UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II

Facoltà di Ingegneria
Corso di Studi in Ingegneria Informatica

Relazione per il corso di Reti di Calcolatori II

Prof. Giorgio Ventre

Codifiche Audio/Video: Skype

Giuseppe Di Luca - Matr. 885-326
Mariana Esposito - Matr. 885-390

Indice

Capitolo 1

Introduzione al VoIP

1.1 Perchè il VoIP?

Lo sviluppo delle connessioni Internet a banda larga e l'utilizzo di piattaforme tecnologiche innovative hanno comportato negli ultimi tempi uno spostamento del traffico telefonico dalle reti tradizionali a commutazione di circuito verso reti a commutazione di pacchetto. In questo scenario, la telefonia via Internet comunemente definita con il termine VoIP (*Voice over IP*) costituisce uno dei servizi a più alto potenziale. Come detto, essendo questa applicazione essenzialmente basata sulle reti di calcolatori, essa utilizza una tecnologia di trasmissione a commutazione di pacchetto: i segnali vocali vengono convertiti in pacchetti di dati digitali, spediti attraverso la rete e assemblati a destinazione.
Sebbene la più ampia diffusione dei sistemi VoIP si sia avuta nell'ambito dei calcolatori e quindi delle reti di computer, possiamo classificare i sistemi VoIP attualmente esistenti nelle seguenti categorie:

- **Telefono a telefono**: i telefoni tradizionali sono connessi a una rete (basata sul protocollo IP) attraverso router che trasformano i segnali vocali;
- **PC a PC**: è il sistema più utilizzato sulla rete Internet. Gli utenti devono essere collegati online per avviare una connessione e, cosa ancor più importante, devono avere software compatibili;
- **Telefono a PC**: i telefoni tradizionali sono connessi a una rete attraverso specifici gateway. Questo consente agli utenti con un telefono di chiamare utenti con PC connessi a Internet.

Nel vasto settore delle telecomunicazioni, per quanto riguarda la diffusione del VoIP hanno giocato un ruolo importante anche i terminali mobili, che hanno subito un'innovazione tale da portarli, nella concezione della terza generazione, a poter offrire servizi multimediali sempre più complessi (come appunto il VoIP) con sistemi a commutazione di pacchetto anziché di circuito.
Il VoIP costituisce un'innovazione che consente non soltanto di avere servizi di telefonia gratuiti o a basso prezzo, ma soprattutto permette di sfruttare le potenzialità delle reti per offrire servizi integrati. Il successo di nuovi operatori indipendenti che offrono servizi VoIP si fonda essenzialmente su due elementi:

1. la diffusione capillare e quasi pervasiva dell'accesso a Internet a banda larga (su reti fisse e mobili);

2. l’affermazione dei sistemi peer-to-peer nell’offerta di servizi di comunicazione on line.

Nel seguito di questo documento ci riferiremo sostanzialmente ai sistemi VoIP basati sulle reti di calcolatori.

1.2 L’avvento delle applicazioni VoIP

Da qualche anno a questa parte numerose aziende hanno iniziato a fornire servizi di telefonia di tipo VoIP, ponendosi nella maggior parte dei casi in maniera indipendente e concorrente rispetto agli operatori di telefonia fissa. L’affiancamento di tali operatori a quelli di telefonia tradizionale trae le sue motivazioni dal fatto che le caratteristiche tecnologiche dei sistemi VoIP consentono alle imprese (e ai consumatori più in generale) di beneficiare di una serie di fattori vantaggiosi sia da un punto di vista economico che di regolamentazione.
Di seguito si elencano alcuni tra questi:

1. **Costi più bassi**: rispetto ai servizi di telefonia tradizionale, il VoIP è caratterizzato da costi di infrastruttura più bassi e da un utilizzo della banda più efficiente, grazie alla compressione dei segnali vocali (si risparmia il 90% delle risorse per ogni singola chiamata). ciò ha importanti implicazioni sui prezzi dei servizi;
2. **Maggiore efficienza**: il VoIP consente alle imprese di integrare PC e rete telefonica e di semplificare le connessioni tra i vari utenti. Inoltre, tale tecnologia può essere facilmente combinata con applicazioni quali videoconferenza, scambio e condivisione di documenti, facilitando l’interazione fra gli utenti del servizio;
3. **Network management**: il VoIP semplifica la struttura della rete, in quanto può essere implementato facendo ricorso a sistemi più economici rispetto alle reti tradizionali, tali da consentire una riduzione dei costi operativi del 50%-60%. Le reti che integrano servizi vocali e di dati rendono più semplici anche le procedure di fatturazione e il supporto tecnico;
4. **Assenza di barriere geografiche**: diversamente dalle reti telefoniche tradizionali, il VoIP non è vincolato da barriere geografiche. Gli operatori prendono semplicemente le chiamate su una linea locale, le trasferiscono in formato digitale sulla rete e poi le riconvertono in voce all’altra estremità della rete stessa;
5. **Diffusione della banda larga**: la crescita dei servizi VoIP incoraggia la diffusione della banda larga, poiché molti di questi servizi necessitano di reti ad alta capacità di trasmissione (DSL, fibre ottiche, cavi). Alcune imprese già offrono pacchetti di servizi che includono VoIP e accesso a Internet. L’integrazione tra le diverse applicazioni consente agli operatori

Anche se non ci soffermiamo più di tanto in questa disamina "pro/contro" VoIP, occorre notare che non è tutto oro ciò che luccica. Infatti nonostante i vantaggi offerti da questi sistemi innovativi, esistono ancora alcune problematiche prevalentemente di natura tecnologica che frenano di fatto lo sviluppo completo di questi servizi. Nel seguito, per brevità, ne citiamo solo un paio:

1. **Qualità e affidabilità del servizio**: la qualità del VoIP non è sempre ottimale a causa della congestione – legata alla tecnologia di trasmissione a commutazione di pacchetto – e della velocità di trasmissione della rete. In particolare, si possono verificare problemi di interruzione e/o disturbi durante le chiamate. In secondo luogo, date le caratteristiche del VoIP, le imprese spesso non hanno il controllo della rete che viene utilizzata per i servizi voce e questo può compromettere l'affidabilità del servizio;
2. **Sicurezza:** i servizi VoIP sono, in generale, più vulnerabili dei servizi di telefonia tradizionali, in quanto non prevedono linee dedicate alle singole chiamate. Inoltre, attraverso il VoIP non è possibile effettuare chiamate d'emergenza, in quanto è difficile individuare l'esatta posizione degli utenti. Oltre a ciò, molti call center di emergenza si basano su tecnologie tradizionali a commutazione di circuito, incompatibili con quelle a commutazione di pacchetto utilizzate dai sistemi VoIP.

