# VoIP: una interessante novità con molte problematiche di sicurezza

In questo articolo daremo una descrizione molto introduttiva ai protocolli di Voice-over-IP (VoIP) in modo da poter poi introdurre alcuni tra i rilevanti problemi di sicurezza di questa nuova tecnologia. Recentemente VoIP ha ricevuto molto interesse ed una grande diffusione, ma purtroppo ben poco si è fatto per gli aspetti di sicurezza che raramente sono considerati nelle implementazioni pratiche.

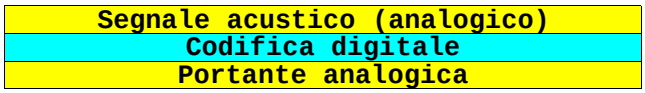
## 1. L'idea di base di VoIP

La principale tecnica delle telecomunicazioni moderne è di trasmettere informazioni modulando un segnale ondulatorio di base (detto anche *portante*), di solito con modulazioni di ampiezza o di frequenza. In altre parole, data una onda di base la codifica dell'informazione è nella variazione dell'ampiezza o della frequenza. Questa modalità di trasmissione, detta *analogica*, è certamente semplice ed efficace, ma ha almeno due problemi principali:

1. l'utilizzo di banda: la banda utilizzata è costante, indipendentemente dal fatto che vi siano o meno informazioni che passano sul canale in quel momento; per una comunicazione telefonica la banda (in Europa) di solito è di 64Kbps;
2. la correzione di errori: è difficile se non impossibile correggere possibili errori di trasmissione dei segnali, infatti il ricevente non ha maniera di distinguere in modo automatico un segnale corretto da uno distorto.

Le trasmissioni *digitali* codificano invece le informazioni da trasmettere in bit, ovvero due segnali differenti facilmente distinguibili: questi possono anche essere due frequenze diverse come 1200 e 2200Hz (in questo caso note sonore). Pertanto chi riceve i dati può correggere errori di trasmissione che risultano in una modifica non eccessiva della frequenza. Questo è il principio di funzionamento dei *modem*, parola che deriva da *mod*ulatore-*dem*odulatore, apparecchi che modulano una portante analogica per trasmettere delle informazioni digitali.

Veniamo infine a VoIP (e protocolli simili). Un segnale acustico e perciò analogico, la voce od un suono, viene prima digitalizzato e pacchettizzato, poi trasmesso tramite un modem su di una portante analogica. Si ha quindi un doppio passaggio, da analogico iniziale a digitale che poi è modulato sulla portante analogica:




Certamente questa non è la tecnica più semplice, visto che il segnale deve essere trasformato due volte, ma allevia notevolmente i due problemi principali della trasmissione analogica: la pacchettizzazione quello dell'utilizzo di banda, la digitalizzazione quello della correzione degli errori.

Vediamo un poco più in dettaglio come la voce o un qualunque segnale acustico viene trasformato in una trasmissione VoIP:

1. Prima di tutto il segnale analogico in arrivo viene **digitalizzato**. Questo è fatto di solito utilizzando una **campionatura**, ovvero dividendo la continuità del segnale analogico in (piccole) bande ed assegnando un numero ad ognuna di esse. Ad esempio in telefonia si assume di solito che la voce umana copra uno spettro dai 300 ai 3400Hz, questo spettro viene diviso in bande ed ad ognuna di essa si associa un numero. Ovviamente la campionatura viene fatta anche sul tempo. La digitalizzazione usuale dei canali telefonici è a 64Kbps utilizzando il CODEC (COmpressor/DECompressor) lineare G.711 PCM (Pulse Code Modulation).
2. Oltre alla digitalizzazione, si può **comprimere** il segnale ed anche scartare parti del segnale alle quali l'orecchio umano non è molto sensibile, in questo caso però l'onda acustica finale è diversa fisicamente da quella iniziale, anche se l'orecchio umano non dovrebbe accorgersene. Inoltre si possono sopprimere i silenzi ed altri segnali non interessanti. I CODEC più utilizzati sono G.729, G.723, G.722, G.728 le cui caratteristiche principali sono riportate in Tabella 1.
3. Infine l'informazione viene **pacchettizzata**: in pratica di solito si raccoglie il segnale digitale che corrisponde a 20ms o 30ms di segnale acustico e lo si trasmette come un unico pacchetto digitale. La trasmissione del segnale non è più continua, e questo permette un miglior utilizzo della banda. Ovviamente i pacchetti devono essere frequenti, altrimenti si introdurrebbero dei ritardi inaccettabili, e quindi sono piccoli. La dimensione del payload del pacchetto di fonia può variare in funzione del CODEC utilizzato e della frequenza di trasmissione, ma è comunque compreso fra 20 Byte (default per G.729) e 60 Byte.

A questo punto abbiamo un pacchetto digitale contenente 20ms di segnale acustico. Questo pacchetto può essere inviato a destinazione utilizzando uno tra i vari protocolli di networking packet-switching. Ultimamente molte implementazioni adottano IP fondamentalmente per la sua duttilità, la sua diffusione, la facilità e basso costo odierno di implementazione. IP d'altra parte ha dei seri problemi di garanzia del servizio, che illustreremo più avanti.

A modo di esempio, consideriamo una telefonata che utilizza IP: il funzionamento di base dei due

telefoni IP è molto semplice. Essi effettuano, in un senso e nell'altro, le operazioni appena descritte, traducono poi il numero telefonico digitato dall'utente nel numero IP del telefono corrispondente, formano il pacchetto IP e lo inviano all'altro telefono. In pratica però le cose non sono così semplici come possono sembrare.

## 2. Traffico, problemi Real Time e QoS

E' ovvio che per poter effettuare una telefonata il segnale acustico deve raggiungere l'orecchio del nostro corrispondente in un tempo *ragionevole*. Questo tempo è stato stimato in 150ms altrimenti l'orecchio umano può rendersi conto dei ritardi. Vi sono vari tipi di ritardi da prendere in considerazione:

1. il tempo di transito sulla linea di trasmissione che dipende solamente dalla distanza da percorrere ad una velocità di solito dell'ordine dei 200.000Km/sec (questo è il motivo dei ritardi nelle conversazioni via satellite);
2. la velocità di trasmissione di una linea, o meglio dell'interfaccia a livello di rete, e la lunghezza di un pacchetto: poiché un pacchetto è trasmesso sequenzialmente bit per bit, la scheda di rete impiega del tempo ad inviare tutto il pacchetto, ed il ricevente di solito attende che sia arrivato tutto il pacchetto prima di esaminarlo; per distanze brevi ed interfacce inferiori al Mbps questi ritardi sono superiori ai precedenti;
3. il tempo che il mittente impiega a codificare il segnale (di solito la decodifica è molto più rapida della codifica) tenendo conto anche della bufferizzazione, ovvero del fatto che in un pacchetto vi sono i dati che riguardano un periodo di 20ms o 30ms.

