# Förmedling av tal över IP

kirei

**Förmedling av tal över IP**



F. Ljunggren, S. Kerker, J. Schlyter

# Innehåll

# 1 Introduktion

På senare tid har det skett en stark konvergens mot IP-baserade kommunikationsnät inom alla sektorer i samhället. Standardiseringen på IP-teknik har fört med sig kostnadseffektiva lösningar med god interoperabilitet, men också krav på att de IP-baserade kommunikationsnäten ska kunna bära en lång rad kommunikationstjänster med vitt skilda kapacitetsbehov och kvalitetskrav.

Internetprotokollet – *Internet Protocol* (IP) – är i grunden en paketförmedlad teknik som utgår från vad som kallas "bästa förmåga" (*best effort"*), det vill säga att IP-tekniken i sig själv inte garanterar en viss tjänstekvalitet. Olika tillämpningar som används i samma IP-nät konkurrerar om de tillgängliga resurserna och kan på så sätt påverka varandras tjänstekvalitet på ett sätt som är svårt att förutse och styra.

För en realtidskänslig tjänst som just telefoni ansågs länge att IP som bärare, utan garanterande kvalitetsfunktioner i näten, inte kunde leva upp till de kvalitetskrav som tjänsten fordrar. Traditionella telefoninät har en garanterad tillgänglig resurs som, under förutsättning att nätet accepterar och etablerar ett samtal, garanterar kapaciteten under hela samtalet oavsett hur många andra användare i nätet som försöker etablera ytterligare samtal. En nackdel med detta traditonella tillvägagångssätt, och med kretskopplad infrastruktur i allmänhet, är att det är mycket svårt att använda de totalt tillgängliga resurserna på ett effektivt sätt. Ett IP-nät använder de tillgängliga resurserna dynamiskt, så när en viss tjänst inte använder nätet kan andra tjänster nyttja den för stunden lediga kapaciteten. Detta är i sig en fundamental skillnad mellan de paketförmedlade nätverken och den traditionella telefonins kretskopplade kommunikationstekniker. I dagsläget är dock IP som bärare av telefoni mycket vanligt och dagens IP-nät hanterar väl de krav som ställs på denna typ av tjänster.

På internet har IP-telefoni i olika former används en längre tid. Tack vare nyttjandet av avancerade talkodtyper och signaleringsfunk-

tioner fungerar flertalet IP-telefonitjänster i de allra flesta fall utmärkt, trots avsaknaden av garanterad tjänstekvalitet i det underliggande IP-nätet. Kvaliteten på den infrastruktur som bär upp internet varierar givetvis stort mellan olika geografiska områden och beroende på de anslutningsformer som används. Utvecklingen går emellertid fort. Parallellt med teleoperatörernas utbyggnad av de senaste mobilnätsteknikerna blir IP, och till stor del även internet, den naturliga bäraren av all typ av trafik. Även för telefoni.

*Målgrupp och syfte*

Denna publikation vänder sig i första hand till ansvariga för införande av IP-telefonilösningar, och särskilt de med tidigare erfarenheter av traditionell telefoni. Handledningen syftar till att ge vägledning vid utformning, kravställning och införande av IP-baserade talkommunikationstillämpningar samt ge en grundläggande förståelse för de utmaningar och tekniska frågeställningar som måste hanteras. Handledningen tar upp och beskriver ett flertal olika signaleringsprotokoll för IP-telefoni som direkt har inverkan på utformning och systemarkitektur.

På liknande sätt beskrivs även ett brett urval av olika tal- och mediakodtyper med olika kvalitetsegenskaper under olika förutsättningar. Flertalet sådana aspekter gällande just tjänste- och talkvalitet behandlas samt olika förutsättningar för detta vid införande av IP-baserad telefoni i ett nytt eller befintligt IP-nät.

Vidare behandlas ett flertal olika säkerhetsrelaterade aspekter att ta hänsyn till där just telefoni från ett IT-säkerhetsperspektiv ofta är ett eftersatt område jämfört med andra internetbaserade tjänster. I traditionella telefoninät betraktas ofta nätet i sig själv som säkert, eftersom det är teleoperatörens nät och som i någon mån står under dennes kontroll. Om telefonitjänsterna levereras över internet eller andra IP-nät ökar ofta exponeringsgraden och andra typer av IT-säkerhetsrelaterade hotbilder växer fram. Dessa hot måste identifieras och sårbarheterna hanteras, så att riskerna vid införandet av IP-telefoni kan lindras på samma sätt som andra tjänster i nätet.

Telefoni visar tydligt hur två olika teknikvärldar möts. Dels den traditionella telefonivärlden där telefoninätet är en integrerad del av tjänsten och dels internet-världen där nätet är den gemensamma bäraren av samtliga tjänster och telefoni bara är ytterligare en tjänst.

Att förstå helheten är en förutsättning för att kunna utforma, kravställa och införa IP-baserade telefonitillämpningar på ett tillfredställande och säkert sätt.

# 2 Teknikintroduktion

I telefonins begynnelse fanns ett begränsat antal telefoner som var direkt anslutna till de motabonnenter de kunde kommunicera med. Snabbt uppstod naturliga problem med skalbarheten i denna anslutningsform, varför det istället etablerades lokala telefonväxelpunkter dit områdets telefoner anslöts. Samtal kunde sedan kopplas mellan alla telefoner som var anslutna till samma växlingspunkt. Snart kopplades närliggande växlingspunkter samman med större stamförbindelser så att abonnenter inom dessa områden kunde kommunicera med varandra, varefter fler långväga förbindelser kopplades in. Idag är de allra flesta telefoninät globalt sammankopplade och bildar det vi i dagligt tal benämner *det publika telenätet*, eller *Public Switched Telephony Network* (PSTN).

I takt med att den digitala tekniken gjorde sitt intåg digitaliserades även telefoninäten. Först digitaliserades de större kärnnoderna och nätcentrala växlar, och senare även mindre telestationer och kundanslutningar. Idag är allt digitaliserat utom i många fall den sista förbindelsen ut från närmaste telestation till abonnenten, den så kallade *local loop*.

Tjänster som fax och dataöverföring via modem blev med tiden viktiga parallellt med vanliga röstsamtal. *Integrated Services Digital Network* (ISDN), som till stor del fortfarande används idag, fick sitt stora genomslag. Den digitala tekniken förbättrade i första hand nätens kapacitet, kvalitet och kostnadseffektivitet, samt möjliggjorde introduktion av nya tjänster.

I samband med IP-nätens utbredning, då de flesta teleoperatörers teknikutveckling gick i riktning mot att ett och samma transmissions- och kommunikationsnät skulle bära samtliga tjänster, blev det naturligt att även telefonitjänsterna flyttades till denna infrastruktur. Oavsett om de IP-baserade nätverken är en del av internet eller inte så är tekniken densamma. Det är i första hand affärsmodellerna som

skiljer. Vissa teleoperatörer väljer att fortsätta att produktifiera telefoni på samma sätt som tidigare och ser enbart det privata IP-nätet som en annan teknisk överföringsteknik.

## 2.1 Traditionell telefoni

Med traditionell telefoni avses klassisk kretskopplad telefoni. Ofta går denna typ av telefoni under benämningen *Plain Old Telephony System* (POTS). Ditt egna telejack är anslutet till din lokala telestation och kopplas sedan vidare in i teleoperatörens hierarki av nätväxlar. Nätlogiken är helt och hållet koncentrerad till de centrala delarna av nätet. Som abonnent har du enbart möjlighet att välja de tjänster som din teleoperatör erbjuder dig. Tjänsten telefoni och telefonnätet är integrerade.

Din anknytning adresseras genom ett telefonnummer (*A-identitet*) och du kan via din terminal ringa andra abonnenter via deras telefonnummer (*B-identitet*). Med andra ord, telefoni så som de flesta av oss känner det. A-identitet är kopplat till din anknytning och oavsett vilken terminal du ansluter så följer identiteten ditt telejack.

Traditionell telefoni använder vanligtvis olika signaleringsprotokoll för olika nivåer i nätet. Mot slutkund används vanligtvis vanlig analog förbindelse mot enstaka terminaler och ISDN PRI mot företagsväxlar. I de centrala delarna av nätet och vid samtrafik med andra teleoperatörer används vanligen *ISDN User Part* (ISUP) som definieras genom *Signaling System 7* (SS7).

Det är även möjligt att för IP-telefoni använda signaleringsprotokoll som används inom traditionell telefoni, till exempel ISUP. IP används då enbart som bärare av till exempel ISUP och IP transporterar ISUP-meddelanden mellan växlarna. *Signaling Transport* (SIGTRAN), som är resultatet av en arbetsgrupp inom *Internet Engineering Task Force* (IETF), definierar hur dessa traditionella SS7-protokoll kan överföras på ett tillförlitligt sätt över ett IP-nät. SIGTRAN-gruppens främsta arbete bestod i att ta fram ett nytt transportprotokoll lämpligt för just detta, nämligen *Stream Control Transmission Protocol* (SCTP). Även andra signaleringsprotokoll för IP-telefoni kan användas för att innesluta och bära ISUP änd-till-änd över ett IP-nät.

## 2.2 IP-telefoni

Med IP-telefoni avses i princip traditionell telefoni med IP-nät som bärare. Det är alltså i första hand en liknande tjänst som i dess mest grundläggande form endast byter underliggande nät från kretskopplad teknik till pakeförmedlad bärare. I och med övergången till IP öppnas emellertid en mängd nya möjligheter för att på ett enkelt sätt utöka telefonitjänsten med till exempel nya talkodtyper, video och datatjänster. Då fortfarande en stor andel av den totala telefonitrafiken förmedlas via kretsförmedlad nätinfrastruktur försvinner i praktiken mervärdet av dessa nya möjligheter. Så länge samtalet inte förmedlas änd-till-änd via IP är det svårt att hävda att IP-telefoni tillför något nytt ur ett tjänsteperspektiv. Det är snarare i fördelarna med att samutnyttja infrastruktur och IP-resurser som vinsterna och besparingarna för teleoperatörer och tjänsteleverantörer ligger.

För IP-telefoni finns ett flertal olika signaleringsprotokoll definierade. Dels finns de fritt tillgängliga och öppet standardiserade protokollen och dels finns ett antal proprietära och leverantörsspecifika protokoll. Till de vanligaste bland de öppna protokollen hör *Session Initiation Protocol* (SIP) och H.323 som belyses mer i detalj senare i denna handledning. De proprietära protokollen är oftast protokoll som enbart används mellan terminal och lokal växel där flera IP-telefoniväxelleverantörer har olika egenutvecklade lösningar.

## 2.3 VoIP

Med VoIP avses de mer internet-anpassade varianterna av IP-telefoni, där de traditionella affärsmodellerna och tjänsteutbudet inte längre är styrande. För signalering används typiskt de standardiserade och öppna protokollen, men det finns även proprietära och mer slutna varianter av sådana system, som till exempel *Skype*. VoIP-tjänster är också i allmänhet globalt tillgängliga internettjänster och har på många sätt mer gemensamt med sådana tjänster än med traditionella telefonitjänster. Till skillnad från traditionella telefonitjänster och i många fall även IP-telefonitjänster, så krävs ingen tilllgång till någon särskild och teleoperatörsägd infrastruktur för att nå åtkomst till telefonitjänsterna, utan en internetanslutning av god kvalitet är tillräcklig.

Denna modell, där telefonitjänsten är frikopplad från leverantören

av IP-tjänsten, benämns ofta *Over the Top* (OTT) och betraktas i vissa fall av traditionella telefonoperatörer som en sämre typ av IP-telefoni, med argumentet att dessa leverantörer inte kan garantera några kvalitetsegenskaper, då de inte har kontroll över IP-nätet. I samband med den kraftiga utbyggnaden av internetinfrastrukturen i kombination med smarta talkodtyper blir sådana problem i praktiken mindre och mindre, åtminstone från ett tekniskt perspektiv. VoIP-tjänster som Skype, Facetime och Google Hangouts har sedan länge visat att det går alldeles utmärkt att använda tal- och videokommunikation med god kvalitet över internet.

I sammanhanget bör även nämnas öppen och federerad SIP-telefoni så som protokollet och arkitekturen egentligen utformades. Med federerad avses att tjänsten istället för att enbart hantera samtal mellan abonnenter direkt anslutna till den specifika SIP-leverantören, även tillåter ändnoderna att direkt kommunicera över internet med andra federerade SIP-system. De flesta SIP-leverantörer väljer emellertid att terminera all trafik som inte kan styras till egna abonnenter direkt till PSTN. Dessa teleoperatörer faller ofta inom kategorin för *IP-telefoni*, så som beskriven ovan, då de också ofta låser in IP-telefonitjänsten i sitt eget stängda nät, så kallad *Walled Garden*.

För att en federerad SIP-tjänst ska fungera måste leverantören publicera sina SIP-servrar i domännamnssystemet (*Domain Name System* (DNS)) och använda ett adresseringsformat som medger samtalsstyrning till externa SIP-system, till exempel URI-baserad adressering. Ofta är det just här problemet uppstår då de flesta IP-telefoner har samma användargränssnitt som de traditionella telefonerna haft sedan länge, nämligen ett tastatur för att ange ett telefonnummer. Ett telefonnummer går inte att använda för federerade SIP-tjänster såvida inte en global adressöversättningsfunktion används där telefonnummer översätts till ett URI-baserat adresseringsschema. Ett exempel på en sådan adressöversättningsfunktion är *E.164 to URI DDDS Application* (ENUM) som också beskrivs mer i detalj senare i denna handledning.

## 2.4 Skillnader mellan traditionell telefoni och IP-baserad telefoni

Den största skillnaden mellan traditionell telefoni och IP-baserad telefoni är givetvis dess bärare och den logiska frikopplingen av underliggande infrastruktur samt separationen i signaleringplan och mediaplan. Även om denna separation även finns i till exempel SS7, är separationen tydligare med IP-nätet som bärare. Dessutom är det inte ovanligt att den IP-baserade infrastrukturen bär upp en mängd andra tjänster och applikationer parallellt med telefonin. På så sätt kan infrastrukturen nyttjas på ett mycket mer effektivt sätt och endast *ett* nät behöver byggas.

Inom både traditionell telefoni och IP-baserad telefoni är det vanligt att signaleringen behöver gå genom flertalet förmedlingsnoder för att ett samtal ska etableras. Inom den IP-baserade telefonin är det däremot sällan önskvärt att mediaströmmarna följer samma mönster, då det ofta bidrar till fördröjningar och påverkar till exempel talkvaliteten på ett negativt sätt. Mediatrafiken följer istället IP-nätets trafikvägar, och eventuella problem som kan uppstå förväntas lösas genom IP-nätets omdirigeringsfunktioner.

Det finns dock flertalet tjänster som kräver att media antingen följer signaleringens väg eller åtminstone ankras i vissa punkter. Ett exempel på detta är så kallad *roaming* och *handover* inom *Voice over LTE* (VoLTE), fjärde generationens mobiltelefoni. I dessa nät finns enbart IP som bärare vilket medför att telefonitjänsten måste realiseras som en IP-telefonitjänst. En överflyttning av samtal mellan VoLTE och till exempel 3G eller GSM innebär således inte bara att mobiltelefonen måste byta radiogränssnitt, den måste även byta från IP-förmedlad tjänst till kretskopplad. Samtidigt behåller arkitekturen den traditionella mobiloperatörsmodellen med hemmanät (*home network*) och besökt nät (*visited network*), vilket gör trafikflödet mer komplicerat än vid till exempel internetkommunikation, där det bara finns *ett* internet. För att möjliggöra snabb omkoppling från till exempel VoLTE till *Global System for Mobile Communications* (GSM), i de fall en VoLTE-terminal förlorar 4G-täckning, så ankrar VoLTE-arkitekturen mediaströmmen i en för operatören lokal nod. När VoLTE-telefonen byter från 4G-nätet till 3G- eller GSM-nätet (*handover*) behöver mobilnätet enbart styra om mediatrafiken från den lokala ankringspunkten i mobiloperatörens eget nät, istället för att förhandla nya ändpunkter

änd-till-änd, vilket kan ta för lång tid i ett pågående samtal. Sådan ankring medför därför kortare överlämningstider mellan de olika mobilnäten och en bättre tjänst för slutanvändaren.

## 2.5 Skillnader mellan IP-telefoni och VoIP

De främsta skillnaderna mellan IP-telefoni och VoIP ligger i affärsmodeller och hur tjänsterna tillgängliggörs. Även om det finns en mängd tjänster och applikationer med proprietära signaleringsprotokoll bland VoIP-leverantörerna är det i första hand just kopplingen till traditionella affärsmodeller och leveranssätt som skiljer. Som beskrivet ovan är IP-telefoni i dess enklaste form enbart en förflyttning av den traditionella tjänsten telefoni från ett kretskopplat nät till ett IP-nät. Ofta innebär det att en slutkund måste vara kund hos just den teleoperatören som tillhandahåller IP-nätet för att få tillgång till de telefonitjänster som finns i detta nät. VoIP-leverantörer levererar däremot sina tjänster över internet där vilken internetanslutning som helst ger tillgång till telefonitjänsterna.

Fördelen med att köpa telefonitjänsten av samma leverantör som även levererar IP-anslutningen är att operatören i någon mån kan garantera tjänstens kvalitet på ett sätt som inte går att åstadkomma över ett publikt kommunikationsnät som internet. Fördelen med VoIP-tjänster är emellertid att de är enkla att flytta, ofta erbjuder möjlighet till bättre talkodtyper och video, samt även kryptering ändtill-änd så länge samtalet enbart går mellan internetanslutna abonnenter.

Nyckeln till VoIP-tjänster och federering är möjligheten att adressera användare med ett globalt adresseringsschema.

# 3 Signaleringsprotokoll

I denna handledning beskrivs de vanligast förekommande och öppet standardiserade signaleringsprotokollen. Som tidigare berörts förekommer även flertalet alternativa lösningar som används både för IP-telefoni och VoIP, såväl öppna som proprietära. En kortare sammanfattning av ett urval av sådana VoIP-tjänster görs mot slutet av detta kapitel.

## 3.1 SIP

*Session Initiation Protocol* (SIP) är ett signaleringsprotokoll som används för att etablera, förändra och avsluta IP-baserade sessioner av olika slag, till exempel sådan dubbelriktad realtidskommunikation som IP-telefonsamtal. IP-telefoni är dock inte det enda användningsområdet för SIP då även andra typer av tillämpningar och tjänster, till exempel direktmeddelanden (*instant messaging*), tillgänglighetsindikering (*presence*) och olika sensorbaserade tjänster kan implementeras med protokollet. SIP används enbart för att hantera sessioner och för att adressera och lokalisera användare med hjälp av till exempel domännamnssystemet (*Domain Name System* (DNS)).

SIP är ett textbaserat protokoll, framtaget och standardiserat inom *Internet Engineering Task Force* (IETF). En rad IETF-dokument – *Request for Comments* (RFC) – definierar och uppdaterar dels själva grundprotokollet SIP, dels en mängd tillägg och utökningar. Vanligen refereras emellertid till [RFC3261] för de fundamentala grunderna i SIP. Utöver detta grunddokument är även [RFC3263] och [RFC3264] av stor betydelse för beskrivningen av protokollets tillämpade funktionalitet. [RFC3263] beskriver hur SIP med stöd av DNS ska kunna lokalisera SIP-resurser, välja transportprotokoll och fördela belastning mellan olika noder. [RFC3264] beskriver hur SIP med stöd av *Session Description Protocol* (SDP) ska kunna förhandla fram gemensamma

media- och talkodtyper för de kommunicerande ändpunkterna.

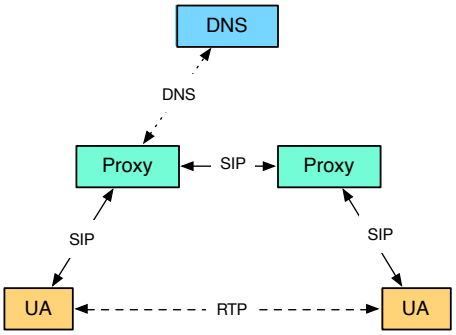
Då SIP enbart är ett signaleringsprotokoll betyder det att det är frikopplat från den eller de data- och mediaströmmar som kan komma att upprättas genom användandet av protokollet. Det innebär att SIP i sig inte är begränsande i vilka typer av informationsströmmar som kan etableras, oavsett om det är ljud, video, eller andra typer av data. Det kan vara samma ändpunkter som signalerar med SIP som även är ändpunkter för den eller de dataströmmar som etableras, men det måste inte med nödvändighet vara fallet. Vid till exempel tredjepartsstyrning (*Third Party Call Control*) hanteras oftast sessionssignaleringen av en applikationsserver som baserat på applikationens logik styr uppkoppling och mediaströmmar till olika ändpunkter.

Från arkitektursynpunkt jämförs SIP ibland med *Simple Mail Transfer Protocol* (SMTP), protokollet som används för att förmedla e-post mellan användare på internet. SIP-baserad IP-telefoni kan liknas vid *interaktiv e-post*, i det hänseendet att arkitekturen, bestående i förmedlingsnoder, ändnoder och "brevlådor", har sina förklarliga likheter mellan de två protokollen. I SIP-modellen har varje användare ett konto i en förmedlingsnod (en "brevlåda"). Varje förmedlingsnod eller grupp av förmedlingsnoder tilldelas det administrativa ansvaret för att ta emot och förmedla in- och utgående SIP-signalering för en eller flera DNS-domäner, på samma sätt som för förmedling av e-post. Ändnoderna registrerar, och kopplar på så sätt upp sig mot förmedlingsnodernas "brevlådor". Signaleringen mellan parternas respektive förmedlingsnoder kan gå i flera steg via andra förmedlingsnoder som används för att knyta samman olika adressrymder eller för att upprätthålla särskilda regelverk för signalering och informationsströmmar inom en viss administrativ domän.

Det kan emellertid skönjas ett antal tydliga skillnader i jämförelsen mellan SIP och e-post. Det första, och kanske mest fundamentala, är att i användningsfallet för SMTP följer informationsöverföringen samma väg och sker med samma steg som signaleringen. När väl den initiala änd-till-ändkommunikation är etablerad genom SIP behöver förmedlingsnoderna inte vara kvar i signaleringsflödet. I praktiken är det däremot vanligt att åtminstone några av dessa noder i kedjan stannar kvar i signaleringsflödet, för att på så sätt kunna hantera funktioner som NAT-traversering, samtalsdebitering och upprätthållande av samtalsregelverk.

En kanske mer uppenbar skillnad är förstås är att kommunika-

tionen mellan avsändare och mottagare, åtminstone i användningsfallet telefoni, avses ske i realtid. Begreppet "brevlåda" för knappast tankarna till realtidskommunikation. Istället för att lagra inkommande signaleringsmeddelanden är förmedlingsnodernas uppgift att så snabbt som möjligt vidarebefordra signaleringen till mottagarens enhet (ändnoden).




**Figur 3.1** – SIP-arkitektur

För att förmedlingsnoden ska kunna adressera mottagarens SIP-enhet på nätverksnivå måste information om detta lagras och hållas aktuell i förmedlingsnoden. Den informationen kan vara antingen statiskt konfigurerad (*statisk registrering*) eller uppdateras genom *dynamiska registreringar* från ändnoderna. Alla registreringar är associerade med en giltighetstid. Inom giltighetstiden måste dynamiska registreringar uppdateras av ändnoden, varefter de annars automatiskt avregistreras och blir onåbara för inkommande SIP-signalering.

Inom ramen för SIP finns som framgått ett antal olika funktioner och typer av SIP-noder definierade. Ofta talas lite slarvigt om SIP-klienter och SIP-servrar, men det finns en tämligen strikt definition av olika typer av signaleringsnoder. Samtidigt ska det påpekas att ju mer applikationslogik som tillförs en signaleringsnod, desto mer flyter de olika typerna och funktionerna ihop.

I princip finns tre olika grundtyper av signaleringsnoder (SIP-noder):

- *SIP User Agent* (UAC/UAS)
- *SIP Proxy Server* (Stateless, Stateful, Transaction Stateful)

- *SIP Back-to-Back User Agent* (B2BUA)

### 3.1.1 User Agent

En *SIP User Agent* (*UA*) är en ändnod som även brukar kallas en SIP-klient. Per definition implementerar alltid en UA både en klientsida – *User Agent Client* (UAC) – och en serversida – *User Agent Server* (UAS). Klientsidan är den del som initierar utgående SIP-begäran (*SIP Requests*) och serversidan är den del som tar emot inkommande SIP-begäran och besvarar dem (*SIP Responses*).

### 3.1.2 Proxy

En *SIP Proxy Server* är en förmedlingsnod som agerar mellanhand och kan förekomma i tre olika konfigurationer: *Stateless, Stateful och Transaction Stateful*. Samtliga dessa konfigurationer har sina för- och nackdelar och skillnaden är i huvudsak hur mycket och hur länge förmedlingsnoden är inblandad i SIP-signaleringen under en session, vilket också kan få till följd olika begränsningar gällande vilken logik och vilket tjänsteutbud en sådan förmedlingsnod kan tillhandahålla.

För att se dessa skillnader är det viktigt att förstå två grundläggande begrepp i SIP: dels det som benämns SIP-transaktion och dels det som benämns SIP-dialog. En transaktion består av en SIP-begäran (*request*) och det eller de svar (*response*) som följer av begäran. Det kan till exempel vara en ändnod som avslutar ett samtal och då skickar en *BYE*-begäran. *BYE* är en *request*, och för att *BYE*-transaktionen ska anses slutförd krävs av ändnoden ett svarsmeddelande (förhoppningsvis *200 OK*) som terminerar transaktionen. Om den klient som skickade *BYE*-begäran inte får ett svar inom givna tidsramar kommer klienten att skicka om begäran ett antal gånger. En SIP-transaktion är alltså minnet av en SIP-begäran och dess svar.