Di certo si può osservare come con il VoIP tramontino alcuni paradigmi dei tradizionali sistemi di telecomunicazioni. Basti osservare, infatti, che la telefonata in commutazione di circuito impegna la linea per tutta la durata della comunicazione ed entrambe le parti utilizzano il tempo della connessione per comunicare, cosa che invece non è detto avvenga sempre nei sistemi VoIP. Il modello economico con il quale si è costruito il mondo delle telecomunicazioni in oltre cent'anni di storia è stato di fatto ancorato al paradigma "*canone più tariffa a tempo*", a testimonianza di un "consumo" dovuto all'impegno della linea (commutazione di circuito) per la durata della chiamata. Con la commutazione di pacchetto, invece, il flusso di dati vocali impegna in modo diverso la linea di comunicazione consentendo così l'esplorazione di forme di tariffazione innovative per questo tipo di servizio. Ciò si traduce, come si può ben comprendere, in nuovi aspetti per quanto riguarda i rapporti di servizio tra gli operatori telefonici ed i relativi utenti.

Capitolo 2

Un particolare sistema VoIP: Skype

2.1 Panoramica generale sul software

Skype è l'applicazione (ma è anche un operatore) VoIP maggiormente diffusa ed utilizzata attualmente in Internet. Fattori che hanno contribuito alla sua grande popolarità sono senza dubbio da ricercare nella semplicità del software offerto e nell'eccellente qualità audio delle conversazioni. Esso unisce caratteristiche presenti nei client più comuni (chat, salvataggio delle conversazioni, trasferimento di file) ad un sistema di telefonate basato su un paradigma di comunicazione di tipo *peer-to-peer*. Questo significa che gli utenti della rete creano una *overlay network* che viene utilizzata per stabilire collegamenti logici tra i vari nodi, ovvero gli utenti stessi della rete. A differenza di altri servizi, Skype funziona senza problemi anche in presenza di NAT e firewall, ed ha una qualità audio sensibilmente migliore rispetto a quella offerta da altri software concorrenti (come MSN Live Messenger, Yahoo Messenger e affini). Il protocollo di comunicazione è stato progettato in maniera tale da cifrare le chiamate e memorizzare le informazioni sugli utenti in modo decentralizzato. In particolare la cifratura utilizzata per criptare le informazioni è *AES* (*Advanced Encryption Standard*) [3], con chiavi di 256 bit, utilizzando in aggiunta l'algoritmo *RSA* [8] con chiavi da 1536 a 2048 bit per negoziare le chiavi simmetriche AES.

2.2 La rete Skype

La rete Skype è costituita essenzialmente da due tipi di nodi: *host normali* e *supernodi* [1].
Un **host normale** è un'applicazione Skype che può essere utilizzata per effettuare chiamate vocali o inviare messaggi di testo, mentre un **supernodo** costituisce l'end-point, cioè il nodo con cui gli host normali comunicano direttamente. Un qualsiasi nodo con un indirizzo IP pubblico può diventare un supernodo, a meno che non abbia risorse – dal punto di vista fisico – insufficienti. Un host normale deve connettersi a un supernodo e deve registrarsi sullo *Skype Login Server* per poter accedere ai servizi della rete Skype. Nonostante non sia un nodo Skype in senso stretto, il *login server* è un'entità importante all'interno della rete Skype. Esso, infatti, viene utilizzato sia come repository per le username e le password di tutti gli utenti della rete che per l'autenticazione di ciascun utente all'atto dell'accesso alla rete stessa garantendo, inoltre, anche l'univocità dei *login name* all'interno di tutta la rete Skype. La figura 1 a pagina seguente illustra la relazione tra gli host, i supernodi e il login server. A parte quest'ultimo, non esiste alcun altro server centrale all'interno della rete Skype.

Le informazioni sugli utenti sono mantenute e propagate in modo decentralizzato, così come tutte le operazioni legate alle funzioni di ricerca degli utenti.
Skype utilizza il protocollo TCP per la sincronizzazione, e sia TCP che UDP per il livello trasporto: si fa notare che queste due attività non avvengono sulla stessa porta.

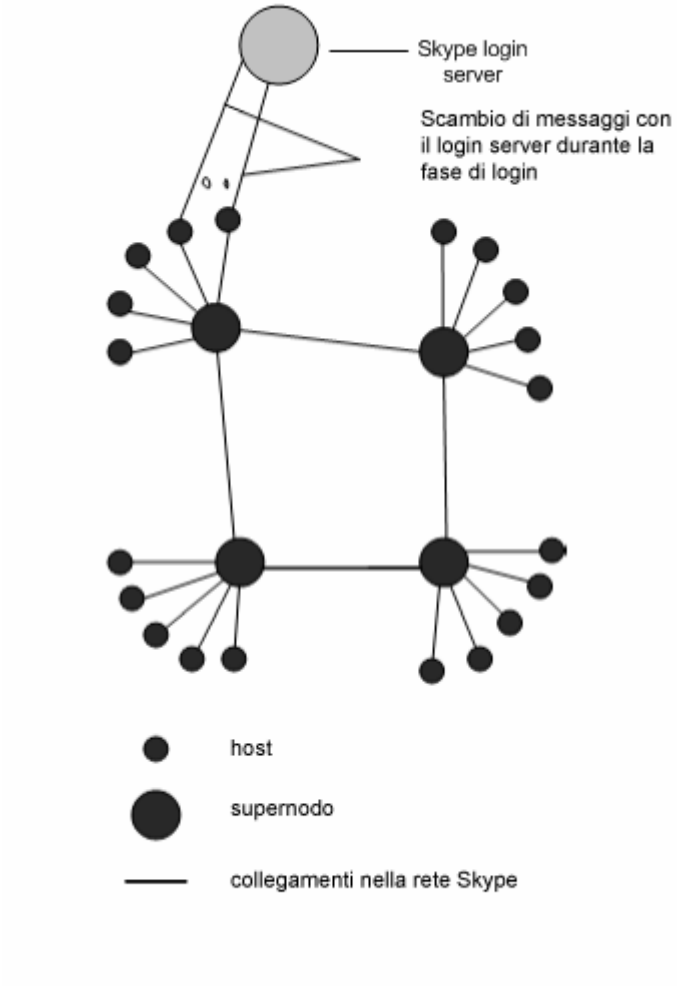

Figura 1 - Rete Skype

Dato che la rete Skype è una *overlay network*, ogni client deve costruire e aggiornare continuamente una tabella di tutti gli altri host raggiungibili chiamata *host cache*, che contiene gli indirizzi IP e i numeri di porta dei supernodi. Nelle versioni del client per Microsoft Windows, ciascun nodo Skype conserva la propria host cache all'interno del registro di sistema.