Se sul primo fattore vi è di solito poco da fare, per il secondo e terzo bisogna trovare un giusto equilibrio. Infatti per ottimizzare il secondo punto bisogna avere pacchetti piccoli e frequenti con interfacce ad alta velocità. Ma per avere pacchetti piccoli bisogna comprimere molto, il che richiede più tempo al terzo punto. Inoltre se i pacchetti sono frequenti e piccoli l'overhead dovuto agli header dei vari protocolli di rete (IP, UDP, RTP ecc.) risulta molto più grande del payload stesso. Infatti di norma l'header di un pacchetto IP+UDP+RTP è di 40 (20+8+12) Byte[1] il doppio dei dati di un pacchetto G.729 che contiene 20ms di segnale acustico! Se si trasferisse invece un pacchetto grande con dati poco compressi corrispondenti ad un campione più lungo nel tempo, si ottimizzerebbe la codifica del segnale e l'overhead dei pacchetti a scapito del tempo di bufferizzazione e del tempo di trasmissione delle interfacce. In questo caso i 150ms di limite sono facilmente superati. Pertanto di solito la soluzione scelta è di avere piccoli e frequenti pacchetti a

1 Per le connessioni IP point-to-point si può adottare C-RTP che comprime in media gli header IP+UDP+RTP sino a 2-4 Byte.

scapito dell'overhead degli header e del tempo di compressione. Si noti che la Compressed Rate in Kbps indicata in Tabella 1 non include gli header IP, per cui tenendo conto anche di questi si ottiene ad esempio che G.729 utilizza 24Kbps e non solo 8Kbps.

Un altro problema nel trasporto della voce in pacchetti è il *jitter*, ovvero la differenza del tempo di tragitto tra due pacchetti. Idealmente due pacchetti di dimensione praticamente identica dovrebbero impiegare lo stesso tempo a percorre il tragitto, ovvero il *jitter* dovrebbe essere zero. Poiché i pacchetti di tipo voce sono inviati ad intervalli regolari, ad esempio ogni 20ms, dovrebbero arrivare ad intervalli regolari. Questo purtroppo non è vero per il traffico IP che non da garanzie sul tempo e neanche sul tragitto del percorso. Pertanto è importante che la rete sulla quale si fa transitare traffico voce sia configurata in modo tale da ridurre al minimo il *jitter*, altrimenti le comunicazioni risulterebbero spezzettate, con distorsioni e perdita di suoni. Va notato comunque che il problema non è particolarmente importante, in condizioni normali, su linee veloci, di almeno 2Mbps.

Il problema del *jitter* ci porta a considerare la *Quality of Service* (QoS). Come è ben noto, IP è un protocollo best-effort, ovvero non da alcuna garanzia sulla consegna di un pacchetto, nè sul tragitto che questo effettuerà o sui tempi che impiegherà. Sono i protocolli a livello superiore che si devono occupare di garantire queste caratteristiche se richieste. Da questo punto di vista, TCP offre la garanzia della consegna del pacchetto, ma nessuna garanzia sui tempi. Inoltre TCP aggiunge parecchio overhead, e quindi ritardi, cosa non ammissibile per il traffico voce. La soluzione di VoIP è di trasportare i pacchetti voce utilizzando UDP e RTP. Il pacchetto voce codificato con uno dei CODEC descritti precedentemente, viene incapsulato in RTP, a sua volta incapsulato in UDP e trasportato da IP. RTP è il Real-Time Transport Protocol definito negli RFC 3550 e 3551. RTP aggiunge 12 Byte di header nei quali sono specificati, tra altri dati, il tipo di encoding del payload (il segnale sonoro), un timestamp del momento di invio del pacchetto ed il numero di sequenza del pacchetto. Questi dati servono ovviamente al ricevente a riordinare pacchetti che arrivassero fuori ordine, a scartare pacchetti arrivati troppo tardi ed ovviamente a decodificare con il corretto algoritmo i segnali. RTP ha un protocollo associato chiamato RTCP (RTP Control Protocol) che svolge delle funzioni di controllo della comunicazione e può essere utilizzato per monitorare la qualità del servizio ed i partecipanti alle comunicazioni. Questo protocollo non ha richieste di real-time.

Bisogna sottolineare che RTP non introduce nessun metodo per controllare ed imporre la QoS nel network attraversato dai pacchetti VoIP. Pertanto ogni qual volta si realizzano reti VoIP è

necessario implementare opportune garanzie di QoS per garantire che i pacchetti VoIP arrivino al destinatario nei tempi stabiliti. In particolare bisogna tenere conto dei seguenti fatti:

1. i pacchetti VoIP devono avere la precedenza su qualunque altro tipo di traffico non real-time
2. vi deve essere sempre una quantità sufficiente di banda a loro riservata
3. i pacchetti VoIP non devono essere mischiati a traffico con grandi pacchetti, soprattutto su linee con velocità inferiori ai 2Mbps.

Supponiamo di avere una linea a 64Kbps ed un transito ftp in corso. Di norma i pacchetti ftp sono di 1500Byte, quindi un pacchetto ci mette 187ms per essere trasmesso dal router. Supponiamo inoltre che ad un router arrivi un pacchetto ftp da una interfaccia, e che il router incominci ad inviarlo dall'interfaccia di uscita. Subito dopo arriva al router un pacchetto VoIP, e questo pacchetto deve essere inviato in uscita sulla stessa interfaccia di uscita del pacchetto ftp. Ora, anche se garantiamo i punti 1 e 2 precedenti, il pacchetto voce dovrà aspettare sino a 187ms prima di essere inviato, arrivando sicuramente troppo tardi a destinazione. Con una linea a 2Mbps il pacchetto voce aspetta invece al più solo 6ms, e quindi potrebbe ancora arrivare in tempo.

