per applicazioni VoIP. Negli ultimi anni, con la diffusione delle tecnologie a larga banda, il mercato è cresciuto esponenzialmente; nel 2008 è previsto che il fatturato dell'industria superi i tre bilioni di dollari con una crescita di oltre il 20 per cento delle utenze. Il futuro punta decisa-

mente alla realizzazione di una 'next generation network', una rete a pacchetti, secondo la definizione riportata nelle raccomandazioni ITU, in grado di fornire accesso dovunque e senza restrizioni a servizi di telecomunicazione basati su diverse tecnologie di trasporto a larga banda ma indipendenti da queste.

Diversi sono in effetti i vantaggi della tecnologia VoIP. Tra questi vi è certamente una significativa semplificazione delle reti per il trasporto dati con la convergenza delle diverse tecnologie che assicura riduzione di costi e maggiore flessibilità. La condivisione dell'infrastruttura di trasporto consente

facilmente la condivisione dei contenuti oltre a favorire lo sviluppo di nuovi servizi a valore aggiunto; si pensi ad esempio alla possibilità di uniformare server e indirizzi per i servizi di messaggiera vocale ed e-mail. La natura stessa della trasmissione dati a pacchetti su rete consente la delocalizzazione dei servizi da un lato e la portabilità di questi dall'altro. Un utente può accedere alla propria linea telefonica indipendentemente dalla sua locazione; un call center può essere strutturato in più uffici dislocati sul territorio anche in aree geografiche agli antipodi l'una dall'altra, rispondendo però allo stesso numero telefonico e assicurando così il servizio

LA QUALITÀ DELLA VOCE

Uno degli aspetti chiave nell'affermazione della tecnologia VoIP è certamente la qualità della voce; del resto, se si esclude il risparmio in bolletta promesso dalla nuova tecnologia, è probabilmente il principale termine di paragone con le tradizionali soluzioni Pstn che l'utente assume come riferimento. Tra i diversi fattori che influenzano la qualità della voce in una trasmissione su rete, vi sono, da un lato, il tipo di codifica adottato nella digitalizzazione del segnale, dall'altro, le prestazioni della rete. In precedenza abbiamo già accennato al processo di campionamento e codifica della voce per la trasmissione: per ridurre la quantità di informazioni da trasmettere è necessario applicare algoritmi di compressione con perdita che conducono a un degrado del segnale ricostruito. Il risultato è un compromesso tra la qualità che si intende ottenere e il costo in termini di banda allocabile alla trasmissione. In termini, invece, di prestazioni della rete i fattori che intervengono a determinare la qualità del servizio sono principalmente la latenza di trasmissione, il jitter, gli errori e le perdite di pacchetti: la latenza massima ammissibile, ad esempio, per riuscire ad avere una conversazione senza fastidiosi tempi di attesa deve essere tipicamente inferiore ai 300 ms. Si consideri, come riferimento, che la trasmissione su fibra ottica da un capo all'altro del mondo richiede circa 70 ms. La pre-

senza invece di jitter nella trasmissione (inteso come variazione del tempo medio di attesa nella ricezione di un pacchetto) richiede che il nodo ricevitore abbia una capacità di 'buffering' adeguata per consentire di riprodurre la voce con una precisa sincronizzazione. Tale sincronizzazione è del resto inficiata dalla presenza di errori nella trasmissione e dalla perdita di pacchetti, benché si sia osservato che perdite entro il cinque per cento permettono di ricostruire un audio di qualità accettabile. In alcuni casi i pacchetti persi vengono semplicemente sostituiti replicando quelli precedenti; in altre applicazioni più sofisticate possono essere adottati algoritmi di interpolazione del segnale per ricostruire le informazioni corrotte.

Altro punto importante, infine, è la cancellazione dell'eco, sia quella acustica dovuta all'interazione tra microfoni e altoparlante o cuffie, sia quella di linea legata al disadattamento del canale di trasmissione. Tipicamente un'eco superiore ai 25 ms crea fastidio nella comunicazione; per questo esistono standard che supportano la cancellazione di eco con code fino a 128 ms.