2.2.1 Le porte utilizzate

Un client Skype apre una porta TCP e una UDP in ascolto in base alla configurazione specificata: inizialmente tali porte vengono scelte in maniera casuale durante l'installazione. In aggiunta, esso apre anche porte TCP in ascolto (porta 80 - HTTP e porta 443 - HTTPS).

A differenza di molti protocolli Internet, come SIP e HTTP, non viene specificata alcuna porta TCP o UDP di default.

2.3 Componenti chiave

Come si può evincere da quanto detto nei paragrafi precedenti, il software che costituisce un client della rete Skype è costituito da diversi componenti. Nel seguito, per brevità, descriveremo nel dettaglio solo alcuni di essi, ovvero quelli che abbiamo ritenuto maggiormente interessanti ai fini della nostra trattazione: l'host cache e i codec.

2.3.1 La Host Cache

Come già detto nel paragrafo 2.2, la host cache è una tabella contenente una lista di indirizzi IP e numeri di porta dei supernodi che ogni client costruisce e aggiorna continuamente. L'operazione di aggiornamento è una fra le più critiche tra quelle svolte da Skype perché, come vedremo, il corretto funzionamento del client dipende soprattutto dalla affidabilità delle entries presenti nella host cache. Essa, infatti, deve contenere almeno una entry valida, ovvero un elemento costituito dall'indirizzo IP e dal numero di porta di un supernodo Skype online.

2.3.2 Codec

Skype utilizza i codec proprietari iLBC, iSAC e un terzo codec non indicato esplicitamente. Questi permettono la trasmissione di segnali a frequenze comprese tra 50-8.000 Hz [2], tipiche dei codec a banda larga. Anche riducendo la banda disponibile per Skype ad un minimo di 1.5 Kilobytes/s il valore minimo e quello massimo delle frequenze udibili che i codec utilizzati lasciano passare sono rispettivamente all'incirca pari a 50 Hz e 8000 Hz. Rimandando ai paragrafi che seguono una trattazione più estesa su tali codec, in questa sede facciamo notare che essi sono utilizzati in modo tale che se entrambi i client Skype coinvolti nella comunicazione hanno IP pubblico, il traffico viene inviato direttamente tra loro su UDP, nello specifico esso transita da e verso le porte UDP configurate sui client stessi. La grandezza del pacchetto contenente i dati relativi alla conversazione è 67 bytes, pari alla grandezza del *payload* UDP. Alcuni studi [1] hanno mostrato che per due utenti connessi ad una LAN a 100 Mb/s senza congestioni di alcun tipo, vengono scambiati circa 140 di questi pacchetti al secondo. L'uplink e il downlink totali consumano in tale circostanza una banda di circa 5 Kilobytes/s per il traffico vocale, corrispondente sostanzialmente a quello

ufficialmente dichiarato da Skype (3-16 Kilobytes/s). Se sia il chiamante che il chiamato stanno dietro a un NAT sulle porte, essi inviano il traffico vocale ad un altro nodo online Skype mediante protocollo UDP. Questo nodo funziona da *proxy* inoltrando il traffico dal chiamante al chiamato e viceversa. Anche in questo caso la grandezza del pacchetto risulta essere di 67 bytes, il che corrisponde ancora una volta ad un utilizzo di 5 Kilobytes/s di banda. Se entrambi gli utenti coinvolti nella comunicazione si trovano dietro a un NAT sulle porte e a un firewall che blocca il traffico UDP, allora chiamante e chiamato inviano e ricevono il traffico dati mediante il protocollo TCP da un nodo Skype online, mantenendo sempre un consumo in banda limitato a circa 5 Kilobytes/s. È stato verificato sperimentalmente [1] che il protocollo Skype sembra preferire, ove possibile, l'utilizzo del protocollo UDP per il traffico vocale: il client utilizza prevalentemente il protocollo UDP per la trasmissione del traffico se si trova dietro a un NAT o ad un firewall che non blocca il transito di pacchetti UDP.

Per quanto concerne la codifica delle informazioni vocali provenienti dagli interlocutori, occorre notare che Skype non supporta alcuna soppressione di silenzio. In particolare se i soggetti coinvolti nella conversazione non parlano per un certo periodo di tempo, vengono comunque inviati ugualmente pacchetti "vocali". Trasmettere questi pacchetti "di silenzio" ha un duplice vantaggio. Per prima cosa ciò consente di mantenere il binding UDP nel NAT e inoltre questi pacchetti possono essere usati per generare del rumore di fondo evitando che la conversazione risulti completamente muta negli intervalli di tempo in cui i due interlocutori effettivamente non producono alcun segnale sonoro. Anche nel caso in cui il traffico venga inviato mediante protocollo TCP i pacchetti "di silenzio" sono ugualmente inviati.

2.4 Funzioni principali

Le principali funzioni di Skype possono essere classificate in avvio (startup), login, ricerca utenti, istituzione e chiusura di chiamata.

2.4.1 Startup

Quando il client Skype viene avviato per la prima volta dopo l'installazione, esso invia una richiesta http 1.1 GET allo Skype server (skype.com). La prima linea di questa richiesta contiene la parola chiave "*installed*". Durante la fase di avvio susseguente, il client invia solamente una richiesta http 1.1 GET allo Skype server (skype.com) per determinare se è

disponibile una versione aggiornata. La prima linea di questa richiesta contiene la keyword "*getlatestversion*".

