ERICSSON
TAKING YOU FORWARD

TUDelft
Delft University of Technology

Ericsson, Netherlands

Faculty of Electrical Engineering, Mathematics and Computer Science

Network Architectures and Services

Video over Internet:

Analysis using SIP, RTP/RTCP Protocols

A.O. Ajagbe
(1535617)

Committee members:

Mentor	: Rogier Noldus, Msc.
Chairman	: dr. ir. F.A. Kuipers
Member	: Dr. E. Onur

October, 2010
M.Sc. Thesis No: PVM 2010 – 067

Acknowledgements

While working on my thesis, I have received immense help and support from staff at Ericsson, TU Delft as well as from my family and friends. From Ericsson, I would like to thank my thesis supervisor, Rogier Noldus for his guidance and advice and for always having time to answer all my numerous questions. I would also like to thank my parents, Mr. & Mrs. S.O. Ajagbe, for their encouragement and financial support throughout the course of my programme at TUDelft. My sincere gratitude also goes to my siblings, Opeyemi and Damilola and extended family for their encouragement and support. My appreciation also goes to Chief Hezekiah Adediji for his support. Lastly but definitely not the least, I thank my wife, Oluwatomilayo, for her love, support and patience that gave me the energy to complete the work presented in this report.

To my beloved ones, with whom I share my challenges and rewards

Abstract

The goal of this project is to analyze video calls over the Internet Protocol (Video over IP) using a simulated IMS environment (IMS-in-a-box). The thesis presents an overview of IMS, its purpose and how it has evolved. The thesis also attempts to give the reader a full and comprehensible understanding of what the IMS is all about, its purpose and why it came into existence.

The thesis considers a number of different video call case scenarios that are most likely to be encountered when making video calls over the Internet using the Internet protocol. This analysis focuses mainly on the SIP, RTP/RTCP protocols and how these three protocols are related and synchronized in order to actually know what is happening during the course of call set up and media exchange between the various end callers involved.

The report looks into aspects of bandwidth consumption by the exchanged media (RTP and RTCP), jitter and its variation over the duration of the calls and the cooperation between the control plane and the user plane in order for a smooth call set up, media exchange and release of network resources.

The thesis focuses on the areas of interest mentioned above, as these protocols have been identified as being of high significance in media transfer during video calls over IP.

Keywords: Voice over IP, IP Multimedia Subsystem, Video over IP, SIP, RTP/RTCP.

Contents

1. Introduction

1.1 Background

The concept of video telephony has been around for quite a number of years but only recently has it come to fruition. The basis technology required to transmit images and sound over the global communications network (Internet) was feasible, but the infrastructure required to support practical video telephony over this network was inadequate.

In recent years, there has been a remarkable achievement in availability of broadband Internet solutions such as DSL (Digital Subscriber Line, cable and land-based access networks which allows people to send massive amount of information in real time.

Also, the ability to compress and decompress video and audio data (in real time with perfect synchronization) over the present-day broadband networks using low-cost computer chips has also been achieved due to technological advancements in the world of telecommunications.

In this age of e-mail, instant and text messaging, video telephony shares the personal nuances that only come from experiencing face-to-face communications. Inflections, expressions, and other non-verbal cues that are lost in cyberspace are preserved with video telephony, helping reconnect people during life's important moments.

All these facts coupled together have actually made video telephony to be well embraced by the end-users thereby making the demand for this service to be on a steady increase.

Due to the increase in the demand for this service by the end users, the service providers must develop a way through which the quality of experience of the end users must be monitored in real-time in order to achieve a satisfactory delivery of service as evaluated by the customers.

1.2 Objective of the thesis and scope of work

In order to achieve this real-time monitoring of the video telephony network within the framework of the IMS (IP Multimedia Subsystem), an in-depth study of Voice over IP, IMS and Video over IP has been done. The three main protocols, SIP (Session Initiation Protocol), RTP (Real-time Protocol) and RTCP (Real-time Transport Control Protocol) have been chosen for data analysis since the RTP/RTCP actually convey the media information that are being transported over the networks with SIP being used as the protocol to provide signaling and session initiation between users who want to engage in media exchange.

The main goal of the thesis is to show that through the analysis of the mentioned protocols, network condition in real-time can be determined thereby showing the quality of user experience. Also, the effects of these protocols being out of sync are analyzed and possible solutions are given. Additionally, the best possible test network set-up is adopted in order for measurements to be done across the network and between entities involved.

1.3 Methodology and thesis structure

We start our research with a thorough study of theoretical background information about Voice and Video over IP, IMS and the various related protocols. Based on the knowledge from this background study, a test network was setup in order to run test calls between end-users. A number of test case scenarios were chosen in order to have as near as possible real life situations of video telephony sessions. By capturing all data traffic involved in each of the test case scenario using Wireshark (a network protocol analyzer), measurements were made, data filtered and analyzed.

The thesis is organized as follows. In chapter 2 we explain the basics of voice over IP which is actually the foundation of media exchange over the Internet. Here the various domains involved and the basic requirements in order to exchange media are explained. Chapter 3 provides insight into the IMS network which is the framework within which our measurements were made. The various protocols and entities involved in call establishment and media transportation over the Internet are explained. Chapter 4 focuses on Video telephony over the Internet, with emphasis on RTP/RTCP, which are the protocols used in transporting media and control information between end users. Chapter 5 describes the test bed setup, test cases used in the analysis and the method used to filter data traffic captured during measurements. Results and analysis are shown in chapter 6. Finally, chapter 7 contains the conclusions based on the research work and suggestions for future work.

2. Introduction to Voice over Internet Protocol (VoIP)

Video call establishment over IP and voice call establishment over IP have great similarity, so when understanding Voice-over-IP, we can understand Video-over-IP. Therefore in order to understand the basic concepts of Voice over Internet Protocol (VoIP), a step-by-step approach is adopted and gradually adding other functional elements needed for establishing a fully developed IP-based communication network.

This chapter focuses on the basics of VoIP with the various domains and entities involved. Furthermore, the registration procedures, signaling using SIP, the control plane and the user plane are described with their functionalities in session establishments and media exchange during call sessions.

2.1 Definition of the domains for communication over the Internet

Users that intend to establish communication through voice or other methods need to have Internet access as defined in figure 2.1, i.e. have internet connectivity through a service provider. The Internet service provider can be the VoIP operator, but this is not compulsory.

The internet access may be fixed or mobile e.g. General Packet Radio System (GPRS), Universal Mobile Telephony System (UMTS), High Speed Packet Access (HSPA) and Long Term Evolution (LTE). In order to have an efficient VoIP system, some conditions need be met like minimum bandwidth, maximum transfer latency, guaranteed data throughput and network resilience. These conditions constitute the Quality of Service (QoS) for (mobile) VoIP.

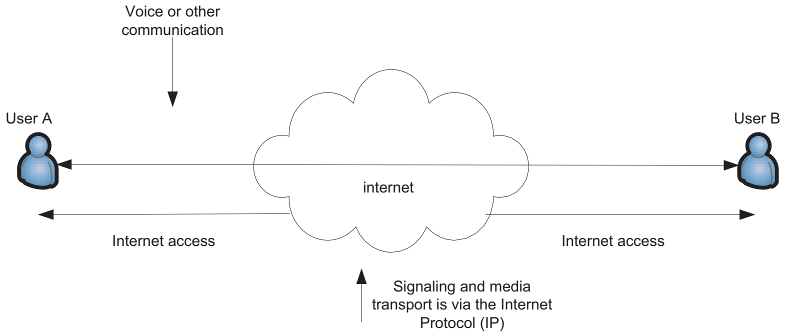

Voice or other communication
User A
User B
internet
Internet access
Internet access
Signaling and media transport is via the Internet Protocol (IP)

Figure 2.1: Communication over the Internet

2.2 Basics of VoIP

In general, VoIP communication takes place between two end users (peers), i.e. peer-to-peer communication, despite the fact the network infrastructure connecting these end users consists of a variety of nodes (registrars, proxies, location servers etc). VoIP communication involves both signaling and media transfer. The signaling runs over TCP/IP or UDP/IP while the media is most of the time transported over UDP/IP therefore, signaling and media parts may traverse different paths through the Internet.

2.3 Media streams for VoIP

As depicted in figure 2.2, signaling and media take different pathways as they both traverse the Internet. The reason being, among others, the fact that VoIP related signaling messages will be routed through a set of designated nodes in the Internet, for the purpose of delivering a service to that call/call establishment. For example, a call establishment message directed to segun.adewale@ericsson.nl will have to be routed through a node that is aware of the current IP address of Segun Adewale and can forward the message to such destination.

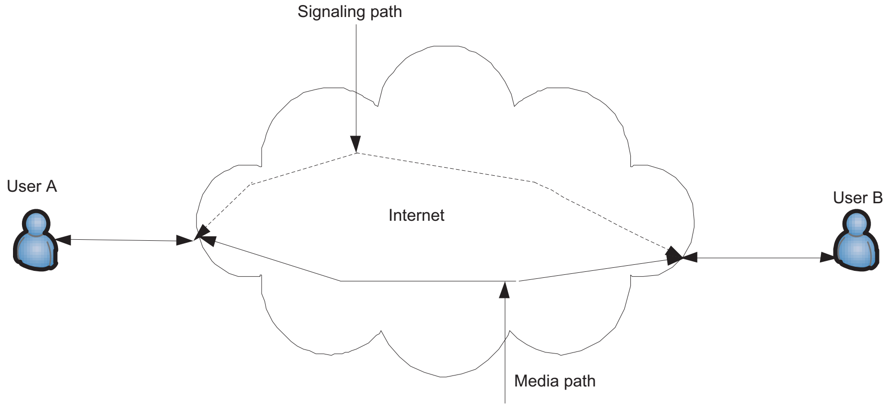


Figure 2.2: Signaling and Media traversing different paths through the Internet

Typically, the media stream undergoes less processing as it traverses the Internet from user A to user B. For this reason, media transfer between these two users may take a different and shorter path.