Così come diversi sono i fattori che influenzano la qualità del servizio, allo stesso modo esistono diversi strumenti per misurare la qualità della voce a seconda dei parametri presi come riferimento e dei metodi adottati. I metodi di tipo intrusivo, ad esem-

pio, si basano sull'invio di campioni vocali sulla rete da un nodo all'altro; il nodo ricevente confronta il segnale ricostruito con il valore atteso per stimare il degrado del segnale. L'elevata complessità degli algoritmi di stima nei sistemi embedded, tuttavia, non consente l'utilizzo di tali tecniche in applicazioni in tempo reale. Tra i metodi di questo tipo i più diffusi sono: Psqm (Perceptual Speech Quality Measurement) rivolto principalmente alla valutazione delle prestazioni degli algoritmi di codifica e compressione; Pesq (Perceptual Evaluation of Speech Quality), estensione del precedente; Pams (Perceptual Analysis and Measurement System) orientato alla trasmissione analogica del segnale. I metodi passivi, invece, non utilizzano un segnale di riferimento ma finiscono per essere legati a valutazioni soggettive. Il MOS (Mean Opinion Score) è forse il più diffuso: si basa sul giudizio collettivo di un gruppo di ascoltatori. Più o meno simile è l'E-Model, che deriva da un parametro di riferimento R compreso tra 1 e 100 prendendo in considerazione fattori diversi quali il ritardo della rete e il degrado del segnale lungo la trasmissione con la stessa accuratezza del MOS ma cercando di ridurre il numero di soggetti coinvolti nel test. Dall'E-Model è derivato il VQMon principalmente rivolto all'analisi delle prestazioni della rete in termini di errori, perdita di pacchetti, jitter e latenza.

24 ore su 24 senza impiegare personale durante l'orario notturno. Tra gli svantaggi principali della nuova tecnologia vi sono, invece, la non sempre buona qualità della trasmissione, inficiata da una serie di fattori tra i quali la latenza di trasmissione, la perdita di pacchetti, il jitter del canale; la stretta dipendenza della disponibilità del servizio dalla banda disponibile per la trasmissione dati e dalla presenza dell'alimentazione elettrica; la mancanza di supporto presso alcuni provider delle chiamate ai numeri di emergenza; la difficoltà di assicurare sicurezza e riservatezza delle conversazione a causa del rischio concreto che la propria trasmissione possa essere intercettata da altri utenti.

Architettura di una rete VoIP

La figura 1 mostra la possibile architettura di una generica rete per applicazioni VoIP in grado di offrire funzionalità e servizi di base analogamente a una tradizionale piattaforma Pstn e sulla stessa scala, seppure con i vantaggi discussi in

precedenza. La rete si basa, come si vede, su tre distinti provider che forniscono le porte di accesso alla rete mediante i 'gateway' e le infrastrutture di connessione tra le diverse componenti. Le funzioni base di controllo e gestione delle chiamate sono realizzate dai 'call agent', in altro modo indicati come 'media gateway controller', 'softswitch' o 'call controller'. Inclusi nella rete del provider, tali servizi sono demandati a tenere traccia dello stato delle chiamate, segnalare, terminare o inoltrare messaggi, originare il flusso di controllo dei dispositivi; funzionalità diverse non direttamente supportate sono trasferite invece sotto il controllo dei call agent agli 'application server'. Diversi sono i protocolli attualmente implementati, tra i quali, come vedremo in seguito, SIP, SIP-T, H.323, Biss, H.248, Mgcp/NCS, SS/, AIN e Isdn. I 'media server' svolgono invece funzioni di mixing nel caso, ad esempio, di chiamate a tre vie, annuncio di chiamata, traduzione delle codifiche e rilevamento dell'attività vocale, generazione e riconoscimento di tono, elaborazione