2.4.2 Login

La funzione di Login è forse in assoluto la più critica di Skype. E' durante questo processo che il client Skype autentica username e password con il login server (mediante protocollo TCP), avverte della propria presenza gli altri *peers* e i contatti, determina il tipo di NAT e firewall (se presenti), e scopre nuovi nodi Skype online dotati di IP pubblico. Questi nuovi nodi vengono utilizzati per mantenere le connessioni con la rete Skype nel caso i supernodi già conosciuti risultino non disponibili.
Come visto nel paragrafo 2.3.2, la host cache deve contenere almeno una entry valida per consentire al client di connettersi alla rete Skype. Se la host cache viene riempita con una sola entry non valida, il client non potrà connettersi alla rete Skype, avvisando del fallimento del processo di login. Di seguito descriviamo dettagliatamente l'algoritmo di login.
Il client invia un pacchetto UDP alla prima entry disponibile nella host cache. Se non ottiene risposta per più di 5 secondi, esso cerca di stabilire una connessione TCP con questa entry, cioè all'indirizzo e al numero di porta specificati. Se la connessione non va a buon fine il client cerca di stabilire una connessione TCP verso l'indirizzo IP della entry sulla porta 80 (HTTP). Se la connessione non va a buon fine ancora una volta, il client cerca di connettersi all'IP della entry sulla porta 443 (HTTPS), con un timeout di 6 secondi. Questo processo appena descritto viene ripetuto per altre quattro volte, al termine delle quali viene riportato il fallimento del processo di login per la mancata connessione. Da quanto appena detto si evince che il client deve stabilire una connessione TCP con un supernodo per connettersi alla rete Skype. Se non può connettersi a un supernodo, esso riporta il fallimento del login. Si fa notare che la maggior parte dei firewall sono configurati per consentire il traffico TCP verso la porta 80 e 443 in uscita. Un client dietro a un firewall che blocca il traffico UDP e permette selettivamente un certo tipo di traffico TCP si avvantaggia di questo fatto. Al login, infatti, stabilisce una connessione TCP con un altro nodo Skype dotato di IP pubblico sulla porta 80 o 443.
Dopo aver effettuato il login per la prima volta dopo l'installazione, la host cache viene inizializzata con sette coppie IP\porte relativi ad altrettanti nodi, definibili con il termine *supernodi di avvio*.
A partire dal primo login il client invia pacchetti UDP ad almeno 4 dei supernodi di avvio della lista precedente. Ciò significa che, molto probabilmente, sia gli indirizzi IP che le porte

dei supernodi di avvio sono cablati dentro il codice del client. Se la host cache viene svuotata dopo il primo login, il client non è più in grado di connettersi alla rete Skype.

Dato che il protocollo di comunicazione di Skype non è aperto, non è ancora chiaro come il client scelga esattamente i supernodi ai quali inviare i pacchetti UDP per effettuare l'accesso tra quelli i cui indirizzi sono presenti nella host cache. Dopo di ciò il client stabilisce una connessione TCP con il supernodo d'avvio che ha risposto. Dato che più di un nodo può rispondere, un client può stabilire una connessione TCP con più di un supernodo d'avvio. Tuttavia esso può comunque mantenere una connessione TCP con almeno un supernodo d'avvio chiudendo le connessioni TCP con gli altri nodi. Una volta stabilita correttamente la connessione con un supernodo d'avvio, viene stabilita una connessione TCP con il login server, che consente lo scambio delle informazioni di autenticazione, e per finire si chiude la connessione TCP con il login server stesso. La connessione TCP con il supernodo persiste fino a quando il supernodo rimane attivo. Quando il supernodo non è più raggiungibile, il client stabilisce una connessione TCP con un altro supernodo. Il totale dei dati scambiati tra client, supernodo, login server e altri nodi durante il login è di circa 9 Kilobytes.

Per quanto riguarda un client dietro a un NAT sulla porta, il flusso di messaggi per il login è praticamente lo stesso di quello del client con IP pubblico, ma vengono scambiati più dati. In media in questa circostanza il client scambia 10 Kilobytes di dati con il supernodo, il login server e con gli altri nodi Skype. Un client che sta dietro a un NAT sulla porta e a un firewall che non fa passare UDP è incapace di ricevere i pacchetti UDP da tutte le macchine poste oltre il firewall: esso deve essere in grado di inviare e ricevere traffico TCP. In tale scenario il client effettuerà una connessione TCP con un supernodo e con il login server scambiando con essi informazioni via TCP. In uno scenario del genere il client scambia in media 8.5 kilobytes di traffico con il supernodo, con il login server e con gli altri nodi Skype.

Alcune applicazioni sperimentali [1] hanno mostrato la capacità del client Skype di determinare in fase di login se questo si trova dietro a un NAT o a un firewall. Ciò è essenzialmente possibile in due modi differenti. Una possibilità è che il client provi a scambiarsi i messaggi con i supernodi utilizzando una variante del protocollo STUN[1]. L'altra possibilità è che durante il login il client invii e possibilmente riceva dati da alcuni nodi dopo aver fatto una connessione TCP con il supernodo, utilizzando poi la sua variante

[1] **STUN** è l'acronimo di *Simple Traversal of User Datagram Protocol(UDP) Through Network Address Translators (NATs)*: è un protocollo client-server che permette a particolari applicazioni di scoprire la presenza ed i tipi di NAT e firewall presenti tra loro e la rete pubblica. STUN permette a questi programmi di conoscere gli indirizzi con cui il dispositivo NAT li sta rendendo visibili alla rete pubblica.

del protocollo STUN per determinare la tipologia di NAT o di firewall dietro cui è posto. Una volta determinata, tale informazione viene custodita all'interno del registro di Windows. In particolare un client dietro firewall invia pacchetti UDP alle sue prime 30 entries della host cache. Non ricevendo alcuna risposta (poiché il firewall scarta in automatico i pacchetti UDP) capisce a questo punto di trovarsi dietro a un firewall bloccante e prova di conseguenza a stabilire direttamente una connessione TCP con le entries presenti nella host cache.

Skype è un client peer-to-peer e le reti peer-to-peer sono, in generale, molto dinamiche. Il client, quindi, deve tenere traccia dei nodi online all'interno della rete Skype, in modo da potersi connettere ad essi se i supernodi, per un qualsiasi motivo, diventano irraggiungibili. A tale scopo, il client invia pacchetti UDP a 22 nodi diversi alla fine del processo di login, e possibilmente riceve risposte da essi a meno che non ci sia di mezzo un firewall che non consente traffico UDP. È possibile ipotizzare che il client utilizzi questi messaggi per avvisare del proprio arrivo sulla rete, sfruttando le risposte a questi stessi messaggi per costruire una tabella dei nodi online: la *tabella dei nodi alternati*. È con questi nodi che il client può connettersi se il suo supernodo diventa irraggiungibile.