E' possibile comunque ottenere delle prestazioni più che soddisfacenti utilizzando VoIP. Se ad esempio si considera l'affitto di una connessione WAN punto-punto (ad esempio Frame-Relay o ATM) a 2Mbps con banda garantita del 75% e QoS ben configurato, è possibile avere, in condizioni normali, contemporaneamente traffico dati a 1Mbps e 20 (ma teoricamente sino a circa 40) conversazioni telefoniche di ragionevole qualità con CODEC G.729. Si noti come con la fonia tradizionale per 20 conversazioni contemporanee sarebbe necessaria una linea ad almeno 1.3Mbps riservata al solo traffico voce.

## 3. H323 ed i protocolli VoIP

Consideriamo un esempio pratico molto diffuso. Abbiamo due sedi di un'azienda con una linea dedicata che le connette. All'estremità della linea vi sono due router sui quali transita sia traffico voce che dati. Ciascun router ha delle schede VoIP[2] che effettuano la codifica/decodifica del segnale acustico. Queste schede sono connesse con il centralino telefonico tradizionale di ciascuna sede (vedi Figura 1). Al centralino telefonico le schede VoIP nel router appaiono come dei telefoni. Pertanto quando viene fatta una telefonata da una sede all'altra, il centralino nella prima sede invia la chiamata a quello che crede essere un telefono ma è invece una scheda VoIP. La scheda codifica il segnale e prepara un pacchetto IP da inviare al router corrispondente. E' importante notare che nella configurazione del router è specificato a chi devono essere inviati i pacchetti destinati ad un

2 Ad esempio per router Cisco le schede più utilizzate sono le FXO/FXS.

certo numero telefonico, o in arrivo su di una certa porta voce. Il primo router invia il pacchetto al secondo router che lo passa alla scheda VoIP. La seconda scheda VoIP decodifica il segnale acustico e lo passa al secondo centralino, che a sua volta lo invia al telefono a cui la chiamata è destinata.

Sin qui è tutto semplice, ma cosa succede se introduciamo anche telefoni IP in LAN oltre alle connessioni WAN? Come minimo c'è bisogno di un centralino VoIP che faccia la traduzione tra numeri di telefono ed indirizzi IP, e poi magari offra anche gli altri servizi ormai abituali di un centralino telefonico quali: la segreteria, music-on-hold, ridirezione delle telefonate ecc. E' ovvio che la situazione si complica di parecchio e per questo sono stati introdotti molti protocolli per gestire le telefonate. Negli ultimi anni sono stati proposti 4 principali protocolli per il controllo della segnalazione (signaling) e delle chiamate (call-control):

1. Session Initiation Protocol (SIP)
2. Media Gateway Control Protocol (MGCP)
3. H.248/Media Gateway Control (MEGACO)
4. H.323

La principale differenza fra questi protocolli è che MGCP e H.248/MEGACO sono protocolli centralizzati, e quindi molto simili al centralino telefonico tradizionale (PBX), mentre SIP e H.323 sono decentralizzati, e quindi più simili ad una rete dati ove vi sono molti agenti intelligenti (si veda la Tabella 2 per un paragone fra i protocolli). E' possibile comunque utilizzare H.323 in modo centralizzato adottando un schema chiamato *gatekeeper routed call signaling* invece dell'usuale *direct signaling*, si vedano le Figure 2 e 3 per la differenza fra i due modelli.

Il protocollo che al momento riscuote più interesse e maggiore diffusione è H.323. Probabilmente il motivo è che H.323 è riuscito a trovare un bilanciamento ragionevole tra l'approccio più tradizionale e conservativo, ma quindi anche riduttivo, di MGCP, che viene pertanto spesso adottato in implementazioni non troppo avanzate, e quello invece molto avanzato di SIP, più vicino alla logica Internet e poco a quella telefonica. Pertanto nel proseguio di questo articolo, a mo' di esempio, ci occuperemo più dettagliatamente di H.323.

H.323 fu originariamente creato per trasportare applicazioni di multimedia, in particolare videoconferenze, su reti locali (LAN). Dalla versione 2 ha incorporato anche le specifiche necessarie al trasporto della voce sia in reti locali che geografiche.
H.323 è ora utilizzato da carriers nazionali ed internazionali per trasportare le chiamate telefoniche

su brevi e lunghe distanze. H.323 è in pratica un *protocollo ombrello*, che specifica come integrare molti altri protocolli per gestire tutti gli aspetti della trasmissione di voce, ma anche fax, video e dati associati, dallo stabilire la connessione iniziale (*call establishment*), alla selezione delle *capabilities* e delle risorse di rete necessarie per la comunicazione. In particolare H.323 utilizza H.225-RAS (Registration, Admission and Status) per il routing delle chiamate, H.225-Signalling per il setup delle chiamate (*call signalling*)[3] e H.245 per lo scambio delle *capabilities (call control)*. Nella Figura 4. è presentato uno schema d'insieme dei protocolli di H.323 in relazione alla pila ISO/OSI. Nella Figura 5. è schematicamente indicata la sequenza di una chiamata VoIP H.323. Cerchiamo ora di capire perché fare una semplice telefonata risulta in un processo così complicato.

Perché un telefono IP possa inoltrare una chiamata è necessario che avvengano alcune cose:

1. come prima cosa è necessario tradurre il numero di telefono in formato usuale (E.164) nell'indirizzo IP del telefono IP o device H.323 destinatario; questo è gestito da H.225-RAS
2. è necessario poi gestire la segnalazione tra i due device terminali della chiamata, ad esempio i segnali di libero, occupato, cornetta alzata ecc.; per compatibilità con la telefonia tradizionale, H.225-Signalling che definisce queste segnalazioni è preso da Q.931, ovvero il ben noto ISDN
3. infine i due device comunicanti possono scambiarsi un elenco di capabilities, servizi aggiuntivi ecc., e controllare il traffico fra di loro.