En SIP-dialog är ofta konceptuellt enklare att ta till sig, och kan i dess enklaste form likställas med ett samtal. En SIP-dialog etableras i samband med att en SIP-session initieras och avsändande ändnod skickar en SIP-begäran (*INVITE*) till mottagande ändnod, som förväntas svara med ett *200 OK*. Just etableringen av sessionen med *INVITE* kräver en trevägshandskakning på applikationsnivå för att på ett tillförlitligt sätt överföra och förhandla fram de mediaparametrar som krävs. Därför används i samband med just *INVITE*

en särskild SIP-metod kallad *ACK*. *INVITE* är den enda SIP-begäran enligt [RFC3261] som etablerar en SIP-dialog. Denna dialog existerar till dess att någon av ändpunkterna begär att pågående session termineras. Till skillnad från vissa typer av förmedlingsnoder måste en ändnod alltid vara dialog-medveten (*dialog aware*) för att kunna hålla isär flera parallellt pågående dialoger/samtal. Dialogen definieras av ett unikt sessions-ID (*Call-Id*) tillsammans med två etiketter (*Local, Remote* respektive *To, From*) som skapas av avsändande respektive mottagande ändpunkt. Denna trippel identifierar unikt varje SIP-dialog.

En helt tillståndslös (*stateless*) förmedlingsnod har inget "minne", utan varje enskilt SIP-meddelande hanteras oberoende av SIP-transaktioner eller SIP-dialoger. Denna typ av förmedlingsnoder är relativt enkla i sin uppbyggnad. De kan av denna anledning ofta hantera mycket stora mängder SIP-trafik, och lämpar sig att använda som en enkel förmedlare med grundläggande filtrering baserat signaleringen. En förmedlingsnod av denna typ saknar transaktionslager i applikationslogiken, och kan därför till exempel inte förmedla en inkommande SIP-begäran till flera destinationer (så kallad *forking*).

För en tillståndshållande (*stateful*) förmedlingsnod är förutsättningarna annorlunda. En förmedlingsnod av denna typ hanterar både SIP-transaktioner och SIP-dialoger och registrerar både samtalsinitiering, samtalsterminering samt hela den i övrigt pågående sessionen. Det innebär att en tillståndshållande SIP-nod är den mest resurskrävande komponenten i modellen. Inblandningen i hela sessionen har dock sina fördelar från ett applikationslogiskt perspektiv då SIP-noden kontextuellt kan agera på händelser under sessionens gång.

Den tredje varianten, en tillståndshållande förmedlingsnod som enbart agerar på transaktionsnivå (*transaction stateful*), är den vanligaste implementationen av en förmedlingsnod och kan anses vara en bra kompromiss mellan de två tidigare beskrivna varianterna. En transaktionsbaserad tillståndshållande förmedlingsnod är bara inblandad i initieringen av sessionen, eventuella förändringar av sessionen och terminering av sessionen. Däremellan har förmedlingsnoden inget minne av den pågående SIP-dialogen. Detta sparar resurser och möjliggör ändå flertalet applikationstillämpningar, och möjliggör framför allt den tidigare nämnda funktionen som är tämligen unik för SIP, nämligen *forking*. Forking innebär att en förmedlingsnod som håller transaktionstillståndet kan, baserat

på en inkommande SIP-begäran, parallellt skicka utgående SIP-meddelanden till flera potentiella mottagare. I det enklaste fallet har samma användare registrerat flera ändnoder. När inkommande session ska etableras kan transaktionstillståndshållande förmedlingsnoder parallellt kontakta samtliga dessa registrerade ändnoder och koppla sessionen dit användaren svarar först.

### 3.1.3 Back-to-Back User Agent

Den sista typen av SIP-enheter är egentligen en kombination av två ändnoder och benämns *Back-to-Back User Agent* (B2BUA). Precis som namnet antyder handlar detta om två ryggkopplade ändnoder med någon form av applikationslogik knuten till sig. Till skillnad från en förmedlingsnod, som i princip enbart förmår använda informationen i SIP-kommunikationens meddelandehuvuden, agerar en B2BUA på samma sätt som en ändpunkt. En förmedlingsnod agerar endast på meddelandehuvudet, och inte på information i själva meddelandet (*Message body*), beroende på att informationen i denna del enbart är avsedd för ändpunkterna och kan vara krypterad änd-till-änd. Det innebär till exempel också att en förmedlingsnod aldrig kan tolka eller påverka informationen i SDP som skickas i innehållsdelen av SIP-meddelandet. En B2BUA har däremot denna förmåga. Därför är B2BUA en attraktiv och vanlig implementationsform för signaleringsnoder där applikationslogik är av stor vikt, till exempel företagsväxlar, policy- och säkerhetsprodukter (*Session Border Controller* (SBC)) samt konferens- och mediatillämpningar. Det är emellertid viktigt att notera att en B2BUA oftast *inte*, till skillnad från ovanstående beskrivna förmedlingsnoder, är en transparent signaleringskomponent. En B2BUA är snarare, beroende på applikationslogik och konfiguration, en SIP-enhet i vilken man kan tillämpa regelverk baserat på nät- eller tjänsteleverantörens villkor.

### 3.1.4 SDP

SDP[RFC4566] används av bland annat signaleringsprotokollen SIP och *Media Gateway Control Protocol* (MGCP), samt även inom första generationens *Web Real Time Communication* (WebRTC) (se vidare avsnitt 3.4), för att möjliggöra beskrivning och förhandling av mediaspecifika parametrar mellan de inblandade ändpunkterna. Para-

metrar som hanteras av SDP gäller till exempel beskrivning av tillgängliga mediatyper, tillgängliga talkodtyper, vilka IP-adresser och portar som används för olika mediatyper samt en rad andra änd-till-änd-attribut. SDP i sig innehåller inte någon förhandlingmekanism för mediaparametrar, utan medger endast en grundläggande textbaserad beskrivning av vad en viss ändpunkt stöder.

I samband med dubbelriktad sessionsetablering måste dock ändpunkterna kunna förhandla fram lämpliga media- och talkodtyper som ska användas för sessionen. För att göra detta med SIP använder man ett eller flera utbyten av SDP-meddelanden mellan ändpunkterna, där ena ändpunkten i första steget föreslår en uppsättning parametrar (*SDP Offer*). Den andra ändpunkten väljer, baserat på den föreslagna uppsättningen, en eller flera media- och talkodtyper som båda ändpunkter har förmåga att hantera. Hur detta utbyte kan göras med just SIP beskrivs i [RFC3264]. Det vanligaste förfarandet är att A-parten inkluderar föreslagen SDP-information vid initial sessionsetablering och att B-parten inkluderar besvarad SDP-information i samband med att B-parten svarar. Det finns emellertid flera alternativa förfaranden och regler som kan vara av betydelse gällande utbyte av SDP-information mellan ändpunkterna med SIP. Dessa finns beskrivna i [RFC3264] och dess uppdateringar.

Förutom att överföra sessionsbeskrivning gällande media- och talkodtyper kan SDP användas till att bära information mellan ändpunkter för en mängd tilläggstjänster, till exempel typ av DTMF-signalering, kapacitetskrav och nycklar för mediakryptering, samt ligga till grund för påförande av QoS-parameterar.

## 3.2 H.323

H.323 är en paraplystandard framtagen inom ITU-T och liknas ofta vid *ISDN över IP*. H.323 är inte direkt jämförligt med andra signaleringsprotokoll som till exempel SIP och MGCP, då H.323-standarden innefattar mer än enbart signaleringsprotokollet. De två grundläggande delstandarderna som är av störst vikt inom detta område är dels H.225.0, som motsvarar signaleringsprotokollet, och H.245 som motsvarar mediaförhandlingen mellan H.323-kompatibla ändpunkter. Vid en grov jämförelse kan H.225.0 alltså sägas motsvara SIP och H.245 motsvara SDP, då de i stort fyller motsvarande funktioner. Likt SIP och MGCP nyttjar även H.323 RTP som applika-

tionsprotokoll för att överföra mediaströmmarna mellan ändpunkter. H.323 är definierat som ett binärt protokoll och använder *Abstract Syntax Notation One* (ASN.1)-syntax.

H.323 publicerades första gången 1996 med grundtanken att användas för tal- och videokommunikation över lokala paketförmedlade nätverk. Protokollet kom dock snabbt att användas av produkt- och tjänsteleverantörer för att leverera IP-telefoni över globala IP-baserade nät. Applikationer som *Netmeeting* kom att få stor spridning, och H.323 blev det första riktiga standardiserade protokollet för telefoni över IP. Till H.323-svitens fördel hör att standarden tidigt definierade inte bara vilka protokoll som skulle användas utan även vilka tjänster som ingick och hur de skulle implementeras. H.323 blev därmed attraktivt att implementera för tillverkare av IP-baserad telefoniutrustning, med syftet att kunna ersätta befintliga *Public Branch eXchange* (PBX)-system och tjänster med en IP-baserad lösning. H.323 var även det första signaleringsprotokollet som definerade användandet av IETF-protokollet RTP för överföring av mediaströmmar.

På liknande sätt som SIP definierar H.323 ett antal olika komponenter och nätelement. Dessa innefattar både signalerings- och medianoder då H.323 är en större övergripande standard än SIP, som lite skämtsamt brukar sägas innefatta allting, inkluderande färgen på H.323-telefonen på skrivbordet.

De fyra komponenter som vanligen förekommer i H.323-strukturen är:

- terminal,
- förmedlingsnod (*Gateway*),
- *Multipoint Control Unit* (MCU), samt
- *Gatekeeper*.

De tre förstnämnda, terminaler, förmedlingsnoder och MCU refereras ofta till som ändpunkter medan en *Gatekeeper* är ett nätelement med liknande funktionalitet som en SIP-proxy. På liknande sätt som i fallet med SIP kan två H.323-terminaler kommunicera direkt med varandra förutsatt att de har den adresseringsinformation som krävs för att kunna etablera en session mellan varandra. I en vanlig H.323-installation tillförs emellertid åtminstone en *Gatekeeper* för att hantera adressering- och autentiseringsfunktioner.

*H.323 terminal*

En terminal är i dess enklaste form en IP-telefon eller mjukvarubaserad klientprogramvara i en dator. Likt SIP-ändpunkter kan en H.323-terminal hantera ett stort antal mediatyper och därmed kan en H.323-ändpunkt vara allt ifrån just denna enkla telefon till ett komplett videokonferenssystem. Arkitekturmässigt är det samma sak. En H.323-ändpunkt implementerar hela eller delar av den definierade H.323-protokollstacken som är betydligt mer definierad och standardiserad än i SIP. En H.323-ändpunkt måste implementera åtminstone de i H.323-standarden definierade signalerings- och mediahanteringsprotokollen *H.225.0* och *H.245* (se avsnitt 3.2.1 och 3.2.2). Utöver dessa krävs även att lämpliga data- och mediaöverföringsprotokoll implementeras, till exempel *V.150* (för dataöverföring), *T38* (för faxöverföring), *Real-time Transport Protocol* (RTP) och *Real-time Control Protocol* (RTCP). En terminal kan registrera sig till en eller flera *Gatekeepers* för att tillkännage sin tillgänglighetsstatus.

*H.323 förmedlingsnod*

En förmedlingsnod skiljer sig principiellt inte nämnvärt från övriga bryggningsenheter beskrivna för andra signalerings- och mediaprotokoll. När en H.323-ändpunkt ska kommunicera med någon abonnent eller mottagare som antingen inte finns nåbar via samma IP-nät eller som använder sig av ett annat signaleringsprotokoll, måste en förmedlingsnod användas för att utföra konverterings- och bryggningsfunktioner mellan de olika näten eller protokollen. En mycket vanlig förmedlingsnod kopplar samman ett lokalt H.323-baserat företagsnät med PSTN via till exempel en kretskopplad *Integrated Services Digital Network* (ISDN) *Primary Rate Interface* (PRI)-trunk. Alternativt kan en förmedlingsnod översätta till ett annat IP-baserat signaleringsprotokoll, till exempel SIP, för att förmedla samtalet vidare till PSTN via en SIP-trunk. Från H.323-perspektiv agerar en förmedlingsnod, likt en terminal, ändpunkt. En förmedlingsnod kan också registrera sig till en eller flera *Gatekeepers* för att tillkännage sin tillgänglighetsstatus.

*H.323 MCU*

En *MCU* är i princip en konferensbrygga eller mediaserver som har förmåga att fläta samman flertalet parallella ljud- och video-

strömmar från och till flera olika deltagare. Det innebär att varje deltagande terminal i konferensen enbart får en färdigblandad ljud- och videoström som representerar samtliga deltagare, istället för att mottagande ändpunkt ska behöva ta emot samtliga media- och dataströmmar och på egen hand hantera dessa. Vissa implementationer av MCU har även stöd för att blanda och förmedla datakanaler mellan aktiva deltagare så att olika dataöverföringbaserade konferenstjänster kan erbjudas. En MCU agerar också från ett H.323-perspektiv som en ändpunkt.

*Gatekeeper*

Som tidigare nämnts kan två H.323-terminaler (*ändpunkter*) protokollmässigt kommunicera och etablera sessioner direkt med varandra utan inblandning av några andra komponenter. För att förenkla adressering och för att inblandade H.323-komponenter ska kunna ansluta till varandra, används dock i praktiken nästan alltid åtminstone en *Gatekeeper* i kommunikationsflödet. De vanligaste tjänster som en *Gatekeeper* tillhandahåller är registrering av ändpunkter, autentisering, adressering och auktorisation (*admission control*). Likt en SIP-proxy kan en *Gatekeeper* agera i ett mer eller mindre tillståndslöst läge. Antingen är en *Gatekeeper* enbart inblandad i initieringen av sessionen, då ändpunkten använder utökningen av signaleringsprotokollet H.225.0, *Registration, Admission and Status* (RAS), för att begära information från en *Gatekeeper* gällande motparten. Denna *Gatekeeper* förmedlar sedan dessa kontaktuppgifter till uppringande terminal och det är upp till terminalen att själv etablera sessionen direkt med motparten. Alternativt stannar denna *Gatekeeper* kvar i signaleringsflödet under hela sessionen, vilket medför att den har full kontroll över hela sessionshanteringen. RAS används av både ändpunkter och *Gatekeepers*, för att kommunicera både med varandra och mellan olika *Gatekeepers*. De ändpunkter som registrerar sig till en viss *Gatekeeper* anses tillhöra samma *zon*, vilket innebär att en *Gatekeeper* är administrativt ansvarig för ändpunkterna inom en viss zon. Olika zoner kan sammankopplas och samtalsdirigering mellan dem styrs i vanliga fall med statiskt konfigurerade prefix- och nummerplaner i respektive *Gatekeeper*.

*Standardisering*

Intresset för H.323-standarden har med tiden klingat av och utvecklingen i någon mån avstannat. Den senaste versionen av standarden är från 2009. Till viss del kan detta förklaras av den rigorösa och därmed långsamma publiceringsprocess som råder inom standardiseringsorganet ITU-T. Men samtidigt har andra betydelsefulla standardiseringsorgan som *3rd Generation Partnership Project* (3GPP) och *European Telecommuniations Standards Institute* (ETSI) valt att för nästa generations fast- och mobiltelefoninät grunda sina specifikationer på SIP-standarden, vilket minskat incitamenten för vidareutveckling av H.323.

Den strikta styrningen av H.323-standarden och dess innehåll, som ledde till genombrottet för IP-baserad telefoni, kan paradoxalt nog också tillskrivas orsaken till att standarden är på väg att marginaliseras. Den anses i många fall inte vara flexibel nog för att kunna följa utvecklingen och möta framtidens krav på IP-telefoni och relaterade tjänster. H.323 fortsätter vara ett protokoll som ofta stöds av framför allt IP-baserade företagsväxlar.

### 3.2.1 H.225.0

H.225.0 är sessionssignaleringsprotokollet i H.323-sviten. Som redan nämnts är detta den del av H.323 som är ungefär jämförlig med SIP. I princip är H.225.0 protokollet Q.931 över IP. Q.931 är det signaleringsprotokoll som ofta används i kretskopplade ISDN-anslutningar med meddelandetyper som *Setup, Call Proceeding, Connect, Alerting, Informational,* och *Release Complete*. H.225.0 används av ändpunkter för att etablera och terminera en session och signaleringen kan skickas direkt mellan ändpunkterna eller via en *Gatekeeper*, beroende på den konfiguration som tillämpas.

RAS är som framgått den del av H.225.0 som används för signalering med *Gatekeepers*. Protokollet används av ändpunkter för att registrera dessa till en viss *Gatekeeper*, eller för kommunikation mellan *Gatekeepers*, och innehåller funktioner för auktorisation (*admission control*), samt adresserings och kapacitetsreservationsbegäran.

Registreringar används för att informera en *Gatekeeper* om hur en viss ändpunkt kan nås, så att inkommande samtal kan etableras. *Gatekeeper*en håller alltså information hur mottagaren av inkommande samtal kan adresseras på nätverksnivå. Auktorisation används för

att efterfråga adresseringsinformation och behörighet att etablera en utgående session mot efterfrågad adress. Även mottagande ändpunkt kommunicerar med dess *Gatekeeper* innan samtalet accepteras, i syfte att verifiera att ändpunkten får acceptera samtalet. Kapacitetsförfrågningar kan göras för att låta lämpliga reserveringsmekanismer i nätet, till exempel *Resource Reservation Protocol* (RSVP) [RFC2205], boka kapacitet för överföringen av önskat samtal. Denna typ av reservation fungerar dock i praktiken enbart inom en och samma administrativa domän, då det kräver stöd i hela IP-nätinfrastrukturen. RSVP fungerar således inte änd-till-änd över ett nät inkluderande flertalet av varandra oberoende autonoma system, som till exempel internet.

### 3.2.2 H.245

Likt SIP använder sig H.323-ändpunkter av ytterligare ett protokoll för att hantera och förhandla de mediaströmmar som den etablerade sessionen ska inkludera. H.245 är H.323-svitens motsvarighet till SDP i SIP och MGCP. När H.225.0 har används för att initiera sessionen kan de inblandade ändpunkterna i sessionen börja utbyta H.245-information för att indikera vilka media- (*ljud, bild, text och data*) och talkodtyper som respektive ändpunkt stöder. Likt SIP och MGCP hanterar H.323-sviten ett flertal olika tal- och videokodtyper. Även om det inte är unikt för H.323, då även SIP har mediatyper för data- och textöverföring, finns standardiserat stöd för mer avancerade konferens- och datatjänster innefattande skärmdelning, elektronisk tavla, filöverföring och textmeddelanden. Dessa tjänster ryms inom T.120-standarden, och är en definierad delmängd inom H.323-sviten.

Till skillnad från hur förhandling av media- och talkodtyper fungerar med SIP, där *SDP Offer* och *SDP Answer* används för att utbyta nödvändig information, etablerar de kommunicerande H.323-ändpunkterna ett primär/sekundär-förhållande där det är primärändpunkten som styr valet av media- och talkodtyp baserat på tillgängliga gemensamma alternativ. De alternativ som finns tillgängliga har indikerats och utbytts mellan respektive ändpunkt under inledningen av H.245-kommunikationen. När väl dessa steg är genomförda kan ändpunkterna öppna logiska kanaler sinsemellan och etablera önskade mediasessioner.

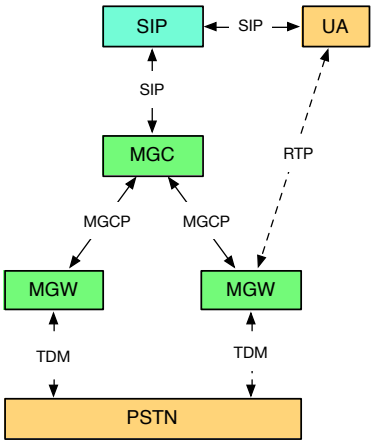
### 3.2.3 H.235.6

H.235.6 definierar en H.323-profil för krypterat tal mellan olika H.323-ändpunkter. Profilen innefattar även hur nyckelhantering ska ske och definierar vidare ett antal olika krypteringsalgoritmer, inkluderat AES. I ett första skede utbyter H.323-ändpunkterna en delad hemlighet genom ett underhandsförfarande baserat på Diffie-Hellman via H.225-signaleringen. Denna delade hemlighet används sedan för att framställa nycklar för kryptering av mediaströmmen. Detta utbyte sker via H.245-signaleringen och nycklarna används sedan för att kryptera mediaströmmen på motsvarande sätt som för SRTP (se avsnitt 4.2). Eftersom nyckelframställningen sker oautentiserat bör H.225-signaleringen skyddas mot janusangrepp genom lämpligt transportskydd, till exempel *Transport Layer Security* (TLS) eller *Internet Protocol Security* (IPsec).

## 3.3 MGCP

*Media Gateway Control Protocol* (MGCP) är ett textbaserat signaleringsprotokoll som till skillnad från SIP och H.323 bygger på en starkt centraliserad modell. Grundarkitekturen är definierad i [RFC3435] och beskriver en separation av signalerings- och medianoder. Enskilda signaleringsnoder, *Media Gateway Controller* (MGC), kan styra en stor mängd distribuerade medianoder, *Media Gateway* (MGW), vilket medför att MGCP-arkitekturen är tämligen lik de mer traditionella kretskopplade strukturerna i PSTN och andra TDM-baserade telefoninät. MGCP använder SDP för att, på liknande sätt som SIP, förmedla information om tillgängliga mediatyper, tal- och videokodtyper, samt portar och IP-adresser för kommunikation. Användandet av SDP förenklar integration av MGCP-baserade lösningar med SIP-baserade lösningar, då SDP kan föras transparent mellan de olika signaleringsprotokollen. MGCP nyttjar även RTP som applikationsprotokoll för överföring av förhandlade mediaströmmar.

Till skillnad från SIP-baserade IP-telefonilösningar, där samtals- och sessionshanteringslogiken implementeras i ändpunkterna, implementerar MGCP en strikt centraliserad sessionshantering. En *Media Gateway* är helt och hållet kontrollerad av en signaleringskontrollnod, även kallad *Softswitch*, där all logik gällande samtalshantering finns koncentrerad. En MGCP-ändpunkt är med andra ord relativt enkel.




**Figur 3.2** – MGCP-arkitektur

En vanlig tillämpning där MGCP används är sammankopplingslösningar av olika slag, till exempel mellan SIP och SS7/ISUP. En sådan lösning kan då konstrueras med en central signaleringsnod (MGC) för kontroll och samtalshantering, men med en distribuerad utbyggnad av multipla medianoder (MGW) för att på bästa sätt ansluta och överföra mediatrafiken på lämpliga platser. Signaleringsnoden kan kommunicera med övriga telefonikomponenter med SIP, men kontrollerar multipla medianoder med hjälp av MGCP. Från ett SIP-perspektiv agerar signaleringsnoden ändnod (*User Agent*), men från ett MGCP-perspektiv agerar signaleringsnoden MGC (*Media Gateway Controller*). Detta medför att en ändnod i detta trafikfall kommunicerar över SIP med signaleringskontrollnoden, men skickar och tar emot mediaströmmar från någon av de medianoder som denna signaleringskontrollnod hanterar och styr över. Vilken medianod som ska användas avgörs genom logiken i signaleringskontrollnoden. Ett sådant val kan göras på baserat på en mängd olika kriterier, men vanligt är detta sker på grundval av A- eller B-nummer.

MGCP är i första hand en arkitektur även om [RFC3435] också definierar det specifika protokollet. Det finns, förutom MGCP, ett mycket snarlikt protokoll som utarbetades som ett samarbete mellan IETF och ITU-T. Detta protokoll är baserat på samma arkitekturdokument som MGCP men är inte kompatibelt, bland annat beroende

på annat meddelandesyntax. Detta protokoll definieras i [RFC3525] och har IETF-namnet *MEdia GAteway COntroll Protocol* (MEGACO), medan dess ITU-T benämning är H.248.

## 3.4 Internetapplikationer och signaleringsprotokoll

Utöver de standardiserade och öppna protokolluppsättningar som beskrivits ovan förekommer flertalet mer eller mindre stängda VoIP-tjänster på internet, som i många fall helt eller delvis använder dessa protokoll i grunden.

Exempel innefattar *Apple Facetime* och *Viber*. Facetime använder sig till stor del av SIP som signaleringsprotokoll medan Viber använder en helt proprietär lösning för signalering och talkodning.

*Skype*

En av de mest tongivande aktörerna inom VoIP-tjänster är *Skype* som med smarta talkodtyper och särskilda signaleringsfunktioner tar sig igenom de flesta nät och brandväggar med gott resultat. Då all signalering och media är krypterad i Skype är det svårt att finna information om hur signaleringen fungerar i detalj, men i princip har Skype lagt grunden till de traverseringstekniker som numera är standardiserade inom IETF, nämligen tekniker som *Session Traversal Utilities for NAT* (STUN), *Traversal Using Relays around NAT* (TURN) och *Interactive Connectivity Establishment* (ICE). Skype bidrog även till arbetet med standardiseringen av den första talkodtypen inom IETF, OPUS.

Inledningsvis var Skype baserat på *peer-to-peer*-teknik där alla användare som installerade en Skype-klient blev en del av nätverket. Beroende på internetanslutningens karakteristik (konnektivitet genom publika IP-adresser, tillgänglig kapacitet, med mera) så blev noden antingen ansluten som en löv-nod till en eller flera *supernoder*, eller så blev noden själv en supernod. En supernod förmedlar trafik från andra noder och behöver ofta både ha publika IP-adresser och god tillgång till nätresurser. En supernod ankrar till exempel mediaflöden från andra noder som befinner sig bakom brandväggar för att på så viss få media och signalering att fungera änd-till-änd mellan samtliga noder i nätverket. För att behålla tillfredställande kvalitet och kapacitet på det överlagrade Skype-nätet driver Skype själva supernoder. Skype ägs sedan 2011 av *Microsoft*.