I processi di login successivi sono sostanzialmente simili al primo login. La host cache viene aggiornata periodicamente con l'indirizzo IP e i numeri di porta dei nuovi peers. Durante i login successivi al primo, il client utilizza l'algoritmo di login per determinare almeno un peer raggiungibile tra i nodi presenti nella host cache, stabilendo quindi una connessione TCP con questo nodo.

2.4.3 Ricerca utenti

Skype utilizza la sua tecnologia *Global Index* per effettuare la ricerca di utenti. Tale ricerca è distribuita e garantisce di trovare un utente, ammesso che esista e che abbia effettuato l'accesso almeno una volta nelle ultime 72 ore.

Un client Skype ha una finestra di ricerca. Dopo avere inserito lo user id da ricercare e aver premuto il bottone "cerca", il client inizia la sua ricerca. Un client con IP pubblico invia un pacchetto TCP al proprio supernodo. Il supernodo risponde con l'indirizzo IP e il numero di porta di quattro nodi da interrogare. Il client, quindi, invia pacchetti UDP a questi nodi. Se non trova l'utente, il supernodo viene informato via TCP, e risponde al client con l'indirizzo IP e il numero di porta di altro otto nodi da contattare. Il processo continua fino a quando il client trova il suo utente, oppure ha determinato che non esiste. In media, durante questo processo, il client contatta otto nodi.

Un client dietro a un NAT scambia dati con il proprio supernodo e alcuni dei nodi che hanno risposto alle richieste UDP durante la fase di login. Se invece oltre al NAT è presente un firewall che non consente traffico UDP, il client invia la richiesta di ricerca via TCP al proprio supernodo che si fa carico di effettuare la ricerca informando successivamente il client sui risultati ottenuti. A differenza della ricerca effettuata nel caso con IP pubblico, il client in questo caso non contatta altri nodi oltre al supernodo: ciò consente di dedurre che in questa fase il client è a conoscenza del fatto di trovarsi dietro a un firewall che blocca il traffico UDP. In ultimo si osservi che per velocizzare il processo di ricerca uno Skype client effettua caching delle informazioni relative agli utenti sui nodi intermedi.

2.4.4 Inizio e fine chiamata

La segnalazione di una chiamata in Skype viene sempre effettuata via TCP. Per gli utenti che non sono nella lista dei contatti, la chiamata è uguale all'operazione che viene effettuata per la ricerca degli utenti con in più la segnalazione della chiamata. Quindi discuteremo dell'inizio chiamata nel caso in cui il chiamato si trova nella lista contatti del chiamante, analizzando separatamente diversi scenari possibili.

Se entrambi gli utenti (chiamante e chiamato) hanno IP pubblico, sono online e risiedono nelle rispettive liste contatti, alla pressione del bottone "chiama", il client chiamante stabilisce una connessione TCP con il client chiamato: tutta l'informazione di segnalazione viene scambiata mediante protocollo TCP. Il chiamante invia anche alcuni messaggi su UDP per alternare i nodi Skype, ovvero i nodi online al momento del login. In questo scenario vengono scambiati in media 3 Kilobytes di dati.

In un secondo scenario possibile il chiamante si trova dietro un NAT sulle porte e il chiamato ha IP pubblico. In tal caso il traffico di segnalazione non fluisce direttamente tra chiamante e chiamato. Il chiamante, invece, invia informazioni di segnalazione su TCP a un nodo online intermedio, che lo inoltra verso il chiamato via TCP. Questo nodo intermedio inoltra anche i pacchetti "vocali" dal chiamante al chiamato via UDP, e viceversa.

Un ultimo caso è quello in cui entrambi gli utenti si trovano dietro un NAT sulle porte e a un firewall che non permette traffico UDP. In tale circostanza sia il chiamante che il chiamato scambiano le informazioni di segnalazione su TCP con un altro nodo online. Il client chiamante invia le informazioni via TCP ad un nodo intermedio, il quale a sua volta le inoltra via TCP al client chiamato e viceversa.

Sono molti i vantaggi nell'avere un nodo che inoltra i pacchetti "vocali" dal chiamante al chiamato e viceversa. Per prima cosa è possibile osservare che tale meccanismo consente

agli utenti dietro NAT e firewall di parlare tra loro. In secondo luogo, se gli utenti dietro NAT o firewall vogliono partecipare ad una conferenza cui aderiscono anche utenti dotati di IP pubblico, il nodo intermedio funge da “mixer” propagando il traffico della conferenza a tutti i partecipanti. Il lato negativo è che questo nodo deve sobbarcarsi una grande quantità di traffico. Durante la chiusura della chiamata, l’informazione di segnalazione viene scambiata mediante protocollo TCP tra il chiamante e il chiamato se entrambi sono su IP pubblici, o tra chiamante, chiamato e rispettivi supernodi altrimenti.

2.5 Codifica Audio

Per trasmettere le chiamate Skype fa uso di un protocollo VoIP (Voice Over IP) proprietario, cioè non formalizzato in alcuno standard internazionale. I dati, trasmessi in forma digitale, vengono cifrati tramite algoritmi non divulgati pubblicamente. L'azienda produttrice del programma assicura un grado di protezione della comunicazione comparabile con quello dei più diffusi standard crittografici. Occorre notare - tuttavia - che questo livello di sicurezza (benché superiore a quello del tutto nullo offerto dalla telefonia tradizionale) non è verificabile a causa della riservatezza dell'algoritmo crittografico.
I codec utilizzati da Skype per comprimere l’audio sono proprietari della Gobal IP Solutions e sono descritti in dettaglio di seguito.