Poiché questi compiti non possono essere gestiti tutti direttamente dai device terminali delle comunicazioni, H.323 definisce le seguenti componenti:

**1.** un **Terminale** è un qualunque device in grado come minimo di essere il punto terminale di una comunicazione voce H.323; opzionalmente può anche accettare video e dati; tipici esempio sono PC con software di VoIP o videoconferenza, telefoni IP ecc.
2. un **Gateway** è un device in grado di connettere reti diverse, sia da un punto di vista della gestione che da un punto di vista del protocollo, ovvero reti VoIP con reti di telefonia tradizionale (come nell'esempio citato all'inizio di questa sezione ed in Figura 1.)
3. un **Gatekeeper** è il cuore di una rete H.323, anche se è un elemento opzionale nel senso che le sue funzioni possono essere svolte ad esempio da un Gateway; i servizi principali offerti da un Gatekeeper sono la risoluzione degli indirizzi, l'autorizzazione ed autenticazione dei terminali e dei Gateway, la gestione della banda, accounting e billing, call-routing ecc.
4. Un **MCU** (MultiPoint Control Unit) che offre supporto per conferenze a tre o più terminali o Gateway.

E' da notare che le funzioni di Gateway, Gatekeeper a MCU possono opzionalmente essere

3H.225-Signalling è derivato da Q.931, pertanto in varie figure è indicato come Q.931.

implementate sullo stesso hardware.

In pratica, per effettuare una telefonata un telefono IP (vedi Figura 5):

1. cerca il Gatekeeper locale con un protocollo di discovery, si registra, viene riconosciuto e ottiene le proprie credenziali, questo usando H.225-RAS
2. ottiene dal Gatekeeper la traduzione del numero telefonico nel numero IP del destinatario, o del Gateway in grado di connetterlo al destinatario
3. invia al destinatario, od al Gateway, un messaggio H.225-Signalling per l'apertura di una connessione (questi messaggi possono essere inviati direttamente al destinatario come in Figura 2, oppure via il Gatekeeper, in questo secondo caso nella modalità *Gatekeeper routed signalling* come in Figura 3);
4. i due terminali della telefonata (od il telefono IP ed il Gateway) a questo punto si scambiano messaggi H.245 con i quali si informano delle proprie capabilities e decidono le porte UDP da utilizzare per RTP e RTCP
5. a questo punto i pacchetti voce, codificati con il CODEC scelto, possono essere inviati da un terminale all'altro.

E' ovvio che se il corrispondente sta su di un'altra rete, il Gateway fa tutte le traduzioni necessarie in modo che la conversazione ed i segnali possano arrivare al destinatario nei protocolli che il destinatario supporta.

In Figura 4. sono citati alcuni altri protocolli che non abbiamo ancora menzionato. H.450.(x) specificano dei servizi supplementari dei telefoni, quali call-transfer, call-hold,. Call-park&pickup ecc. mentre H.235 specifica metodi sicuri per l'autenticazione. H.261 e H.263 specificano i CODEC per il video nel caso di videoconferenze, X.224 e T.12(x) specificano la trasmissione di dati in videoconferenze, quali ad esempio lo scambio di file o le whiteboard interattive.

## 4. Aspetti generali di sicurezza

Dopo avere dato una veloce introduzione ai sistemi VoIP, riconsideriamo questa infrastruttura e protocolli dal punto di vista delle loro problematiche molto generali di sicurezza.

### 4.1 Sicurezza di base

Diversity

Se sono ovvi i vantaggi da un punto di vista economico e di gestione della convergenza dei servizi voce e dati, è anche ovvio che così facendo si contraddice uno dei primi principi della sicurezza,

ovvero avere sistemi il più possibile diversi e ridondati. Quando voce e dati transitano su infrastrutture indipendenti e con tecnologie diverse, un problema o guasto ad uno di essi non implica alcun deterioramento di servizio dell'altro. Unificando l'infrastruttura ed i protocolli di voce e dati, un guasto od un attacco andato a buon fine su un qualsiasi dispositivo di rete, oltre che al traffico dati, può arrivare a bloccare anche quello voce: una accoppiata non ancora nota a tutti coloro i quali sono abituati al fatto che le due cose sono indipendenti.

Alimentazione elettrica

Va tenuto presente come nella telefonia tradizionale non sia necessaria l'alimentazione indipendente dei telefoni e quindi anche in caso di mancanza di corrente elettrica, se il centralino telefonico è dotato di un Uninterrruptable Power Supply (UPS), la rete voce continua a funzionare, o per lo meno è in grado di garantire le telefonate di emergenza. Invece in una rete VoIP tutti i device, e quindi anche i telefoni, per funzionare hanno bisogno di una alimentazione indipendente[4] e non sono in grado di garantire neppure i servizi di emergenza in assenza di corrente elettrica. Questo richiede la progettazione della rete con una copertura quasi totale dell'alimentazione da parte di UPS, che comunque di norma garantiscono una autonomia limitata in caso di blackout.

Eavesdropping e VLAN

Nelle implementazioni attuali di VoIP il traffico voce è di solito in chiaro (discuteremo più avanti dei problemi legati alla cifratura del traffico voce), pertanto è molto facile ascoltare le telefonate IP utilizzando comuni tecniche di eavesdropping, e questo è sicuramente un problema, a maggior ragione quando le conversazioni sono su argomenti *sensibili.* Per prevenire ciò, ed anche ottimizzare la Qualità del Servizio (QoS), si tende a creare nelle LAN delle Virtual-LAN (VLAN) riservate al traffico telefonico. D'altronde anche le VLAN, se non opportunamente configurate, possono avere dei problemi come il *VLAN hopping*, ovvero la possibilità di far passare traffico da una VLAN ad un'altra, e vanno pertanto prese tutte le necessarie precauzioni per prevenirlo.

DoS nelle reti VoIP: Signaling, Media e Gateway/Trunk

Le reti IP sono vulnerabili ad attacchi di tipo Denial-of-Service (DoS), e questi per la telefonia IP si tramutano in difficoltà, abbassamento della qualità o addirittura impossibilità di fare e ricevere telefonate. In particolare, il DoS può interessare i protocolli di *segnalazione* causando ritardi nelle chiamate o impossibilità a stabilire la chiamata; può essere un *media DoS* vale a dire un Denial-of-Service che coinvolge tutta la rete, rendendo impossibile oltre che lo stabilire la chiamata anche il

4 L'alimentazione dei telefoni IP può anche essere fornita dagli switch ai quali sono connessi tramite la rete ethernet.

solo uso del telefono e degli altri device connessi alla rete. Nella rete telefonica tradizionale questo si traduce nel *telefono isolato*. Un DoS può interessare anche il gateway di collegamento con la rete telefonica, che può essere saturato e rendere impossibile la comunicazione con l'esterno malgrado la connettività interna non ne risenta. Va infine sempre considerato il caso più semplice di DoS, ovvero quando la rete non ha implementata una configurazione sufficiente di QoS per il traffico voce. In questo caso, anche solo il contemporaneo accesso ad un sito Web con lo scarico di un filmato in flash od il trasferimento di grandi documenti con ftp, può portare ad un effettivo DoS sul traffico VoIP con la conseguente quasi impossibilità di effettuare qualunque telefonata.