For the transfer of signaling messages and media packets (IP packets containing RTP media packets), the following distinction can be made:

- VoIP directs the routing of signaling messages and media streams between users and nodes on the Internet based on host name (IP address or domain name). The control protocol, Session Initiation Protocol (SIP) includes mechanisms to ensure the routing of signaling messages through designated nodes on the Internet.
- For IP based routing, i.e. routing on OSI layer 3, the VoIP operator shall have in place adequate infrastructure, to guarantee reliable transfer of signaling messages and media streams between IP addresses on the Internet.

2.4 Defining Internet domains for VoIP

In VoIP deployment, different domains with each of the domain playing different roles have to be defined. VoIP is basically a service on the Internet just like e-mail service, FTP, HTTP etc. In order to use any service on the Internet, the end users must have internet access. The VoIP subscriber will be a subscriber to the VoIP service which may be offered and operated by e.g. KPN in Netherlands. The VOIP subscriber’s Realm (Home

VOIP domain) will in that case be e.g. ims.kpn.nl. The realm is used to associate the subscriber with the VOIP service, not to identify the subscriber.

The VoIP subscriber herself has a set of public user identities. These identities may be identities within a virtual domain, such as ericsson.nl. Ericsson.nl is the enterprise name, which is used for identifying the VoIP subscriber. The subscriber may be a user within this enterprise domain, e.g. segun.adewale@ericsson.nl. The actual VoIP services are hosted by the IMS operator (explained in chapter 3), e.g. ims.kpn.nl. The relation between ims.kpn.nl and ericsson.nl does not have to be visible for the outside world.

2.4.1. Roles of various domains

- VoIP Operator

The VoIP operator is the one offering the VoIP service. The VoIP operator owns/operates designated VoIP nodes on the internet. These VoIP nodes (service hosts) are typically located in an IP sub-domain, with strict security control with respect to incoming and outgoing IP traffic. This sub-domain is often configured as Demilitarized Zone (DZ), i.e. Local Area Network (LAN) or Wide Area Network (WAN) protected by Firewall and Proxies. Once users subscribe to the VoIP service from an operator, they will be provisioned in the user database of the operator. On registration with the VoIP network, a message is sent to designated node in the VoIP network to mark the user as being registered.

Adequate selection of host names is important to provide selective DNS resolving.

- End-user domain

The VoIP user is allocated to a VoIP domain. The identity of the end-user is known as Address of record (AoR), the name / number one would put on one’s business card, i.e. publicly available user identification. The AoR is used to establish communication with a VoIP subscriber. The AoR in VoIP typically has the form of a Universal Resource Identifier (URI), consisting of a user part and a domain part, in the form of a Fully Qualified Domain Name (FQDN).

The user part of the AoR may be a phone number. Usage of phone numbers in VOIP is widespread. Whereas VoIP started out with names only, the use of phone numbers (E.164 numbers, named after ITU-T Recommendation specifying the format of these numbers) is introduced to facilitate that people may continue to be reachable under their existing phone numbers. ENUM database contains mapping between E.164 number and URI.

The URI for a VoIP subscriber is formally preceded by the scheme for the identifier. Hence, the AoR of subscriber may be sip:alice.jones@volvo.se etc. The ‘sip:’ has dual meaning in this case: (1) it indicates the format (scheme) of the URI and (2) it indicates that that subscriber is contactable through SIP as signaling protocol.

To contact a VoIP subscriber through her phone number, the tel: scheme may be used, e.g. tel:+46705453600. A VoIP subscriber would normally have both a SIP URI and a Tel URI. The latter is needed not only for the purpose of being contactable under a phone number, but also for establishing a call towards a non-VoIP user, such as a GSM subscriber residing in the PSTN/PLMN domain.

2.5 First step in communicating over the Internet (peer-to-peer media exchange).

The exchange of media packets between two parties having IP connectivity requires that the two parties must have beforehand informed each other about their respective IP addresses, coding details as well as port numbers on which they will be listening for media of agreed type. The two parties install a listener on the indicated port (and have opened the port for incoming UDP traffic), with the said listener being able to receive and decode IP packets containing the agreed media stream. The listener is further configured to provide the decoded media to a Digital to Analogue converter, for producing the voice.

The media is transported over IP using Real-Time Transport protocol (RTP, IETF 3550) [12]. This is done without any form of flow or congestion control, hence, quality of the voice is not guaranteed.

2.5.1 Media exchange with flow control

RTP Control Protocol (RTCP) may be used by the communicating parties to control the flow of media between them. RTCP, which is also defined in IETF RFC 3550 [12], provides control mechanism to monitor and control the quality and reliability of the transfer of the RTP packets. RTP messages and RTCP messages are carried over separate virtual connections, typically distinguished through different port numbers.

Since RTCP signaling is used to control the RTP media streaming, the RTP listener will be a combined RTP-RTCP listener.

2.5.2 Media exchange with the addition of control plane

The addition of the phone application (control plane) allows the communicating parties to place a call, i.e. establish a communication session with each other.

The addition of the control plane allows for the following functions:

- calling and alerting the peer;
- accepting or declining an incoming call;
- terminating an established call;
- providing auxiliary information during call establishment, such as Calling party name or number, Subject (as in e-mail);
- exchange data w.r.t. the user plane to be established, such as IP address & port number, codec (and sampling rate etc.).

Based on the example used here, routing of the call is still IP-based i.e. at this point, the destination subscriber is still identified with IP address and not a URI.

2.6 User-to-user SIP signaling

The capabilities related to the establishment and termination of a call, as described in the last section, are offered through the Session Initiation Protocol (SIP). SIP is layer 5 application (OSI reference model), running on UDP/IP, TCP/IP or SCTP/IP. It is also sometimes placed on layer 6 and 7. SIP is specified by the Internet Engineering Task

Force (IETF), RFC 3261 [9]. SIP is functionally a successor (and improvement) of the ITU-T VoIP recommendation H.323.

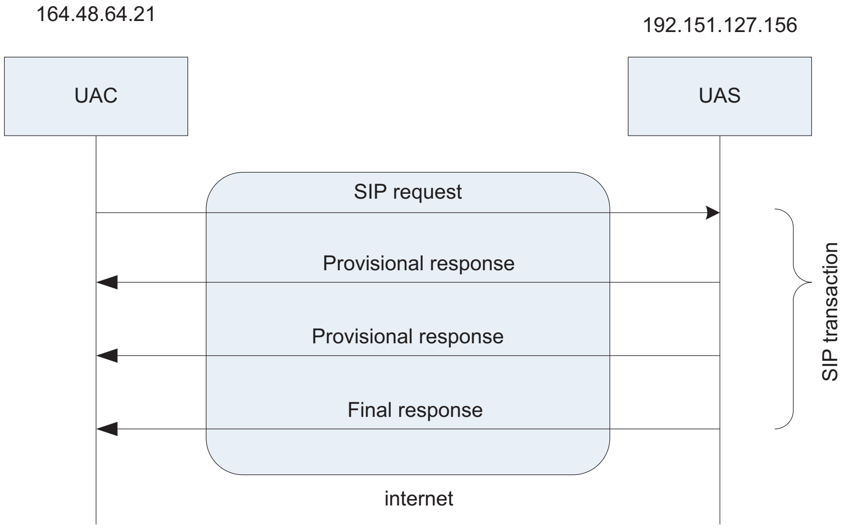


Figure 2.3: SIP transaction between two end users of VoIP

SIP is a transaction based protocol and it adopts the client-server transaction model. In order to establish communication session, a transaction has to be initiated towards a peer by an entity. This is depicted in figure 2.3. An invitation together with the associated response(s) makes up a transaction. A transaction request results in one final response with zero or more provisional responses. These provisional responses inform the initiator of the transaction about the progress of the processing of the request.

Entities that make use of SIP are called as User Agent (UA). A UA assumes the role of a User Agent Client (UAC) when it initiates a SIP request and User Agent Server (UAS) when it receives a SIP request. When a call is established and the called peer terminates the call, the UA that initiated the call acts as UAS because it’s the peer receiving a termination request from the called peer.

2.6.1 SIP signaling (voice call establishment)

SIP signaling as well as the concept of User Agent is clearly introduced in figure 2.4. The signaling depicted in the figure only reflects layer 5 signaling.

The SIP method used in initiating this request is called an INVITE one of the six basic requests SIP uses together with their respective responses in order to establish a call session). Here, one entity requests a service from another entity by sending a request message towards that other entity.