2.5.1 iLBC

Il termine iLBC [5] è un acronimo che sta per *"internet Low Bitrate Codec"* (codec per internet a basso bitrate). iLBC è un codec offerto gratuitamente (free software) dalla Global IP Solutions ed è stato studiato specificamente per comprimere segnali vocali. Questo codec è stato realizzato per supportare un range ristretto di frequenze, e lavora con una frequenza di campionamento pari ad 8 kHz. Esso supporta essenzialmente due tipologie di frame di lunghezza base, la prima con un bit-rate di 13.3 kbps e tempo di codifica pari a 30 ms, mentre la seconda con bit-rate di 15.2 kbps e tempo di codifica pari a 20 ms. L’algoritmo implementa un controllo sui pacchetti persi durante la conversazione, e ciò garantisce una notevole robustezza nei confronti del fenomeno del *packet loss* (perdita di pacchetto). Il codec iLBC tratta ogni pacchetto in maniera indipendente dagli altri, e tale caratteristica lo rende pertanto ideale per le comunicazioni a commutazione di pacchetto. Esso mostra un degrado della qualità dell’audio proporzionale alla perdita o al ritardo dei pacchetti. Ciò

contrasta con il comportamento di altri codec basati sul paradigma CELP[2] [4], progettati per commutazioni di circuito e per essere resistenti ad errori nei bit più che alla perdita di pacchetti. Un parametro rilevante per il codec audio utilizzato, nel caso in cui siano presenti perdite di pacchetti, è il numero di frame necessarie per recuperare il singolo pacchetto perso. Nel caso del codec iLBC questo numero è nullo. Il primo pacchetto seguente un pacchetto perso è esattamente codificato come previsto. Siccome iLBC è un codec audio a banda stretta, esso utilizza tutti i 4 kHz di frequenza di banda disponibili, mentre molti codec standard a basso bit rate utilizzano solo le frequenze comprese nella banda che va dai 300 Hz ai 3,4 KHz: ciò ha un chiaro effetto sulla qualità dell'audio. In aggiunta alle caratteristiche dello spettro dell'audio codificato, iLBC imita accuratamente le caratteristiche del segnale originale dando luogo ad un suono più limpido e naturale rispetto agli altri codec standard a basso bit rate. Sono state effettuate diverse prove per valutare le prestazioni di questo codec. La figura 2 mostra i risultati del test Dynastat dove il codec iLBC è stato confrontato con gli altri due codec standard G.729 e G.723.1 [5].

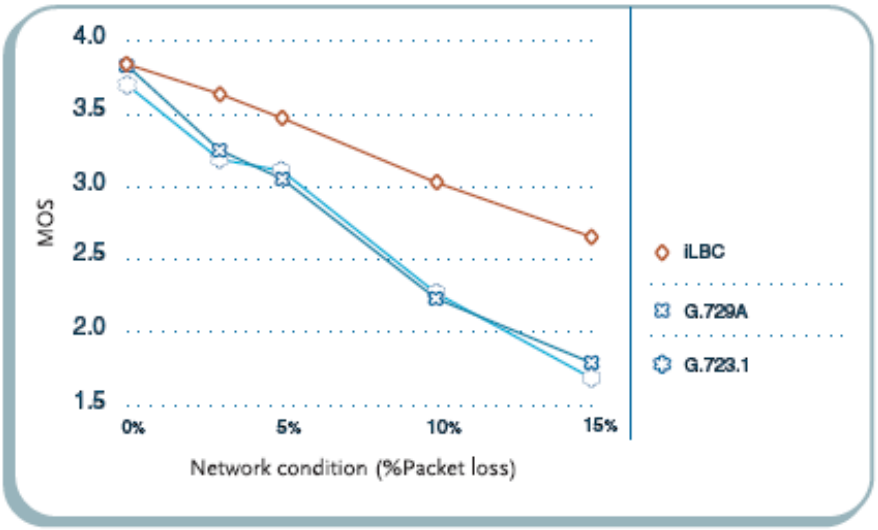

Figura 2 - Test di confronto

I risultati mostrano chiaramente la superiorità di iLBC quando è utilizzato in un ambiente reale, dove le proprietà intrinseche dei pacchetti risultano di alta qualità rispetto anche alle condizioni avverse della rete. Il test mostra inoltre non solo un miglioramento significativo delle prestazioni rispetto ai codec standard correnti (come G.723.1, G.728, G.729, GSM etc.) in condizioni di perdita di pacchetti, ma la qualità risulta equivalente o migliore rispetto agli standard anche in un canale pulito cioè senza perdita di pacchetti. Inoltre si è visto che

[2] **CELP** sta per **Code Excited Linear Prediction** ed è un algoritmo di codifica vocale originariamente proposto da M.R. Schroeder e B.S. Atal nel 1985.

non vi sono significative differenze di prestazioni nell'utilizzo con tempo di codifica pari a 20 ms o 30 ms, ad eccezione del caso in cui si abbia perdita di pacchetti: in tal caso, infatti, la rete risulta essere più robusta con le frame con tempo di codifica pari a 20 ms. Riassumendo i principali benefici apportati da questo codec possono essere così elencati e riassunti dalla tabella 1:

- non è standardizzato da IETF;
- brevetto free con una qualità audio superiore rispetto ai codec G.729 e G.723.1;
- maggiore robustezza rispetto ai codec standard per quanto riguarda la perdita di pacchetti;
- supporto per dimensioni multiple delle frame, che apporta una maggiore flessibilità che va incontro alle differenti applicazioni per le tecnologie VoIP;
- alta resistenza alla perdita di pacchetti rispetto ai tradizionali codec a basso bit-rate;
- prestazioni in ogni caso non inferiori ai codec standard;

Specifiche	
Implementazione	Fixed point ANSI C code
Dimensione frame	20 e 30 ms
Bit rate	13.3 Kbps (30 ms frame) e 15.2 Kbps (20 ms frame)
Campionamento	8 kHz
Qualità di base	Migliore di G.729A e comparabile con la G.729E
Resistenza a perdite di pacchetti	Migliore di G.729 e superiore a G.729E
Ridimensionamento pacchetti persi	Soluzioni integrate

Tabella 1 - Specifiche iLBC

2.5.2 iSAC

iSAC [9] è un codec adattativo utilizzato da Skype e fa parte della suite software MediaWare prodotta da Global IP Solutions. Esso è stato progettato esclusivamente per la qualità del suono in banda larga in applicazioni sia a basso che ad alto bit-rate. Offre una qualità migliore rispetto alla codifica PSTN (*Public Switched Telephony Network*), ottenuta regolando la frequenza di trasmissione al fine di garantire migliori prestazioni audio per le

connessioni attualmente esistenti. Proprio per la caratteristica di alta qualità sperimentata dagli utenti, iSAC è divenuto il codec in assoluto più utilizzato per le comunicazioni VoIP a banda larga, venendo implementato approssimativamente da circa 100 milioni di end-point. Il codec regola automaticamente la velocità di trasmissione partendo da un minimo di 10 kbps fino ad un massimo di 32 kbps. Questa flessibilità rende iSAC adatto per applicazioni VoIP a banda larga, applicazioni multimediali in tempo reale, conferenze, o ad esempio per giochi che utilizzano connessioni di rete prevedendo trasmissioni di flussi vocali. Esso è in grado di gestire anche applicazioni senza audio parlato, come ad esempio la musica e il rumore di fondo, comportandosi bene anche in condizioni di *packet loss*. Risulta pertanto in grado di offrire un'elevata qualità vocale, ad alto jitter[3] in ambienti con elevate perdite di pacchetti, garantendo un miglioramento delle prestazioni con una notevole riduzione del ritardo. È anche disponibile in modalità a bassa complessità, risultando ideale anche per dispositivi con risorse limitate, come telefoni cellulari e PDA, o in generale anche per applicazioni di audio-conferenza su dispositivi mobili. Tale versione a bassa complessità rende dunque appetibile l'utilizzo di iSAC in qualsiasi scenario, permettendo in definitiva la comunicazione ad alta qualità su piattaforme multiple.