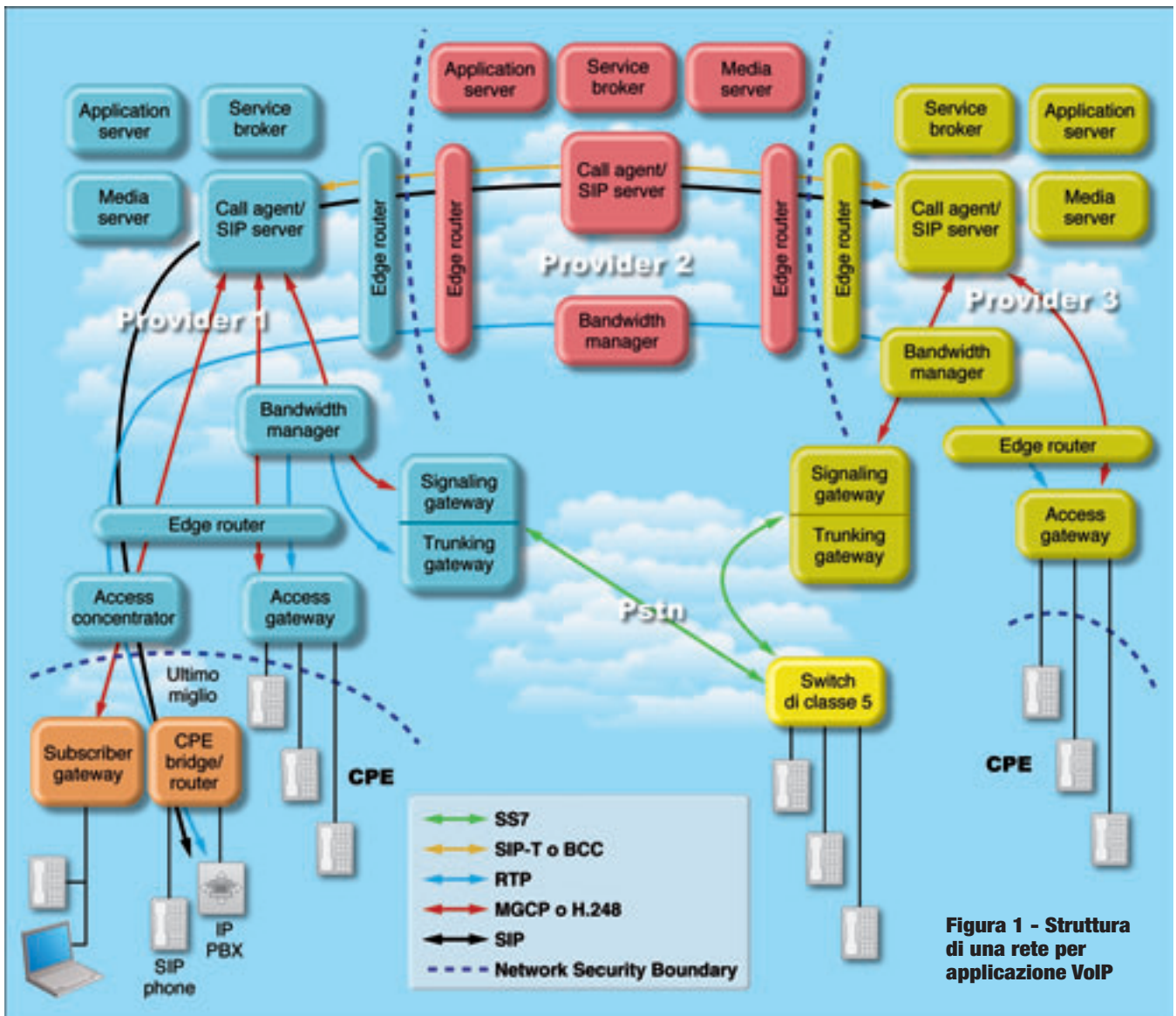


Figura 1 - Struttura di una rete per applicazione VoIP

dei fax. Il controllo della distribuzione dei servizi tra application server, media server, call agent ed eventualmente altre tecnologie quali SCP o 'parlay gateway' è gestito dal 'service broker'.