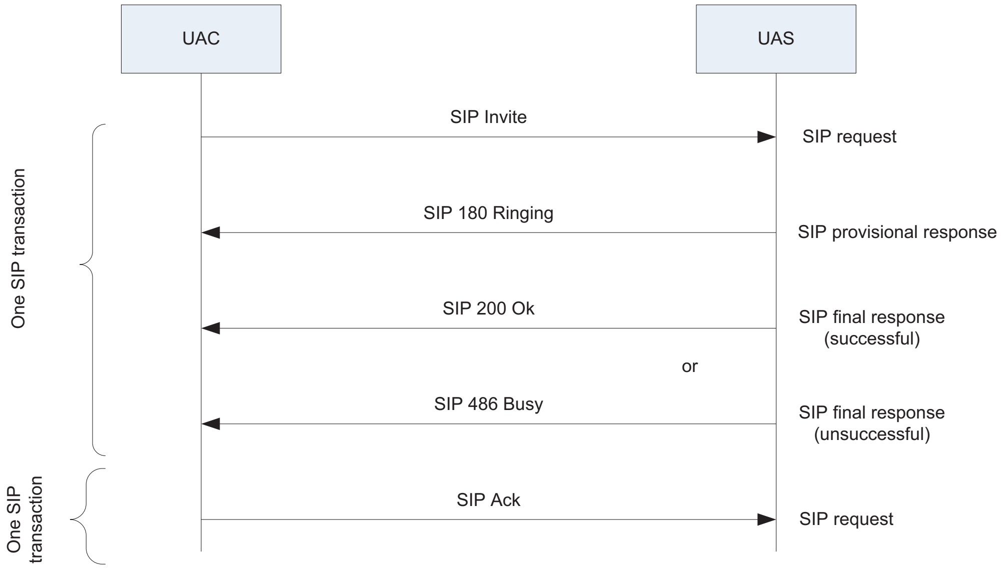


Figure 2.4: Example of a typical SIP transaction between end users

The other entity executes the request and sends a response to the initiator of the request. This INVITE method can be compared with a Remote Procedure Call (RPC). Two basic principles are related to the usage of RPC:

1. State model – an entity sending a RPC can send the request message from within a state model instance. The state model is a model describing a process, such as establishing a voice call. These series of messages exchanged between the sender and receiver result in state transitions. As a result of the exchange of messages, the state model of the sender and the receiver remain synchronized.
2. Relation – a relation (session) may be established due to the execution of a RPC between the sender and responder of the RPC. This relation can be likened to an established voice call.

For this reason, the two entities involved in a SIP based voice call need to keep state models related to SIP requests (transactions) synchronized and need to keep their relation synchronized too.

2.7 Changing of roles between UAC and UAS

As depicted in figure 2.5, during a sip session, a UA can assume a client or server role depending on which of the two UAs is initiating or receiving transaction request. In order for a UA e.g. a SIP phone, to act as a User Agent Server, it has to fulfill the following conditions:

1. A listener must have been installed on the designated port number on which it wants to receive SIP requests over TCP/IP or UDP/IP.
2. Its IP address and port number must have been sent towards the peer (or intermediate node) in order for the peer to send SIP requests to that UA.

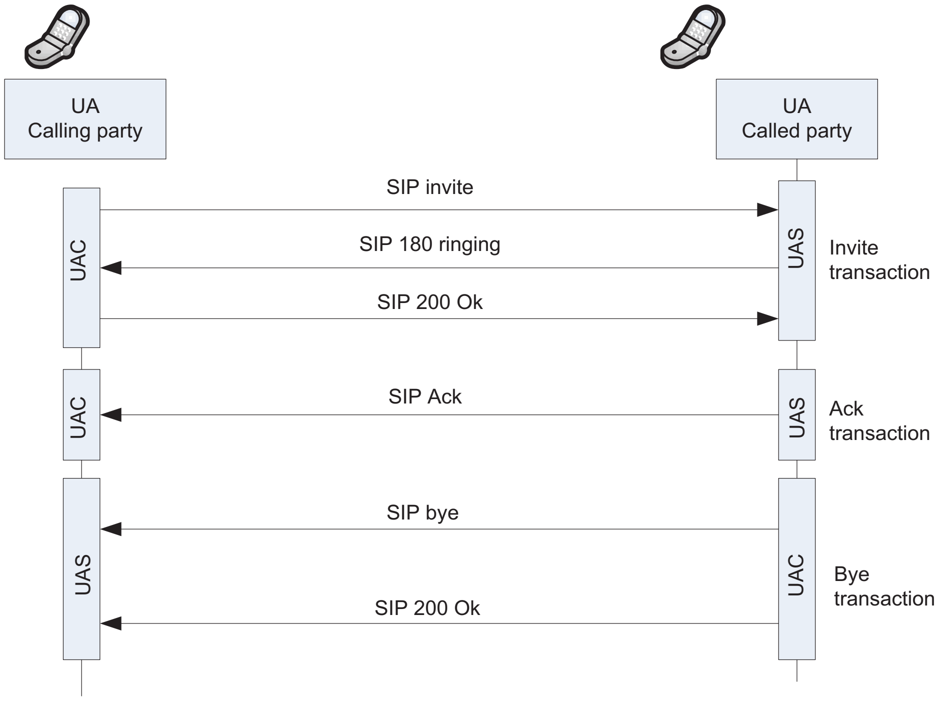


Figure 2.5: Switching of roles between UAC and UAS within the same UE

2.8 The Registrar

The registrar is an important entity in the Voice over IP network. The registrar maintains binding between a VOIP user's IP address and that person's public user identity. This binding is not a permanent one because the IP address of a user may change over time, in which case the registrar will have to be updated by the VoIP phone. The VoIP operator offers the registrar as a service to its subscribers. With this binding, the user's IP address is shielded from the outside world. Calls to this public user identity are routed to the registrar, which forwards the call to the IP address currently bound to this public user identity.

Resulting from the registration with a registrar, a relation is established between the VOIP user and the registrar and this makes it possible for the registrar to know where the subscriber is reachable. The VOIP user needs to refresh the binding periodically in order to maintain this relation with the registrar.

The role of the registrar is completed once the call establishment request has reached the destination terminal. The destination terminal responds towards the sender of the request when the call establishment request has arrived at the destination. These request and response messages contain enough information about the two parties so that the two can send subsequent request and responses to each other's IP address bypassing the registrar. The registrar is not involved in the exchange of media between the two parties. The IP addresses to be used for media transfer are exchanged in the SIP signaling flow.

2.8.1 Registration

This involves placing one's contact in an allocated registrar. This is initiated by the SIP terminal especially when switched on. First of all, there must be IP connectivity for the SIP phone and also there must exist an administrative relationship between the VoIP user and the VoIP operator. The operator must have assigned the user a public user identity as well as its home realm name. The registration message is sent towards the registrar within the VoIP domain home realm.

Without re-registration, the binding in the registrar expires so the binding has to be refreshed periodically. How often this is done is negotiated during the registration process.

2.8.2 The inbound proxy

In order for a service to be provided by a VoIP operator, which includes registration (binding of IP address and public user identity) or establishing a call towards a user via that user's registrar, the request message has to be directed towards the designated service node (registrar) that will provide the service (forwarding to IP address). The inbound proxy selects a registrar, in a case where multiple registrars are in the VoIP network, which will execute the service.

As long as the subscriber remains registered in the VOIP network, she will be served by that registrar. Hence, operational condition of that specific registrar is crucial for the service to that subscriber.

2.8.3 Signaling for inbound proxy (registration)

The VoIP subscriber deposits (registers) her contact address (address where she is contactable for the establishment of VoIP sessions) in the registrar using the inbound proxy to select the registrar as depicted in figure 2.6.

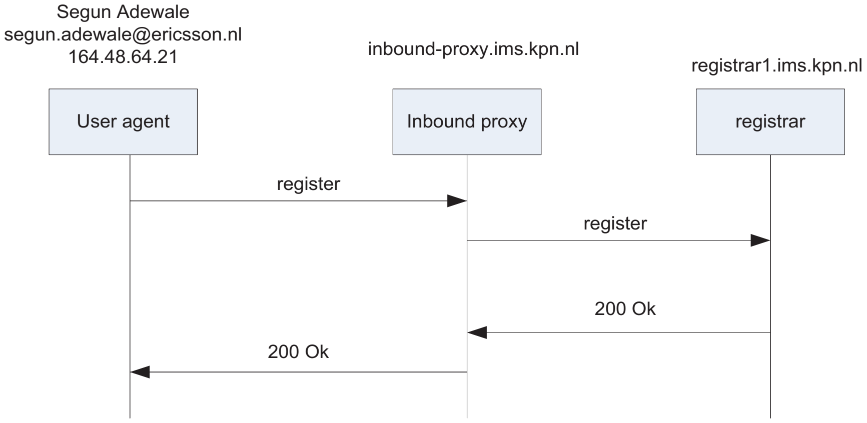


Figure 2.6: Registration of user's contact address through the inbound proxy

The User agent directs the registration message to ims.kpn.nl; the inbound proxy takes care that the registration message is forwarded to e.g. registrar1.ims.kpn.nl.

Resulting from the registration, the User agent and registrar have established a relation ('registration relation'). The registrar has stored the subscriber's public user identity and the subscriber's current contact address. The User agent has stored the address of the registrar where it is registered. The inbound proxy itself does not maintain user data; user data is maintained in a separate database.

2.8.4 Signaling for inbound proxy (call establishment)

A similar role as the one performed in selecting registrar during registration is done during terminating call establishment. As depicted in figure 2.7, a call that is destined for bob.smith@peugeot.fr is routed to the VOIP network where Bob Smith is a subscriber of. This routing of a call for Bob Smith to the VOIP network where he is a subscriber of is based on DNS domain name resolving. When DNS is queried to provide the name (and IP address) of a host that can handle a call for Bob Smith, DNS will return the host name of an inbound proxy of the VOIP network of Bob Smith.