*WebRTC*

WebRTC är ett tämligen nytt initiativ med målet att integrera realtidsmedia, så som ljud och video, i webbläsaren utan att behöva använda insticksprogram. WebRTC är ett standardiseringssamarbete mellan IETF och *World Wide Web Consortium* (W3C), där IETF ansvarar för de protokollrelaterade delarna genom IETF-gruppen *rtcweb*, och W3C definierar de API:er som webbutverklare avses använda sig av. WebRTC utvecklades inledningsvis av *Google* men öppnades 2011 upp för samtliga intressenter att ta del av och bidra till dess utveckling, och därmed påbörjades även standardiseringsarbetet. Detta arbete är pågående och flertalet implementationer av olika delar av WebRTC-teknikerna används redan på internet.

I första hand definierar WebRTC tre olika API:er. Dessa API:er tillåter en webbutvecklare att på ett enkelt sätt få tillgång till en enhets mikrofon och kamera, och för att kunna etablera en direktkontakt mellan två olika kommunicerande parter. Det är viktigt att notera att WebRTC *inte* definierar något signaleringsprotokoll, utan bör snarare betraktas som ett ramverk för att etablera realtidskommunikation änd-till-änd mellan webbläsare som agerar klienter.

Det finns emellertid möjlighet att använda SIP för signalering inom ramen för WebRTC. SIP använder då websocket[RFC6455] för transport. De SIP-specifika områdena vid användandet av websocket som transport finns definierade i [RFC7118]. Att använda SIP som signaleringsprotokoll för WebRTC har sina fördelar i vissa typer av tillämpningar, till exempel då andra IP-telefontjänster och traditionella telefoniabonnenter ska integreras i samtal med WebRTC-baserade klienter. Används då SIP över websocket som signaleringsprotokoll är det tämligen enkelt överbrygga signaleringen till andra transportprotokoll för att kommunicera med andra typer av SIP-ändpunkter.

Andra exempel på signaleringsprotokoll som kan användas för WebRTC är *Extensible Messaging and Presence Protocol* (XMPP), som lämpar sig väl för tillämpningar där direktmeddelanden (*Instant Messaging*) och tillgänglighetsstatus (*presence*) är av stor betydelse.

I andra tillämpningar där interoperabilitet med annan telefoni eller direktmeddelanden inte är prioriterat kan dock SIP eller XMPP som signaleringsprotokoll vara onödigt komplext. Verkligheten har visat genom en smärre uppsjö av mer eller mindre standardiserade lösningar, att en mer strömlinjeformad metod för signalering,

anpassad för just WebRTC, kan vara att föredra.

I samband med att standardiseringsarbetet för WebRTC inleddes, gjordes även ett antal principiella antaganden som syftade till att höja skyddsgraden för realtidskommunikation. Till exempel är det ett krav att alla mediaströmmar ska vara krypterade med SRTP och använda DTLS för nyckelhanteringen. Detta innebär att nyckeluppsättningar för mediakrypteringen aldrig skickas via några mellanhänder, utan etableras direkt mellan de kommunicerade parterna.

Man har även inom ramen för standarden definierat talkodtypen OPUS som obligatorisk att implementera i klientdelarna. I skrivande stund finns inga obligatoriska videokodtyper definierade. Kandidaterna för video är i första hand VP8 och H.264, men det finns argument både för och emot att göra någon eller båda dessa obligatoriska. Diskussionen rör bland annat hårdvarustöd samt patent och andra immateriella rättigheter.

# 4 Mediaöverföring

Gemensamt för i princip samtliga IP-telefoni- och VoIP-protokoll är att signalering och mediaöverföring hanteras separat. Detta innebär att det är olika protokoll som hanterar dessa funktioner. Samtalssignalering måste i många fall, antingen beroende på protokollets utformning eller beroende på tjänstens utformning, traversera flertalet signaleringsnoder på vägen mellan avsändare och mottagare av samtalet. När det gäller mediaöverföring är det dock oftast önskvärt att kommunikationen går direkt mellan de kommunicerande parterna, i första hand för att minimera fördröjningseffekter. I vissa fall är det emellertid inte möjligt att skicka mediaströmmen direkt mellan de kommunicerande parterna. Detta kan dels ha tekniska orsaker, till exempel adressöversättning *Network Address Translation* (NAT), brandväggar eller behov av omkodning (*transcoding*), men kan också drivas av rättsliga krav på möjlighet till hemlig teleavlyssning och dekryptering.

För signalering har redan ett antal olika alternativa protokoll presenterats. Flertalet av dessa signaleringsprotokoll använder emellertid ett och samma protokoll för mediaöverföring; RTP.

## 4.1 RTP

*Real-time Transport Protocol* (RTP) [RFC3550] är ett applikationsprotokoll som använder *User Datagram Protocol* (UDP) som underliggande transportprotokoll.

I varje RTP-paket transporteras ett antal tidsbundna mediafragment. Beroende på val av talkodtyp (Coder Decoder *(CODEC)*) och vilket paketeringsintervall som används, representerar varje RTP-paket en viss tidsenhet av mediaströmmen. Att avgöra vilket paketeringsintervall som ska användas är en avvägning mellan att minimera andelen överskottsdata (*overhead*) genom att skicka längre medi-

afragment i varje RTP-paket, och att minimiera fördröjningen mellan de kommunicerande parterna genom att använda mindre RTP-paket innehållande kortare mediafragment.

Ökas paketeringsintervallet ökar fördröjningen då avsändaren måste invänta inspelningen av hela mediafragmentet innan det kan kodas och förmedlas till mottagaren. Risken för bitfel som resulterar i att ett RTP-paket måste kasseras ökar med paketets storlek, samtidigt som konsekvenserna av ett förlorat paket innebär att ett längre mediafragment gick förlorat. Olika tal- och videokodtyper är olika känsliga för förluster av mediafragment. Vissa har förmåga att återskapa mediaströmmen även med relativt hög andel förlorad information, medan andra är mer känsliga.

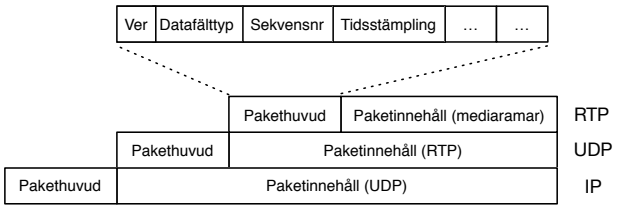
Som redan nämnts använder RTP UDP som underliggande transportprotokoll. Detta innebär att RTP skickas som datagram och att det inte finns någon inbyggd mekanism i transportprotokollet att hantera omsändning av förlorade paket. Eftersom RTP i normala fall, särskilt i tillämpningen telefoni, hanterar dubbelriktade realtidsströmmar av media, är det föga intressant att skicka om ett tappat RTP-paket. Ett tappat paket är ett förlorat paket.

RTP tillför ett antal parametrar och funktioner som inte finns i UDP. Dessa finns i protokollets pakethuvud. De viktigaste funktionerna som RTP tillför UDP är:

- datafält/talkodtyp,
- sekvensnumrering, samt
- tidsstämpling.

Tack vare denna information kan mottagaren av RTP-paketet avgöra vilken tal- eller videokodtyp som används för kodning av den information som finns i RTP-paketets datafält (*payload*), så att applikationen kan välja rätt kodtyp för avkodning. För de statiskt definerade datafältstyper som är definierade i [RFC3550] finns en direkt koppling mellan datafältsvärde och den tal- eller videokodtyp som används. För dynamiskt definierade datafältstyper saknas denna direkta koppling till använd talkodtyp och mottagande applikationer måste ha tillgång till ytterligare information som beskriver kopplingen mellan dynamisk datafälts- och talkodtyp för varje givet tillfälle. Denna information överförs i vanliga fall mellan ändpunkterna i samband med

samtalsetablering och mediaförhandling, till exempel som ett mediaattribut i SDP.




**Figur 4.1** – RTP-struktur

Sekvensnumreringen används för att säkerställa att mottagna RTP-paket, med hjälp av en jitterbuffert, kan spelas upp i rätt ordningsföljd, och tidsstämplingen kan användas av applikationen för att spela upp mottagna tidsfragment med rätt intervall.

## 4.2 SRTP

*Secure RTP* (SRTP) [RFC3711] tillför äkthetskontroll och sekretesskydd genom kryptografiska mekanismer och skyddar på så vis mediaströmmens innehåll mot ett flertal olika typer av hot och avlyssningsmöjligheter. SRTP förutsätter också ett nyckelhanteringsprotokoll, till exempel *SDP Security Descriptions* (SDES) [RFC4568], *Multimedia Internet KEYing* (MIKEY) [RFC3830] eller *Datagram Transport Layer Security* (DTLS) [RFC4347], för att utbyta den nyckelinformation som SRTP använder för att kryptera och autentisera varje paket i mediaströmmen mellan ändnoderna. Nyckelhanteringsprotokollet är inte definierat inom ramen för SRTP, utan sker via någon form av utombandssignalering. Valet av nyckelhanteringsprotokoll är således av största vikt för att uppnå erforderlig kryptografisk skyddsnivå (se vidare avsnitt 7).

Krypteringsfunktionerna med SRTP är inte är avhängigt av vald tal- eller videokodtyp. Det innebär att oavsett vilken talkodtyp som används så kan en och samma krypterings- och autentiseringsteknik användas på samtliga mediaströmmar. SRTP påför särskilt små mängder överskottsdata och lämpar sig därför väl för att sekretesskydda och autentisera mediaöverföringar av såväl låg som hög bitströmshastighet, inklusive videoströmmar.

## 4.3 RTCP

För varje RTP-paketerad mediaström kan en tillhörande kontrollkanal enligt protokollet *Real-time Control Protocol* (RTCP)[RFC3550] användas. Om mediaströmmen använder SRTP ska även den krypterade och autentiserade varianten av RTCP, *Secure RTCP* (SRTCP) [RFC3711] användas för detta.

Den primära funktionen för RTCP är att förmedla kvalitetsparametrar, *Quality of Service* (QoS), mellan de kommunicerande ändpunkterna. RTCP tillhandahåller änd-till-änd-relaterad information om den mediaström som RTCP-strömmen är associerad med. De värden som RTCP rapporterar innefattar bland annat jitter, fördröjning och förlorade paket. Applikationen i respektive ändpunkt kan med stöd av denna information göra aktiva val gällande till exempel byte eller anpassning av aktuell tal- eller videokodtyp.

RTP använder vanligtvis jämna UDP-portnummer för att kommunicera. RTCP nyttjar i normalfallet det udda UDP-portnumret närmast över dess associerade RTP-ström. Alternativt kan portnummer för RTCP-strömmen förhandlas fram under sessionsetableringen. Till skillnad från RTP innehåller RTCP även en absolut tidsstämpling. Detta gör att applikationer kan använda RTCP för att synkronisera multipla mediaströmmar.

För att möjliggöra utökad rapportering av kvalitetsparametrar kan *RTCP Extended Reports* (RTCP XR)[RFC3611] användas. RTCP XR är en utökning av RTCP, och definierar ytterligare ett antal paketformat och rapporter. Stöd för RTCP XR förhandlas fram mellan ändpunkterna vid sessionsetablering. RTCP XR tillför bland annat detaljerad information gällande samtalskvalitet samt utökad statistik gällande pakethantering, fördröjning och jitter.

## 4.4 Översikt av talkodtyper

Då fokus för denna handledning rör just tal behandlar detta avsnitt enbart några av de talkodtyper som är mest vanligt förekommande i denna tillämpning. Videokodtyper faller således utanför avgränsningarna för handledningen. Generellt kan talkodtyper indelas i två grupper: TDM-optimerade (statiska) och IP-optimerade (dynamiska).

De statiska talkodartyperna har en fast bitströmshastighet och mediaströmmen utgörs av kontinuerliga mätvärden (*sample based*

*encoding*). De förutsätter att den överföringskapacitet som krävs för mediaströmmen alltid finns tillgänglig i kommunikationsnätet. Detta fungerar väl i traditionella kretskopplade telefonisystem, där den överföringskapacitet som krävs allokeras i samband med samtalsetableringen. Statiska talkodtyper gör det också enkelt att beräkna den överföringskapacitet som krävs i kommunikationsnätet för att hantera ett visst antal samtidiga samtal över specifika anslutningar. Då förluster av mediafragment är relativt ovanliga i sådana synkrona kretskopplade kommunikationsnät är dessa talkodtyper sällan anpassade för att hantera någon större andel förlorad information. Redan vid tämligen blygsamma andelar paketförluster kan mediaströmmen blir svår att återskapa med godtagbar kvalitet.

Eftersom det paketförmedlade kommunikationsnätets kvalitetsegenskaper över tid kan variera på ett betydligt mer påtagligt sätt än motsvarande kretskopplade kommunikationsnät, kan användning av sådana statiska talkodtyper i paketförmedlade nät därför medföra kvalitetsproblem.

Andra talkodtyper som är särskilt framtagna för att användas i paketförmedlade nät har ofta en dynamisk karaktär vilket innebär att de över tid kan anpassa mediaströmmens kvalitetsegenskaper efter rådande omständigheter i det paketförmedlade nätet. Detta har givetvis sina gränser, men dynamiska talkodtyper tillsammans med olika tekniker för att hantera större mängder förlorade paket kan göra det möjligt att avkoda mediaströmmarna med godtagbar kvalitet, även vid brist på överföringskapacitet eller i anslutningsformer där betydande mängder paketförluster är ofrånkomliga.

Till skillnad från de statiska talkodtyperna, där bitströmmen utgörs av kontinuerliga mätvärden, kodar dynamiska komprimeringsalgoritmer som regel informationen i ett visst tidsintervall blockvis (*frame based encoding*). Blockstorleken hänger alltså samman med och påverkar paketeringsintervallet, och får således också en direkt påverkan på till exempel RTP-paketets minsta möjliga storlek samt på vilket sätt eventuell ompaketering kan göras utan omkodning (se vidare avsnitt 4.8).

### 4.4.1 G.711

[G.711] är en ITU-T-standardiserad talkodtyp och är den absolut vanligaste kodtypen för tal i traditionella telefoninät. G.711 är

en smalbandig okomprimerad talkodtyp och är optimerad för att hantera och representera just tal i frekvensomfånget 300-3400 Hz. Den analoga signalen mäts och kvantiseras till ett 8-bitarsvärde med taktfrekvensen 8 kHz, innebärande att G.711 producerar en konstant bitströmshastighet på 64 kbps.

Två olika varianter av G.711 förekommer: *A-law* och *μ-Law*. *μ-Law* används i första hand i USA och *A-law* är den variant som är vanligast förekommande i resterande delar av världen, Sverige inkluderat. Skillnaden mellan *A-* och *μ-law* ligger i första hand i kvantiseringen av olika frekvensområden, där *μ-law* ger något högre upplösning i de högre frekvensområdena medan *A-law* ger något högre upplösning i de lägre frekvensområdena.

Då fortfarande i princip hela det traditionella PSTN-systemet använder sig av G.711 är det naturligt att denna talkodtyp även förekommer i samband med telefoni som transporteras över IP-baserade kommunikationsnät. Merparten av världens telefoniabonnenter är fortfarande anslutna via *Public Switched Telephony Network* (PSTN), och använder då per automatik G.711. Om inte samma talkodtyp används eller stöds av båda kommunicerande parter måste någon, i vanliga fall telefonioperatören, omkoda mediaströmmarna så att olika talkodtyper används av den paketförmedlade respektive kretskopplade sidan av samtalet. Storskalig omkodning är i många fall en relativt resurskrävande och dyr process, varför G.711 även ofta används på den paketförmedlade sidan av samtalet.

### 4.4.2 G.722

[G.722] är också en talkodtyp standardiserad av ITU-T. Till skillnad från G.711 är G.722 en bredbandig talkodtyp som hanterar ett betydligt vidare frekvensomfång med en markant bättre ljudupplevelse för lyssnaren som resultat. G.722 är standardiserad för tre olika bitströmshastigheter: 48, 56 och 64 kbps, där 64 kbps är den hastighet som i praktiken används. Detta betyder att G.722 jämfört med G.711 levererar väsentligt bättre talkvalitet med samma bitströmshastighet. G.722 omfattar frekvensomfånget 50-7000 Hz och varje mätvärde av den analoga signalen kvantiseras till ett 16-bitarsvärde med taktfrekvensen 7 kHz, innebärande både ett bredare frekvensomgång och högre upplösning än G.711.

Inom IP-telefoni talas ibland om begreppet *HD Voice*, där HD står

för *High Definition*. G.722 är en av de talkodtyper som innefattas av begreppet, och i kommersiellt tillgänglig hård- och mjukvara för IP-telefoni är det vanligen G.722 som stöds om produkten är märkt *HD Voice*.

### 4.4.3 G.729

[G.729] är ytterligare en talkodtyp standardiserad av ITU-T. Till skillnad från G.711 är G.729 en smalbandig och komprimerande talkodtyp som kräver väsentligt lägre bitströmshastighet; omkring 8 kbps. G.729 är tack vare sin låga bitströmshastighet en av de talkodtyper som är bäst lämpade att användas i miljöer där det förekommer starka begränsningar i anslutningars dataöverföringskapacitet, och där en försämrad talkvalitet är en acceptabel avvägning till förmån för till exempel ett större antal samtidigt pågående samtal.

Det finns ett flertal varianter, eller utökningar (*Annex*), till G.729. Dessa olika Annex beskriver olika tillägg eller extrafunktioner. De vanligast förekommande varianterna är G.729a, G.729b och G.729d. G.729a är en förenklad variant som kräver mindre beräkningskraft men gör också ytterligare avkall på ljudkvaliteten. G.729b tillför funktioner som taldetektering, *Voice Activity Detection* (VAD), och bakgrundbrusframställning, *Comfort Noise Generation* (CNG). G.729d tillför möjlighet att använda en lägre bitströmshastighet om 6.4 kbps.

Då G.729 är en starkt komprimerande talkodtyp är det inte lämpligt att använda inombandssignalering, så som DTMF och fax-toner, då kodning och återskapande av dessa inte kan garanteras med tillräcklig kvalitet. Används G.729 eller andra starkt komprimerande talkodtyper för tal, bör DTMF och fax-toner skickas utanför ljudströmmen, till exempel som RTP-händelser enligt [RFC4733].

### 4.4.4 OPUS

Till skillnad från ovan beskrivna talkodtyper är OPUS framtagen och standardiserad av *Internet Engineering Task Force* (IETF) genom [RFC6716], och är en talkodtyp som lämpar sig väl för realtidskommunikation över internet och andra IP-baserade kommunikationsnätverk. OPUS är en dynamisk talkodtyp vilket innebär att den har förmåga att anpassa bitströmshastighet och kodningsalgoritmer

baserat på rådande omständigheter i underliggande kommunikationsnät. OPUS stöder bitströmshastighet från så lågt som 6 kbps med 8 kHz taktfrekvens och 4 kHz frekvensomfång, upp till 510 kbps med 48 kHz taktfrekvens innefattande fullt 20 kHz frekvensomfång, innebärande att hela det frekvensspektrum som ett mänskligt öra kan uppfatta tas upp och överförs.

På detta vis kan OPUS, inom rimliga gränser, dynamiskt växla mellan högkvalitativ högupplöst media när kapaciteten i kommunikationsnätverket så tillåter, till lågupplöst smalbandigt media när kapacitetsbrister uppstår.

### 4.4.5 AMR och AMR-WB

*Adaptive Multi-Rate* (AMR), även benämnt *Adaptive Multi-Rate Narrowband* (AMR-NB) eller *GSM-AMR* är en smalbandig talkodtyp som i första hand används i mobilnät. AMR antogs 1999 av *3rd Generation Partnership Project* (3GPP) som standardkodtyp gällande tal och används idag till stor del i mobilnät baserade på *Global System for Mobile Communications* (GSM) och *Universal Mobile Telecommunications System* (UMTS). AMR är anpassad för kodning av tal och verkar i frekvensomfånget 200-3400 Hz med en variabel bitströmshastighet i nio olika bitströmshastigheter mellan 4,75 och 12,2 kbps.

*Adaptive Multi-Rate Wideband* (AMR-WB) är en bredbandig variant av AMR och används i första hand för telefonitjänster i fjärde generationens mobilnät och dess arkitektur, *Voice over LTE* (VoLTE). AMR-WB har även stor potential att bli den taltyp som i första hand används även mellan olika tele- och mobiloperatörer för att erbjuda så kallad *HD Voice*, då mobiltelefoner i allmänhet har inbyggt hårdvarustöd för denna talkodtyp.

Likt G.722 är frekvensomfånget som AMR-WB hanterar betydligt bredare än vad som gäller för vanlig AMR. AMR-WB omfattar frekvensomfånget 50-7000 Hz med 16-bitars kvantisering och har likt AMR en variabel biströmshastighet men med nio olika steg mellan 6,6 och 23,85 kbps. I dess högsta bitströmshastighet, 23,85 kbps, är AMR-WB kvalitetsmässigt likvärdig med G.722. AMR-WB finns även definierad som talkodtyp av ITU-T och benämns då G.722.2.

AMR är utvecklat och framtaget av bland andra Nokia och Ericsson och det finns flertalet patent- och licensrelaterade frågor som måste tas hänsyn till vid användandet av dessa talkodtyper.

### 4.4.6 MELP och MELPe

*Mixed-excitation linear prediction* (MELP) är en talkodtyp framtagen av Amerikanska försvarsdepartementet och används i första hand i militära taktiska miljöer med fokus kring mycket smalbandiga talkodtyper med låga bitströmshastigheter. MELP definieras som talkodtyp att användas inom den militära standarden MIL-STD-3005. I dess första version genererade MELP en bitströmshastighet av 2,4 kbps, det vill säga betydligt lägre än till exempel G.729 beskrivet ovan. Under slutet av 1990-talet utvecklades en ännu effektivare variant, MELPe, där *e* står för *enhanced*, som medger bland annat ännu lägre bitströmshastighet (1,2 kbps), förbättrad kodning och förbättrad hantering av bakgrundsljud, en mycket viktig detalj i militära tillämpningar och miljöer. MELPe antogs av MIL-STD-3005 och erbjuder motsvarande talkvalitet med halva bitströmshastigheten som MELP. MELPe är dessutom bakåtkompatibel med MELP. År 2002 antog även NATO MELPe som standardkodtyp för tal enligt NATO-standard STANAG-4591. MELPe, som tillsammans med G.729d, är de talkodtyper som framför allt används i samband med säkert tal (SCIP, se avsnitt 7.2.5). År 2005 tillfördes ytterligare en MELPe-variant till STANAG-4591 med ännu lägre bitströmshastighet, 600 bps. Denna variant levererar en ljudbild av tämligen metallisk karaktär som likväl kan vara fullt tillräcklig för vissa tillämpningar och vissa miljöer.

MELPe är utvecklat av Texas Instruments, Microsoft, Thales med flera, innebärande att användning av denna talkodtyp utanför Amerikanska försvaret och NATO kan kräva ett hänsynstagande till ett antal patent- och licensrelaterade frågor.

## 4.5 Val av talkodtyp

I samband med att fler och fler telefonisamtal förmedlas med hjälp av IP-baserade kommunikationsnätverk änd-till-änd, öppnas möjligheten för att på ett bättre sätt använda moderna talkodtyper anpassade för paketförmedlade kommunikationsnät. Då till exempel SIP erbjuder en förhandling i samband med sessionsetableringen, är det möjligt för de inblandade parterna att på ett så optimalt sätt som möjligt välja media- och talkodtyper under sessionetableringen.

I kommunikationslösningar där ena ändpunkten är ansluten via ett kretskopplat telefoninät finns som i regel ingen praktisk möjlighet

att använda någon annan talkodtyp än G.711, då det är den talkodtyp som används i de traditionella telefoninäten. Den delen av sessionen som transporteras över det IP-baserade kommunikationsnätet skulle kunna använda en talkodtyp bättre anpassad för det paketförmedlade nätet, till exempel OPUS, men detta skulle innebära att den utrustning som terminerar IP-telefonisamtalet och konverterar till det kretskopplade nätet skulle vara tuvngen att göra omkodning. All potentiell förbättrad talkvalitet och bredbandig kvantisering av den IP-förmedlade delen av talet skulle ändå gå förlorad när konverteringen till G.711 i det traditionella telefoninätet är tvingande.

I en helt IP-baserad omgivning där ändpunkterna transparent har möjlighet att indikera stöd för sina respektive mediatyper, och dessa kan flöda transparent änd-till-änd genom nätet, finns dock stor vinning att använda talkodtyper konstruerade för att transporteras i paketförmedlade nät. OPUS är ett bra exempel på detta och är, som beskrivits ovan, en dynamisk talkodtyp som kan anpassas efter rådande nätförhållanden. OPUS är ofta implementerat i mjukvaruklienter som körs i generiska plattformar, men däremot ännu tämligen sällsynt i terminaler med särskilt tillägnad hårdvara. En förklaring kan vara att OPUS är en talkodtyp som kräver förhållandevis mycket tillgänglig beräkningskapacitet, vilket inte alltid finns tillgänglig i sådan hårdvara.

## 4.6 Tonval - DTMF

För att överföra tonval (*Dual-tone Multi-frequency* (DTMF)) genom IP-telefoni finns i princip tre olika tillvägagångssätt:

- inbandssignalering i ljudströmmen
- utombandssignalering i mediaströmmen
- utombandssignalering i samtalssignaleringen

Om inbandssignalering i ljudströmmen (*inband audio*) används mellan mediaändpunkterna transporteras tonerna direkt i ljudströmmen på liknande sätt som tonerna framförs i traditionella telefoninät. Nackdelen med att använda inbandssignalering via ljudströmmen för DTMF är att tonerna riskerar att förvanskas om någon typ av omkodning av mediaströmmen sker i nätet. Det går heller

sällan särskilt bra att använda denna metod i samband med nyttjandet av komprimerande talkodtyper.

Den vanligaste metoden att överföra DTMF-toner mellan ändpunkter är att använda separata RTP-paket som utombandssignalering i mediaströmmen enligt [RFC4733]. Stödet och användandet av dessa *RTP Events* förhandlas mellan ändpunkterna i samband med samtalsetablering. Fördelen med att använda denna DTMF-metod är att DTMF-tonerna, oavsett vald talkodtyp och eventuell komprimering och omkodning i nätet, kan transporteras oförvanskade mellan ändpunkterna. Så länge DTMF-tonerna enbart behöver transporteras änd-till-änd mellan mediaändpunkterna är detta den rekommenderade metoden att använda.