In definitiva iSAC è ideale per:

- dispositivi e applicazioni a banda limitata;
- dispositivi e applicazioni caratterizzate da audio a banda larga;
- provider e utenti che lavorano con reti real time richiedendo una forte affidabilità;

Le caratteristiche principali di questo codec possono essere così elencate nella tabella 2:

- eccellente *trade-off* tra bit-rate/qualità audio;
- regolazione automatica della velocità del bitstream per migliorare la qualità;
- possibilità di utilizzo di un bit rate impostato su un valore fisso;
- disponibilità in modalità a bassa complessità;
- alta resistenza alla perdita di pacchetti (maggiore degli standard attuali);
- utilizzo della piena larghezza di banda disponibile a 8 KHz con frequenza di campionamento pari a 16 KHz;
- alta interoperabilità con centinaia di milioni di punti terminali VoIP esistenti.

A pagina seguente riportiamo due tabelle riassuntive che riportano le caratteristiche del codec iSAC.

[3] Differenza di ritardo osservato tra due pacchetti appartenenti allo stesso flusso di comunicazione.

Specifiche	
Implementazione	Fixed point ANSI C code
Dimensione frame	30 e 60 ms
Bit rate	Tra 10 Kbps e 32 Kbps
Campionamento	16 kHz
Qualità di base	Migliore di G.722.2
Banda audio	8 kHz
Complessità	Comparabile con G.722.2
Ritardo algoritmo	Per i frame maggiori di 3 ms

Caratteristiche in modalità a bassa complessità	
Bit rate	In media 40 kbps
Qualità	Migliore di G.722.2
Complessità	6-10 MIPS

Tabella 2 - Caratteristiche del codec iSAC

2.5.3 Codec audio alternativi

Skype utilizza anche codec alternativi a iLBC e iSAC come *G.729* e *Speex*. Con G.729 si indica un algoritmo di compressione audio per la voce descritto nella specifica G.729 dell'ITU-T [6]. L'algoritmo divide l'audio della voce in parti da 10 ms. Musica o toni come il DTMF o i toni del fax non possono essere trasmessi fedelmente con questo codec, e quindi per questi si utilizza il G.711 o il metodo *out-of-band*. G.729 è utilizzato soprattutto nel Voice over IP (VoIP) qualora si abbiano scarsi requisiti in termini di banda. Lo standard G.729 opera a 8 kbit/s, ma vi sono estensioni che permettono di trasmettere anche a 6.4 kbit/s e 11.8 kbit/s per segnali di qualità leggermente inferiori o superiori. È molto comune anche il *G.729a* che è compatibile col G.729, ma richiede risorse computazionali leggermente inferiori rispetto a quest'ultimo: risulta ovvio che la minore complessità va a discapito della qualità del suono. Recentemente G.729 è stato esteso per fornire un supporto ai codec per il parlato a banda larga per cui, ad esempio, le frequenze acustiche trasmesse

vanno ora da 50Hz a 7KHz. Questa estensione viene chiamata *G.729.1*: essa è organizzata in maniera gerarchica, ovvero il suo bit rate e la qualità ottenuta si possono modificare semplicemente troncando il bitstream.

Speex è un codec open source per la compressione dell'audio parlato sviluppato allo scopo di non infrangere alcun brevetto software. La sua distribuzione è regolata dalla Licenza BSD e può essere utilizzato in flussi formato *Ogg* oppure trasmesso direttamente usando i protocolli UDP/RTP. Il codificatore Speex usa il formato bitstream Ogg, ed i suoi progettisti vedono il proprio progetto come complementare a *Vorbis*, un codec di compressione audio di uso generico. A differenza di molti altri codec per il parlato, Speex non è mirato espressamente all'utilizzo nelle applicazioni di telefonia, ma piuttosto alla compressione basata su file. Gli obiettivi del progetto sono stati quelli di creare un codec che permettesse sia un'ottima qualità che bassi bit rate, il che ha condotto allo sviluppo di un codec con bit rate multipli (VBR – *Variable bit rate*). Una qualità molto buona significa inoltre il supporto per le larghe frequenze (frequenza di campionamento pari a 16 kHz) oltre al supporto per quelle strette (qualità telefonica, frequenza di campionamento pari a 8 kHz). I progettisti di Speex, tenuto conto che il codec sarebbe stato utilizzato anche in contesti VoIP, hanno fatto in modo che questo avrebbe garantito una certa robustezza in caso di pacchetti persi, ma non in caso di pacchetti corrotti, dato che l'UDP assicura che i pacchetti o arrivino integri o non arrivino. Tutto questo ha portato alla scelta della tecnica di encoding chiamata *Code Excited Linear Prediction* (CELP). Una delle molte ragioni è che CELP ha dimostrato di essere adatto allo scopo e soprattutto scalabile sia a bassi bitrate (e.g. 4.8 kbit/s) che ad alti bitrate (come con *G.728* a 16 kbit/s). Le caratteristiche principali di Speex possono essere riassunte come segue:

- software libero/open source, libero da brevetti e royalties;
- integrazione di banda stretta e larga nello stesso bit-stream;
- ampia gamma di bitrate disponibile (da 2 kbit/s a 44 kbit/s);
- cambio dinamico di bitrate (VBR);
- complessità variabile.