The inbound proxy knows from its associated database in which registrar Bob Smith is registered. So, the call for Bob Smith can be forwarded to the appropriate registrar. The inbound proxy forwards the call, destined for bob.smith@peugeot.fr, to registrar1.ims.orange.fr. The inbound proxy is never involved in media exchange between calling and called party.

The registrar is only involved in the forwarding of the call establishment request. Subsequent VOIP messages can be exchanged between the entities involved in the voice call, by exchanging their contact addresses end-to-end.

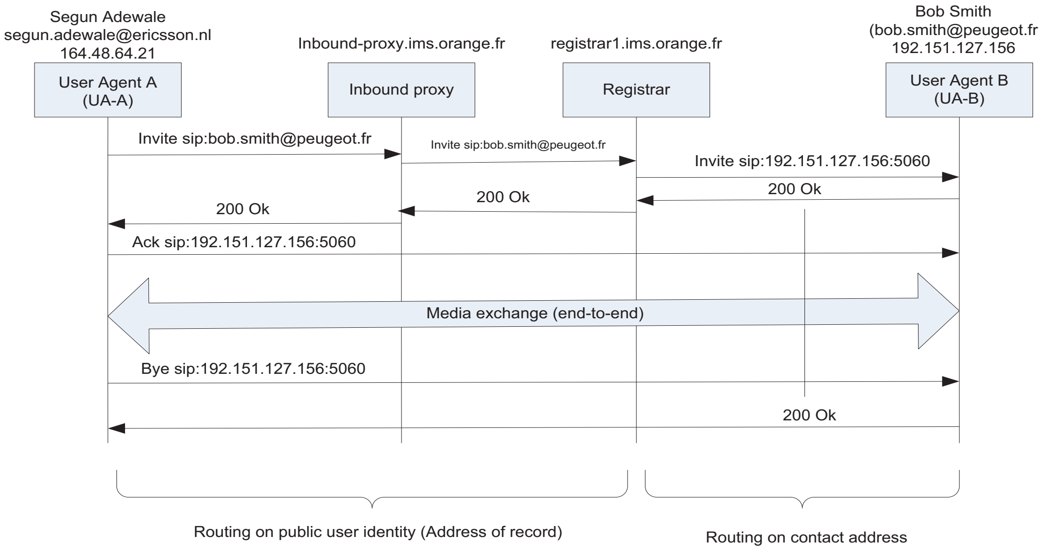

Segun Adewale
segun.adewale@ericsson.nl
164.48.64.21
Inbound-proxy.ims.orange.fr
registrar1.ims.orange.fr
Bob Smith
(bob.smith@peugeot.fr
192.151.127.156
User Agent A
(UA-A)
Inbound proxy
Registrar
User Agent B
(UA-B)
Invite sip:bob.smith@peugeot.fr
Invite sip:bob.smith@peugeot.fr
Invite sip:192.151.127.156:5060
200 Ok
200 Ok
200 Ok
Ack sip:192.151.127.156:5060
Media exchange (end-to-end)
Bye sip:192.151.127.156:5060
200 Ok
Routing on public user identity (Address of record)
Routing on contact address

Figure 2.7: Routing of a terminating call to the called party through the inbound proxy resolved by DNS

2.9 The Location Register

The inbound proxy doesn't include any subscriber database. The location register helps the inbound proxy in directing incoming service requests to the registrar. The location registrar contains the subscriber data hence, the inbound proxy queries it to retrieve subscriber information.

During a VoIP subscriber registration process, the registration request message traverses an inbound proxy, which queries the Location register to get directive regarding the registrar to which the registration request message shall be forwarded. The registrar receives and processes the registration request message and informs the Location register about the address of the registrar where the subscriber is now registered.

In VoIP network, each subscriber is serviced by a particular Location register, meaning that that subscriber has his/her subscription data provisioned in that Location register. Entities in the VoIP network that want to use the service of a Location register in order to obtain the address of the registrar for that subscriber must be configured with the address of the location register serving that subscriber. In a VoIP network of multiple Location registers, a Subscriber locator function (SLF) is present, containing a database with the public user identities of all subscribers in the network and an indication in which location register each subscriber is provisioned.

2.9.1 Signaling for inbound proxy and Location server (call establishment)

The inbound proxy queries the Location register (for a terminating call establishment request message for a VOIP subscriber) in order to get directive to the registrar to which the request message shall be forwarded. The communication between inbound proxy and Location register is done using Diameter (transmission interface between the two entities) and not SIP.

One could visualize the Diameter communication between inbound proxy and Location server as being perpendicular to the SIP signaling that traverses the inbound proxy. This leads to the following comparison between inbound proxy and registrar:

- when the inbound proxy has to forward a terminating call establishment request message to the registrar for the destination party of this call, the inbound proxy makes use of external database, to receive data that is needed to forward the call establishment request message to the registrar; the inbound proxy does not keep subscriber data;

- when the registrar has to forward a terminating call establishment request message to the contact address of the destination party of this call, the registrar uses internally stored data for this purpose; the registrar keeps subscriber data.

2.10 User-to-network proxy (protected access to the network)

As stated earlier, the VoIP network of an operator resides in a demilitarized zone, i.e. the operator's own sub-network, therefore the signaling messages entering or leaving the operator's VoIP domain shall pass through a network gateway. The user-to-network proxy ensures that all signaling from/to the user is verified and corrected, if needed. For example, when a VOIP user establishes a call, the user-to network proxy verifies the identity of the calling user. Another example is that for a terminating call, the user-to-network proxy removes the calling party's number from the signaling message, if the calling party had indicated that her number should not be revealed to the called party.

Also, the VOIP network may contain multiple user-to-network proxies and a VOIP subscriber has to be assigned to a particular user-to-network proxy. When the subscriber registers to the VOIP network, the VOIP phone selects a user-to-network proxy. When the subscriber registers at multiple terminals, then each terminal selects a user-to-network proxy.

The subscriber needs to know the following during registration:

1. His home network realm e.g. ims.kpn.nl and
2. His user-to-network proxy e.g. user-proxy1.ims.kpn.nl

All service requests to the network traverse his user-to-network proxy. The user-to-network proxy forwards the registration request to the inbound proxy during registration, which also forwards it to the registrar and so on.

Due to the registration procedure, the user agent (terminal), user-to-network proxy, registrar and location server contain some data related to the registration, registration refresh and data needed for call establishment.

In order to secure the data message exchange between the VoIP terminal and the user-to-network proxy, a Security Association may be established between these two entities, signaling between the two can also be compressed and encrypted.

During registration, authentication may be applied. A VoIP terminal and the Location register, acting as subscriber persistent data storage, may share Long term secret. This information, known only to these two entities, is used to have the network authenticate the user. This information may also be used to establish the Security Association.

Figure 2.8 gives a full illustration of the end-to-end signaling flow for a typical VoIP call using SIP. Call signaling continues to traverse user-to-network proxy as well as the registrar once the call is established. As long as the registrar acts as service node during a call, it has to stay in the signaling path.

The registrar of the calling party acts as service node for the originating call. As shown in the figure, the call establishment request message from the calling party does not traverse an inbound proxy in that calling party's VoIP domain. The reason for this is that the address of the registrar and service node of the calling party is reported to and stored by that calling party (internal in phone) during registration. This forms part of the registration binding between the VoIP terminal and the registrar. So, the terminal can direct the service request, for establishing a call, directly towards the service node. The service request will be routed through the user-to-network proxy of this calling party.