I vissa trafikscenarion behöver signaleringsnoder som inte hanterar mediaströmmarna ändå få tillgång till DTMF-tonera som skickas mellan ändpunkterna. Vanligt är till exempel att en företagsväxel som släpper kontrollen över mediaströmmen, ändå måste få tillgång till DTMF-tonerna för att aktivera och tillhandahålla tjänster till slutanvändaren. För att detta ska fungera måste DTMF-signaleringen använda signaleringsvägen istället för mediavägen. Beroende på vilket signaleringsprotokoll som används så sker detta på olika sätt. För används till exempel SIP-metoden *INFO*.

Det är inte ovanligt att noder i nätet eller ändsystem måste hantera flera metoder för att kunna överföra DTMF-toner på ett tillförlitligt sätt. Att konvertera mellan olika metoder är också möjligt men förutsätter ofta att sådana noder i nätet har tillgång till både media och signalering för att kunna göra denna översättning.

## 4.7 Modem och fax

Traditionella telefoninät används även för att överföra andra mediatyper än tal mellan anslutna abonnenter. Till exempel var modemtrafik en tämligen viktig del då internet tog fart på riktigt, och teleoperatörerna började sälja internettillgång via uppringda modempooler. Även telefax har varit och är fortfarande en viktig tjänst som traditionellt levereras som överlagrade dataförbindelser i de kretskopplade telefoninäten.

I samband med övergången till IP-nät som bärare för traditionella telefonitjänster ställer dessa mediatyper särskilda krav. Det går naturligtvis att starkt ifrågasätta det lämpliga i att överlagra en modem-

förbindelse med låg kapacitet över ett IP-nät som definitionsmässigt måste tillhandahålla en väsentligt högre kapacitet för samma överföring. Det naturliga vore att ersätta faxen med en tjänst anpassad för informationsöverföring via IP. Men då utrustningen finns kvar och används behöver ofta även dessa tjänster hanteras. De flesta signaleringsprotokoll för IP-telefoni har därför stöd även för att etablera sådana typer av överlagrade dataförbindelser och förhandlar på liknande sätt som för andra mediatyper enbart fram en datakanal mellan ändpunkterna och en lämplig kodningstyp.

De mest använda kodningstyperna för att överlagra sådana överföringar i ett IP-nät är de två ITU-standarderna T.38 för fax och V.150.1 för modem.

### 4.7.1 Faxrelä - T.38

T.38 är en standard från 1998 framtagen av ITU för att kunna överföra fax över IP-nät. Fax är konstruerat för att användas i synkrona och pålitliga kretskopplade nät med begränsade fördröjningar och små informationsförluster. Ett IP-nät är asynkront och variationer i fördröjningar och enstaka paketförluster är inte ovanliga. För att kunna överföra fax i ett IP-nät på ett tillförlitligt sätt måste därför dessa skillnader hanteras och det är det som T.38 är konstruerat att göra.

För att ansluta en fax till ett IP-nät används en T.38 gateway. För faxutrustningen ser anslutningen ut att vara en vanlig PSTN-anslutning. Faxutrustningen kommunicerar med mottagande fax över T.30-protokollet som om de vore anslutna till ett traditionell kretskopplad telefoninät. T.38 gateway överlagrar T.30-ramarna i IP-paket och hanterar synkronisering, buffring och omsändning av IP-paketen för att säkerställa att all information överförs på ett tillförlitligt sätt.

En stor mängd ändutrustning för IP-telefoni stödjer idag T.38 för fax. Det är däremot inte alltid säkert att IP-telefonileverantören gör det, och fax är därför fortfarande en källa till problem inom IP-telefonin. Alternativet till att använda T.38 för att överföra fax i ett IP-nät är att förlita sig på att IP-nätet har så låg variation av fördröjning (*jitter*) och så få paketförluster att fax-överföringen kan bäras av en okomprimerad talkodtyp, vanligtvis G.711. Det är dock inte ovanligt att hastigheten på faxöverföringen måste ställas ned för att få ett bra resultat.

### 4.7.2 Modemrelä - V.150.1

V.150.1 är en ITU-standard från 2003 för att bära modemtrafik över IP-nät. Likt fax finns det i princip två sätt att överföra sådan modemtrafik, antingen en reläfunktion likt V.150.1, eller genom en okomprimerad talkodtyp, vanligtvis G.711. Likt fax är modem konstruerade att överföra data i synkrona och tillförlitliga kretskopplade nät och lider därför av liknande problem när trafiken ska bäras av ett IP-nät. En V.150.1 gateway används därför att att hantera dessa skillnader och erbjuder en överlagrad modemförbindelse över IP.

Modemtrafik är antagligen mindre vanlig i rena IP-nät då de allra flesta datatjänster givetvis använder IP-nätet direkt. Att överlagra en dataförbindelse på 9.6kbps över ett IP-nät med hög kapacitet är sällan befogat. Däremot, i vissa trafikfall när en modemansluten utrustning i ett kretskopplat nät vill etablera en datakanal till en IP-ansluten utrustning, kan denna typ av modem-relä vara den enda möjliga lösningen. Ett exempel på detta är när krypterade talförbindelser mellan enheter anslutna till kretskopplade och paketförmedlade nät ska etableras, till exempel med *Secure Communiations Interoperabillity Protocol* (SCIP).

## 4.8 Omkodning och ompaketering

I de flesta fall är det önskvärt att ändpunkterna sinsemellan förhandlar fram de media- och talkodtyper som ska användas. Olika signaleringsprotokoll har olika metoder för att underlätta denna förhandling, vilket beskrivs mer i detalj i avsnitt 3. Grundprincipen är som regel att avsändande ändpunkt meddelar de media- och talkodtyper som denna har stöd för, och att mottagande ändpunkt väljer bland dessa för att komma fram till en gemensam uppsättning samtalsparametrar.

Det kan emellertid förekomma trafikfall där det av olika anledningar inte är möjligt eller ens önskvärt att media- och talkodtyper förhandlas transparent änd-till-änd. I dessa fall är det möjligt att trafiknoder, ofta placerade i teleoperatörers nät, gör antingen omkodning (*transcoding*), det vill säga konvertering mellan olika tal- och videokodtyper eller ompaketering (*transrating*), det vill säga uppdelning eller sammanslagning av mediafragment i större eller mindre paketstorlekar.

Omkodning kan vara aktuellt av flera anledningar. Det finns fall när ändpunkterna helt enkelt inte har stöd för någon önskvärd gemensam talkodtyp eller där nätägaren eller ägaren av telefonitjänsten vill tvinga en viss talkodtyp mot vissa abonnenter eller system. Ett vanligt exempel är samtrafik mellan fast- och mobilnät där det inte är ovanligt att olika talkodtyper används i respektive nät.

Ompaketering är mindre vanligt då mediaändpunkter i regel har förmåga att hantera en rad olika paketeringsintervall. Varje RTP-paket innehåller dessutom tidsangivelser för de mediafragment som transporteras, just för att mottagande ändpunkt ska kunna återskapa mediaströmmen i den takt som den kodades.

Det förekommer dock utrustning som enbart hanterar ett specifikt paketeringsintervall, som således kräver att mediasegmenten som anländer i RTP-paketen har just detta paketeringsintervall. I dessa fall kan det bli nödvändigt att införa ompaketering, till exempel för att konvertera varje inkommande 20 ms mediafragment till två utgående 10 ms mediafragment.

Vid användning av mer avancerade talkodtyper kräver denna funktion ofta kostnads- och resursdrivande *Digital Signal Processor* (DSP)-utrustning för att kunna dela upp paket så att inte enskilda mätvärden eller ramar förstörs.

Både omkodning och ompaketering har viss negativ påverkan på mediaströmmen. Dels måste mediaströmmen förmedlas via den eller de noder som hanterar respektive funktion. Detta medför ofta suboptimala trafikflöden för mediaströmmen som annars hade kunna gå direkt mellan ändpunkterna. Den andra delen av påverkan är den fördröjning och det potentiella jitter som denna nod kan tillföra mediaströmmen. Fördröjning är ofrånkomlig, då utrustningen måste hantera mediafragmenten i buffertar för att sedan utföra själva omkodningen, steg som sammantaget tar minst lika lång tid som det paketeringsintervall som används.

# 5 Adressering

För att upprätta kommunikationen mellan två eller fler olika ändpunkter behövs någon form av identifierare för att på så sätt adressera och dirigera trafik till respektive ändpunkt. Inom traditionell telefoni är den självklara identifieraren det välkända och globalt unika telefonnumret enligt den internationella nummerplanen, *E.164* [E.164], standardiserad av ITU-T.

Principiellt kan man skilja på privata och publika adresseringsplaner. För traditionell telefoni är telefonnummer enligt *E.164* den publika adressplanen. Samtidigt existerar ett närmast obegränsat antal privata nummerplaner för till exempel företagsinterna telefonisystem, lokala anknytningar och slutna telefoninät. Även teleoperatörer använder ett flertal olika prefix och nummerkombinationer för att särskilja och styra trafik, både internt i sina egna nät samt vid samtrafik med varandra. Inom traditionell telefoni finns enbart telefonnumret som adresseringsfält, och det är naturligt att det är detta som modifieras och justeras för att möjliggöra extra funktionalitet i operatörernas nät.

Adresseringsmöjligheterna för IP-baserad telefoni ser emellertid annorlunda ut. Både SIP och H.323 använder sig på liknande sätt som för elektronisk post av *Uniform Resource Identifier* (URI) för adressering. Detta medför att det i första hand används *Fully Qualified Domain Name* (FQDN) för att styra trafik och beskriva vilken mottagare som adresseras. Likt e-post adresseras en SIP eller H.323-användare genom formatet *user@domain*, till exempel *sip:user@example.com* för SIP.

## 5.1 Publik nummerplan

Den globala standardiserade nummerplanen för publik telefoni, *E.164*, är den vi alla är vana vid när vi traditionellt har telefonerat. Det

är ur den nummerplanen alla publikt nåbara telefoner får sina telefonnummer och adresseringsuppgifter. Nummerplanen är hierarkisk och varje telefonnummer inkluderar som mest 15 sifferpositioner. Varje land är tilldelad ett eget nationellt prefix, även benämnt landskod, och de nationellt signifikanta delarna av telefonnumret hanteras av respektive nations regultoriska myndighet för telekommunikation. I Sverige är denna myndighet *Post- och Telestyrelsen* (PTS). Som indikation att det är fråga om just ett *E.164*-nummer, brukar det kompletta telefonnumret föregås av ett plus-tecken, till exempel +46317878000.

## 5.2 Privata nummerplaner

Till skillnad från den publika nummerplanen för internationell telefoni finns en mängd privata nummerplaner som tillämpas i olika slutna system och telefoninät. I dess enklaste och vanligaste form används en privat nummerplan av i princip varje företagsväxel – *Public Branch eXchange* (PBX) – där interna anknytningar, kortnummer och andra specialfunktioner finns representerade. Vanligast är att företagsväxelkunden tilldelas en nummerserie ur den publika nummerplanen av teleoperatören, som kunden i sin tur själv kan koppla mot interna anknytningar och växelfunktioner.

## 5.3 URI-baserade adresseringsplaner

För URI-baserade adresseringsplaner används som tidigare nämnts ett format likt e-postadressering, till exempel *sip:user@example.com* för SIP. Detta medför att användarnamnet, informationen till vänster om skiljetecknet "@", på liknande sätt som för en e-postserver bara har betydelse för den eller de SIP-noder som administrativt ansvarar för SIP-domänen *example.com*. Användarnamnet behöver alltså enbart vara unikt i den kontexten. Detta innebär att mellan olika IP-telefonisystem, till exempel olika SIP-domäner, styrs trafiken baserat på vad som står i domändelen av en SIP-URI. För att trafiken sedan verkligen ska kunna förmedlas till mottagande system krävs givetvis att avsändande och mottagande system kan kommunicera med varandra över ett och samma IP-nät, och samtidigt använder samma namnrymd. Abonnenter som använder publika adresser på internet kan adressera varandra, och använder i praktiken också en

gemensam namnrymd genom domännamnssystemet. Det vi kallar domännamnssystemet, *Domain Name System* (DNS)), knyts samman genom den globala rot-zon som administreras av *Internet Assigned Numbers Authority* (IANA). En gemensam rot-zon är en förutsättning för att alla parter ska kunna adressera varandra på internet, och att alla adresser i namnrymden ska vara unika.

## 5.4 URN-baserade adresseringsplaner

Till skillnad från URI-baserade adresseringsplaner, som anger identifierare för en specifik resurs, så representerar en *Uniform Resource Name* (URN) [RFC2141] istället en tjänst. Hur denna identifierare sedan slutligen löses upp till en eller flera specifika destinationer är något som URN-schemat i sig inte har någon koppling till.

Ett exempel på en URN-baserad adressering för en tjänst som slutligen ska resultera i en samtalsstyrning är hantering av nödnummer och SOS-tjänster. [RFC5031] definierar ett antal SOS-relaterade URN-baserade namnrymder innefattande en mängd olika tjänster som till exempel brandförsvar, polis, ambulans, och så vidare. Exempel på adressering med URN innefattar:

- *urn:service:sos*
- *urn:service:sos.police*
- *urn:service:sos.poison*

För att kunna styra samtal baserat på en URN måste först den beskrivna tjänstens resurser identifieras och lokaliseras. Det innebär att ett första steg i framkopplingen av ett samtal är att lösa ut vilken eller vilka adresseringsdestinationer som är kopplade till denna tjänst i den givna kontexten. I fallet nödnummer är kontexten ofta kopplad till geografiskt område; baserat på var användaren som vill nyttja tjänsten befinner sig ska möjligen olika resurser anropas. Om användaren som vill etablera ett nödsamtal finns i Sverige bör samtalet styras till ett svenskt svarsställe för SOS. Om användaren till exempel vill nå polisen är kanske ytterligare finmaskighet önskvärd, så att samtalet styrs till den lokala polismyndigheten.

Hur mekanismen är utformad som, baserat på en anropad URN och andra för tjänsten relevanta parametrar, omvandlar URN till

lämpligt resultat, är inte definierat av adresseringsformatet i sig. Hur omvandling sker och till vad (till exempel en SIP URI) styrs av tillämpning och tjänst. Ett exempel på protokoll för att efterfråga resurser för en viss tjänst baserat på lokaliseringsinformation är *Location-to-Service Translation Protocol* (LoST) [RFC5222].

För SOS-tjänsten skulle resultatet av en URN-definierad tjänsteadressering (*urn:service:sos*) för en abonnent som befinner sig i Sverige kunna resultera i att samtalsetableringen styrs till den returnerade SIP URI:n (*sip:112@sos.se*).

En fördel med URN-baserad adressering är att globala namnrymder och identifierare för olika tjänster alltid kan användas på samma sätt oavsett var abonnenten befinner sig, men baserat på till exempel lokaliseringsinformation kan samtalen styras till lokala eller regionala tillhandahållare av just den aktuella tjänsten.

## 5.5 Telefonnummer, ENUM och Internet

Telefonnumret är den djupast rotade och mest använda identifieraren för telefoni. För traditionell telefoni är det en självklarhet att använda telefonnumret, men som tidigare nämnts är enbart telefonnumret som identifierare för URI-baserade adresseringsplaner inte särskilt lämpligt. Telefonnumret används i traditionell telefoni för samtalsstyrning medan i IP-telefonisystem med URI-baserad samtalsstyrning tillhör telefonnumret användardelen av URI och har därmed ingen betydelse förrän samtalet har nått det lokala IP-telefonisystem där mottagande abonnent finns. Först då är användardelen aktuell för att styra inkommande IP-telefonisession till en specifik slutanvändare eller anknytning i systemet. Detta medför att telefonnumret får en helt annan betydelse i URI-baserade IP-telefonisystem. I princip kan inte enbart telefonnumret användas för samtalsstyrning utan annan information måste tillföras på något sätt.

Det finns olika sätt att hantera denna omständighet och det finns även andra faktorer med påverkan på adresseringen som har med telefonnummer att göra. För att i Sverige kunna beställa nummerserier ur den publika nummerplanen, *E.164*, måste organisationen vara en registrerad teleoperatör hos PTS. För att kunna beställa motsvarande adresseringsuppgifter för att styra IP-telefonisamtal på internet krävs enbart en DNS-domän och att beställaren uppfyller domänregistratorns villkor och i övrigt gör rätt för sig inom ramen för de

kommersiella villkoren för internetanslutning och domänregistrering. Således finns ingen direkt inblandning av myndigheter och andra överstatliga organ.

Det finns heller inte något som hindrar ett IP-telefonisystem att använda *numeriska användarnamn* som tillsynes är identiska med motsvarande *E.164*-anknytningar, men det är viktigt att inse att användandet av dessa numeriska användarnamn på internet inte har någon automatisk koppling till dess *E.164*-motsvarighet i telefonivärlden. En teleoperatör som både har publika telefonnummer ur *E.164*-trädet associerat till sig och även tillhandahåller en IP-telefonibaserad tjänst kan självklart välja att använda samma numeriska användarnamn för IP-telefoni och URI-baserad samtalsstyrning som operatören använder för traditionell telefoni. Detta numeriska användarnamn är dock egentligen fortfarande bara en pseudonym och det är upp till terminerande system mellan den IP-baserade och traditionella telefoninätet att korrekt översätta mellan dessa två adresseringsformat.

### 5.5.1 Telefonnummer som URI

Något som ytterligare ökar komplexiteten vad gäller hantering av telefonnummer i ett IP-baserat telefonisystem är möjligheten att representera telefonnummer i URI-format. [RFC3966] beskriver en *tel:*-uri som gör just detta. *tel:*-URI har egentligen fler likheter med URN än med URI, då den likt URN snarare beskriver en tjänst – telefonnummer – än en direkt adressering därav. Nedan visas exempel på representation av ett telefonnummer ur publika *E.164*-nummerplanen i URI-format:

- *tel:+46317878000*

Samma typ av URI kan även användas för telefonnummer som inte är fullständiga *E.164*-nummer, till exempel telefonnummer ur privata nummerplaner eller lokala anknynningar. För att skilja på dessa måste telefonnummer som representeras av en *tel:*-URI som inte är fullständiga *E.164*-nummer, måste URI:n kompletteras med information som beskriver i vilken kontext det angivna telefonnumret är giltigt. Följande exempel beskriver en lokal anknytning (8001) som enbart är giltig inom domänen *example.com*:

- *tel:8001;phone-context:example.com*

Som tidigare beskrivits använder de flesta IP-telefonisystem URI-baserad samtalsstyrning, vilket medför att det krävs ett domännamn (eller en IP-adress) för att IP-telefonisystemet ska kunna lösa ut vart trafiken ska styras om abonnenten i fråga inte finns tillgänglig inom den lokala domänen. Som visas i ovanstående exempel av *tel:*-URI så finns där inget domännamn och ingen sådan information att hämta. IP-telefonisystemet måste på något sätt tillföra information för att ha förmåga att ta beslut gällande hur en session med denna adresseringsform ska hanteras. Denna information måste ofta tillföras IP-telefonisystemet i form av statisk konfiguration. Ett exempel på sådan statisk samtalsstyrningsinformation är en nummertabell, en *dialplan*, som till exempel styr samtliga trafikflöden av denna typ till en förmedlingsnod (*gateway*), som i sin tur ansluter IP-telefonisystemet till det publika telenätet för vidare förmedling av samtalet till andra system.

### 5.5.2 ENUM

För att lösa den tekniska problematiken kring samtalsstyrning på internet med telefonnummer som identifierare finns en DNS-baserad lösning vid namn *E.164 to URI DDDS Application* (ENUM). ENUM definierar bland annat en algoritm för att på ett strukturerat sätt transformera ett *E.164*-formaterat telefonnummer till en DNS-fråga, och på så sätt möjliggöra att genom DNS tillföra nödvändig information för att kunna styra samtalet direkt till mottagaren över till exempel internet, istället för att terminera samtalet via PSTN för transit. ENUM definieras av [RFC6116] och [RFC6117]. Viktigt att notera är att benämningen *ENUM* gäller enbart om det är just *E.164*-telefonnummer som avses. Samma teknik kan emellertid användas även för privata och interna nummerplaner, men går då under benämningen *Private ENUM* eller *Infrastructure ENUM*.

Algoritmen som definieras av ENUM genomför följande transformering av den siffersträng som ENUM-frågan ska ställas till:

- Ta bort alla tecken förutom siffror ⇒ 46317878000
- Vänd ordning på siffrorna ⇒ 00087871364
- Infoga en punkt mellan varje siffra ⇒ 0.0.0.8.7.8.7.1.3.6.4

- Addera domän-suffix *e164.arpa* ⇒ 0.0.0.8.7.8.7.1.3.6.4.e164.arpa.

Informationen kan nu användas för att ställa en *Name Authority Pointer* (NAPTR)-fråga via DNS där en ersättningssträng returneras, under förutsättning att det efterfrågade telefonnumret finns provisionerat. Ersättningssträngen tillför den information som behövs för att kunna styra samtalet på internet, till exempel en komplett SIP-URI.

För att tydliggöra med ett exempel: Utgångsläget är ett E.164-telefonnummer som ett IP-telefonisystem har mottagit i en *tel:*-URI. Då det saknas URI-baserad samtalsstyrningsinformation i denna typ av URI måste IP-telefonissystemet antingen förlita sig på statisk konfiguration eller, som i detta fall, göra en ENUM-förfrågan.

- *tel:+46317878000*

I DNS finns följande NAPTR-post som motsvarar det inkomna telefonnumret korrekt provisionerat:

```
0.0.0.8.7.8.7.1.3.6.4.e164.arpa. 86400 IN NAPTR 0 0 "u" "E2U+sip"
                                        "!^.*$!sip:info@pbx.kirei.se!" .
```

Det reguljära uttrycket som finns definierat i NAPTR-postens ersättningssträng appliceras på ursprungssträngen och resulterar i ovanstående exempel att en SIP URI skapas med användardelens telefonnummer ersatt med användarnamnet *info* och ett domännamn för samtalsstyrning, *@pbx.kirei.se*, adderad. Resultatet blir en SIP URI som IP-telefonisystemet kan använda för att direkt styra samtalet till mottagande system, *pbx.kirei.se*, över till exempel internet.

ENUM förutsätter att E.164-nummerplanen läggs in i *e164.arpa*-trädet i DNS-infrastrukturen. Under *e164.arpa* är avsikten att respektive nations regulatoriska myndighet ska utse en organisation ansvarig för hanteringen av de egna telefonnummerserierna på motsvarande sätt som toppdomänadministratörer och internetleverantörer hanterar uppgifter gällande delegerade IPv4- och IPv6-prefix i *in-addr.arpa* respektive *ip6.arpa*.

Även om ENUM som teknik ofta används flitigt av teleoperatörerna som ett gränssnitt för uppslag för samtalsstyrning, till exempel vid kontroll av porterade telefonnummer, görs informationen sällan publikt tillgänglig. Det finns både regulatoriska och framför allt affärsmässiga anledningar och motiv till att ENUM med publik E.164-information i DNS inte har fått det globala genomslag som varit

avsikten. För IP-telefoni bör man därför se telefonnumret som en pseudonym, och istället använda alfanumeriska användarnamn och adressera IP-telefonianvändare på samma sätt som man gör för andra URI-baserade adresseringstjänster, till exempel e-post.

# 6 Förmedlingsnoder

På liknande sätt som system för traditionell telefoni innefattar olika växlar och system för att förmedla samtalstrafik, finns det ett antal motsvarande nodtyper för IP-baserade telefonisystem. I tidigare kapitel har de olika signaleringsprotokollen och deras respektive funktioner diskuterats ingående. Vad som återstår gäller sammankopplingen av dessa olika nät för telefoni, det vill säga sammankoppling av traditionell telefoni och IP-telefoni, men även sammankoppling av olika typer av IP-telefoni. Denna sammankoppling måste i båda fall ske både i signaleringsplanet och i mediaplanet.

## 6.1 Signaleringskonvertering

För att kunna sammankoppla olika telefonilösningar som använder olika tekniker, oavsett om det är kretskopplad telefoni eller olika varianterer av IP-telefoni, krävs konvertering av signaleringsprotokollet. I de fall då båda signaleringsprotokollen är IP-baserade kan denna konvertering vara enklare, men en konverteringsnod måste också ta hänsyn till respektive systems säregna egenskaper och funktioner. I många fall gör dessa konverteringen mer komplicerad och tillåter ofta enbart grundläggande funktionalitet – systemspecifika tjänster går sällan att konvertera på ett fullvärdigt sätt.

För traditionell kretskopplad telefoni är de två vanligaste gränssnitten *Integrated Services Digital Network* (ISDN) och *ISDN User Part* (ISUP). ISUP används på nätnivå och ISDN är i första hand ett kundgränssnitt. Om en konverteringsnod ska användas för att koppla samman en SIP-baserad telefonilösning och en ISUP-baserad traditionell telefonilösning måste en protokollkonverteringsfunktion givetvis implementera båda protokollen och korrekt konvertera dessa så att respektive systems grundläggande funktionalitet uppnås, det vill säga att samtal kan etableras och hanteras korrekt. ISUP används

i första hand när en SIP-lösning integreras med en teleoperatörs huvudnät och hanteras ofta av ett fåtal centrala konverteringsnoder för signalering och ett flertal distribuerade konverteringsnoder för mediabryggning. ISDN-baserade konverteringsnoder integrerar oftast både signalerings- och mediakonvertering i samma utrustning och används i större omfattning vid till exempel kundanslutning av ett företagstelefonisystem mot en teleoperatör eller tvärt om, där kundens utrustning enbart har ISDN-stöd men teleoperatören använder till exempel SIP i sitt eget nät.