La connessione di reti IP con tradizionali reti Pstn è gestita mediante 'signaling gateway' e 'trunking gateway'; questi ultimi, in particolare, traducono i pacchetti dati della rete VoIP in formato compatibile con sistemi TDM tipicamente sotto il controllo dei call agent/media gateway controller e mediante protocolli tipo H.248 e Mgcp. Uno dei vantaggi principali delle reti VoIP, come abbiamo visto, è la gestione ottimale della banda disponibile per la trasmissione: il 'bandwidth manager', in questo senso, installato all'interno della rete del provider dei servizi, è responsabile della fornitura del QoS richiesto; controlla, in particolare, l'accesso alla rete delle singole chiamate in corso oltre a definire la politica di controllo dei flussi di informazioni sulla base delle singole chiamate all'interno degli Edge router. Questi implementano il routing del traffico IP sulla dorsale di distribuzione primaria. Access gateway e access concentrator, infine, consentono all'utente finale l'accesso alla rete IP presso la propria residenza per la fruizione dei servizi di telefonia. Gli access gateway, in particolare, interfacciano gli apparecchi telefonici tradizionali Pots (Plain Old Telephone Service). Gli access concentrator, invece, realizzano l' 'ultimo miglio' della rete IP consentendo l'accesso per applicazioni dati, video e audio.

A essi si connettono i tipici bridge/router che supportano, seppure in modo non nativo, applicazioni VoIP mediante impiego di apparecchi quali telefoni SIP o IP phone e i subscriber gateway in grado di ricevere, invece, simultaneamente dati e flussi voce da sorgenti tradizionali.

Lo stack di protocollo delle applicazioni VoIP

La figura 2 mostra uno schema semplificato dei principali protocolli utilizzati in applicazioni VoIP. Come si vede, la tecnologia poggia su reti IP implementate mediante architetture Sonet/SDH, ATM o Ethernet e utilizza i protocolli UDP e TCP a seconda del servizio che deve essere fornito. Lo standard TCP, in particolare, realizza una trasmissione 'full duplex', con 'acknowledge' e controllo di flusso che garantiscono, eventualmente mediante ritrasmissione, l'affidabilità della connessione; considerata tuttavia la struttura, il protocollo non è adatto alla trasmissione dei pacchetti vocali. Per tali applicazioni viene piuttosto utilizzato il protocollo UDP, caratterizzato da minore complessità e ideale per comunicazioni nelle quali la riduzione della latenza è prioritaria. La figura 2 mostra i tre servizi principali che concorrono a realizzare un'applicazione VoIP, ovvero il 'signaling' vero e proprio, che consente la gestione della connessione con la localizzazione dei nodi da connettere, la negoziazione delle risorse richieste e da utilizzare nella comunicazione e l'inizio e la