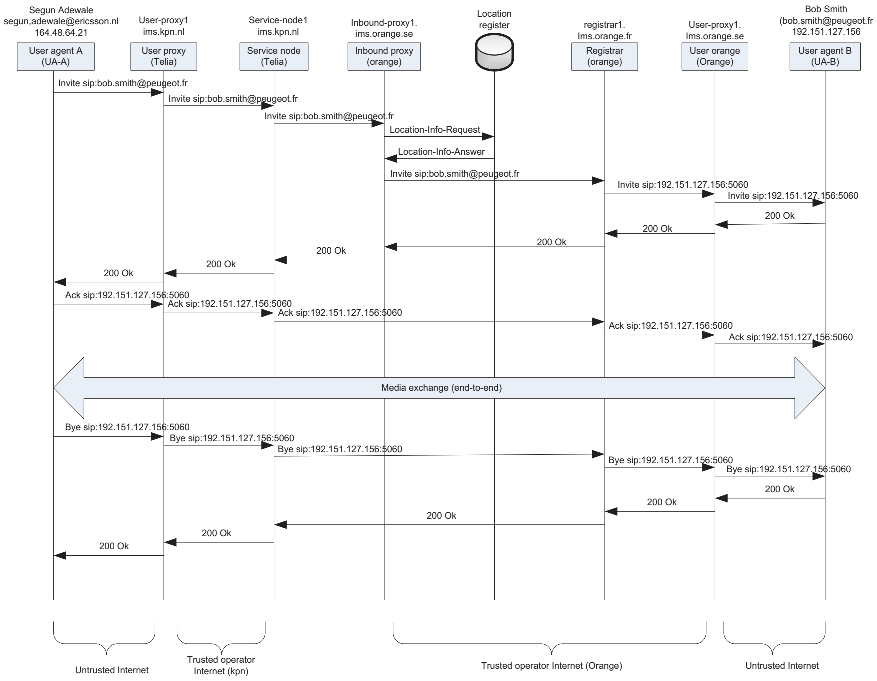

Segun Adewale
segun,adewale@ericsson.nl
164.48.64.21
User-proxy1
ims.kpn.nl
Service-node1
ims.kpn.nl
Inbound-proxy1.
ims.orange.se
Location
register
registrar1.
Ims.orange.fr
User-proxy1.
Ims.orange.se
Bob Smith
(bob.smith@peugeot.fr
192.151.127.156
User agent A
(UA-A)
User proxy
(Telia)
Service node
(Telia)
Inbound proxy
(orange)
Registrar
(orange)
User orange
(Orange)
User agent B
(UA-B)
Invite sip:bob.smith@peugeot.fr
Invite sip:bob.smith@peugeot.fr
Invite sip:bob.smith@peugeot.fr
Location-Info-Request
Location-Info-Answer
Invite sip:bob.smith@peugeot.fr
Invite sip:192.151.127.156:5060
Invite sip:192.151.127.156:5060
200 Ok
200 Ok
200 Ok
200 Ok
200 Ok
200 Ok
Ack sip:192.151.127.156:5060
Ack sip:192.151.127.156:5060
Ack sip:192.151.127.156:5060
Ack sip:192.151.127.156:5060
Ack sip:192.151.127.156:5060
Media exchange (end-to-end)
Bye sip:192.151.127.156:5060
Bye sip:192.151.127.156:5060
Bye sip:192.151.127.156:5060
Bye sip:192.151.127.156:5060
Bye sip:192.151.127.156:5060
200 Ok
200 Ok
200 Ok
200 Ok
200 Ok
Untrusted Internet
Trusted operator
Internet (kpn)
Trusted operator Internet (Orange)
Untrusted Internet

Figure 2.8: Full illustration of the end-to-end signaling flow for a typical VoIP call using SIP

2.11 Layered signaling: control plane and user plane

The concept of the layered architecture, constituted by separate control plane and user plane is the contemporary digital communication networks approach. The present section presents some aspects resulting from this architecture in Voice over IP networks, especially where it concerns the synchronization of control plane and user plane. This separation of control plane from user plane allows for a very efficient media routing.

The control plane signaling and user plane signaling are routed via different paths through the access network, through the VOIP network and through the public internet or operator internet. The VOIP operator may steer the control plane through designated (operator controlled) interconnect network infrastructure. For routing a call between different IMS networks, the call may take a breakout to Circuit switched (CS) network, routed via the CS network towards the destination IMS network, where break-in takes place.

SIP, as used in IMS, allows for routing signaling messages through selected designated SIP proxies. The media may be routed through the same interconnect network infrastructure. Even with control plane and user plane steered through the same interconnect network infrastructure, there is the issue of synchronization of the two planes. A media stream (RTP message transfer) is functionally associated with a call (SIP session).

2.11.1 Synchronization of the control plane and the user plane

Voice clipping is an effect whereby the first part of the voice signal (e.g. 1 second) is muted due to non-synchronized arrival of control plane message and user plane media.
In an ideal case, media transfer shouldn't start until control plane signaling is completed. When a call traverses the boundary between VoIP network and GSM network, this issue becomes more pronounced. A media gateway control function MGCF) can be used to check this effect. The MGCF may defer the through-connection of media between the two networks until the call has reached the active state, i.e. the receiving of the 200 Ok message over the control plane. The 200 Ok message is the final response sent by the called party to inform the calling party that it is ready to receive media.

3. IP Multimedia Subsystem (IMS)

In order for the provision of a network within whose framework the thesis would be carried out, the IMS framework was adopted. For this reason, this chapter gives an insight into the IMS framework and its importance and benefits, the architecture and protocols involved in its operations, the layered approach adopted in the operation of the IMS and the core functional elements which are building blocks of the whole system.

3.1 Introduction to IMS

IP Multimedia Subsystem (IMS) is a global, access-independent and standards-based IP connectivity and service control architecture that describes the Next Generation Networking (NGN) architecture for implementing IP based telephony and multimedia services. IMS defines a complete architecture and framework that enables the convergence of voice, video, data and mobile network technology over an IP-based infrastructure. It is an enabler of true convergence and interworking in several dimensions, independent of access or network, and provides users with the ability to access services irrespective of their location and device. It fills the gap between the two most successful communication paradigms, cellular and Internet technology. The vision for IMS is to define a model that separates the services offered by telephone service providers from the access networks used to receive those services and to provide cellular access to all the services that the Internet provides [2].

3.2 History of IMS

IMS was defined by the Third Generation Partnership Project (3GPP), www.3gpp.com, which is a collaboration among a number of telecommunications standards bodies, as part of their standardization work for supporting GSM networks and radio technology evolution. In 2002, IMS was introduced in 3GPP Release 5, where "Session Initiated Protocol" (SIP), defined by the Internet Engineering Task Force (IETF), was chosen as one of the main traffic protocols (session control) for IMS (control plane). IMS has been further enhanced in Releases 6, of 3GPP and onwards to include additional features like presence and group management, interworking with WLAN and CS based systems, and Fixed Broadband access [17].

Another standard body, The Third Generation Partnership Project 2, (3GPP2) www.3ppp2.com, standardized their own IMS. 3GPP2 was born to evolve North American and Asian Cellular Radio-telecommunication Intersystem Operations into a third-generation system. The initial release of 3GPP2 specifications on IMS largely adopts from 3GPP Release 5. The two IMS networks defined by the two organizations are fairly similar but not exactly the same. 3GPP2 added appropriate adjustments for their specific issues in order to support CDMA2000 access. These adjustments include mobility management (use of mobile IP), smart card (an optional smart card, which acts as the counterpart of the UICC (Universal Integrated Circuit Card, is referred to as the removable user identity module, R-UIM), access security (alternative authentication methods other than IPSec are supported), P-CSCF location (the P-CSCF and the PDSN-

Packet Data Serving Node, a component in the CDMA2000 mobile network acting as a connection point between the radio access and IP networks, in 3GPP2 may not need to be in the same network), database (in contrast to 3GPP, 3GPP2 separates the functions of HSS into authentication, authorization and accounting, AAA server, and other databases). Nevertheless, the purpose of both organizations is to ensure the IMS applications will work consistently across different network infrastructures [1].

In addition to the 3GPP and 3GPP2, Open Mobile Alliance (OMA), www.openmobilealliance.org, plays an important role on specifying and developing IMS service standardization. The services defined by OMA are built on top of IMS infrastructure, such as Instant Messaging (IM), Presence service, and Group Management Service [17].

3.3 Benefits of IMS

The idea of IMS has been discussed as a way to offer multimedia services everywhere using different access technologies. Having already been very familiar with accessing Internet services like Web access, email, or instant messaging via a 2.5G and 3G cellular phone, one may start wondering why the need for IMS.

The benefits of IMS can be demonstrated in the following four aspects [16]:

- IMS provides a common platform to reduce time-to-market for rolling out new multimedia services: One of the biggest challenges in today's communication network is to improve the long and costly process for creating a new service. Service providers are looking for ways to reduce the time-to-market for rolling out new multimedia services. The IMS infrastructure addresses this issue by providing a standardized service platform (SIP application server) and reusable components like the CSCFs, HSS etc, so that service providers are not tied to the timescales and functions of their primary Network Equipments Providers. The standardized interfaces provided by IMS infrastructure enables service providers to easily adopt a service created by third parties and create a service that integrates with many services effectively. In addition, with the standardized interface (interfaces connecting the various entities of the IMS network) provided by IMS, the service is no longer solely provided by a single provider; any provider who implements the standardized interface can provide the service. The multi-vendor service creation industry leads to an open market, and allows service providers to choose the most effective way to roll out new services [5].
- IMS provides multimedia services with Quality of Service (QoS) enablement: Although the dramatically increased bandwidth in 3G cellular networks provides a much faster and more reliable Internet access compared with 2.5G cellular network, there are no guarantees about the quality of the services. A 3G cellular network provides what is known as "best effort", which means the network will do its best to ensure the required bandwidth (this pertains to the IP transport in the radio network), but there is no guarantee it will remain at the same level. Consequently, the bandwidth of a particular connection can vary significantly over time. In order to solve this problem, Quality of Service (QoS) mechanisms

were developed in order to provide certain guarantee levels of network bandwidth during transmission instead of the so called "best effort". IMS specifies enablement of Quality of Service within the IP network and takes advantage of the QoS mechanism like Integrated and Differentiated Services (intserv and diffserv) to improve and guarantee the transmission quality [2] by reserving resources which may be implemented with Resource Reservation Protocol (RSVP) defined in RFC 2205 [43].

- IMS allows operators to charge multimedia session appropriately: If a user uses videoconference over the 3G cellular network, there is usually a large data transfer that consists of audio and video. This is usually expensive since the operator may charge by the number of bytes transferred. On the other hand, if the operator is willing to provide a different charging scheme based on the actual service type, it may be more beneficial to the users. The advantage of IMS is that it provides information about the service type being invoked by the user and thus allows the operators to determine how to charge the users based on service types, i.e. they can choose to charge user by the number of bytes transferred, by the session duration (time-based), or perform any new type of charging [2].
- IMS allows all services to be available irrespective of the users' location and access network: A typical and particularly annoying problem when working with cellular technology is that some of the services will not be available when the user is roaming in another country. To resolve this problem, IMS uses Internet technology and protocols in order to allow users to move across the countries and still be able to have access and execute all the services as if they were from their home networks [7].
- Security is a fundamental requirement in every telecommunication system and the IMS is not an exception. The IMS provides at least a similar level of security as the corresponding GPRS and circuit-switched networks: for example, the IMS ensures that users are authenticated before they can start using services, and users are able to request privacy when engaged in a session [17]. The 3GPP has defined new security functions as part of the IMS model, and a base function of the call session control function that establishes the core of the IMS.