Det är även vanligt att konverteringsfunktioner måste användas mellan olika IP-baserade telefonisystem då, som tidigare tydliggjorts, flertalet olika signaleringsprotokoll förekommer. Även om SIP idag är det dominerande signaleringsprotokollet finns fortfarande stora H.323-lösningar i drift, framför allt inom företags- och videosystem. Dessa måste konverteras till exempelvis SIP för att kunna integreras i teleoperatörernas nät. Även om både SIP och H.323 är IP-baserade signaleringsprotokoll och använder samma mediaöverföringsprotokoll (RTP) måste en konverteringsnod på liknande sätt som för ISDN och ISUP hantera olikheterna och signalera på korrekt sätt genom respektive protokoll. Media kan dock i detta fall förmedlas direkt mellan ändpunkterna vilket minskar resursbehovet på konverteringsnoden.

En mindre vanlig konverteringsfunktion, likväl populär hos vissa teleoperatörer, är en variant av signaleringsprotokollet SIP som benämns SIP-I. SIP-I är egentligen vanlig SIP men med den funktionen att ISUP-meddelandet från ISUP-nätet bakom konverteringsnoden skickas med i SIP-meddelandet på liknande sätt som SDP. Detta innebär att om mottagande ändpunkt också kan tolka SIP-I så kan denna välja att använda ISUP-informationen i SIP-meddelandet istället för SIP-informationen direkt. I de fall där SIP-infrastrukturen enbart används som transitnät för att förmedla trafik mellan två olika ISUP-baserade nät är det rimligt att använda SIP-I, förutsatt att den enkapsulerade ISUP-information kan tolkas av respektive mottagande förmedlingsnod. Om mottagande ändpunkt inte har förmåga att tolka sådan ISUP-information är det naturligtvis av föga intresse att använda SIP-I.

## 6.2 Mediabryggning

För att konvertera och brygga media mellan olika nättyper och telefonisystem behövs i många fall *Digital Signal Processor* (DSP)-resurser då en konverteringsnod för media kanske måste konvertera både talkodtyp och paketeringsintervall. En mediabrygga som ska konvertera tal mellan ett kretskopplat och ett paketförmedlat nät måste paketera mediafragment inkommande från TDM-nätets tidsluckor till IP- och RTP-paket för att förmedla ut på IP-nätet, och vice versa för inkommande trafik från IP-nätet. På IP-sidan måste dessutom viss buffert användas för att kunna hantera variation av paketleverans vilket ökar den totala fördröjningen för ompaketeringen. Olika talkodtyper brukar sägas ha olika *komplexitet* vilket styr behovet av DSP-resurser för att hantera och konvertera kodningen. I de fall då mediakonvertering mellan olika talkodtyper måste göras styrs DSP-åtgången av de inblandade talkodtypernas beräkningsmässiga komplexitet och därmed också resursåtgången för konverteringen. Sådan resursåtgång kan bli betydande i större installationer, med följden att de investeringar som krävs för DSP-utrustning blir förhållandevis dyra.

Att brygga media mellan olika IP-baserade telefonisystem är oftast betydligt enklare då de flesta använder RTP. Så länge samma talkodtyp stöds av båda parter och ingen omkodning behöver göras kan RTP-paketen ofta förmedlas direkt mellan ändpunkterna även om dessa använder olika signaleringsprotokoll.

Det förekommer förstås även att olika IP-nät inte är direkt anslutna till varandra, och att förmedlingsnoder måste traverseras för att tillåta samtrafik mellan dessa IP-nät. Vanliga exempel innefattar då teleoperatörer önskar hålla sina interna nät, där interna system och andra telefoniresurser är lokaliserade, separerat från kunders och andra publika nät. I dessa fall används ofta en IP-till-IP-konverteringsnod som i SIP-fallet agerar B2BUA, vilken terminerar både signalerings-, media- och IP-strömmar på respektive sida och i respektive nät. Baserat på konfiguration och regelverk hanterar sedan konverteringsnoden trafiken mellan dessa nät och kopplar samman de parter som ska kommunicera med varandra enligt denna policy. En mycket vanlig och kommersiellt tillgänglig produkt som implementerar sådan funktionalitet är *Session Border Controller* (SBC), som både konverterar signalering och media och agerar likt en brandvägg

för IP-telefoni.

## 6.3 Regelverkstyrd bryggning

Regelverkstyrd bryggning av IP-till-IP-trafik är mycket vanlig hos teleoperatörer som levererar IP-telefoni. Då den i dagsläget största delen av alla telefonsamtal terminerar antingen i mobil- eller fastnät som inte är baserade på IP-telefoni, är de tjänster och fördelar som till exempel SIP medför av underordnat intresse. Istället är det en grundläggande taltjänst som ska erbjudas, som i mångt och mycket liknar den telefonitjänst som funnits allmänt tillgänglig i närmare 100 år. I dessa nät och tjänster är det ofta inte flexibilitet och valmöjlighet som eftersträvas, utan istället en god kvalitet på en grundläggande taltjänst. Därför används ofta regelverkstyrd bryggning, ofta med hjälp av en SBC, för att på ett kontrollerat sätt hantera kunder och samtrafik. Då samtal oftast terminerar på PSTN eller i mobilnät är det inte heller av intresse att erbjuda andra talkodtyper än de som stöds av PSTN, det vill säga G.711. Därför tillämpas ofta en policy i bryggningsfunktionen som ser till att enbart ett fåtal talkodtyper och tjänster i signaleringsprotokollet släpps igenom. På detta sätt garanterar teleoperatören ett stringent förfarande vid signalering vilket gör att tjänsten är lättare att driva och förvalta, och förutsättningarna för att enkelt hantera olika typer av kundutrustning ökar.

Då en SBC som agerar B2BUA per definition inte är en transparent komponent, så som en *SIP Proxy Server* är, är SBC en högst omdebatterad komponent i sammanhanget som betraktas som innovationshämmande och konceptuellt felaktig, då den bryter de grundläggande principerna om internets änd-till-änd-konnektivitet. SBC har i många sammanhang kommit att bli en funktion där operatören snabbt kan göra justeringar och modifieringar av signaleringen för att lösa interoperabilitetsproblem som tillstött på grund av bristande kvalitet i SIP-implementationer. Detta hävdar många har haft en negativ påverkan på kvaliteten i SIP-implementationer generellt, då det blivit mindre viktigt att de fungerar korrekt från början, eftersom det alltid går att ordna upp i efterhand med en SBC. I realiteten har det dock visat sig att en SBC och dess funktionalitet, på liknande sätt som brandväggar, har kommit att bli en de facto standard både hos teleoperatörer och i större kundanslutningar.

# 7 Informationssäkerhetsskydd

Informationssäkerhet är ett vitt begrepp som inbegriper skydd mot en rad olika hotbilder och angreppsvektorer. För telefoni och förmedling av tal över IP kan en viss generell indelning göras: skydd av signalering respektive skydd av media. Olika tekniska lösningar är tillämpliga för respektive trafiktyp och några av dessa beskrivs i det följande.

För IP-telefoni finns även två generella tillvägagångssätt för att skydda signalering och media. Det ena, *Tal över säkert nät*, innebär att all IP-trafik som förmedlar telefonitjänsten redan är skyddad och möjligen krypterad av underliggande anslutningar på ett sådant sätt att ytterligare skydd av telefonitrafiken inte är nödvändig. Detta är till exempel tillämpligt i krypterade VPN-miljöer där all IP-trafik mellan de kommunicerande parterna redan är skyddad med för ändamålet lämpliga tekniker.

Den andra modellen, *Säkert tal – Secure Voice over IP* (SVoIP) – innebär att skyddet av IP-telefonitjänsten *inte* förlitas på underliggande nätverk. IP-telefonitjänsten behöver då innefatta lämpligt kryptografiskt skydd för kommunikationen oavsett vilka tekniker som används för förmedling av trafiken. Det är i första hand den modellen som diskuteras i denna skrift då den första modellen, *Tal över säkert nät – Voice over Secure IP* (VoSIP) –, är generell och inte har direkt bäring på tjänsten IP-telefoni.

Utöver de IP-telefonispecifika delarna är det även viktigt att se säkerheten för IP-telefonitjänsten från ett helhetsperspektiv. Till exempel är infrastrukturella tjänster som *Domain Name System* (DNS) och *Network Time Protocol* (NTP) viktiga delar i hur förmedlingsnoder för IP-telefoni loggar, styr och förmedlar trafik. I fallet DNS är det till exempel lämpligt att DNS-informationens integritet skyddas genom användning av *Domain Name System Security Extensions* (DNSSEC).

På samma sätt som för andra typer av tjänster i IP-förmedlade nät finns givetvis även hot och attackvektorer riktade mot underliggande infrastruktur som får direkt påverkan på IP-telefonitjänsten. En störning av förmedling av IP-paket i underliggande nät påverkar naturligtvis även tjänsten IP-telefoni. Överbelastningsangrepp, antingen den är riktad mot IP-nätets förmedlingsnoder, infrastrukturella tjänster eller på IP-telefoniinfrastrukturen i sig, påverkar direkt IP-telefonitjänsten. Det är därför viktigt att ta hänsyn till och ha förståelse för *hela* infrastrukturen och dess beroenden.

## 7.1 Skydd av signalering

Skydd av signalering kan ske mellan ändpunkterna (skydd på meddelandenivå) eller i överföringen av signaleringen (transportskydd). Beroende på vilket signaleringsprotokoll som används finns olika aspekter att ta hänsyn till.

För SIP gäller principiellt att förmedlingsnoder som hanterar SIP-meddelanden längs signaleringsvägen mellan de inblandade ändpunkterna *enbart* ska behöva läsa och hantera SIP-meddelandets huvud (*eng. header*) där nödvändig adresseringsinformation finns. Resterande del av SIP-meddelandets innehåll (*eng. body*) kan därmed krypteras mellan ändpunkterna. Denna del innehåller i vanliga fall *Session Description Protocol* (SDP) och information om mediatyper, tal- och videokodtyper, krypteringsnycklar för mediaströmmar, med mera. Det är också värt att notera att vid användning av vissa telefonirelaterade tjänster så som tillgänglighetsindikering (*eng. presence*) och direktmeddelanden (*eng. instant messaging*) är det inte ovanligt att denna information förmedlas via signaleringsprotokollet och inte via en separat mediaström. Detta medför att informationen som överförs via signaleringsprotokollet kan vara av känsligare karaktär än enbart information gällande själva sessionsetableringen. Detta måste givetvis tas hänsyn till vid val av säkerhetsmekanismer.

Brandväggar, adressöversättningsfunktioner *(NAT)* och andra icke-transparenta noder i trafikvägen mellan ändpunkter och förmedlingsnoder har ofta problem att hantera IP-telefonisamtal där signaleringen skyddas. Dessa noder agerar ofta dynamiskt, genom att baserat på information i signaleringsprokollen förmedla eller blockera trafik. När dessa noder inte längre kan tolka signaleringsinformationen kan det innebära störningar i telefonitjänsten.

*Skydd på meddelandenivå*

Skydd av signaleringsprotokoll på meddelandenivå är inte beroende av transportskyddet i mellanliggande noder i signaleringskedjan, förutsatt att de skyddade meddelandena kan transporteras obehindrat genom dessa. Beroende på hur signaleringsprotokollet är konstruerat kan emellertid uppgifter om signaleringen exponeras för mellanliggande noder, till exempel identiteterna bakom de som kommunicerar, även om meddelandeskyddet i sig garanterar sekretess, riktighet och äkthet. I praktiken används denna typ av skydd mycket sällan i IP-telefonisystem.

*Transportskydd*

Vid skydd på transportnivå upprättas skyddet vid etablering av kommunikationskanalen mellan varje par av noder i signaleringskedjan. Signaleringen exponeras därmed i klartext utan sekretesskydd eller skydd mot förändring i signaleringsnoderna. Ofta måste signaleringsnoderna modifiera delar av meddelandet som en naturlig del i förmedlingen av det samma.

I det enklaste signaleringsfallet för ett vanligt IP-telefonisamtal mellan två användare i två federerande system innebär detta minst tre olika kommunikationspar:

- *klient A - server A*
- *server A - server B*
- *server B - klient B*

Den första kommunikationskanalen etableras mellan respektive klient och server i samband med att klienten *registrerar* sig. I samband med registrering *autentiseras* användaren och respektive server associerar en identitet med klienten för att kunna förmedla inkommande samtal.

Kommunikationskanalen mellan servrarna etableras vanligen först då dessa två federerande system initierar kontakt med varandra för att förmedla ett samtal. Här kan tillkännagiven information i DNS så väl som tillämpat regelverk för samtalsstyrning i respektive server styra valet av transportprotokoll. För SIP måste denna kommunikationskanal använda *Transport Layer Security* (TLS) om *sips:* används som adresseringsmetod. Däremot innebär inte detta att

den sista kommunikationssträckan, den mellan server B och klient B, använder det efterfrågade transportskyddet. Det finns heller inte någon garanti för att den efterfrågade kryptografiska styrkan gällande nyckellängder, algoritmer (med mera) används i samtliga kommunikationskanaler.

*Identitet*

Det finns ett antal standardiserade tekniker för att koppla en autentiserad identitet till en signaleringssession. För SIP är det i första hand *SIP Identity* [RFC4474], för vilket det endast finns begränsat stöd implementerat i dagens kommersiellt tillgängliga produkter.

*SIP Identity* fungerar i princip så att den lokala SIP-proxyn, efter autentisering av användaren, signerar delar av det utgående meddelandet och inkluderar en referens till utfärdaren av det signerande certifikatet. På så vis kan en mottagande förmedlingsnod använda denna referens för att genom en lista av betrodda certifikatutgivare verifiera att avsändande SIP-proxy är den förmedlingsnod som den utger sig för att vara. Om signaturen stämmer kan också mottagande SIP-proxy välja att lita på att avsändande förmedlingsnod har autentiserat sin lokala användare, och kan då med förtrolighet presentera den angivna identiteten för mottagaren.

Att verifiera avsändande användares identitet mellan olika autonoma och federerande IP-telefonisystem är komplicerat och ofta är de tekniker som finns tillgängliga inte heltäckande. På många sätt kan IP-telefoni även i detta fall jämföras med e-post, där motsvarande problem finns. För e-post har detta varit en bidragande faktor till de stora mängder skräppost som förekommer. För IP-telefoni är det ännu inte ett lika stort problem, även om förekomsten av regelmässiga angrepp mot framför allt SIP-baserade IP-telefonisystem på internet idag är en realitet. En stor skillnad mellan att skicka skräppost och att ringa skräpsamtal är att samtalen fortfarande i de flesta fall kräver att någon lyfter på luren och svarar, vilket gör metoden tämligen ineffektiv.

Att däremot försöka angripa en IP-telefon eller ett IP-telefonisystem för att komma över inloggningsuppgifter, som sedan kan användas för att etablera samtal till det publika telenätet via detta system, är mycket vanligt. Ofta är dessutom IP-telefonisystemen som ansluts mot internet och andra publika IP-nät otillräckligt skyddade

och ofta saknas förståelse för hotbilden och hur säkerheten i dessa system kan kompromitteras. I många fall upptäcks intrånget först då ekonomiavdelningen vid det utsatta företaget tar emot onormalt höga leverantörsfakturor.

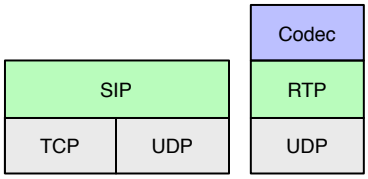
Ofta lyfts sådana incidenter fram som anledning att enbart erbjuda IP-telefonitjänster inom de egna näten och separat från internet. Dessa nät har ansetts vara säkrare och mindre utsatta för externa hot då exponeringsgraden är mer begränsad, men ger också inlåsningseffekter. I samband med nätens och internettjänsternas utveckling, inte minst på mobilsidan, är detta dock en allt mindre hållbar strategi. Användare behöver kommunicera över internet i allt högre grad, då det är där många tjänster realiseras. Därmed måste näten öppnas upp för trafik till och från internet, vilket också ökar antalet angreppsvektorer och ställer högre krav på IP-telefonisystemets robusthet och förmåga att stå emot angreppsförsök.

De angreppsvektorer som här lyfts fram har emellertid ofta sin motsvarighet även i traditionella telefoninät, där det under närmare ett halvt sekel utvecklats en hel kultur kring att utforska och genom olika metoder manipulera dessa nät och tjänster (*phreaking*). Företeelsen att *hacka* lokala telefonväxlar är alltså inte något nytt för just IP-telefoni, men i och med anslutningen av växeln mot internet ökar hotet genom högre komplexitet och en högre exponeringsgrad. Även manipulation av avsändarens identitet (A-identitet) förekommer i de traditionella telefonisystemen. I många tjänster som teleoperatörer erbjuder sina kunder görs sådan manipulation medvetet, varvid frågan uppstår vad en så kallad verifierad A-identitet egentligen innebär. Till sist faller det ofta tillbaka på att du som användare känner igen rösten på den du talar med.

## 7.2 Skydd av media

För att skydda mediaöverföring, till exempel talströmmar, som transporteras med *Real-time Transport Protocol* (RTP) används vanligen *Secure RTP* (SRTP) (se vidare avsnitt 4.2). SRTP inkluderar emellertid inte någon mekanism för nyckelutbyte, varvid detta måste ske på annat sätt. Nyckelutbyte kan vara integrerad del av signaleringsprotokollet via SDP (avsnitt 3.1.4), hanteras som en del av mediaströmmen, eller ske helt manuellt.

Vilken typ av skydd som kan etableras mellan signaleringsno-

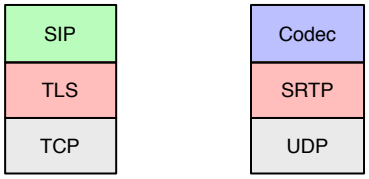



**Figur 7.1** – Media utan skydd

derna kan variera och bero på ett flertal olika faktorer. Det faktum att ändnoderna i många fall inte har någon information om att starkt transportskydd verkligen upprättats och upprätthållits under signaleringen, medför att ändnoderna har svårt att indikera och säkerställa det totala skyddet för kommunikationen. Detta medför i sin tur att skydd av mediaströmmar där nyckelmaterial utbyts via signaleringskanalen, och således är helt avhängigt av signaleringskanalens sekretesskydd, innebär betänkliga svårigheter att realisera sådant skydd på ett tillförlitligt sätt. I dagsläget är likväl den vanligaste metoden för skydd av mediaströmmen som stöds av kommersiellt tillgängliga produkter just *SDP Security Descriptions* (SDES), som använder signaleringsprotokollet för att överföra nyckelinformationen mellan ändpunkterna. SDES kräver att signaleringskanalen är skyddad på ett tillförlitligt sätt, men som ovan beskrivits är detta svårt att garantera i många situationer.

Istället är metoder där nyckelinformationen utbyts direkt mellan de kommunicerande ändpunkterna att föredra. Dessa metoder är inte beroende av att signaleringskanalen är skyddad. Ofta skickas enbart publik nyckelinformation eller fingeravtryck i signaleringskanalen för att koppla samman signalerings- och mediasession eller för att kunna presentera information för användaren. Metoder som bygger på dessa principer är bland andra DTLS-SRTP, ZRTP och SCIP.

Det är värt att notera att det är vanligt att *Dual-tone Multi-frequency* (DTMF)-toner förmedlas i mediaströmmen. Detta medför att en okrypterad mediaström givetvis förmedlar denna information i klartext. Många telefonitjänster där till exempel telefonkoder och PIN-koder knappas in via telefonen överförs som just DTMF-toner och därmed i klartext om mediaströmmen inte är krypterad. Om någon obehörig får tillgång till mediaströmmen kan denne på ett mycket




**Figur 7.2** – Skydd av signalering och media med TLS och SRTP

enkelt sätt tillgodogöra sig informationen och koderna, och därmed på ett obehörigt sätt få tillgång till tjänsterna.

### 7.2.1 SDES

Med SDES [RFC4568] överförs nyckelmaterialet i klartext som attribut i SDP (se vidare avsnitt 3.1.4) och förutsätter alltså att signaleringskanalen är skyddad på annat sätt. Respektive ändpunkt använder nyckelinformationen för att skapa sessionsnycklar som sedan används för att kryptera innehållet i varje SRTP-paket som skickas. Samma nyckelmaterial används för att skapa sessionsnycklar för autentisering av varje SRTP-paket. Autentiseringen av SRTP-paketet används för att säkerställa att en tredje part inte ska kunna injicera information in i mediaströmmen.

### 7.2.2 MIKEY

Med *Multimedia Internet KEYing* (MIKEY) [RFC3830] används SDP som en underliggande transportkanal för att förhandla och utbyta kryptografiska parametrar via signaleringsprotokollet. MIKEY kan i sig använda flertalet olika mekanismer för att åstadkomma nyckelutbytet och nya metoder kan adderas till de befintliga. Exempel på definierade mekanismer är *Pre-shared key*, *Public key*, *Diffie-Hellman key exchange* och *Sakai-Kasahara Key Encryption* (SAKKE)[RFC6509].

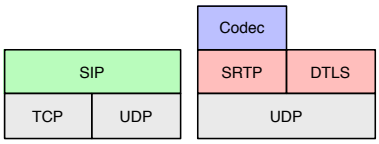
MIKEY är ett generellt ramverk för att tillåta olika nyckelhanteringsmetoder och har nått en tämligen god implementationsgrad i kommersiellt tillgängliga produkter. Det förekommer emellertid kompatibilitetsproblem mellan olika tillverkare och deras implementationer, delvis beroende på de generella definitionerna och de många

olika tillgängliga mekanismer och dess metoder för själva nyckelutbytet.

### 7.2.3 DTLS-SRTP

Genom att kombinera *Datagram Transport Layer Security* (DTLS) och SRTP så som beskrivs i [RFC5764], åstadkoms en metod för att transportera krypterad RTP som SRTP med hjälp av det nyckelmaterial som framförhandlats genom användningen av DTLS.

Denna metod är bland annat tvingande att implementera och använda i *Web Real Time Communication* (WebRTC), men lämpar sig även för användning med andra protokoll, till exempel *Session Initiation Protocol* (SIP)[RFC3261]. För SIP har dock denna teknik ännu inte nått någon större spridning i kommersiellt tillgängliga produkter, men kan troligen få ett större genomslag framöver allt eftersom kundernas krav på säker talkommunikation över internet ökar.

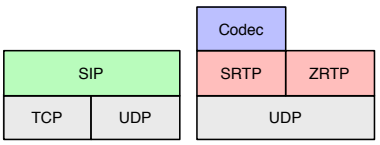



**Figur 7.3** – Skydd av media med DTLS-SRTP

### 7.2.4 ZRTP

*Zimmerman Real-time Transport Protocol* (ZRTP)[RFC6189] transporteras precis som DTLS-SRTP över RTP, och bygger i första hand på framställande av nyckelmaterial genom ett underhandsförfarande baserat på Diffie-Hellman.

Som ett resultat av nyckelutbytet presenteras en autentiseringssträng – *Short Authentication String* (SAS) – för användaren. Genom att utbyta denna sträng över den framförhandlade talkanalen kan man, förutsatt att man känner igen motpartens röst, säkerställa att nycklarna framställts direkt av de kommunicerande parterna utan att en tredje part känner till det framförhandlade nyckelmaterialet.




**Figur 7.4** – Skydd av media med ZRTP

### 7.2.5 SCIP

*Secure Communiations Interoperabillity Protocol* (SCIP) är en NATO-standard som beskriver en änd-till-ändlösning för krypterad talkommunikation. SCIP definierar både själva kodningen av talströmmen och nyckelförhandlingsmekanismer för framställning av det nyckelmaterial som används för kryptering av själva mediaströmmen. Flera olika typer av krypteringsalgoritmer kan användas och det finns både internationella (EU, NATO) och nationella varianter. SCIP kan definieras för flera olika talkodtyper och den enda talkodtyp som är tvingande att implementera för en SCIP-godkänd produkt är *Mixed-excitation linear prediction* (MELP). MELP är, som tidigare beskrivits, en talkodtyp med mycket låg bitströmshastighet som är framtagen på uppdrag av amerikanska försvarsdepartementet och används i första hand just i militära och taktiska miljöer där tillgänglig dataöverföringskapacitet kan vara starkt begränsad. Kravet på att implementera åtminstone MELP finns av interoperabilitetsskäl, för att en sessionsetablering alltid ska kunna falla tillbaka på denna talkodtyp och för att på så sätt kunna etablera en session även under särskilt ogynnsamma förhållanden i underliggande bärarnätverk.

SCIP är i första hand ett enkapsuleringsprotokoll då handskakning och nyckelhantering sker direkt änd-till-änd över en etablerad kommunikationskanal. På något sätt måste dock ändpunkterna etablera denna kommunikationskanal och för att göra det behövs i allmänhet något signaleringsprotokoll. För detta kan till exempel SIP användas, för att som för vilket vanligt telefonsamtal som helst etablera sessionen mellan två SCIP-kapabla ändpunkter. För signaleringsprotokollet ter sig SCIP som vilken samtalsetablering som helst. Det är först när ändpunkterna är sammankopplade som de har möjlighet att börja utbyta SCIP-ramar och förhandla fram lämplig talkodtyp och lämpliga krypteringsfunktioner. Beroende på hur SCIP-

stödet är implementerat i dessa ändpunkter kan de antingen fortsätta en oskyddad talförbindelse eller så kan de inleda SCIP-handskakning, utbyta sessionsnycklar och övergå till en krypterad förbindelse. Det finns även implementationer av SCIP som enbart stödjer det krypterade läget, det vill säga att det inte finns något stöd för okrypterat tal, vilket då måste indikeras vid sessionsetablering i SDP.