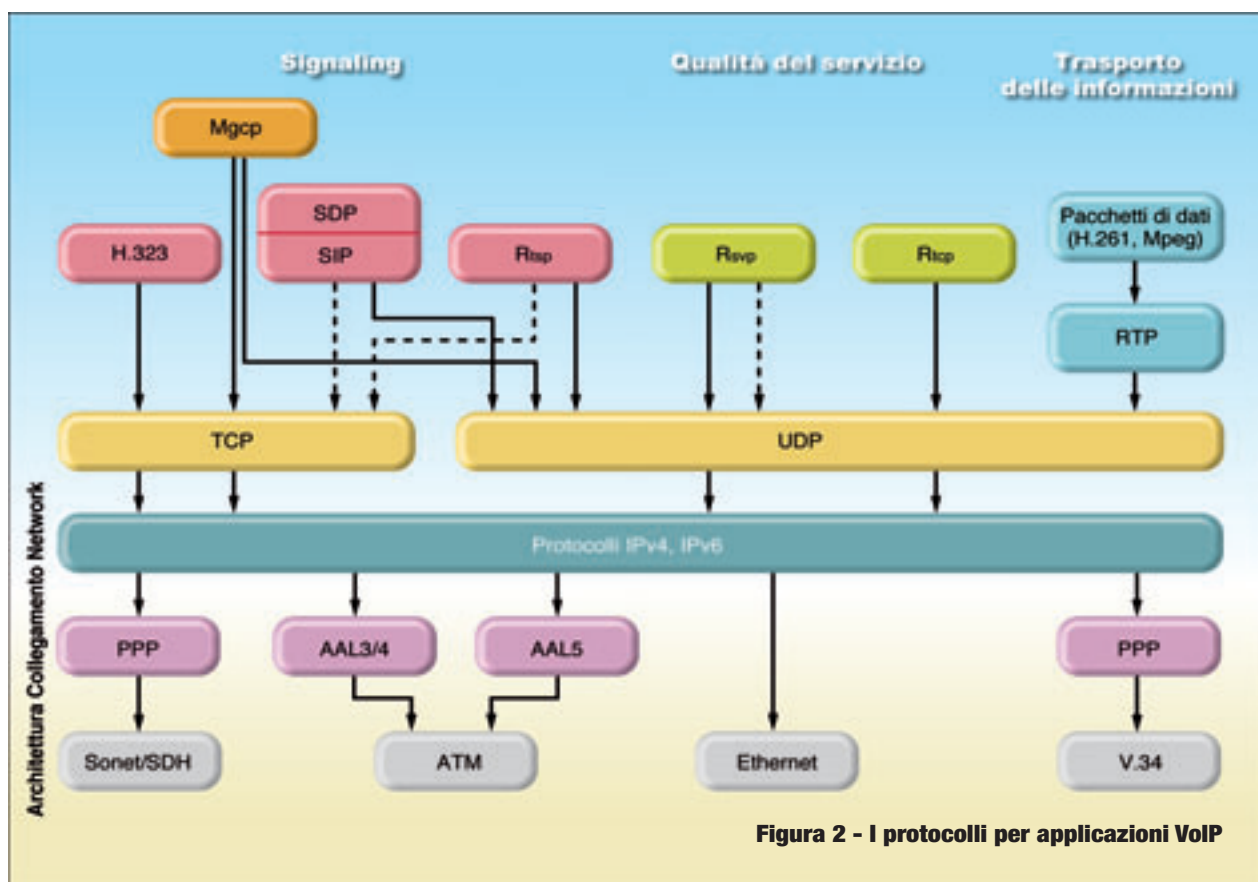


Figura 2 - I protocolli per applicazioni VoIP

chiusura della sessione; la qualità del servizio; il trasporto delle informazioni dati, audio e video. Per quanto concerne la parte di signaling i due principali standard sono H.323 (o più propriamente le funzionalità di RAS e le sezioni H.245 e H.225 del protocollo) e, più recente, SIP. RAS (Registration, Admission and Status), ad esempio, è l'insieme di servizi per le trasmissioni nell'ambito della specifica H.323 tra nodi comunicanti e 'gatekeeper' prima di attivare la chiamata; la connessione viene stabilita su UDP. Scopo principale è la registrazione della chiamata per individuare localizzazione e indirizzi IP dei nodi interessati, la verifica dell'ammissibilità della stessa e il controllo della banda e dello stato dei nodi da mettere in comunicazione. L'H.245, invece, gestisce la creazione del canale logico per la chiamata; la connessione per i messaggi di controllo viene stabilita su TCP. Durante la negoziazione i due 'end point' scambiano informazioni su capacità e preferenze consentendo, in particolare, di selezionare il codec per la trasmissione audio più adatto. L'H.225, infine, implementa la connessione su TCP per controllare l'inizio dalla chiamata, segnalarne una, verificarne lo stato, terminare la comunicazione. Più o meno simili sono le funzionalità del protocollo SIP (Session Initiation Protocol). Inizialmente definito per creare un meccanismo di connessione su larga scala, è fortemente orientato all'interoperabilità e basato sul riutilizzo di standard esistenti. Tra le caratteristiche principali vi sono uno schema di indirizzamento URL che assicura la portabilità dell'indirizzo indipendentemente dalla locazione fisica e la gestione di flussi multimediali diversi quali giochi interattivi, chiamate vocali o servizi di messaggia istantanea durante la stessa connessione. User agent e network server sono i due componenti principali: i primi includono gli utenti in connessione e il server che inizia la chiamata; gli altri sono invece i server proxy per la registrazione e redirezione della chiamata. Il meccanismo di comunicazione è basato su un semplice schema richiesta/risposta, tra le richieste consentite vi sono messaggi di invito da client a client, di fine della chiamata, per ottenere informazioni su altri client o server, di cancellazione di messaggi in invio, di registrazione; i tipi di risposte previsti sono invece indicazioni di successo della richiesta, errore del client, del server o di sistema, informazioni richieste.