3.4 IMS Protocols

When 3GPP was developing the IMS specifications, it first considered the existing protocols developed by IETF and ITU-T. This led to the use of SIP (Session Initiation Protocol) as its signaling protocol. The Session Initiation Protocol (SIP) is the basic underlying protocol in IMS [2] for session control.

3.4.1. Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP) is a very important part of IMS, providing the basis that makes the entire system work. The IETF Session Initiation Protocol (SIP) as described in RFC 3261 is an application layer 5 for initiating and controlling multimedia sessions within an IP network [9]. SIP has several features but it doesn't carry the content of the session, instead, this content is streamed over protocols and between destinations that are negotiated using SIP [6]. The fact that SIP was chosen as the core protocol, 3GPP used a layered approach in the IMS architectural design; hence the transport and bearer services are separated from the IMS signaling network and session management services (SIP). Further services are run on top of the IMS signaling network. The layered approach aims at a minimum dependence between layers. A benefit is that it facilitates the addition of new access networks to the system later on. The IMS is designed to be access-independent so that IMS services can be provided over any IP connectivity network.

SIP is a text-based, human readable format protocol and following standard internet conventions. It's also a peer-to-peer protocol (communication between two users), in that a user agent will typically switch between client and server modes as the situation demands and the protocol works with both IPv4 and IPv6. User agent refers to both end points of a communication session.

SIP uses SCTP, TCP and UDP as its transport layers, the choice between these being made on a stage-by-stage basis as a SIP message traverses user agents and proxies. SIP includes its own reliability mechanisms, which makes the lighter-weight UDP the usual choice, but for technical networking reasons (including message size limits, firewall and NAT traversal) TCP is sometimes preferred or required.

SIP provides a suite of security services, which include denial-of-service, authentication (both user to user and proxy to user), integrity protection, and encryption and privacy services [9].

Much like HTTP, the basic units of interaction in SIP are the client request and the server response, with single transaction being composed of a request and one or more responses.

3.4.2. Session Description Protocol

When initiating multimedia teleconferences, voice-over-IP calls, streaming video or other sessions, there is a requirement to convey media details, transport addresses, and other session description metadata to the participants.

Session description protocol (SDP) provides a standard representation for such information, irrespective of how that information is transported. SDP is purely a format for session description -- it does not incorporate a transport protocol, and it is intended to use different transport protocols as appropriate, including the Session Initiation Protocol [9], Real Time Streaming Protocol (RTSP) [11], electronic mail using the MIME extensions, and the Hypertext Transport Protocol (HTP) [10].

Session Description Protocol (SDP) gives a standardized way to define the media capabilities of a proposed session, and as such usually forms the payload of SIP INVITE, OPTIONS, UPDATE and ACK requests and responses.

SDP is a simple, compact, text-based protocol. The purpose of SDP is to convey information about media streams in multimedia sessions to allow recipients of session description to participate in the session. In each SDP description, three different types of information are being conveyed: the session level description, one or more time descriptions and zero or more media descriptions which include session name, time(s) the session is active, the media comprising the session, information needed to receive those media (addresses, ports, formats etc) [6], [10].

3.4.3. Real-time Transport Protocol

Real-time transport protocol (RTP), specified in RFC 3550 [12], provides delivery services to transport real-time multimedia traffic including video and audio over unreliable transport mediums such as User Datagram Protocol (UDP). Those services include payload type identification, sequence numbering, time stamping and delivery monitoring. RTP may be used with other suitable underlying network or transport protocols. RTP contains the necessary attributes to ensure correct media buffering and jitter management by providing a timing relationship between source and sink of the media session [2].

RTP itself does not provide any mechanism to ensure timely delivery or provide other quality-of-service guarantees, but relies on lower-layer services to do so. It does not guarantee delivery or prevent out-of-order delivery, nor does it assume that the underlying network is reliable and delivers packets in sequence. The sequence numbers included in RTP allow the receiver to reconstruct the sender's packet sequence [12].

3.4.4. The AAA Protocol

The Diameter protocol specified in RFC 3588 [13] was chosen to be used in IMS as the AAA (Authentication, Authorization and Accounting) protocol.

It is an evolution of RADIUS (RFC 2865), which is a protocol that is widely used on the Internet to perform AAA. Diameter consists of a base protocol that is complemented with Diameter applications. These applications are extensions to Diameter to suit a particular application in a given environment [3].

IMS uses Diameter in a number of interfaces, although not all interfaces use the same Diameter application.

3.4.5. H.248 MEGACO

Media Gateway Control/H.248 (MEGACO) protocol (ITU-T Recommendations H.248.1) [14] and its packages are used by signaling nodes to control nodes in the media plane (for example, a media gateway controller controlling a media gateway) within the IMS (IP network) environment and PSTN (Public Switched Telephone Network). It is a (master/slave) protocol for control of gateway functions at the edge of the packet network e.g., IP-PSTN trunking gateways. The main function of Megaco is to allow gateway decomposition into a call agent (call control) part (known as Media Gateway Controller, MGC) – master, and a gateway interface part (known as Media Gateway, MG) – slave. Megaco defines the protocol for Media Gateway Controllers to control Media Gateways for the support of multimedia streams across IP and PSTN networks.

H.248 was co-developed by ITU-T and IETF (RFC 3525) [15], [3].

3.5 IMS architecture

A layered approach to architectural design is applied by 3GPP. Media transport and bearer services are separated from the IMS signaling network and session management services. Further services are run on top of the IMS signaling network. The layered approach aims at minimum dependency between layers. One of the benefits is that it facilitates the addition of new access networks to the system later on. Another important benefit is that it allows for separate path for the media stream, compared to control plane (this is a fundamental difference with circuit-switched networks). The layered approach increases the importance of the application layer. When applications are isolated and common functionalities can be provided by the underlying IMS network, the same applications can run on user equipment (UE) using diverse access types [17].
IMS architecture supports a wide range of services that are enabled based on SIP. As can be seen from Figure 3.1 below, IMS architecture delivers multimedia services that can be accessed by a user from various devices via an IP network or traditional telephony system (PSTN). The underlying network architecture can be divided into three layers (Device Layer, Transport/Media Layer, and Control / Session Layer) plus the application / service layer, and will be introduced from bottom to top respectively.

- **Device Layer**: The IMS architecture provides a variety of choices for users to choose end-point devices. The IMS devices such as computers, mobile phones, PDAs, and digital phones are able to connect to the IMS infrastructure via the access network applicable for that device. Other types of devices, such as traditional analog telephones, although they are not able to connect to an IP network directly, are able to establish a connection with these devices via a Public Switched Telephone Network (PSTN) Gateway. The PSTN gateway acts as an adapter between the traditional telephones that use the IP based IMS network.
- **Transport / Media Layer**: The transport layer is responsible for initiating and terminating SIP sessions and providing the conversion of data transmitted between analog/digital formats and an IP packet format. IMS devices connect to the IP network in the transport layer via a variety of transmission media, including Wi-Fi (a wireless local area networks technology), DSL, Cable, GPRS (General Packet Radio Service is a mobile data service), and WCDMA (Wideband Code Division Multiple Access, a type of 3G cellular network).
- **Control / Session Layer**: The Call Session Control Function (CSCF), which is a general name that refers to SIP servers or proxies, is one of the core elements in the control layer. CSCF handles SIP registration of the end points and processes SIP signaling messages of the appropriate application server in the service layer. Another element in the control layer is the Home Subscriber Server (HSS) database that stores the unique service profile for each end user. The service profile may include user identities, registration information, access parameters and service triggering information. By centralizing a user's information in HSS, service providers can create unified personal directories and centralized user data administration across all services provided in IMS.

- **Service / Application Layer:** On top of the IMS signaling network we have the service layer. The three layers described above provide an integrated and standardized network framework to allow service providers to offer a variety of multimedia services in the service layer. The services are all provided by service logic executed in application servers. The application servers are not only responsible for hosting and executing the services, but also provide the interface against the control layer using SIP. A single application server may host multiple services, for example, telephony and messaging services run on one application server; one advantage of this flexibility is to reduce the workload of the control layer. There are many application servers providing different services, and three standard services of IMS will be highlighted below.

Presence server: A "Presence server" provides the services to collect manage and distribute the real time availability and the means for communicating among users. It allows users to both publish their presence information and subscribe to the service in order to receive notification of changes by other users.

Group List Management server: A "Group List Management server" provides services that allow users or administrators the ability to manage, create, modify, delete and search the network-based group definition and the associated lists of members. It also maintains the access permissions and other specific properties associated with the groups and the members. It is also used to provide buddy lists for instant messaging or other services.