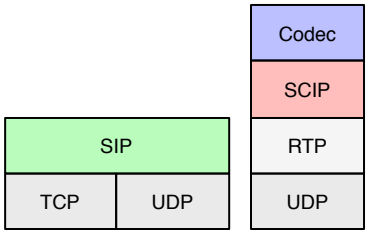
SCIP kan användas både i kretskopplade och paketförmedlade nät. I kretskopplade nät transporteras SCIP ovanpå en etablerad kommunikationskanal och i paketförmedlade nät transporteras SCIP-ramarna antingen som en överlagrad datakanal i lämpligt transportprotokoll eller som en definierad datafältstyp (scip) i RTP. Om båda SCIP-ändpunkter använder IP som bärare är det lämpligt att använda RTP för att enkapsulera SCIP-ramarna. Om SIP används som signaleringsprotokoll definieras SCIP-kanaler som en egen datafältstyp, med tillhörande mediatyper i SDP. När SIP väl har etablerat sessionen utbyter ändpunkterna SCIP-ramar över den etablerade mediaströmmen och SCIP-handskakningen startar. Om båda SCIP-ändpunkter *inte* använder IP som bärare används istället en modemöverlagringsstandard, *V.150.1*, för att transparent bära SCIP-ramarna mellan ändpunkterna.

En stor fördel, i framför allt militära och andra högrisktillämpningar, är att SCIP-standarden möjliggör änd-till-ändkrypterade samtal mellan SCIP-ändpunkter ansluta till både kretskopplade och paketförmedlade nät. Dessa nät kan bryggas samman via en förmedlingslösning som transparent för över SCIP-ramarna mellan IP-paketen och den överlagrade modemförbindelsen. SCIP är ett av mycket få förekommande protokoll som hanterar säkert tal med änd-till-ändkryptering över olika typer av bärarnät, inklusive radionät.

SCIP utvecklades inledningsvis av *National Security Agency* (NSA) i USA för att ersätta äldre talkryptosystem. År 2011 släpptes huvuddelen av specifikationerna gällande protokollet [SCIP] fritt och idag finns flertalet tillverkare av såväl civila som miltärt SCIP-godkända produkter på marknaden.

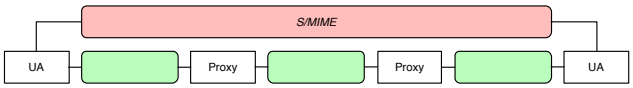
## 7.3 Säkerhet med SIP

I den ursprungliga versionen av SIP [RFC2543] specificerades att *Pretty Good Privacy* (PGP)[RFC2015][RFC4880] skulle användas för skydd på meddelandenivå. Detta ändrades i den version av stan-




**Figur 7.5** – Skydd av media med SCIP

darden som publicerades år 2002 [RFC3261], i vilken PGP är ersatt av *Secure/Multipurpose Internet Mail Extensions* (S/MIME) [RFC5750][RFC5751]. Det bör dock noteras att stödet för SIP med S/MIME i princip är obefintligt i de implementationer och produkter som finns kommersiellt tillgängliga på marknaden idag. Skydd på meddelandenivå är i praktiken inte implementerat i SIP-baserade produkter.

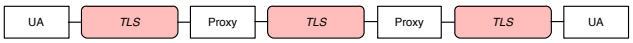



**Figur 7.6** – SIP med S/MIME

För skydd av SIP-signalering på transportnivå används TLS [RFC5246]. Om adresseringsmetoden *sip:* används kan signaleringsnoderna på eget initiativ välja att etablera transportskydd med TLS. Används däremot adresseringsmetoden *sips:* [RFC5630] krävs att transportskydd etableras för all signalering. Några krav på kryptostyrka eller autentisering ställs emellertid inte genom specifikationen av protokollet och kommuniceras inte heller till ändnoderna, vilket gör att den totala skyddsnivån kan vara svår att fastställa för de kommunicerande parterna.

Det är därför svårt att som användare veta att till exempel TLS använts och används av samtliga inblandade förmedlingsnoder hela vägen mellan ändpunkterna. En indikering i SIP-telefonen till användaren, på liknande sätt som det välbekanta hänglåset i webbläsaren, invaggar användaren i falsk trygghet. Då skydd av mediaströmmen dessutom är frikopplad från skyddet av signaleringen skulle en sådan

indikering vara mycket svår för användaren att värdera.




**Figur 7.7** – SIP med TLS


**Figur 7.8** – SIP med partiell TLS

Ofta finns såväl praktiska som regulatoriska aspekter att ta hänsyn till när det gäller skydd av signalering och media, speciellt i rollen som teleoperatör. Då de största mängderna telefonitrafik fortfarande på något sätt traverserar *Public Switched Telephony Network* (PSTN) och det traditionella kretskopplade telefoninätet måste skyddade signalerings- och mediaströmmar ändå termineras och avkrypteras innan trafiken kan förmedlas vidare. På samma sätt som omkodning av tal- och videoströmmar kräver resurser i teleoperatörernas nät krävs resurser för terminering och avkryptering, vilket gör att teleoperatören ofta, med hänsyn tagen till kostnaderna, ser en begränsad nytta av skydd.

Ur regulatoriska perspektiv måste även teleoperatören till exempel kunna fånga signalering och mediaströmmar för brottsbekämpande ändamål. Dessa kan då inte vara skyddade. För IP-telefonitrafik som enbart förmedlas i paketförmedlade nät är det förstås ändå önskvärt att skydda både signalering och mediaströmmar.

# 8 Kvalitet

De kvalitetskrav som ställs på IP-telefonitjänsten varierar beroende på vilka egenskaper som är prioriterande. Till exempel kan ett röstsamtal med mycket smalbandig talkodtyp väl fylla sitt syfte då viktig information ska framföras i vissa lägen, men kanske knappast är något som upplevs som *god* kvalitet för en flerparts röstkonferens i företagsmiljö. Således är kvalitetsbegreppet svårt att tillämpa universellt, då det måste kopplas till den situation och den tillämpning i vilket samtalet genomförs.

Inom traditionell telefoni där statiska talkodtyper med fast bitströmshastighet vanligen används är det tämligen enkelt att beräkna den kapacitet som transmissionsnätet måste erbjuda för att tjänsten ska kunna levereras på ett godtagbart sätt. I kretskopplade nät finns en reserverad resurs som garanterar att ett samtal som etableras erhåller tillräcklig kapacitet i nätverket för att levereras. I ett IP-nät finns inte detta på samma sätt, även om olika tekniker och protokoll kan användas för att till viss del efterlikna och möjliggöra resursallokering. Dessa olika tekniker och protokoll går under den samlande benämningen *Quality of Service* (QoS), och handlar om både statisk och dynamisk resursallokering i IP-nätet för viss typ av IP-trafik, till exempel IP-telefoni. Gemensamt för de olika QoS-teknikerna är att de i praktiken enbart är tillämpbara inom en och samma administrativa domän, då dessa QoS-parametrar sällan implementeras mellan olika autonoma system eller mellan olika administrativa domäner.

Till skillnad från traditionell telefoni där statiska talkodtyper oftast används är det mer och mer vanligt att dynamiska talkodtyper används för IP-telefoni. Detta innebär att mediaströmmen i sig anpassar sig dynamiskt till rådande nätverkstillstånd. Detta gör QoS-implementationen svårare då det inte blir lika enkelt att beräkna de resurser som behöver reserveras för att möjliggöra samtalet, men även

mindre nödvändigt då – åtminstone till viss del – bristande överföringskapacitet kan hanteras direkt av talkodtypen.

## 8.1 Kvalitetsparametrar

Generellt finns ett antal parametrar som kan mätas och som i sin tur påverkar olika talkodtyper på olika sätt. De parametrar man oftast avser i samband med kvalitet i relation till IP-telefoni är:

- fördröjning,
- jitter, och
- paketförluster.

**Fördröjning** inom telefoni är den totala tid det tar från det att ursprungssignalen fångas upp och kodas genom lämplig talkodtyp, till dess att den kan spelas upp av mottagande ändpunkt. Detta innefattar hela kedjan, från fördröjningen som följer genom valet av talkodtyp, tiden det tar att paketera och överföra RTP-paketet innehållande mediaramarna, jitterbuffrar i mottagande ändpunkt, avkodning och slutligen uppspelning. För dubbelriktade realtidskommunikationer, så som tal, är det viktigt att den totala fördröjningen inte blir för stor då det medför att de kommunicerande parterna lätt pratar i munnen på varandra. Beroende på underliggande transmissionsnät finns det naturligtvis fysiska begränsningar gällande minsta möjliga fördröjning. Används till exempel en satellit-länk för transmission av ett samtal går det inte heller att undvika den fördröjning som tillförs beroende på det avstånd signalen måste färdas.

**Jittervärdet** beskriver *variationen* i fördröjning mellan olika RTP-paket när de anländer till mottagaren. Mottagande ändpunkt behöver erhålla de olika mediasegmenten som är förpackade i RTP-paket i ett så pass jämnt flöde att de kan avkodas och spelas upp korrekt. Om variationen i tid mellan de olika RTP-paketen (och därmed mediasegmenten) är för stor, klarar inte mottagande ändpunkt att korrekt spela upp den mediaström som skickas. Samtliga mottagande mediaändpunkter implementerar därför en jitterbuffert av något som slag, som har förmåga att hantera en viss variation. Men ju större variation ändpunkten måste kompensera för, desto större jitterbuffert behövs, vilket har en direkt inverkan på den totala fördröjningen av uppspelningen änd-till-änd.

**Paketförluster** uppkommer då IP-paket förloras av det underliggande nätverket under transport mellan avsändare och mottagare. För applikationer som använder ett tillförlitligt transportprotokoll, till exempel *Transmission Control Protocol* (TCP), utgör paketförluster inom rimliga gränser inget större problem då transportprotokollet säkerställer att dessa sänds om och levereras. För RTP, som använder *User Datagram Protocol* (UDP) som transportprotokoll, finns inte denna funktion inbyggd. Det är oftast heller inte önskvärt att skicka om ett förlorat RTP-paket, då den tid som ljudsegmenten i RTP-paketet representerade, med stor sannolikhet redan är passerad.

Dessa ovan nämnda parametrar rapporteras kontinuerligt ändtill-änd mellan mediaändpunkterna med hjälp av *Real-time Control Protocol* (RTCP). Detta innebär att respektive ändpunkt har kontinuerlig information om hur respektive RTP-ström hanteras av underliggande nätverk. Baserat på denna information kan ändpunkten göra aktiva val. Till exempel kan ändpunkten välja att ändra talkodtyp om kvalitetsparametrarna blir för dåliga, ändpunkten kan avsluta vissa mediatyper, till exempel deaktivera en videoström för att enbart använda ljud, och så vidare. En dynamisk talkodtyp kan även använda denna information för att ändra bitströmshastighet och på så vis, utan den andra ändpunktens direkta inblandning, anpassa förutsättningarna till rådande nätverkssituation.

Utöver dessa mediarelaterade kvalitetsparametrar kan även signaleringsspecifika mätvärden vara av intresse, till exempel *uppkopplingstid* och *tillförlitlighet*. Med uppkopplingstid avses den tid det tar från det att uppringande part har initierat samtalsetablering, tills dess att det verkligen "ringer på" hos mottagaren. Tillförlitlighet handlar om sannolikheten att ett initierat samtal verkligen kopplas fram. I traditionella telenät nämns ofta siffran 99,999%, *the five nines*. För IP-telefoni är detta givetvis kopplat till tillgänglighetssiffror i underliggande IP- och transmissionsnät samt de tjänsteplattformar som tillhandahåller IP-telefonitjänsten.

## 8.2 Tjänstekvalitet

Få IP-baserade tjänster är så hårt kopplade till kvalitets- och tillgänglighetskrav som just IP-telefoni. Den självklara anledningen är att den traditionella kretskopplade telefonin som regel har en mycket hög tillgänglighet och god talkvalitet. Därför ställs också höga krav

på IP-telefoni för att möta användarens berättigade förväntningar beträffande tjänstekvaliteten. En nackdel i sammanhanget med denna hårda koppling till den traditionella telefonitjänsten är att IP-telefoni, åtminstone inledningsvis, i hög grad enbart var en ersättning av transportnätet från kretskopplad infrastruktur till paketförmedlad sådan. Istället för att samtidigt även byta till talkodtyper lämpade för paketförmedlade nät användes (och används fortfarande) till största delen samma talkodtyper som i traditionell telefoni. Istället har utvecklingen av IP-baserad telefoni inriktas på att efterlikna de kretskopplade nätens egenskaper och krav.

*QoS*

Området QoS kan delas upp i olika tekniska delar, men i grund och botten handlar det i IP-nät i praktiken enbart om olika former av kö-mekanismer och regelverk som reglerar vilka IP-paket som ska kasseras i första hand. Det finns i denna del två huvudsakliga tekniska spår:

- *Integrated Services* (IntServ), och
- *Differentiated Services* (DiffServ).

Båda dessa tekniker implementeras i första hand i IP-nätets förmedlingsnoder och handlar i praktiken om olika interna köhanterinsmekanismer. I fallet *IntServ* reserveras en dedikerad resurs för en specifik dataström genom nätet. Det innebär att varje förmedlingsnod i nätet garanterar att noden kommer ha tillräckligt med överföringsresurser tillgängliga för att kunna hantera den för reservationen (tjänsten) överenskomna bitströmshastigheten. Reservationen i förmedlingsnoderna görs av ändpunkterna med hjälp av *Resource Reservation Protocol* (RSVP). RSVP signalerar i dataströmmens framföringsväg och ställer förfrågan till varje förmedlingsnod längs vägen om möjlighet finns att genomföra önskad resursreservation. Om reservationen inte är möjlig måste alternativ förmedlingsnod väljas, eller reservationen i sin helhet nekas. Då stödet för IntServ måste finnas i samtliga inblandade förmedlingsnoder är detta inte en teknik som vanligen finns tillgänglig mellan olika administrativa domäner, utan är mer en funktion som lämpar sig inom kontrollerade och harmoniserade nätmiljöer. Grunderna för IntServ definieras i [RFC1633] och reservationsprotokollet, RSVP, definieras i [RFC2205].

Det andra tekniska spåret, DiffServ, definieras i grunden av [RFC2474] och [RFC2475]. Utöver dessa dokument specificeras sedan två huvudkategorier av kvalitetsfunktioner, *Expedited Forwarding* (EF) [RFC3246] och *Assured Forwarding* (AF) [RFC2597]. Till skillnad från IntServ sker ingen garanterad resursreservation, då det istället är fråga om ett övergripande regelverk för hur olika typer av trafik, efter klassificering, ska hanteras av förmedlingsnoder vid överbelastning. Noterbart är att DiffServ i princip enbart används när en förmedlingsnod är överbelastad på ett sådant sätt att den måste kassera vissa paket. DiffServ använder ett antal bitar i IP-paketets pakethuvud (*Differential Services Code Point* (DSCP)) för köindelning. Baserat på värdet i detta fält kan förmedlingsnoden tillämpa prioriteringar i sin interna köhantering och säkerställa att viss typ av trafik, till exempel fördröjningskänsliga RTP-paket, hanteras genom en prioriterad kö. Annan trafik som är mindre känslig för fördröjning eller paketförluster, kan sorteras i en annan kö som vidarebefordras i mån av tid och tillgänglig kapacitet. En stor mängd mer eller mindre sofistikerade köhanteringsalgoritmer förekommer, vilka implementeras i olika grad av olika tillverkare av utrustning och programvara.

Vidare finns rekommendationer kring hur dessa DSCP-värden ska sättas för att generellt kunna kopplas till en specifik typ av köhantering, till exempel just en prioritetskö. Då detta enbart är rekommendationer är det fortfarande sällan något som transparent överförs mellan olika operatörer och IP-nät vid samtrafik. Däremot är det tämligen vanligt att någon form av DiffServ-implementation finns i operatörernas nät för att säkerställa att till exempel just realtidskänsliga tjänster och inte minst den egna administrativa trafiken inom nätet får företräde vid en eventuell överbelastning.

## 8.3 Tillgångskontroll

Ett svårt område att hantera när det gäller IP-telefoni är just resursutnyttjandet när det handlar om flera samtidigt pågående sessioner inom en och samma IP-förbindelse. Inom traditionell telefoni med ISDN-trunkar fanns ett fast antal tidsluckor som kunde användas för sessioner. När en session var etablerad var resurserna garanterade för det specifika samtalet under hela samtalets gång. Försökte en abonnent etablera ett nytt samtal när samtliga resurser redan var allokerade så levererades en spärrton och samtalet kunde inte

kopplas upp. Motsvarande funktionalitet finns inte naturligt i IP-nätet då det inte förekommer något begrepp som anger när en anslutning är fullbelastad på samma sätt. För IP-telefoni som använder traditionella talkodtyper, till exempel [G.711] (se även avsnitt 4.4.1), går det att på ett tämligen enkelt sätt beräkna hur många samtidigt pågående samtal en viss IP-förbindelse har förmåga att hantera. För fasta nät med hierarkiska strukturer går det därför tämligen enkelt att implementera en samtalsräknare som på signaleringsnivå spärrar nya samtalsförsök när resurspoolen för någon av de inblandade IP-länkarna är fullutnyttjad. Det finns dock ingen egentlig koppling till IP-nätet i sig utan det är enbart en tjänst på applikationsnivå som styr samtalsetableringen på detta sätt. Utan en sådan funktion finns det heller inget som förhindrar att ett ytterligare samtal etableras på en redan fullutnyttjad IP-förbindelse, vilket kan medföra att samtliga pågående sessioner får sämre kvalitet, så till den grad att de kan komma att bli omöjliga att använda. Att införa någon form av begränsning av antalet samtidiga samtal går under benämningen *Call Admission Control* (CAC).

I distribuerade och sammankopplade nät, där det finns flera möjliga transportvägar mellan ändpunkterna, blir det genast än svårare att på ett effektivt och korrekt sätt implementera sådan tillgångskontroll. Ett sätt att åtminstone till viss del hantera problematiken är att använda dynamiska talkodtyper som anpassar sig till rådande omständigheter. I den mån så pass många samtidiga sessioner etableras att det medför att överföringskapaciteten mellan de kommunicerande parterna inte räcker till, kan talkodtypen växla ned till en mer smalbandig variant på bekostnad av upplevd samtalskvalitet.

*Tillgångskontroll i mobila miljöer*

I mobila miljöer är det ofta särskilt komplicerat att på ett korrekt sätt tillämpa tillgångskontroll för samtal. Samtidigt är det också särskilt viktigt i dessa miljöer, som ofta lider av just begränsade resurser i fråga om IP-transport. De skiftande underliggande transmissions- och transportnäten medför att en nod i ena stunden kan ha god tillgång till IP-kapacitet, men i nästa stund måste traversera förbindelser med mycket låg överföringskapacitet, där endast ett starkt begränsat antal samtidiga samtalssessioner kan pågå.

Till viss del kan tjänstekvaliteten förbättras med statisk DiffServ-implementation som mjukt reserverar viss kapacitet till realtidskritiska tjänster som taltrafik. Detta hindrar dock på intet vis att fler samtal än vad som resursmässigt ryms etableras över dessa lågkapacitetslänkar. Tillsammans med dynamiska och adaptiva talkodtyper kan problematiken till viss grad lindras, emellertid utan garantier. För att kunna garantera resurser kan IntServ med RSVP vara en möjlig väg även om det ställer stora krav på transparens mot underliggande förmedlingsnoder.

## 8.4 Att mäta kvalitet

Att mäta kvaliteten i ett IP-telefonisamtal kan göras på olika sätt och är ofta både kompicerat och subjektivt. Inom traditionell telefoni används sedan länge en subjektiv modell för att bedöma kvaliteten på ett samtal, nämligen *Mean Opinion Score* (MOS). MOS är i princip en skala från 1 till 5, där värdet 5 är högsta upplevda talkvalitet och definieras ITU-T. I princip sitter ett antal referenspersoner och lyssnar på lämpliga fraser och gör sedan en bedömning av upplevd talkvalitet.

ITU-T *Perceptual Evaluation of Speech Quality* (PESQ) är en vidareutveckling av MOS-modellen där automatiserade tester ersätter referensgrupper. PESQ använder talströmmar som ingångsvärden då många telefonisystem är anpassade för att hantera enbart de frekvenser som det mänskliga talet omfattar. Att då använda toner och brus istället för riktiga talsekvenser kan ge märkliga resultat som inte heller speglar den upplevda verkliga kvaliteten i nätet. PESQ-modellen är idag en väl vedertagen modell som används av både teleoperatörer och produkttillverkare för att mäta upplevd talkvalitet på nät och utrustning. För att använda PESQ-modellen krävs att tekniken licensieras.

Både MOS-modellen och PESQ-modellen går även att tillämpa på IP-telefoni, men för IP-telefoni kan parametrar från mediaströmmarna enklare användas till att beräkna ett upplevt MOS-värde. Eftersom mediaströmmarna (RTP) kontinuerligt rapporterar kvalitetsparametrar med hjälp av mediaströmmens kontrollprotokoll, RTCP, så kan värden som fördröjning, jitter och paketförlust användas för att beräkna ett MOS-värde. De flesta protokollanalysatorer kan direkt med hjälp av mjukvarustyrda MOS- och PESQ-modeller redovisa ett beräknat MOS-värde på en mediaström under

förutsättning att samtlig trafik för den mediaströmmen finns inspelad.

# 9 Nätverkspåverkan

För traditionell telefoni inom kretskopplade nät är det underliggande transmissionsnätet hårt kopplat till telefonitjänsten. Det är i princip ett och samma nät. För IP-förmedlad telefoni är däremot nätverket betydligt mer frikopplat från själva telefonitjänsten. IP kan transporteras över en rad olika nättyper och skapar en abstraktionsnivå som gör det möjligt för tillämpningar högre upp i tjänstehierarkin att fungera oavsett underliggande nätverkstyper. Från ett konceptuellt perspektiv behöver en applikation eller tjänst som nyttjar IP som bärare inte ha någon kännedom om underliggande infrastruktur. I praktiken förekommer likväl många faktorer som medför att underliggande infrastruktur sätter begränsningar i det IP-baserade nätverkets förmåga att överföra information på ett för varje tjänst tillfredställande sätt.

## 9.1 Anslutningars egenskaper

Olika anslutningar med olika egenskaper kan indelas i fyra huvudkategorier enligt följande [IP-Handl]:

- Höghastighetsförbindelser
- Fjärrförbindelser
- Lågkapacitetsförbindelser
- Satellitförbindelser

Olika förutsättningar råder för var och en av dessa typförbindelser i frågan om att överföra talkommunikation i realtid. Även om en tillämpning inte direkt har någon kännedom om IP-nätets karakteristik och begränsningar är det viktigt att förstå dess inverkan på den tjänst som nätverket är tänkt att leverera. En tillämpning eller tjänst

som är utvecklad och testad i ett lokalnät med mycket hög kapacitetstillgång via höghastighetsförbindelser, kan komma att visa sig vara mindre väl lämpad att användas i miljöer där endast mer begränsad överföringskapacitet finns tillgänglig.

### 9.1.1 Höghastighetsförbindelser

Höghastighetsförbindelserna, som typiskt realiseras genom lokalnät eller med optisk fiber över längre avstånd, håller en god kvalitet där jitter, fördröjning och paketförluster kan förekomma främst som en följd av hög belastning. Dataöverföringskapaciteten varierar från 10 Mbps upp till 40 Gbps, och fördröjning från mindre än 1 ms upp till cirka 10 ms.

För tillämpningen IP-telefoni innebär denna typ av nätverk sällan några problem, och så länge det inte finns länkar i nätet som frekvent är överbelastade krävs sällan någon särskild hantering av IP-telefonitrafik för att tillhandahålla en hög tjänstekvalitet.

### 9.1.2 Fjärrförbindelser

Fjärrförbindelser kännetecknas av att ha något lägre dataöverföringskapacitet, men framför allt längre fördröjning än höghastighetsförbindelsen. Dataöverföringskapaciteten kan även vara asymmetrisk, det vill säga olika överföringskapacitet uppströms jämfört med nedströms. En fjärrförbindelse kan till exempel realiseras genom mikrovågsteknik (radio), hyrda förbindelser med olika typer av digital anslutning över kopparkabel – till exempel *Digital Subscriber Line* (DSL) – eller höghastighetsförbindelser som sträcker över stora geografiska avstånd och därigenom påför längre fördröjning än cirka 10 ms.

För tillämpningen IP-telefoni tillför dessa typer av förbindelser möjliga begränsningar i att leverera en bra kvalitet på tjänsten. Dels genom att begränsad kapacitetstillgång kan medföra att IP-telefonitjänsten måste använda sig av smalbandiga talkodtyper som i många fall direkt innebär sämre ljudkvalitet i talströmmen. Om förbindelsen dessutom påför jitter och en inte försumbar andel paketförluster får många talkodtyper svårare att återskapa talströmmen i mottagaren vilket riskerar att ytterligare påtagligt försämra talkvaliteten. Slutligen medför även den begränsade överföringskapaciteten i

sådana förbindelser begränsningar gällande antal samtidigt pågående samtal, särskilt vid asymmetriska förbindelser (då talkommunikation är en symmetrisk tillämpning). Att implementera köhanteringsmekanismer och prioriteringsfunktioner vid sådana flaskhalsar i nätverket, tillsammans med användandet av dynamiska talkodtyper, kan till viss del lindra sådana kvalitetsproblem inom rimliga gränser.

### 9.1.3 Lågkapacitetsförbindelser

Med lågkapacitetsförbindelser avses digitala förbindelser med en överföringskapacitet från några kbps och upp till några Mbps. Begreppet inkluderar även traditionella modem, det vill säga en uppringd förbindelse där datakommunikationen omvandlas till analoga signaler kapabla att fångas upp och transporteras över ett traditionellt telefonisystem som om det vore röstkommunikation.

Vid användning av den här typen av länkar kan effekter av s.k. serialisering bli påtagliga, och medföra väsentlig påverkan på andra tjänster med vilka anslutningen delas. Paket fördröjs och jitter uppstår. Effekterna är särskilt påtagliga för realtidstrafik där till exempel röstkommunikation blandas med asynkrona och transaktionsbaserade tjänster som tar mycket överföringskapacitet i anspråk.

En annan typ av lågkapacitetsförbindelse med speciell karaktär är olika typer av mobila radiosystem. Förutom att radionät av naturliga skäl ofta har en tämligen hög grad av paketförluster och jitter, åtminstone i mobila miljöer, tillkommer ofta tekniska funktioner för att minska radiosystemets sårbarhet för störningar och öka dess täckningsgrad. Det kan vara fråga om funktioner för frekvenshoppning eller system som kontinuerligt anpassar sin utsändningseffekt till att kunna leverera just den nödvändiga signalen, men inte mer.