Raddoppio dei target IP

Una semplice considerazione statistica: ogni telefono IP su una rete è ovviamente un indirizzo IP utilizzato in più. Da un punto di vista della gestione della rete e della sicurezza questo significa avere una quantità ben più alta, indicativamente il doppio, di device e potenziali bersagli.

Open network authentication (ovvero no authentication)

Nelle configurazioni usuali i telefoni IP non sono autenticati, o l'autenticazione è basata sull'indirizzo MAC Ethernet, pertanto è facile *impersonare* un telefono IP con tutte le conseguenze che questo può portare, dirottamento della chiamata, intercettazione, ecc... .

Interazioni Voce/dati non autorizzate: Covert Channel e Furto di QoS

La separazione del traffico voce da quello dati in rete locale, che, come già detto, avviene solitamente tramite VLAN, è un fattore critico della telefonia IP. Al traffico voce è tipicamente data e garantita una migliore qualità del servizio (QoS) proprio per la criticità del traffico che porta. Essendo tutto logicamente ma non fisicamente separato, è bene prendere in considerazione la possibilità di punti di comunicazione non monitorati fra i due ambienti, genericamente detti *covert channels*, che possono dare la possibilità di tunnelizzare dati attraverso il canale voce e viceversa, aggirando così le politiche di sicurezza studiate per l'uno e per l'altro. Si può arrivare anche ad un innovativo furto di QoS nel quale si sfrutta il canale voce per il trasferimento dei propri dati, usando quindi per questi un canale con un servizio molto migliore e garantito.

Internet Joyriding con VoIP diventa Frode Telefonica

Il Joyriding, ovvero l'utilizzo di internet non autorizzato al semplice fine di avere connettività, sommato a VoIP si tramuta nella vecchia frode telefonica, l'utilizzo di una linea telefonica altrui senza pagare. Mentre finora la frode telefonica è stata relegata a diramazioni abusive o altri trucchi che richiedevano la modifica fisica di apparecchi spesso sigillati, con VoIP la frode telefonica si

prospetta anche a tutti coloro i quali hanno semplicemente accesso ad internet. L'accoppiata VoIP, internet e Wireless-LAN è poi potenzialmente esplosiva, basti pensare alla possibilità di una delle mille cabine per le chiamate telefoniche internazionali connessa via WLAN alla propria rete VoIP!

### 4.2 Sicurezza a livello applicativo

I possibili problemi per l'infrastruttura VoIP non finiscono qui, anzi a livello applicativo sono ancora più preoccupanti. Dato il grande numero di protocolli e le loro diverse funzioni e caratteristiche, non possiamo qui entrare in una analisi delle problematiche di sicurezza di ciascuno di essi. Esporremo quindi solo alcune problematiche generali comuni a tutto il livello applicativo dell'infrastruttura VoIP.

VoIP, applicativi e telefoni IP

Una delle leggi fondamentali della sicurezza informatica dice che più è complesso un sistema, più grande il codice e gli applicativi, più grande è la possibilità che vi siano errori nei protocolli, nelle implementazioni dei protocolli, nella interazione tra i protocolli e nella realizzazione pratica dei programmi. Da tutto quello che abbiamo visto sinora, tenendo inoltre conto del fatto che VoIP è una tecnologia ancora nuova, la possibilità che nel prossimo futuro sia soggetta a problemi di sicurezza a livello applicativo è molto probabile. Quindi *è necessario pensare ad un telefono IP in modo molto simile ad un usuale PC ed aspettarsi che possa essere soggetto ad attacchi, a virus, worm ecc., e di dover intervenire velocemente con patch di sicurezza*. Infatti, a differenza di un telefono usuale, un telefono IP è in pratica un'appliance con il proprio Sistema Operativo e i propri applicativi, e come tale deve essere considerato. Questo richiede anche che l'infrastruttura di rete sia progettata tenendo conto di questi fattori, e che l'infrastruttura amministrativa e di gestione sia in grado di intervenire velocemente all'occorrenza.

VoIP Server

I *call-manager* e *gate-keeper*, vale a dire i server che gestiscono un impianto di telefonia IP, in genere non sono altro che applicazioni che girano su Sistemi Operativi general-purpose spesso su hardware Personal Computer, dai quali ereditano tutte le potenziali vulnerabilità.

Unified-messaging e attacchi VMAIL

Lo *unified messaging* è il punto di contatto fra mondo dati e voce, fra mondo Email e voice-mail. La possibilità di accedere ai file vocali via PC o tramite un programma di posta elettronica e viceversa, con l'utilizzo di text-to-speech per inviare voice-mail scrivendole come Email, unisce i

due mondi, ed ovviamente si apre a ricevere in eredità tutta una serie di attacchi a livello applicativo finora relegati esclusivamente all'uno oppure all'altro mondo.

Softphone (applicazioni VoIP per PC)

Come abbiamo già detto, il telefono IP ha sostanzialmente la forma di un telefono, ma è in fondo solo una applicazione di rete. Questa applicazione può anche essere implementata su di un qualunque Personal Computer che diventa al tempo stesso anche un telefono IP, detto *softphone* . Questo può essere molto utile ad esempio a chi viaggia molto, e può in ogni momento connettersi alla rete VoIP interna dell'azienda con il proprio portatile ovunque sia, in pratica portandosi dietro il proprio telefono. Ma il *softphone* è una normale applicazione e finisce per ereditare tutte le vulnerabilità tipiche dei PC: vulnerabilità del Sistema Operativo, virus, worms ecc., perdita/furto del portatile e così via. Inoltre la stessa presa di rete è utilizzata sia per il traffico dati che per quello voce, e come abbiamo visto la separazione dei due è critica nelle infrastrutture VoIP. Un PC-softphone diventa perciò un punto molto critico nell'infrastruttura VoIP in quanto è in una posizione ottima per effettuare vari attacchi descritti precedentemente, dall'eavesdropping al furto di QoS ecc.