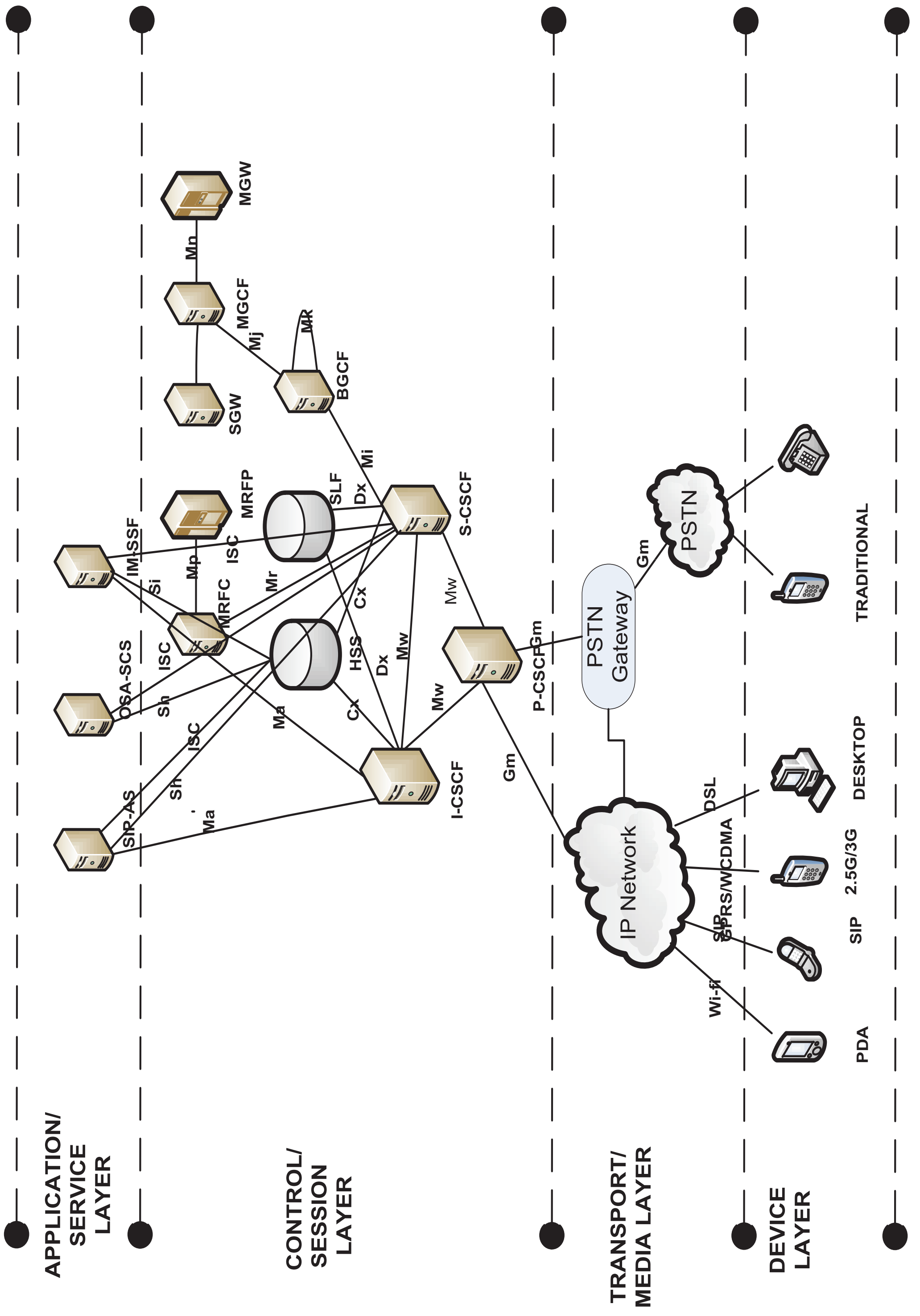


Figure 3.1: IMS architecture

3.6 Core IMS functional elements and related elements

There are a number of core IMS functional elements specified by the 3GPP [16] and IMS-related entities. These functions are not necessarily separate physical functions. These entities can roughly be classified in to six main categories: session management and routing family (CSCFs), databases (HSS, SLF), interworking elements (BGCF, MGCF, IM-MGW, and SGW), services (application server, MRFC, MRFP), support entities (THIG, SEG, PDF) and charging. It is important to understand that IMS standards are set up so that the internal functionality of the network entities is not specified in detail. For example, the Home Subscriber Server (HSS) contains three internal functions: IMS functionality, necessary functions for the circuit switched (CS) domain and necessary functions for the packet switched (PS) domain [27]. 3GPP standards do not describe how IMS functionality interacts with functions designed for Packet Switched (PS); instead, they describe reference points between entities and functionalities supported at the reference points (e.g., how does CSCF obtain user from HSS).
A brief description of the core IMS components:

3.6.1. Call Session Control Function

The Call Session Control Function (CSCF) is a SIP server which processes the IMS signaling traffic and during registration in order to control multimedia sessions. There are four types of CSCF:

- Proxy CSCF (P-CSCF): this is the initial point of contact for signaling traffic in to the IMS (user-to-network interface UNI). All SIP signaling traffic from or to the UE goes via the P-CSCF. The P-CSCF behaves like a proxy as defined in [RFC 3261] [9]. A user is allocated a P-CSCF during the P-CSCF discovery as part of the registration process, and the P-CSCF may provide a two-way IPsec (Internet Protocol security) association with the user; all signaling traffic traverses the P-CSCF for the duration of the session and also outside a session for non-session establishing SIP transactions. In addition, the P-CSCF may behave as a User Agent (UA) also as defined in [RFC3261] [9] for releasing sessions in abnormal conditions and for generating independent SIP transactions which deal with registration. There can be one or many P-CSCFs within an operator's network.
  The functions performed by the P-CSCF are highlighted below as stated in 3GPP [TS 23.228, TS 24.229] [18], [19], [17]:
  - To forward SIP register to the Interrogating-CSCF (I-CSCF) based on a home domain name provided by the UE in the request.
  - To forward SIP requests and responses received from the UE to the Serving-CSCF (S-CSCF).
  - To send accounting-related information to the Charging Collection Function (CCF).
  - To provide integrity protection of SIP signaling and maintain a security association between the UE and the P-CSCF. Integrity protection is by means of IPSec. Release 6 also provides confidentiality protection.
  - To decompress and compress SIP messages from the UE [20, 21, 22, 19].

- To subscribe a registration event package at the user's registrar (S-CSCF).
- To execute media policing. The P-CSCF is able to check the content of the Session Description Protocol (SDP) payload to check whether it contains media types or codecs, which are not allowed for a user. When the proposed SDP does not fit the operator's policy, the P-CSCF rejects the request and sends a SIP non-successful final response.
- P-CSCF plays an important role in IMS emergency session handling as the P-CSCF is tasked to detect emergency requests in all possible cases. P-CSCF is expected to reject emergency attempts based on operator policy (e.g., user attempting to make emergency call via home P-CSCF when roaming) or based on network capability [25].

Just like we have in GSM and GPRS, visited and home networks also exist in the IMS set-up. Most of the core IMS nodes/entities are located in the home network where the UE is registered while some can also be located in either the home or the visited network. One of such nodes is the P-CSCF. Once the P-CSCF is not located in the network where the UE is registered then, it is said to be residing in a visited network. In view of this, IMS supports roaming so that mobile users that are in a visited network can still have access to multimedia services. [3GPP TS 23.221] [37] defines two possible ways to achieve roaming. In the first case, IP-CAN (IP Connectivity Access Network) is provided by the visited network which provides the IP address to the UE. Therefore, the roaming UE contacts the visited IMS network through the visited access network. IMS signaling is passed from the visited IMS network to the home IMS network where the service control resides. The second approach involves the IP-CAN being provided by both visited and home networks. Here, the home network provides the UE with IP address. So, here, the UE contacts the home network directly. The first approach is better based on routing efficiency perspective in the sense that media and signaling traverses a shorter distance instead of being routed all the way to the home network.

The Serving GPRS Support Node (SGSN) links the RAN (Radio Access Network) to the packet core network. It is responsible for performing both control and traffic-handling functions for the PS domain. The control part contains two main functions: mobility management and session management. The control part of session management deals with connection admission control and any changes in the existing data connections. It also supervises 3G network services and resources. The SGSN acts as a gateway for user data tunneling, i.e., it relays user traffic between the UE and the GGSN.

The Gateway GPRS Support Node (GGSN) provides interworking with external packet data networks. The prime function of the GGSN is to connect the UE to external data networks, where IP-based applications and services reside. The external data network could be the IMS or the Internet. The GGSN routes IP packets containing SIP signaling from the UE to the P-CSCF and vice-versa. Additionally, the GGSN takes care of routing IMS media IP packets toward the destination network (e.g., to GGSN in the terminating network). The interworking service provided is realized as access points that relate to the different networks the subscriber wants to connect. In most cases the IMS has its own access point. When the UE activates a bearer towards an access point (IMS), the GGSN allocates a dynamic IP address to the UE. This allocated IP address is used in IMS registration and when the UE initiates a session as a contact address of the UE.

- Serving CSCF (S-CSCF): is the brain of the IMS and it located in the home network [7]. It performs session control and registration services for UEs. While UE is engaged in a session, the S-CSCF maintains a session state and interacts with service platforms and charging functions as needed by the network operator for support of the services. S-CSCF also provides the coordination logic that effectively invokes the application servers in order to deliver the requested service. It is the central node of the signaling plane. The S-CSCF interfaces with the HSS and SLF in order to determine user service authorization by downloading the user profile; the S-CSCF is allocated for the duration of the registration.