När det kommer till mobil IP-radio förekommer både system som är inriktade på att tillhandahålla punkt-till-punktförbindelser och system som är tänkta att fungera mer eller mindre som ett datalänknät mellan en flertal noder. I fallet punkt-till-punktförbindelser är anslutningen i första hand att betrakta som en länk mellan två nätverk eller anslutning av enstaka system till ett nätverk. För mer avancerade radiosystem som är tänkta att sammankoppla flera noder på ungefär samma sätt som ett datalänknätverk, men med den skillnaden att noder hela tiden rör sig, krävs dynamiska protokoll som hanterar trafikstyrningen i nätet. För anslutande enheter kan dessa

radiobaserade IP-nät i princip ses som ett lokalt IP-nät, dock ofta med begränsad kapacitet och en högre grad av intermittens, paketförluster och jitter.

För IP-telefoni påför dessa typer av IP-nät givetvis begränsningar då nätverkens och förbindelsernas karakteristik knappast är optimala för realtidskritisk dubbelriktad strömmande media. På liknande sätt som för övriga lågkapacitetsförbindelser kan vissa köhanterings- och prioriteringsfunktioner tillsammans med dynamiska och smalbandiga talkodtyper lindra vissa sådana problem. Även tekniska lösningar specifika för underliggande radionät, så som nyttjandet av IP-multicast för grupputsändning och andra mer systemspecifika lösningar på olika nivåer, kan medföra förbättrade förutsättningar för just tal.

### 9.1.4 Satellitförbindelser

Satellitbaserad kommunikation används för att från fasta eller mobila anläggningar kommunicera över långa avstånd. De flesta kommersiella kommunikationssatelliterna befinner sig i geostationär bana, *Geostationary Orbit* (GEO). Detta innebär att satelliterna till synes står still över en given punkt på jorden. Geostationära satelliter befinner sig på ungefär 3 500 mils avstånd och i en bana runt ekvatorn. För en geostationär satellit är den totala fördröjningen upp till satelliten och tillbaka omkring 240 ms. För en så kallad VSAT-länk, vilken kräver mindre antenner, men använder en markbaserad relästation ökar svarstiderna upp mot 900 ms.

För IP-telefoni medför den förhållandevis höga graden av fördröjning en försämring av den upplevda kvaliteten och kräver viss tillvänjning av användaren. Om fördröjningen änd-till-änd överstiger ca 200–250 ms är det inte ovanligt samtalsparterna talar i mun på varandra. För att minimera eventuella jitterproblem har satellitsystem ofta förmåga att beakta QoS-märkning på IP-telefonipaketen, för att på så sätt säkerställa att samtliga paket som tillhör en och samma session alltid placeras i jämna tidsluckor i transmissionssystemet. Jitter och paketförluster är då sällan något större problem för IP-telefonitillämpningen i satellitbaserade transmissionslösningar. Satellitförbindelser kan ha en förhållandevis hög tillgänglig överföringskapacitet men som ofta är en kostnadsdrivande faktor, och beror mycket på var i världen användaren befinner sig och när denne behöver ha

täckning.

## 9.2 Filtrering och adressöversättning

En av de absolut största tekniska utmaningarna vid införande av IP-telefoni är att lösa traverseringen av brandväggar och adressöversättningsfunktioner (*Network Address Translation* (NAT)) i IP-näten. Brandväggar och adressöversättning utgör i sig två olika problemområden för IP-telefoni, men i grund och botten handlar det om hur ändpunkter bakom dessa nätverksfunktioner kan nås av inkommande trafik från de externa nätverken. Generellt sett är detta inte något som standardiseringsorgan som IETF tar någon större hänsyn till vid utarbetandet av specifikationer, då de protokoll och tekniker tas fram inom ramen för organisationen förutsätter ett öppet änd-till-änd-transparent internet som bärare. Verkligheten när det gäller just IP-telefoni ser emellertid annorlunda ut.

För att förstå problematiken krävs en grundläggande förståelse för hur brandväggar och adressöversättningsfunktioner fungerar. I detta sammanhang förutsätts för diskussionens skull att en brandvägg enbart filtrerar trafik, det vill säga att den inte gör någon adressöversättning. Detta innebär att en klient som är ansluten bakom brandväggen har en publik IP-adress som är adresserbar från det externa nätverket. Däremot, om adressöversättning används, så har klienten som är ansluten bakom denna funktion en privat IP-adress, det vill säga en IP-adress som *inte* är adresserbar från något externt nätverk.

Av de olika signaleringsprotokoll för IP-telefoni som diskuterats i denna rapport använder samtliga RTP för att transportera media och samtliga lösningar har separata mediaöverförings- och signaleringsplan. Detta innebär att från ett brandväggs- och adressöversättningsperspektiv har de alla likvärdiga problemställningar och lösningar för att fungera genom sådana nätverksenheter. Därför kommer denna handledning belysa problematiken och lösningarna i generella termer och peka på tekniska lösningar endast för i första hand signaleringsprotokollet SIP, då det är det protokoll som troligen är mest förekommande i framtiden.

### 9.2.1 Brandväggssystem

Brandväggssystem är ofta installerade som avskiljare mellan olika interna (privata) och externa (publika) nätverkssegment. En brandvägg konfigureras för att genom filtrering skydda de interna nätverksresurserna från obehörig extern kommunikation, och enbart de tjänster och tillämpningar som behöver kunna ta emot trafik tillåts kommunicera i dessa riktningar. Inom det interna nätverket är det i många fall vanligt att klienter behöver kunna kommunicera med en stor mängd tjänster på externa nätverk, till exempel internet. Vanligen tillåter därför en brandväggslösning i princip all trafik som initieras från sådana klientnätverk mot externa resurser. Brandväggen upprätthåller sedan tillstånd på de från insidan initierade sessionerna och säkerställer att svarstrafik tillåts tillbaka från de kontaktade externa resurserna till klienten som har initierat kommunikationen. Detta innebär att det för varje kommunikationsflöde uppstår ett *titthål* (*pinhole*) i brandväggssystemet som tillåter den kontaktade externa parten att skicka svarstrafik till den initierande interna klienten. Beroende på hur brandväggen är konfigurerad och dess applikationslogik, kan brandväggssystemet analysera trafiken som flödar mellan parterna inte endast på nätverks- och transportnivå, i syfte att avgöra även om innehållet i trafiken är i enlighet med regelverket och på så sätt styra parternas kommunikation baserad på applikationsspecifik information.

### 9.2.2 Adressöversättning

Utöver den filtering som en brandväggsfunktion tillhandahåller, komplicerar adressöversättning ytterligare för IP-telefonitjänsten. Adressöversättningsfunktionen förutsätter att all trafik initieras från de interna nätverkssegmenten för att kommunikation överhuvudtaget ska kunna flöda ifrån externa resurser. Vanligen tilldelas interna nätverksresurser privata IP-adresser som inte är adresserbara från externa nätverk, vilket medför att trafik endast kan flöda i denna riktning då en intern klient först initierat en session. Adressöversättningsfunktionen etablerar då ett tillstånd på liknande sätt som ett brandväggssystem, men översätter såväl avsändande adress som portnummer, innan ett paket vidareförmedlas till den externa resursen. Detta innebär att samtliga externa tjänster som anropas kommer att

skicka svarstrafik tillbaka till adressöversättningsfunktionens publika IP-adress och adressöversättningsfunktionen kommer via en tillståndstabell förmedla svarstrafik till rätt initierande klient.

För många typer av vanligt förekommande transaktionsbaserade internettjänster och protokoll utgör detta inget större praktiskt problem, förutom att en adressöversättningsfunktion på ett effektivt sätt bryter den grundläggande principen om änd-till-änd-transparens som internet bygger på. Så länge tjänsten som anropas är adresserbar via publika adresser kan adressöversättningsfunktionen styra inkommande svarstrafik till rätt klient. Så länge samma sessionstillstånd används för all data som överförs mellan klienten, som till exempel vid användandet av *Hypertext Transfer Protocol* (HTTP), kan adressöversättningsfunktionen fortfarande styra dessa sessioner mellan parterna. Som framgått i tidigare delar av denna skrift fungerar IP-telefoni emellertid inte på samma sätt. Signaleringsprotokollet används för att etablera, modifiera och terminera själva sessionen medan mediaprotokollet används för att överföra mediaströmmen änd-till-änd. Detta komplicerar IP-telefonins tillvaro i samband med brandväggar och adressöversättning.

Beroende på adressöversättningsfunktionens implementation kan en sådan funktion även agera mer eller mindre strikt. Vissa varianter tillåter en klient ansluten via ett internt nätverk, och som har initierat trafik till en specifik IP-adress på ett externt nätverk, att ta emot inkommande trafik även från andra externa adresser. Detta innebär att när väl klienten har kommunicerat med *någon* publik adress så kan den ta emot trafik från *många* publika adresser, det vill säga adressöversättningsfunktionen håller samma association oavsett vilken extern adress som den interna klienten kommunicerar med. Andra mer strikta implementationer håller olika associationer baserat på kombinationen av intern klient och extern adress så att varje kommunikationssession mellan den interna klienten och en specifik extern adress är unik. Detta innebär att en intern klient *enbart* kan ta emot inkommande trafik från just den externa adressen som den har etablerat en session med.

Gemensamt för dessa typer är att varje tillstånd har en giltighetstid associerad med tillståndet, och när denna giltighetstid löpt ut stängs sammankopplingen mellan den interna klienten och de externa resurserna, innebärande att kommunikation inte längre kan flöda från och till klienten förrän den på nytt initierar en kommunikationsväg. Då

skapas ett nytt tillstånd och en ny association.

### 9.2.3 Problembild

Den huvudsakliga problematiken för IP-telefoniprotokollen som uppstår i samband med brandväggar och adressöversättningsfunktioner relaterar till nåbarheten och i separationen av signaleringsoch mediatrafik. Eftersom mediaparametrar förhandlas fram under sessionsetableringen finns ingen praktiskt möjlighet för en brandvägg eller adressöversättningsfunktion att ha statiska regler eller att kunna tolka förhandlingen och därmed dynamiskt öppna de portar för inkommande trafik som krävs för att etablera sessionen. Ändpunkter för IP-telefoni har av naturliga skäl oftast också olika externa ändpunkter för signalering respektive för media, då signaleringen ofta traverserar flertalet noder på vägen mellan avsändare och mottagare, medan mediaströmmen går direkt mellan dessa.

När en SIP-ändnod registreras mot sin respektive proxy anger klienten vilken IP-adress som den för stunden är nåbar via. Om klienten då är ansluten bakom en adressöversättningsfunktion och proxyn via en publik adress, kommer ändnoden att meddela proxyn att den är nåbar via en IP-adress som denna inte kan adressera. Detta innebär att kommunikation som proxyn försöker initiera till ändnoden inte kommer att nå fram. På motsvarande sätt uppstår problem då ändnoden söker etablera ett samtal. I SDP-informationen beskrivs bland annat av ändnoden vilken IP-adress och port som mediaströmmen ska sändas till. Om detta är en privat IP-adress kommer mottagande ändnod inte kunna adressera och skicka mediatrafik till denna.

Då båda kommunicerande parter finns bakom samma adressöversättningsfunktion så kan dessa naturligtvis kommunicera med varandra. Däremot, då till exempel en intern proxyserver ska kommunicera med andra externa proxyservrar, uppstår samma problem. I dessa fall kan brandväggen och adressöversättningsfunktionen konfigureras statiskt med vidarebefordring av portar på sådant sätt att SIP-signalering som inkommer från externa resurser alltid förmedlas vidare till den interna proxyn. För media är det dock svårare, såvida inte servern även agerar mellanhand för alla mediaströmmar, att statiskt vidarebefordra signalerings- och mediatrafik till ett större antal interna ändnoder. Det skulle snabbt bli ett administrativt elände

och svårt att få att fungera tillförlitligt i praktiken.

### 9.2.4 Traverseringstekniker

I princip finns tre olika tillvägagångssätt för att lösa de problem som uppstår i att traversera brandväggs- och adressöversättningsfunktioner.

Den första varianten är att brandväggen eller adressöversättningsfunktionen har inbyggd applikationslogik för att tolka signaleringsprotokollet, så att den har förmåga att etablera de sidotillstånd som förhandlats fram. Denna typ av lösning kallas *Application Layer Gateway* (ALG), och kan anses vara den mest naturliga lösningen på problemet. Det finns dock en del nackdelar förknippade med denna lösning. En första är att applikationslogiken måste utvecklas i takt med att nya funktioner implementeras i till exempel SIP. Då en ALG i princip alltid måste agera som en B2BUA, det vill säga som en ändpunkt, måste den ha stöd för alla de funktioner som förhandlas fram och riskerar därmed att bli begränsande gällande änd-till-änd-funktionalitet. Ett annat problem från säkerhetssynpunkt är att en ALG måste kunna tolka all signaleringsinformation för att ha möjlighet att agera baserat på innehållet. Detta innebär att en ALG effektivt förhindrar all typ av kryptering av signaleringen. Fördelarna med en ALG-baserad lösning, under förutsättning att den fungerar tillfredställande, är att ändpunkterna på de interna nätverkssegmenten inte behöver någon särskild logik eller andra hänsyn till funktionen. Generellt sett, och från ett praktiskt perspektiv, är det inte ovanligt att en ALG-funktion medför lika hög grad av problem som den avsåg att lösa. Telefonioperatörer som tillhandahåller SIP-baserade tjänster ses ofta rekommendera kunderna att slå av lokal ALG, och istället förlita sig på serverbaserad traverseringsteknik.

Serverbaserad traverseringsteknik är vanligt bland tillhandahållare av allmänt tillgängliga telefonitjänster över publika IP-nät som internet. Den serverbaserade traverseringstekniken går även under benämning *Hosted NAT Traversal* (HNT) och implemeneteras ofta i *Session Border Controller* (SBC), då denna funktion ofta verkar i både signalerings- och mediaflödet. Denna lösning kräver resurser från teleoperatörens sida, men anses generellt vara den mest tillförlitliga traverseringstekniken, och framför allt är den under teleoperatörens kontroll. I princip bygger den serverbaserade traverseringstekniken

på att kundens brandväggssystem och adressöversättningsfunktion är konfigurerad på sådant sätt att trafik som initieras från anslutna ändnoder på interna nätverkssegment tillåts att kommunicera med IP-adresser på telefonioperatörens nätverk. Det förutsätts vidare att när sessionen initieras från ändnoderna till de externa resurserna, får dessa sända svarstrafik tillbaka till ändnoderna på de interna nätverken. Den proxy som tar emot SIP-signaleringen kommer att jämföra IP-adresserna i SIP-signaleringen med de adresser som i IP-protokollhuvudet anges som avsändande IP-adresser. Om dessa skiljer sig åt är det en tydlig indikator på att ändpunkten är ansluten bakom en adressöversättningsfunktion och då måste den serverbaserade traverseringstekniken träda in. I ett första steg skickar mottagande SIP-proxy svarstrafik till samma IP-adress och port som meddelandet kom in på *istället* för det som verkligen står i de SIP-huvuden som i vanliga fall ska styra trafiken. På samma sätt måste även SIP-proxyn ändra mediaadresserna i SDP och agera mellanhand för mediaströmmen mellan ändpunkterna. Det innebär alltså att även media måste skickas via dessa publika adresser som agerar mellanhänder (*relay*) för mediatrafiken. Dessa mellanhänder säkerställer även att om en ändpunkt bakom en adressöversättningsfunktion har indikerat i SDP att den avser ta emot mediaströmmen på en viss port och adress, så förmedlar mellanhanden tillbaka mediatrafiken på samma adress och port som den har fått in mediatrafik på. Detta kallas *media latching*. För att detta ska fungera måste ändpunkterna stöda symmetriska mediaströmmar, innebärande just att ändpunkterna stöder att ta emot mediaströmmar på samma adress och port som den skickar ut mediaströmmar på. En ytterligare förutsättning för att serverbaserad traversering för media ska fungera är att ändpunkten ansluten bakom adressöversättningsfunktionen skickar ett första RTP-paket och på så sätt etablerar tillståndet i adressöversättningsfunktionen. Först då kan den serverbaserade lösningen sända svarstrafik till ändpunkten via denna kanal och därigenom upprätta ett dubbelriktat trafikflöde änd-till-änd. Den serverbaserade lösningen har fördelen att ändpunkter och slutkunder inte behöver ha något ytterligare stöd implementerat, annat än just dessa krav på att tillåta utgående trafik och stöd för symmetrisk media. Nackdelen är att media inte kommer att ta en direkt väg mellan ändpunkterna utan kommer att förmedlas via dessa mellanhänder. Å andra sidan är media direkt mellan två ändpunkter bakom adressöversättnings-

funktioner inte möjlig. Den serverbaserade lösningen måste också ha förmåga modifiera SDP-informationen för att på så sätt infoga sig själv som mellanhand. Detta innebär också att det inte går att använda änd-till-änd-kryptering av SIP-signaleringen.

Den sista traverseringstekniken är egentligen den enda standardiserade varianten. Inom IETF har mycket tid ägnats åt att ta fram ett standardiserat ramverk för att lösa problem med filtrering och adressöversättning. Resultatet benämns *Interactive Connectivity Establishment* (ICE)[RFC5245]. ICE är egentligen en kombination av flera olika traverseringstekniker, som i vissa fall liknar den tidigare beskrivna serverbaserade lösningen. De huvudsakliga komponenterna i ICE är *Session Traversal Utilities for NAT* (STUN)[RFC5389] och *Traversal Using Relays around NAT* (TURN)[RFC5766].

STUN används i dess enklaste form av en SIP-ändnod för att ta reda på den publika IP-adress som adressöversättningsfunktionen använder. En SIP-klient som är ansluten bakom en adressöversättningsfunktion kan med hjälp av en STUN-förfrågan till en publik STUN-server få tillbaka ett svar som innehåller information om var SIP-klientens STUN-fråga såg ut att härstamma ifrån. Det vill säga, STUN-svaret innehåller den publika adress och port som adressöversättningsfunktionen associerat med den interna SIP-klienten för att kommunicera med externa resurser. Ändnoden kan då ange denna adress och port direkt i dess SIP- och SDP-huvuden så att externa ändpunkter kan använda dessa uppgifter för att direkt kommunicera med SIP-klienten på det interna nätet. STUN har dock begränsningen att enbart fungera med enklare former av adressöversättning där associationen mellan den interna klienten och externa resursen tillåter att alla externa adresser kan skicka trafik till klienten. Om detta inte är möjligt, det vill säga att adressöversättningsfunktionen är striktare implementerad, kommer STUN inte att fungera.

Istället måste då trafik, likt den serverbaserade lösningen, förmedlas via mellanhänder med publika adresser. Eftersom en striktare adressöversättningsfunktion endast tillåter den interna klienten att ta emot trafik från den externa adress och port som kontaktats, måste all trafik som klienten ska ta emot härstamma från denna adress och port. Det är det som både den serverbaserade lösningen och TURN åstadkommer. Det innebär att all media kommer speglas via TURN-servern på motsvarande sätt som via den serverbaserade lösningen.

Ramverket ICE är ett tillvägagångssätt för ändpunkterna att på ett strukturerat sätt utbyta information om vilka traverseringsalternativ som står till buds. Ändpunkterna utbyter *kandidater* som tjänar som möjliga alternativ för hantering av signalering och media och på så sätt kan ändpunkterna prova sig fram till ett alternativ som fungerar. TURN är ofta sista utvägen och används i princip alltid när båda ändunkter i sessionen befinner sig bakom olika adressöversättningsfunktioner eller brandväggar. Tyvärr kan ICE-förhandlingen leda till förlängda uppkopplingstider, ibland flera sekunder, då samtliga alternativa kommunikationskanaler måste utvärderas innan slutgiltig kandidat används. För WebRTC har därför ett tillägg till ICE tagits fram, Trickle-ICE, som medför ett snabbare förfarande och möjlighet till förkortad etableringstid.

# 10 Rekommendationer

Avsikten med detta avslutande kapitel i handledningen är att ge läsaren ett antal övergripande rekommendationer till ställningstaganden som kan behöva göras inför och under ett implementationsprojekt. Då olika implementationsprojekt kan ha drastiskt skilda förutsättningar ges heller inte rekommendationer för teknikval och liknande. Meningen att de rekommendationer som ges här ska kunna tillämpas vid såväl större som mindre implementationsprojekt.

## 10.1 Kravinhämtning

Ett av implementationsprojektets kanske viktigaste punkter gäller kravinhämtning. De krav som ställs på IP-telefonisystemet gällande både tjänster, användande och inte minst tillgänglighet, är ofta styrande för valet av tekniker, produkter och systemarkitektur. Det är därför av stor vikt att tidigt fånga upp de krav som finns både från användare och systemägare för att dels kunna välja rätt lösning och dels nå önskat resultat vid införandet.

För kapacitetsdimensionering av systemet är det viktigt att ta hänsyn till både antalet användare och antalet enheter. Det är även viktigt erhålla en övergripande förståelse för de trafikmönster som kan komma att uppstå, i syfte att dimensionera systemet rätt. Givetvis innefattar detta även en blick in i framtiden för att fånga upp eventuella möjliga förändringar och trender, så att inte systemägaren kort efter leverans behöver börja med att utöka systemet.

Valet av IP-telefoner och andra enheter är också viktigt att ta med i kravinhämtningen. Till exempel är förutsättningarna olika för fasta bordsplacerade telefoner och mjukvarutelefoner som följer med användarnas datorer. Det är vanligen lättare att hantera fasta telefoner när det gäller centraliserad administration, jämfört med sådana mjukvarutelefoner som är installerade i användarnas datorer. Det kan

också bli fråga om olika förutsättningar kring hur dessa IP-telefoner kommunicerar med telefoniservrarna i systemet. Fasta telefoner kan enklare anslutas till en separat infrastruktur för IP-telefonitrafiken, medan mjukvarutelefoner som körs på användarens vanliga dator medför att IP-telefonitrafik i praktiken måste blandas med övrig datatrafik.

Viktigt är också att tidigt försöka skaffa sig en bild av vilka telefonispecifika tjänster som IP-telefonisystemet behöver stödja. Beroende på produktval implementeras dessa telefonitjänster olika. Vissa traditionella telefonitjänster är svårare att implementera i IP-telefonisystem, åtminstone SIP-baserade sådana, medan andra tjänster är betydligt enklare att realisera i ett IP-telefonisystem. Används SIP som signaleringsprotokoll finns många tjänster implementerade direkt i telefonen, och kräver därför inget direkt stöd i telefoniservrarna.

Ofta behövs emellertid ytterligare applikationslogik för att realisera tjänster och då finns denna logik vanligen implementerad i telefoniservrarna.

För IP-telefoni smälter dessutom fler tjänster samman jämfört med traditionella telefonisystem. Exempel innefattar indikering av tillgänglighet (*eng. presence*), direktmeddelanden, videosamtal, indikering av röstmeddelande, och så vidare. De blir betydligt enklare att integrera då samma protokoll kan användas för flera av dessa tjänster. Värt att notera är att till exempel tillgänglighetsindikeringstjänster kan vara ganska kapacitetskrävande i större system med ett stort antal användare, då det i flera implementationer är tämligen signaleringsintensivt.

Övriga tjänster som telefonist, hänvisningssystem, med flera, kräver ofta särskild granskning och kravanalys. Ofta berör sådana tjänster även externa system som kräver integration. Det är viktigt att skaffa sig förståelse för hur dessa eventuella system interagerar med resterande delar av IP-telefonisystemet, och hur det passar in i vald design.

Det är även viktigt att fånga upp särskilda krav på dataöverföringar via fax och modem. Som tidigare beskrivits i avsnitt 4.7 är detta ofta problemområden inom IP-telefonitekniken, då de vanligaste lösningarna bygger på att överlagra dataförbindelsen över RTP-strömmen. Denna är känslig för störningar i IP-nätet och fax och modem kan därför lätt uppleva tjänstestörningar om hänsyn inte

har tagits till dessa tjänster redan från början. Om möjligt bör fax- och modemrelätekniker användas för att överföra dessa tjänster i IP-nätet. Reläfunktionerna ger ofta bättre kvalitet och högre överföringshastighet, men lider inte sällan av interoperabilitetsproblem mellan olika tillverkares implementationer. Man kan inom ramen för implementationsprojektet även se över om andra modernare tekniker kan användas i arbetsflödena för att helt ersätta fax och modemlösningar.

Slutligen kommer IP-telefonisystemet troligen också behöva kommunicera med externa system. Oavsett om dessa är andra privata eller publika telefonisystem så måste trafiken utbytas via något gränssnitt. Frågan är huruvida detta utbyte ska ske via traditionella kretskopplade gränssnitt eller direkt över IP. Många kommersiellt tillgängliga IP-telefonisystem har funktioner för att omvanda IP-telefonitrafiken till standardiserade traditionella telefonigränssnitt som ISDN PRI eller enkel POTS. Det blir dock allt vanligare att även externa system använder IP och SIP för att förmedla trafiken. Fördelen med detta är ökad tjänstetransparens och att systemägaren slipper införa kostnadsdrivande bryggningsfunktioner. Fördelen med att använda traditionella kretskopplade gränssnitt är att det blir en tydlig gränslinje mellan det interna IP-telefonisystemet och externa system. Det bör nämnas att utvecklingen hos teleoperatörer tydligt går i riktning mot SIP och IP-baserade lösningar vilket medför att traditionella kretskopplade gränssnitt antagligen på sikt kommer att försvinna mer och mer.

## 10.2 Säkerhetsaspekter

När den funktionella kravfångsten är gjord behöver även en hot- och sårbarhetsanalys genomföras, för att på en hög nivå identifiera de risker som införandet av telefonilösningen kan komma att medföra. Riskerna ska sedan lindras till acceptabla nivåer genom utformning av säkerhetskontroller inom ramen för implementationsprojektet.