### 4.3 VoIP, firewall, IDS e crittografia

Per proteggere le infrastrutture di VoIP si può pensare di utilizzare le usuali tecniche di sicurezza informatica adottate per le reti IP, dai firewall alla crittografia. Purtroppo la situazione non è così semplice, poiché VoIP, per le sue specifiche caratteristiche, rende molto difficile implementare queste tecniche.

VoIP versus firewall

Come è gia stato scritto, la telefonia IP è costituita da un grande insieme di protocolli. L'utilizzo di firewall per filtrare e proteggere il traffico VoIP è sicuramente una buona idea, ma si scontra con almeno tre problemi. Il primo è l'ulteriore ritardo aggiunto dai firewall al tempo di percorrenza dei pacchetti VoIP, che però per buoni firewall è praticamente trascurabile. Il secondo è il grande numero di protocolli necessari ad una telefonata IP ed il fatto che molti di essi richiedono l'apertura di porte dinamiche, ovvero non fisse ma concordate di volta in volta tra i device destinatari del traffico. Questo richiede che i firewall siano in grado di ispezionare almeno superficialmente i dati scambiati e capire quali porte debbano essere aperte e chiuse dinamicamente. Ovviamente questa analisi anche se superficiale a livello applicativo dei contenuti dei pacchetti richiede ulteriore tempo, e in alcuni casi ulteriori ritardi possono venire aggiunti. Infine, un firewall

può offrire una protezione elevata solo se è in grado di analizzare completamente il contenuto del traffico, cioè se si comporta come un Application Layer Gateway (o Proxy). A parte la complessità dell'analisi ed il numero di protocolli da considerare, ovviamente questo introdurrebbe ulteriori grandi ritardi e si presenta pertanto di difficile implementazione.

Realtime/QoS versus Sicurezza e Crittografia

Verrebbe naturale pensare che se la voce su IP è intercettabile ed i device sono difficilmente autenticabili, la cifratura sia la naturale soluzione al problema. In realtà la cifratura introduce un problema di latenza (cifrare il dato in tempo reale è un compito gravoso per i telefoni, call-manager, gateway e per i dispositivi di instradamento) che deve quindi essere considerato in fase di progetto, ma soprattutto introduce un nuovo problema nella gestione della qualità del servizio, difatti non appena i dati sono cifrati si perde la possibilità di applicare tecniche di gestione della priorità del traffico. Alcuni vendor hanno soluzioni che indirizzano questo problema, restano comunque da considerare i nuovi tempi di latenza introdotta ed anche la difficoltà di mantenimento del QoS in ambienti multivendor. Ovviamente non aiuta l'assenza di standard per la cifratura del traffico VoIP e l'instradamento del traffico cifrato. Una delle poche situazioni ove la cifratura del traffico VoIP di solito non è problematica, una volta che si tiene conto dei ritardi aggiunti, è quando si instrada traffico VoIP WAN su tunnel IPSEC dedicati.

Cifratura e firewall, IDS ecc.

Si applica anche a VoIP la dicotomia tra cifratura e firewall, IDS ecc.: se il traffico è cifrato, il firewall, il Network Intrusion Detection e le tradizionali tecnologie di analisi e filtraggio non possono fare alcunché per proteggere la rete e gli applicativi poiché non sono in grado di capire cosa viene trasportato sulla rete. Gli Host Based Intrusion Detection System sono gli unici che ben si sposano alla cifratura, ed in questo caso è fortemente consigliato il loro impiego sui gatekeeper, call-manager, softphone ed in generale su qualunque sistema di VoIP implementato su sistemi operativi general-purpose.

## 5. Conclusioni

In conclusione, VoIP offre degli indubbi benefici economici, di gestione e sfruttamento delle risorse della propria rete informatica. D'altro canto è una tecnologia giovane, complessa e purtroppo sviluppata almeno sino a poco tempo fa senza troppa attenzione alle problematiche di sicurezza. Pertanto l'adozione di questa tecnologia senza considerarne le possibili conseguenze e senza le

minime protezioni può portare a delle serie conseguenze. Comunque, con una adeguata implementazione e l'aiuto dei purtroppo ancora troppo pochi esperti del settore, l'adozione di VoIP può essere fatta in ragionevole sicurezza e con indubbi benefici.

Andrea Pasquinucci
Libero Professionista in Sicurezza Informatica
pasquinucci@ucci.it

Marco Misitano, CISSP
marco@misitano.com

Riferimenti Bibliografici:

[1] Articoli introduttivi su VoIP e H323 si possono trovare su protocols.com http://www.protocols.com/pbook/VoIP.htm, e Packetizer http://www.packetizer.com/iptel/h323/

[2] Altri articoli introduttivi sono gli online tutorials di IEC: http://www.iec.org/online/tutorials/

[3] Cisco SAFE, *IP Telephony Security In Depth*, http://cisco.com/go/SAFE

[4] H.323 è una raccomandazione ITU che definisce *packet-based multimedia communications systems*, si veda http://www.itu.int/, il database delle raccomandazioni ITU si può trovare a http://www.itu.int/ITU-T/publications/recs.html, altre informazioni sugli standard si possono trovare ad esempio in http://ftp3.itu.int/av-arch/avc-site/0309_Par/0309_Par.html o http://www.packetizer.com/iptel/h323/doc_status.html

[5] G.7** sono raccomandazioni ITU

[6] Per RTP e C-RTP si vedano gli RFC 3550, 3551, 2508 e 3544

[7] MGCP è definito nel RFC 2705

[8] H.248 è una raccomandazione ITU

[9] MEGACO è definito nel RFC 2885

[10] SIP è definito nel RFC 2543

| Schema di compressione | Velocità in (Kbps) di compressione | Risorse di CPU richieste | Qualità della voce risultante | Ritardo aggiunto |
|---|---|---|---|---|
| G.711 PCM | 64 (senza compressione) | Non richieste | Ottima | N/A |
| G.723 MP-MLQ | 6.4/5.3 | Medie | Buona (6.4) Sufficiente (5.3) | Alto |
| G.726 ADPCM | 40/32/24 | Basse | Buona (40) Sufficiente (24) | Molto basso |
| G.728 LDCELP | 16 | Molto alte | Buona | Basso |
| G.729 CSACELP | 8 | Alte | Buona | Basso |