2.6 Codifica Video

Skype offre prestazioni migliori rispetto a tante altre applicazioni per le chiamate video in Internet, tutto grazie all'utilizzo del particolare codec VP7 sviluppato dalla On2: tale codec rappresenta, infatti, la migliore tecnologia di compressione video attualmente disponibile dal punto di vista del rapporto compressione/qualità.
Di seguito si riportano alcuni motivi per i quali VP7 supera tutti gli altri codec attualmente utilizzati per la compressione video real time:

- test obiettivi utilizzando Peak Signal-to-Noise Ratios (PSNR) e altri parametri mostrano che le prestazioni sono migliori rispetto a tutte le tecnologie concorrenti tra cui *Windows Media 9*, *Real 10*, *H.264* e *MPEG-4* in una vasta gamma di tassi di trasferimento, dal dial-up (28,8 Kbps) a fino alle qualità DVD e HD;
- progettato per essere in grado di lavorare su DSP (*Digital Signal Processors*) poco costosi;
- ideale per i chipset embedded non-PC e dispositivi set-top-box;
- software facilmente aggiornabile;
- compressione ad alta definizione (*HD*) senza restrizioni di codifica;
- possibilità di utilizzo di opzioni del tipo "*fast compress*", modalità che può codificare in tempo reale su una fascia alta con poca perdita di qualità senza eccessive richieste in termini di risorse;
- modalità di codifica che garantiscono la possibilità di riproduzione di file trasmessi a datarate costanti.

Il codec VP7 offre, inoltre, una importante caratteristica di scalabilità. Esso fornisce tre profili:

- **Hi-Def Profile**: progettato specificamente per un rapida visualizzazione su processori poco costosi;
- **Simple Profile**: progettato per assicurare la migliore qualità possibile a bassi datarate su processori poco costosi;
- **Advanced Profile**: progettato per assicurare la migliore qualità possibile a datarate estremamente bassi (e.g. meno di 200 kbps).

Alcuni test hanno mostrato che la qualità delle immagini codificate attraverso VP7 (configurato in maniera tale da essere la massima possibile) è del tutto paragonabile a quella ottenuta mediante la

codifica attraverso compressione MPEG-4 veloce allo stesso datarate. In effetti si stima che la complessità computazionale della codifica mediante questo codec sia circa 1.8 volte la complessità di calcolo per MPEG-2. Inoltre VP7 ha un costo di decodifica sostanzialmente minore rispetto al codec H.264, con un tempo di calcolo di circa 2-3 volte inferiore (alla stessa qualità).

Si fa notare inoltre che i video codificati attraverso VP7 sono più semplici da decodificare per alcuni motivi che non sono direttamente collegati alla complessità dell'algoritmo stesso di codifica. In particolare vi è il fatto che VP7 contiene meno "codice" rispetto ad altri codec concorrenti, e ciò lo rende facilmente trasportabile. VP7 è estremamente conveniente per i flussi video real-time (come per Skype) perché le sue caratteristiche implementative consentono una semplice fase di codifica e decodifica sullo stesso processore, caratteristica questa molto importante nel caso di sistemi real time. Per esempio si può pensare ad una applicazione di videoconferenza che richiede che il processore codifichi il flusso video in uscita decodificando al contempo stesso un flusso video in ingresso. In questo tipo di applicazioni, l'encoder real-time VP7 può essere configurato in maniera tale da lasciare decidere all'applicazione (dinamicamente) quanto tempo spendere per la codifica (fermo restando il prerequisito fondamentale di elaborare al meglio possibile il flusso in ingresso). L'utilizzo dell'encoder VP7 real time consiste, in ultima analisi, solo in un leggero degrado della qualità delle immagini in uscita rispetto al caso in cui si utilizzi l'encoder non real-time. Questo degrado, ovviamente, è tanto più evidente quanto più le immagini da codificare contengono oggetti in movimento, texture complesse e così via.

Il codec utilizzato da Skype, inoltre, possiede nella specifica versione la possibilità di effettuare un controllo dinamico estremamente configurabile ed accurato del datarate. In particolare l'applicazione può scegliere tra un datarate costante o variabile; in effetti dapprima viene provata la trasmissione del flusso a datarate costante. Se il buffering a datarate costante ottempera alle specifiche richieste, il video viene riprodotto senza alcun bisogno di rebuffering. Altrimenti, scelta la trasmissione a datarate variabile, l'applicazione può specificare quanto debba variare il datarate rispetto ad un valore di *bias* che costituisce il valore assunto come datarate costante, assegnando un valore minimo ed uno massimo come estremi del range all'interno del quale far variare il datarate.

Per raggiungere il datarate richiesto, VP7 utilizza tre tecniche separate:

- variazione della qualità delle frame;
- ricampionamento spaziale del video di input;
- temporaneo ricampionamento del video in ingresso.

Il codec gestisce il controllo del datarate attraverso una modellazione sofisticata del buffer del client su cui il video viene riprodotto. In particolare mantiene informazioni:

- sulla specifica quantità di dati che sono stati precaricati nel buffer;

- sulla massima dimensione dei dati che possono essere memorizzati nel buffer;
- sul livello di riempimento del buffer da ritenere critico per iniziare a prendere misure più drastiche per abbassare il datarate (effettuando, ad esempio, un ricampionamento temporale).

In ultimo ci preme far notare che VP7 realizza una compressione a due passi. Durante il primo passo il codec utilizza un set di euristiche molto accurate per produrre il miglior risultato possibile. In particolare genera delle statistiche che vengono poi scritte su di un file. Durante il secondo passo, poi, legge le statistiche generate nella prima fase e prende alcune decisioni come quelle di:

- attribuire un maggior numero di bit ad aree dell'immagine ritenute più critiche;
- scegliere i migliori frame chiave;
- determinare le dimensioni dei frame basandosi sulla rilevanza del frame corrente e predicendo quelle future.

Bibliografia

[1] Salman A. Baset and Henning Schulzrinne, *An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol*, September 15, 2004

[2] S.V Andersen W.B Kleijn R. Hagen, J.Linden, *iLbc a linear predictive coder with robusteness to packet loss.*

[3] http://www.nist.gov/CryptoToolkit

[4] Shoam, *constrained excitation coding of speech*, in abstract IEEE Workshop of speechcoding for telecomm

[5] www.gipscorp.com/files/english/white_papers/iLBC.WP.pdf

[6] www.itu.int/rec/T-REC-G.729/e

[7] J. Rosenberg, J. Weinberger, C. Huitema, and R. Mahy. STUN: Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs). RFC 3489, IETF, Mar. 2003

[8] www.rsa.com

[9] www.gipscorp.com/high-quality-codecs/index.php