Per quanto concerne, invece, la qualità del servizio, due dei protocolli più diffusi sono Rsvp e Rtcp. Rsvp (Resource Reservation Protocol), ad esempio, è un protocollo 'out of band', 'end to end' che serve alla gestione dell'allocatione dei servizi alle connessioni; la stazione ricevente opera come richiedente dei servizi. A ogni nodo che supporti il protocollo vengono verificati il livello di priorità e i privilegi (in termini, ad esempio, di latenza o di banda allocata) che è possibile concedere alla trasmissione richiesta in base sia al carico della rete sia alle altre connessioni contemporaneamente in essere; le applicazioni ricevono quindi un feedback sui servizi di trasporto accreditati. Il protocollo tuttavia non con-

sente di modificare la tabella di routing in caso di congestione del traffico sulla rete ma piuttosto permette in base alle informazioni di servizio associate di individuare la linea di connessione ottimale per la trasmissione. Tra gli svantaggi principali del protocollo vi sono la difficoltà di trasportare il servizio su larga scala e il tempo necessario per l'arbitraggio e l'allocatione dei servizi privilegiati. Rtcp (RTP Control Protocol) è invece il protocollo di supporto alle applicazioni per il trasporto delle informazioni quali RTP che consente di monitorare il traffico dei dati, controllare e identificare i servizi garantiti; alcune delle funzionalità principali sono, ad esempio, l'identificazione della sorgente all'interno dei gateway, il riporto di feedback dal ricevente verso un trasmettitore multicast, il supporto alla sincronizzazione di flussi dati multimediali di tipo diverso. Rtcp XR è una più recente estensione del protocollo Rtcp che definisce un insieme di metriche che possono essere implementate con costi ragionevoli in gateway, call managers e IP phone; utilizzando dei semplici algoritmi embedded, ad esempio, è possibile stimare il MOS (Mean Opinion Score, valutazione soggettiva della qualità della chiamata) o misurare il jitter della trasmissione. Il trasporto delle informazioni avviene invece mediante protocollo RTP (Realtime Transport Protocol) o cRTP. Su di essi si adattano i codec per la codifica dei dati in pacchetti. Nell'ambito delle trasmissioni VoIP alcuni dei codec più diffusi sono, ad esempio, G.711 basato su PCM e che assicura elevata qualità del segnale ricostruito con bit rate 64 kbps; G.723.1, per applicazioni a elevata compressione a 5.3 kbps e 6.3 kbps; G.726, tipicamente utilizzato in connessione PBX con bit rate a 32 kbps; G.728, basato su Celp per applicazioni a bassa latenza a 16 kbps. Il protocollo RTP aggiunge ai dati codificati informazioni per la trasmissione su UDP quali un numero sequenziale di pacchetto, il tempo di campionamento delle informazioni, l'identificativo della sorgente di sincronizzazione, il tipo di dati trasportati; tali informazioni consentono la corretta ricostruzione dell'informazione al nodo ricevente anche in presenza di errori di comunicazioni o perdita di pacchetti sulla rete. Il protocollo cRTP (Compressed RTP) è una variante che consente la compressione degli header per ridurre la quantità di informazioni trasportata da 24 kbps a 11,3 kbps nel caso, ad esempio, di codec G.729; in questo modo su un canale a 56 kbps è possibile stabilire fino a quattro connessioni simultanee.

Verso una nuova generazione

La crescente diffusione della rete su scala mondiale e la definizione di nuovi protocolli di comunicazione consentono la creazione di nuovi servizi a valore aggiunto per gli utenti. VoIP è una delle nuove tecnologie che stanno lentamente affermandosi; con la possibilità di trasferire il traffico vocale su rete, la convergenza delle applicazioni multimediale su una unica piattaforma condivisa di nuova generazione è sempre più vicina. ■