Tabella 1. Alcuni CODEC di compressione del segnale acustico

| | H.323 | SIP | MGCP/H.248/MEGACO |
|---|---|---|---|
| **Standards Body** | ITU | IETF | MGCP ITF<br>MEGACO IETF<br>H.248 ITU |
| **Architettura** | Distribuita | Distribuita | Centralizzata |
| **Versione Corrente** | H.323v4 | RFC2543-bis07 | MGCP 1.0, Megaco, H.248 |
| **Call control** | Gatekeeper | Proxy<br>Redirect Server | Call agent<br>Media Gateway ctrl |
| **Endpoints** | Gateway Terminal<br>IP phones<br>Media servers .. | User agent | Media gateway |
| **Signaling Transport** | TCP<br>UDP | TCP<br>UDP | MGCP UDP<br>Megaco TCP/UDP<br>H.248 TCP/UDP |
| **Multimedia capable** | SI | SI | SI |
| **DTMF relay transport** | H.245(signaling) o<br>RFC 2833 (media) | RFC 2833 (media) o INFO (signaling) | Signaling o<br>RFC 2833 (media) |
| **Fax relay transport** | T.38 | T.38 | T.38 |
| **Servizi Supplementari** | Fornito dagli endpoints o call control | Fornito dagli endpoints o call control | Fornito dai call agents |

Tabella 2. Paragone fra protocolli VoIP.