Det är lämpligt att riskanalysen även omfattar infrastrukturen som ska bära IP-telefonitjänsten, såväl som andra medel för röstkommunikation. Det behöver knappast påpekas att de risker som identifieras även måste ställas i relation till de risker som finns i befintliga och möjligen andra kompletterade medel för röstkommunikation (till exempel mobiltelefoni). Riskanalysen behöver även innefatta de kostnader, både i form av direkta utgifter och minskad användbarhet,

som säkerhetskontrollerna kan komma att medföra, och det är därför viktigt att kriterier för riskacceptans fastställs.

Vid utvärderingen av riskerna är det viktigt att se till hur man hanterar prioriterade områden. Beroende på hur kritisk telefonitjänsten är för verksamheten kan riskerna hanteras på olika sätt. Att planera för alternativa medel för röstkommunikation, som kan tillgodose organisationens primära behov vid svåra störningar i den primära IP-telefonitjänsten, kan ofta visa sig vara en kostnadseffektiv lösning i jämförelse med att försöka höja tillgänglighetsgraden i det egna IP-telefonisystemet. Även kritiska beroenden till externa leverantörer kan behöva lindras. Till exempel kan en organisation som i vanliga fall köper PSTN-anslutning via en *SIP-trunk* som levereras av en specifik operatör ha en alternativ trunk-anslutning, eller enstaka SIP-abonnemang, hos en annan operatör över internet.

Resultatet av riskanalysen ska sedan ligga till grund för val av systemlösning och tillhörande säkerhetsmekanismer. Ett vanligt misstag är att först välja systemlösning, för att sedan genomföra riskanalysen och i efterhand försöka lindra de identifierade riskerna till acceptabla nivåer. Ofta med begränsad framgång, med högre kostnader och för verksamheten förhöjd risk som resultat.

I utformningen av säkerhetsåtgärder måste hänsyn förstås tas till funktionen hos de produkter och tjänster som finns kommersiellt tillgängliga. I praktiken är det till exempel mycket svårt finna leverantörer som har stöd för skyddad signalering på meddelandenivå. En privat IP-förbindelse där IP-telefonitrafiken är logiskt skild från övrig trafik kan ur vissa perspektiv ge något bättre förutsättningar för att skydda IP-telefonitrafiken från obehörig insyn och manipulation, men så länge den inte är krypterad är det i praktiken ett avtalsreglerat skydd, inget tekniskt.

En bra ansats vid säkerhetskravställningen av ett IP-telefonisystem är att systemet ska vara robust nog och ha förmåga att på egen hand stå emot intrångsförsök och andra typer av externa hot, och att inte förlita säkerheten på att systemet är isolerat från alla typer av nätverksdrivna angrepp. Mot bakgrund av de trender som syns, där IP-telefonisystem blir allt mer uppkopplade direkt mot externa parter, är det ett rimligt antagande att ett system som införs relativt isolerat och logiskt separerat med tiden kan komma att behöva kommunicera direkt med andra icke-betrodda parter.

Även kommunikation med sådana parter som betraktas som

betrodda, till exempel produktleverantörer och systemintegratörer, kan kräva kontrollmekanismer som gör att hela systemets säkerhet inte beror på dessa parters egna system och rutiner.

## 10.3 Systemarkitektur

Mycket av utfallet från de tidigare rekommendationspunkterna har betydande inverkan på systemarkitekturen. Men utöver detta tillkommer även specifika produktleverantörers valda lösningar, där systemägaren själv bör ges en förståelse för hur olika arkitekturer påverkar IP-telefonitjänsten både inledningsvis och framgent.

Beroende på lösningens tänkta storlek kan en av de större frågorna vara huruvida samtliga abonnenter ska tillhöra ett och samma system eller om de ska delas in i flera mindre autonoma system. Principiellt kan man jämföra detta beslut med om IP-telefonitjänsten ska implementeras som ett centraliserat eller som ett distribuerat system. Ett centraliserat system har ofta fördelar från drift och underhållsperspektiv, medan ett distribuerat system framför allt har sina fördelar i robusthet genom möjligheten till fortsatt lokal funktionalitet även om andra delar av infrastrukturen är utsatt för störningar. Detta beslut bör dels luta sig mot resultatet från den tidigare genomförda riskanalysen, men kan också vara av mer en principiell natur, till exempel beroende på organisationens uppbyggnad och dess geografiska spridning över olika tidszoner, et cetera. Det är emellertid viktigt att påpeka att ett centraliserat system mycket väl kan uppfylla versamhetens krav på tillgänglighet och robusthet, och att säkerställande av dessa egenskaper kan göras på en rad olika sätt.

Ett annat ställningstagande är valet mellan öppna eller proprietära systemlösningar när det till exempel gäller signaleringsprotokoll. I dag har i praktiken de flesta kommersiellt tillgängliga IP-telefonisystem åtminstone externa gränssnitt som följer de öppna och standariserade protokollspecifikationerna, företrädesvis *Session Initiation Protocol* (SIP). Huruvida de interna protokollen, det vill säga de protokoll och funktioner som används inom IP-telefonisystemet, mellan dess servrar och abonnenter, ska vara öppna eller proprietära är en svårare fråga att besvara. Från ett systemperspektiv kan det vara mindre viktigt om systemägaren vid valet av produkt också har för avsikt att fortsätta använda produkter från samma leverantör, till exempel vid utökning av antalet terminaler och abonnenter. Används

öppna protokoll även internt ges emellertid möjlighet att komplettera med andra produkter som använder dessa öppna protokoll, vilket av naturliga skäl ger en ökad framtidssäkring av IP-telefonilösningen som helhet. Å andra sidan kan det mycket väl finnas specifika tjänster och funktioner som inte kan realiseras på önskvärt sätt med de öppna gränssnitten och kräver då proprietära lösningar.

På samma sätt som till exempel teleoperatörer ofta väljer att begränsa stödet för olika mediatyper (*video, HD Voice, med mera*) i syfte att kunna garantera funktionalitet och viss nivå av kvalitet på sina tjänster, bör systemägaren av ett IP-telefonisystem ta ställning till huruvida liknande begränsning och styrning ska tillämpas. I vissa fall kan det vara önskvärt att begränsningar införs för att kontrollera IP-telefonitjänstens kapacitetsnyttjande i nätet. Till exempel kan videosamtal selektivt behöva blockeras, då videoströmmar är betydligt mer kapacitetskrävande än de flesta vanliga talströmmarna, och samtidigt enbart tillåta ett fåtal talkodtyper med kända kapacitetskrav. Det samma gäller för funktioner som federering, det vill säga möjligheten att över publika IP-nät, till exempel internet, direkt kunna etablera IP-telefonisamtal med andra federerande IP-telefonisystem genom nyttjandet av i första hand URI-baserad adressering. Att öppna upp ett IP-telefonisystem mot internet och andra federerande system är givetvis så som IP-telefoni och SIP är avsett och konstruerat att fungera. Det är då viktigt att riskanalysen tagit hänsyn till denna typ av exponering av systemet och de möjliga externa hotbilder som då aktualiseras.

## 10.4 Infrastruktur och kapacitet

Som flera gånger nämnts i denna handledning är IP-telefoni, till skillnad från traditionell telefoni, en av flera tjänster i en gemensam infrastruktur. Detta innebär att den underliggande infrastrukturen är en delad resurs som sannolikt samtidigt används av flera andra tillämpningar utöver telefonin. Det är därför av största vikt att den underliggande infrastrukturen dimensioneras för *samtliga* tjänster i nätet. Så som beskrivits i kapitel 8 är IP-telefonitjänster förknippade med vissa kvalitetskrav som skiljer sig från övrig datatrafik. Det är av största vikt att dessa krav omhändertas och hanteras på ett ändamålsenligt sätt.

För att kunna säkerställa att tillräcklig kapacitet finns tillgänglig i underliggande infrastruktur bör systemägaren ha en uppfattning om

hur många abonnenter och samtidigt pågående samtal som förväntas. Inom traditionell telefoni, där kapacitetsnyttjandet för varje samtal är ett konstant värde, beräknas mängden samtidigt pågående samtal baserat på antalet abonnenter och uppskattad samtalslängd under dygnets intensivaste timmar (*eng. busy hour*). På så sätt får man enligt vedertagna beräkningsmodeller och formler (*Erlang*) fram värden som infrastrukturen ska dimensioneras för.

Samma typ av beräkningar går att tillämpa för IP-telefoni, men det är viktigt att inse skillnaderna. Framför allt gäller skillnaderna det faktum att IP-telefoni mycket väl kan använda dynamiska talkodtyper, vilket innebär att kapacitetskraven och kapacitetsutnyttjandet för varje etablerat samtal kan variera med tiden. Även möjligheten till användning av flera parallella mediatyper, till exempel genom att en videoström adderas till befintligt röstsamtal, givetvis påverkar kapacitetsberäkningen i allra högsta grad. Dessutom är det viktigt att ta med i beräkningarna att IP-telefonitrafiken endast är en trafiktyp i en gemensam infrastruktur, där andra tjänster kan komma att påverka kapacitetstillgången. Generellt kan dock sägas att Erlang-beräkningar är tillämpbara på IP-telefoni så länge kända talkodtyper av statisk karaktär används (eller då beräkningen utgår ifrån en viss talkodtyps maximala kapacitetsförbrukning).

Ett av de val som kan komma att få störst påverkan på implementationsprojektet gäller hur man avser hantera IP-telefonitrafiken i nätet. Antingen samkörs denna trafik med övriga tjänster, eller så separeras telefonitrafiken logiskt. På lokal nivå kan detta göras genom att till exempel använda *Virtual Local Area Network* (VLAN)-teknik. Trafikprioritering, *Quality of Service* (QoS), kan då relativt enkelt implementeras för att säkerställa att IP-telefonitrafiken ges företräde före transaktionsbaserade tjänster i det lokala nätet. Ju större lokalt nät som används desto mer betydelse kan denna typ av trafikprioritering ha. Då telefonitrafiken flödar utanför det lokala nätet och till exempel över organisationsgränser, kan de olika trafiktyperna behöva blandas och trafikprioritering blir väsentligt svårare att säkerställa änd-till-änd.

Om IP-telefonisystemet ansluter till *Public Switched Telephony Network* (PSTN) med en så kallad *SIP-trunk* kan denna trunk mycket väl levereras över privata förbindelser, logiskt skild från övrig data- och internettrafik, med till exempel *Multiprotocol Label Switching* (MPLS). SIP-trunkar kan även köpas över internet från andra leve-

rantörer än de som levererar organisationens övriga anslutnings- och internettjänster, vilket medför att IP-telefonitrafiken samtrafikerar den gemensamma förbindelsen och traverserar därefter i många fall flertalet operatörer på vägen över internet till leverantören av SIP-trunken. I dessa fall kan det vara lämpligt att försöka införa QoS på IP-nivå (nivå 3), för att på så sätt prioritera IP-telefonitrafiken före övrig datatrafik i de delar av infrastrukturen där så är möjligt. Beroende på hur den bärande infrastrukturen ser ut kan detta vara mer eller mindre effektivt, då prioritering och klassificering av viss trafiktyp inte är något som brukar överföras eller implementeras mellan olika internetoperatörer.

# Förkortningar

**3GPP** 3rd Generation Partnership Project

**AF** Assured Forwarding

**ALG** Application Layer Gateway

**AMR** Adaptive Multi-Rate

**AMR-NB** Adaptive Multi-Rate Narrowband

**AMR-WB** Adaptive Multi-Rate Wideband

**ASN.1** Abstract Syntax Notation One

**B2BUA** Back-to-Back User Agent

**CAC** Call Admission Control

**CNG** Comfort Noise Generation

**CODEC** Coder Decoder

**DNS** Domain Name System

**DNSSEC** Domain Name System Security Extensions

**DSCP** Differential Services Code Point

**DSL** Digital Subscriber Line

**DSP** Digital Signal Processor

**DTLS** Datagram Transport Layer Security

**DTMF** Dual-tone Multi-frequency

**EF** Expedited Forwarding

**ENUM** E.164 to URI DDDS Application

**ETSI** European Telecommuniations Standards Institute

**FQDN** Fully Qualified Domain Name

**GEO** Geostationary Orbit

**GSM** Global System for Mobile Communications

**HNT** Hosted NAT Traversal

**HTTP** Hypertext Transfer Protocol

**IANA** Internet Assigned Numbers Authority

**ICE** Interactive Connectivity Establishment

**IETF** Internet Engineering Task Force

**IP** Internet Protocol

**IPsec** Internet Protocol Security

**ISDN** Integrated Services Digital Network

**ISUP** ISDN User Part

**LoST** Location-to-Service Translation Protocol

**MCU** Multipoint Control Unit

**MEGACO** MEdia GAteway COntroll Protocol

**MELP** Mixed-excitation linear prediction

**MGC** Media Gateway Controller

**MGCP** Media Gateway Control Protocol

**MGW** Media Gateway

**MIKEY** Multimedia Internet KEYing

**MOS** Mean Opinion Score

**MPLS** Multiprotocol Label Switching

**NAPTR** Name Authority Pointer

**NAT** Network Address Translation

**NSA** National Security Agency

**NTP** Network Time Protocol

**OTT** Over the Top

**PBX** Public Branch eXchange

**PESQ** Perceptual Evaluation of Speech Quality

**PGP** Pretty Good Privacy

**POTS** Plain Old Telephony System

**PRI** Primary Rate Interface

**PSTN** Public Switched Telephony Network

**PTS** Post- och Telestyrelsen

**QoS** Quality of Service

**RAS** Registration, Admission and Status

**RFC** Request for Comments

**RSVP** Resource Reservation Protocol

**RTCP** Real-time Control Protocol

**RTCP XR** RTCP Extended Reports

**RTP** Real-time Transport Protocol

**S/MIME** Secure/Multipurpose Internet Mail Extensions

**SAKKE** Sakai-Kasahara Key Encryption

**SAS** Short Authentication String

**SBC** Session Border Controller

**SCIP** Secure Communiations Interoperabillity Protocol

**SCTP** Stream Control Transmission Protocol

**SDES** SDP Security Descriptions

**SDP** Session Description Protocol

**SIGTRAN** Signaling Transport

**SIP** Session Initiation Protocol

**SMTP** Simple Mail Transfer Protocol

**SRTCP** Secure RTCP

**SRTP** Secure RTP

**SS7** Signaling System 7

**STUN** Session Traversal Utilities for NAT

**SVoIP** Secure Voice over IP

**TCP** Transmission Control Protocol

**TLS** Transport Layer Security

**TURN** Traversal Using Relays around NAT

**UAC** User Agent Client

**UAS** User Agent Server

**UDP** User Datagram Protocol

**UMTS** Universal Mobile Telecommunications System

**URI** Uniform Resource Identifier

**URN** Uniform Resource Name

**VAD** Voice Activity Detection

**VLAN** Virtual Local Area Network

**VoLTE** Voice over LTE

**VoSIP** Voice over Secure IP

**W3C** World Wide Web Consortium

**WebRTC** Web Real Time Communication

**XMPP** Extensible Messaging and Presence Protocol

**ZRTP** Zimmerman Real-time Transport Protocol

# Sakregister

# Referenser

[E.164] ITU-T. E.164: The international public telecommunication numbering plan.
http://www.itu.int/rec/T-REC-E.164-201011-I/en.

[G.711] ITU-T. G.711: Pulse code modulation (PCM) of voice frequencies, November 1988.
http://www.itu.int/rec/T-REC-G.711-198811-I/en.

[G.722] ITU-T. G.722: 7 kHz audio-coding within 64 kbit/s, September 2012.
http://www.itu.int/rec/T-REC-G.722-201209-I/en.

[G.729] ITU-T. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). G.729, Januari 2007.
http://www.itu.int/rec/T-REC-G.729-200701-I/en.

[IP-Handl] J. Strömbergsson F. Ljunggren, J. Schlyter. IP-baserade kommunikationsprotokoll. http://www.kirei.se.

[RFC1633] R. Braden, D. Clark och S. Shenker. Integrated Services in the Internet Architecture: an Overview. RFC 1633 (Informational), juni 1994.
http://www.ietf.org/rfc/rfc1633.txt.

[RFC2015] M. Elkins. MIME Security with Pretty Good Privacy (PGP). RFC 2015 (Proposed Standard), oktober 1996. Updated by RFC 3156.
http://www.ietf.org/rfc/rfc2015.txt.

[RFC2141] R. Moats. URN Syntax. RFC 2141 (Proposed Standard), maj 1997.
http://www.ietf.org/rfc/rfc2141.txt.

[RFC2205] R. Braden, L. Zhang, S. Berson, S. Herzog och S. Jamin. Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification. RFC 2205 (Proposed Standard), september 1997. Updated by RFCs 2750, 3936, 4495, 5946, 6437, 6780.
http://www.ietf.org/rfc/rfc2205.txt.

[RFC2474] K. Nichols, S. Blake, F. Baker och D. Black. Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. RFC 2774 (Proposed Standard), December 1998. Updated by RFC 3168, RFC 3260.
http://www.ietf.org/rfc/rfc2474.txt.

[RFC2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang och W. Weiss. An Architecture for Differentiated Services. RFC 2475 (Informational), december 1998. Updated by RFC 3260.
http://www.ietf.org/rfc/rfc2475.txt.

[RFC2543] M. Handley, H. Schulzrinne, E. Schooler och J. Rosenberg. SIP: Session Initiation Protocol. RFC 2543 (Proposed Standard), mars 1999. Obsoleted by RFCs 3261, 3262, 3263, 3264, 3265.
http://www.ietf.org/rfc/rfc2543.txt.

[RFC2597] J. Heinanen, F. Baker, W. Weiss och J. Wroclawski. Assured Forwarding PHB Group. RFC 2597 (Proposed Standard), juni 1999. Updated by RFC 3260.
http://www.ietf.org/rfc/rfc2597.txt.

[RFC3246] B. Davie, A. Charny, J.C.R. Bennet, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu och D. Stiliadis. An Expedited Forwarding PHB (Per-Hop Behavior). RFC 3246 (Proposed Standard), mars 2002.
http://www.ietf.org/rfc/rfc3246.txt.

[RFC3261] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley och E. Schooler. SIP:

Session Initiation Protocol. RFC 3261 (Proposed Standard), juni 2002. Updated by RFCs 3265, 3853, 4320, 4916, 5393, 5621, 5626, 5630, 5922, 5954, 6026, 6141.
http://www.ietf.org/rfc/rfc3261.txt.

[RFC3263] J. Rosenberg och H. Schulzrinne. Session Initiation Protocol (SIP): Locating SIP Servers. RFC 3263 (Proposed Standard), juni 2002.
http://www.ietf.org/rfc/rfc3263.txt.

[RFC3264] J. Rosenberg och H. Schulzrinne. An Offer/Answer Model with Session Description Protocol (SDP). RFC 3264 (Proposed Standard), juni 2002. Updated by RFC 6157.
http://www.ietf.org/rfc/rfc3264.txt.

[RFC3435] F. Andreasen och B. Foster. Media Gateway Control Protocol (MGCP) Version 1.0. RFC 3435 (Informational), januari 2003. Updated by RFC 3661.
http://www.ietf.org/rfc/rfc3435.txt.

[RFC3525] C. Groves, M. Pantaleo, T. Anderson och T. Taylor. Gateway Control Protocol Version 1. RFC 3525 (Historic), juni 2003. Obsoleted by RFC 5125.
http://www.ietf.org/rfc/rfc3525.txt.

[RFC3550] H. Schulzrinne, S. Casner, R. Frederick och V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (Standard), juli 2003. Updated by RFCs 5506, 5761, 6051, 6222.
http://www.ietf.org/rfc/rfc3550.txt.

[RFC3611] T. Friedman, R. Caceres och A. Clark. RTP Control Protocol Extended Reports (RTCP XR). RFC 3611 (Proposed Standard), november 2003.
http://www.ietf.org/rfc/rfc3611.txt.

[RFC3711] M. Baugher, D. McGrew, M. Naslund, E. Carrara och K. Norrman. The Secure Real-time Transport Protocol (SRTP). RFC 3711 (Proposed Standard), mars 2004. Updated by RFC 5506.
http://www.ietf.org/rfc/rfc3711.txt.

[RFC3830] J. Arkko, E. Carrara, F. Lindholm, M. Naslund och K. Norrman. MIKEY: Multimedia Internet KEYing. RFC 3830 (Proposed Standard), augusti 2004. Updated by RFCs 4738, 6309.
http://www.ietf.org/rfc/rfc3830.txt.

[RFC3966] H. Schulzrinne. The tel URI for Telephone Numbers. RFC 3966 (Proposed Standard), december 2004. Updated by RFC 5341.
http://www.ietf.org/rfc/rfc3966.txt.

[RFC4347] E. Rescorla och N. Modadugu. Datagram Transport Layer Security. RFC 4347 (Proposed Standard), april 2006. Updated by RFC 5746.
http://www.ietf.org/rfc/rfc4347.txt.

[RFC4474] J. Peterson och C. Jennings. Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP). RFC 4474 (Proposed Standard), augusti 2006.
http://www.ietf.org/rfc/rfc4474.txt.

[RFC4566] M. Handley, V. Jacobson och C. Perkins. SDP: Session Description Protocol. RFC 4566 (Proposed Standard), juli 2006.
http://www.ietf.org/rfc/rfc4566.txt.

[RFC4568] F. Andreasen, M. Baugher och D. Wing. Session Description Protocol (SDP) Security Descriptions for Media Streams. RFC 4568 (Proposed Standard), juli 2006.
http://www.ietf.org/rfc/rfc4568.txt.

[RFC4733] H. Schulzrinne och T. Taylor. RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals. RFC 4733 (Proposed Standard), december 2006. Updated by RFCs 4734, 5244.
http://www.ietf.org/rfc/rfc4733.txt.

[RFC4880] J. Callas, L. Donnerhacke, H. Finney, D. Shaw och R. Thayer. OpenPGP Message Format. RFC 4880 (Proposed Standard), november 2007. Updated by RFC

5581.
http://www.ietf.org/rfc/rfc4880.txt.

[RFC5031] H. Schulzrinne. A Uniform Resource Name (URN) for Emergency and Other Well-Known Services. RFC 5031 (Proposed Standard), januari 2008.
http://www.ietf.org/rfc/rfc5031.txt.

[RFC5222] T. Hardie, A. Newton, H. Schulzrinne och H. Tschofenig. LoST: A Location-to-Service Translation Protocol. RFC 5222 (Proposed Standard), augusti 2008. Updated by RFC 6848.
http://www.ietf.org/rfc/rfc5222.txt.

[RFC5245] J. Rosenberg. Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols. RFC 5245 (Proposed Standard), april 2010. Updated by RFC 6336.
http://www.ietf.org/rfc/rfc5245.txt.

[RFC5246] T. Dierks och E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246 (Proposed Standard), augusti 2008. Updated by RFCs 5746, 5878, 6176.
http://www.ietf.org/rfc/rfc5246.txt.

[RFC5389] J. Rosenberg, R. Mahy, P. Matthews och D. Wing. Session Traversal Utilities for NAT (STUN). RFC 5389 (Proposed Standard), oktober 2008.
http://www.ietf.org/rfc/rfc5389.txt.

[RFC5630] F. Audet. The Use of the SIPS URI Scheme in the Session Initiation Protocol (SIP). RFC 5630 (Proposed Standard), oktober 2009.
http://www.ietf.org/rfc/rfc5630.txt.

[RFC5750] B. Ramsdell och S. Turner. Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.2 Certificate Handling. RFC 5750 (Proposed Standard), januari 2010.
http://www.ietf.org/rfc/rfc5750.txt.

[RFC5751] B. Ramsdell och S. Turner. Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.2 Message Specifi-

cation. RFC 5751 (Proposed Standard), januari 2010.
http://www.ietf.org/rfc/rfc5751.txt.

[RFC5764] D. McGrew och E. Rescorla. Datagram Transport Layer Security (DTLS) Extension to Establish Keys for the Secure Real-time Transport Protocol (SRTP). RFC 5764 (Proposed Standard), maj 2010.
http://www.ietf.org/rfc/rfc5764.txt.

[RFC5766] R. Mahy, P. Matthews och J. Rosenberg. Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN). RFC 5766 (Proposed Standard), april 2010.
http://www.ietf.org/rfc/rfc5766.txt.

[RFC6116] S. Bradner, L. Conroy och K. Fujiwara. The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM). RFC 6116 (Proposed Standard), mars 2011.
http://www.ietf.org/rfc/rfc6116.txt.

[RFC6117] B. Hoeneisen, A. Mayrhofer och J. Livingood. IANA Registration of Enumservices: Guide, Template, and IANA Considerations. RFC 6117 (Proposed Standard), mars 2011.
http://www.ietf.org/rfc/rfc6117.txt.

[RFC6189] P. Zimmermann, A. Johnston och J. Callas. ZRTP: Media Path Key Agreement for Unicast Secure RTP. RFC 6189 (Informational), april 2011.
http://www.ietf.org/rfc/rfc6189.txt.

[RFC6455] I. Fette och A. Melnikov. The WebSocket Protocol. RFC 6455 (Proposed Standard), december 2011.
http://www.ietf.org/rfc/rfc6455.txt.

[RFC6509] M. Groves. MIKEY-SAKKE: Sakai-Kasahara Key Encryption in Multimedia Internet KEYing (MIKEY). RFC 6509 (Informational), februari 2012.
http://www.ietf.org/rfc/rfc6509.txt.

[RFC6716] JM. Valin, K. Vos och T. Terriberry. Definition of the Opus Audio Codec. RFC 6716 (Proposed Standard), september

2012.
http://www.ietf.org/rfc/rfc6716.txt.

[RFC7118] I. Baz Castillo, J. Millan Villegas och V. Pascual. The WebSocket Protocol as a Transport for the Session Initiation Protocol (SIP). RFC 7118 (Proposed Standard), januari 2014.
http://www.ietf.org/rfc/rfc7118.txt.

[SCIP] Secure Communication Interoperability Protocol (SCIP).
https://www.iad.gov/SecurePhone/.