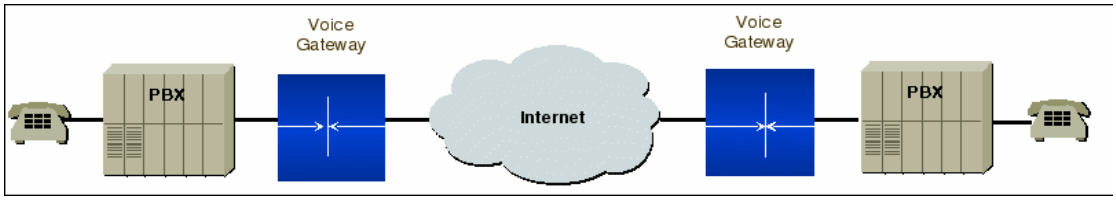



Figura 1. VoIP su di una connessione WAN

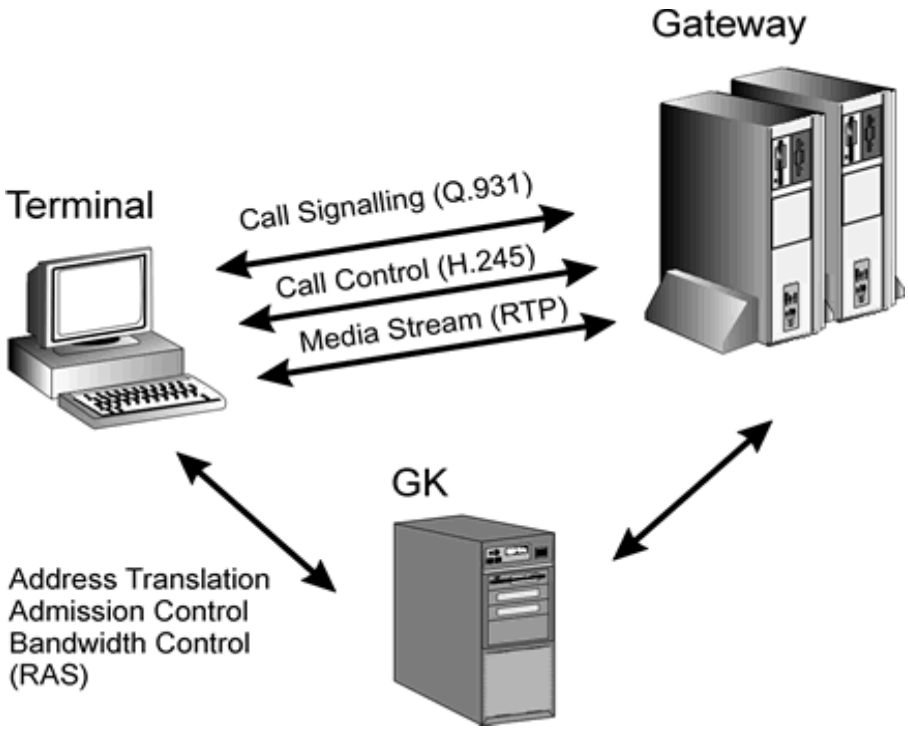



Figura 2. Modello Gatekeeper direct signalling

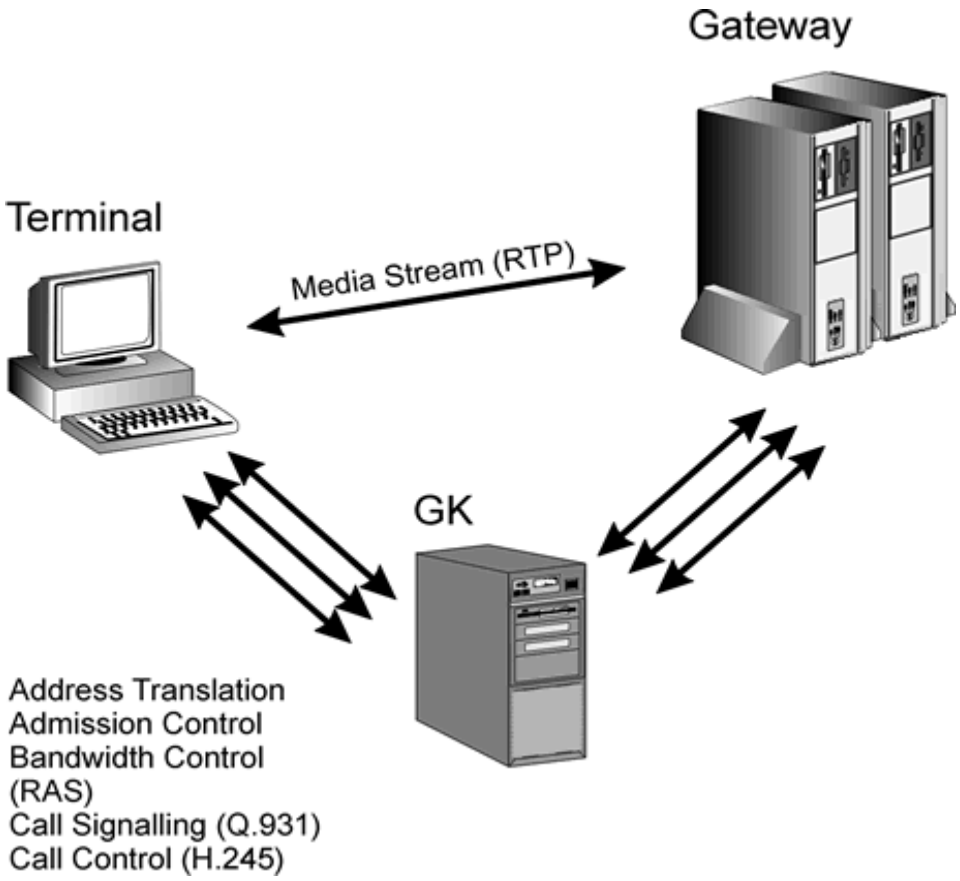



Figura 3. Modello Gatekeeper routed signalling

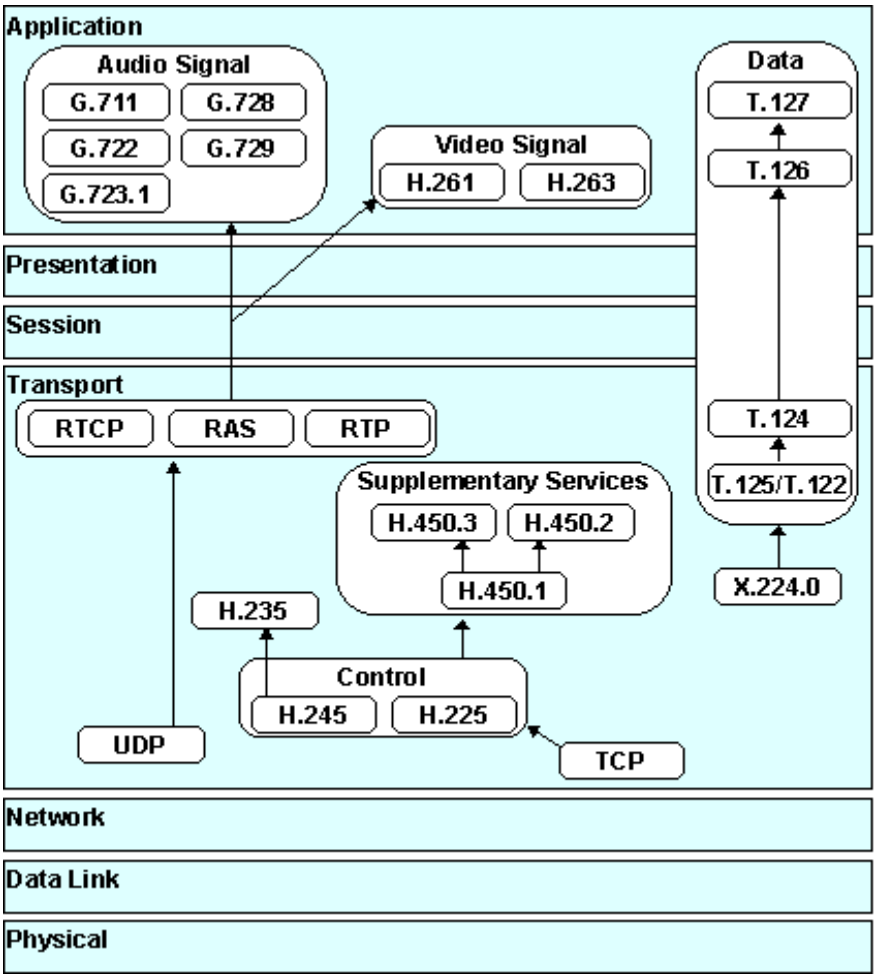



Figura 4. I protocolli H.323 in relazione alla pila ISO/OSI

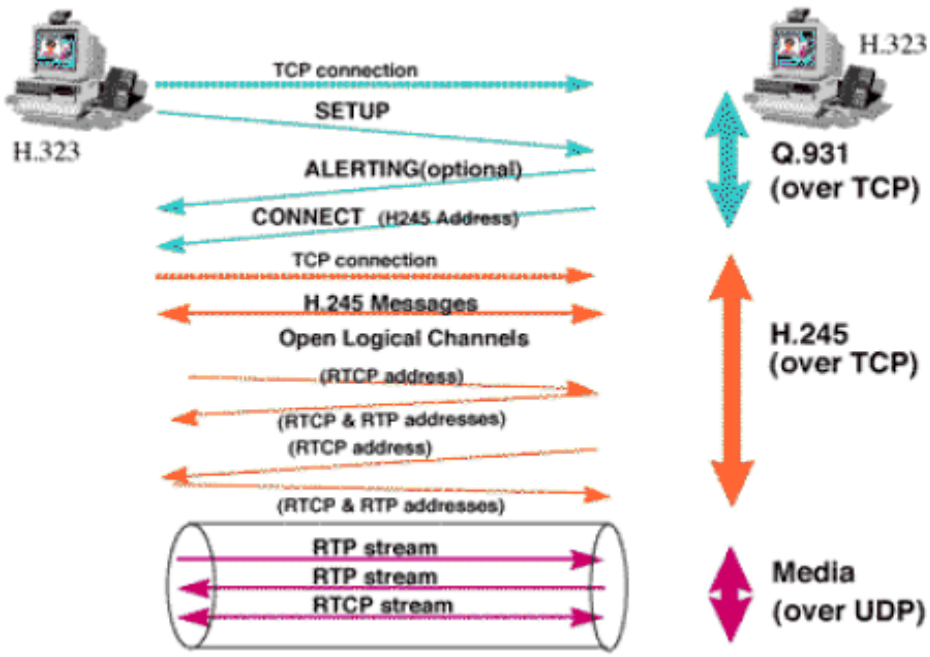



Figura 5. Sequenza di una chiamata VoIP H.323 (da [2])