  There may be multiple S-CSCFs in a network and their functionalities may be different within an operator's network. More specifically, the functions performed by the S-CSCF are [17]:

  - To handle registration requests by acting as a registrar as defined in [RFC 3261][9]. The S-CSCF obtains the UE's IP address and which P-CSCF the UE is using as an IMS entry point during the initial SIP Register request.
  - To authenticate users by means of the IMS Authentication and Key Agreement (AKA) scheme. The IMS AKA achieves mutual authentication between the UE and the home network.
  - To download user information and service related data from the HSS during registration or when handling a request to an unregistered user.
  - To route mobile-terminating traffic to the P-CSCF of called party and to route mobile-originated traffic to the I-CSCF of called party's IMS network, to the Breakout Gateway Control Function (BGCF) or to one or more Application Server (AS).
  - To perform session control. The S-CSCF can act as a proxy and as UA as defined in [RFC3261] [9]. It performs the role of a proxy when services are being requested by the UE because the S-CSCF interfaces with the application servers (ASs). It also sometimes provide responses to requests thereby making it to play the role of a UA i.e. it may terminate and independently generate SIP transactions.
  - To interact with service platforms i.e., the capability to decide when a request or response needs to be routed to a specific AS for further processing.
  - To translate an E.164 number to a SIP universal resource identifier (URI) using a domain name system (DNS) translation mechanism as defined in

[RFC3401, RFC3761 [23, 36]. This translation is needed because routing of SIP signaling in IMS uses only SIP URIs.

- To supervise registration timers and to be able to de-register users when the registration time has expired.
- To execute media policing. The S-CSCF is able to check the content of the Session Description Protocol (SDP) payload to check whether it contains media types or codecs, which are not allowed for a user. Once the proposed SDP does not fit the operator's policy or user's subscription, the S-CSCF rejects the request and sends a SIP error message.
- To send accounting-related information to the CCF for offline charging purposes and to the Online Charging System (OCS) for online charging purposes.

- Interrogating CSCF (I-CSCF): a SIP proxy that provides the next hop to the S-CSCF that serves a user. Its primary role is to forward a request to the S-CSCF, by contacting the HSS. The I-CSCF is a contact point within an operator's network for all connections destined to a subscriber of that network operator. There may be multiple I-CSCFs within an operator's network. It is located in the home network where the user is registered.

  The functions performed by the I-CSCF are [17]:

  - To contact the HSS to obtain the name of the S-CSCF that is serving a user.
  - It also interfaces with the HSS and SLF in order to route incoming SIP message to the destination node within the IMS domain.
  - To assign an S-CSCF based on received capabilities from the HSS. This is done during the registration of a UE. An S-CSCF is assigned if there is no S-CSCF allocated.
  - To forward SIP requests or responses to the S-CSCF.
  - To send accounting-related information to the CCF.

- Emergency CSCF (E-CSCF): is a dedicated functionality to handle IMS emergency requests such as sessions towards police, fire brigade and ambulance. The main task of E-CSCF is to select an emergency centre also known as a Public Safety Answering Point (PSAP) where an emergency request should be delivered. Typically a selection criterion is a calling user's location and possible type of emergency (e.g., police, coast guard). Once an end user needs to reach an emergency dispatcher they are expected to dial one of common emergency numbers such as 911, 112. The UE is expected to translate this dialed emergency number to an emergency service Uniform Resource Name (URN) as specified in RFC5031. The UE places this URN in the Request URI field of the INVITE request. When an INVITE request arrives to P-CSCF, it needs to analyze the received Request-URI in order to detect all emergency session attempts. The P-CSCF stores a configurable list of local emergency service identifiers which are valid for the operator to which the P-CSCF belongs. The P-CSCF then selects E-CSCF to handle the request. Before forwarding the request to E-CSCF, the P-CSCF may optimally request and/or verify UE's location information using Location

Retrieval Information (LRF). The E-CSCF is responsible for routing emergency sessions towards the most appropriate PSAP whose selection is based on user's location and type of emergency. Once the appropriate emergency centre is selected, the E-CSCF routes the request to the emergency centre [25]. The E-CSCF can be located either in the visited or home network depending on the implementation adopted.

3.6.2. Application Servers

An application server (AS) hosts and executes services. Application servers are not pure IMS entities: rather, they are functions on top of IMS. They are entities that provide value-added multimedia services in the IMS.

SIP proxy is one of the components that are used by SIP to perform many of the call set-up functions. As defined in [RFC 3621] [9], SIP makes use of elements, called proxy servers (SIP proxy) "to route requests to the user's current location, authenticate and authorize users for services, implement provider call-routing policies, and provide features to users".

Within the SIP network the SIP proxy actually manages the setup of calls between SIP devices including the controlling of call routing and it also performs necessary functions such as registration, authorization, network access control and in some cases it also handles network security [3].

The SIP standard [RFC 3261] [9] briefly defines a B2BUA as a logical entity that receives a request and processes it as a User Agent Server (UAS) and in order to respond to them, it acts as a User Agent Client (UAC) and generates requests. Additionally it maintains dialog state and participates in all of the requests sent on the dialogs it has established. The standard defines it as a concatenation of two UAs, each playing the role of a UAC or UAS depending on the situation, and therefore doesn't provide additional definition for this entity [3]. Application servers can run in a number of defined SIP operational modes i.e. SIP proxy mode, SIP UA (User Agent) mode (endpoint), or SIP B2BUA (Back-to-Back User Agent) mode (i.e., a concatenation of two SIP User Agents).

The main functions of the AS are [25]:

- The possibility to process and modify an incoming SIP request message from the P-CSCF or I-CSCF.
- The capability to originate SIP requests.
- The capability to send accounting information to the charging functions.

A SIP AS is a SIP-based server that may be dedicated to a single service or host a wide range of value-added multimedia services. A SIP AS could be used to provide presence, messaging, Push to talk Over Cellular and conferencing services [25]. Since an AS may be dedicated to a single service and a user may have more than one service, there may be one or more AS's per subscriber or more than one AS involved in a single session.

The AS can either be located in the home network or in an external third-party network to which the home network maintains a service agreement. The third party here means a network or a stand-alone AS. The AS interfaces with the HSS through the Sh reference point. An AS may need data (related to particular identity of user or related to public service identity) or need to know which S-CSCF to send a SIP request and this type of information is stored in the HSS. Therefore there has to be a reference point between the HSS and the AS. This reference point is called the Sh reference point and the protocol is Diameter.

3.6.3. Service Capability and Interaction Manager

The Service Capability and Interaction manager (SCIM) function emerged in the 3GPP release 5 version of the 3G network [27]. It allows the ability to provide more complex service brokering above and beyond that provided with the S-CSCF [3]. The SCIM is used to invoke multiple IMS services using a single trigger from the core network.

The main functionality of the SCIM is to select, invoke and compose services and service features at the reception of SIP messages from the IMS core network. SCIM, located in the IMS application layer, makes up an additional layer between the S-CSCF and application servers. With the projection that IMS service environment will have a considerable larger number of available services compared with IN and that it would have a more dynamic environment because available services can change more frequently, the use of SCIM will be more flexible and less expensive to use than IFC [38]. Customized service selection and composition is based on user profiles associated to users.

The 3GPP specification of the SCIM leaves the internal structure and details of implementation open and as a result, the range of solutions under this label is broad.

3.6.4. The Databases: the HSS and the SLF

There are two main databases in the IMS architecture: the Home Subscriber Server (HSS) and Subscription Locator Function (SLF).

The HSS is the main central data base for user- and service-related data of the IMS. It contains all the user-related subscription data required to handle multimedia sessions. These data include user identities, registration information, access parameters, location information, security information (including both authorization and authentication information), user profile information, service-triggering information as defined in [3GPP TS 23.228 [18] and the S-CSCF (Serving-CSCF) allocated to the user.

User identities consist of two types: private user identities and public user identities. The private user identity is a user identity that is assigned by the home network operator and is used for such purposes as registration and authorization while the public user identity is the identity that other users can use for requesting communication with the end-user. The HSS also provides user-specific requirements for S-CSCF capabilities. This information is used by the I-CSCF to select the most suitable S-CSCF for a user. IMS access parameters needed to set up sessions which include parameters like user authentication, roaming authorization and allocated S-CSCF names are also stored in the HSS.

One or more HSS may be present in a network depending on the number of IMS subscribers that can be handled by a single HSS, the traffic capacity of the equipment

(e.g. number of interrogations per second) and the organization of the network. There are multiple reference points between the HSS and other network entities.
A network with more than one HSS requires a Subscription Locator Function (SLF). The SLF is a database that maps user identities to HSSs. The SLF enables the I-CSCF, the S-CSCF and the AS to find the address of the HSS that holds the subscriber data for a given user identity when multiple HSSs have been deployed by the network operator. Both the HSS and SLF support the Diameter protocol as defined in [RFC 3588] [13] with Cx and Dx as the respective reference points [2].

3.6.5. The Media Resource Function

The MRF provides source of media in the home network in which it resides. It provides IMS applications the capability to provide announcements, to mix media streams, to transcode between different codecs and do various sorts of media analysis.
The MRF is divided into the signaling plane mode called the MRFC (Media Resource Function Controller) and a media plane node called the MRFP (Media Resource Function Processor). The MRFC is tasked to handle SIP communication to and from the S-CSCF and to control the resources in the MRFP using an H.248 interface. The MRFP implements all the media-related functions, such as playing and mixing of incoming media streams (e.g., for multiple parties) [2], [25], media stream processing (e.g., audio transcoding, media analysis) [3GPP TS 23.228, TS 23.002] [18], [27]. The two together provide mechanism for bearer-related services such as conferencing, announcements to a user or bearer transcoding in the IMS architecture.

3.6.6. MGW/MGCF/SGW/BGCF

The Media Gateway (MGW)/Media Gateway Control Function (MGCF) and Signaling Gateway (SGW) collectively represent equipment that provides interworking with the legacy PSTN (Public Switched Telephone Network) and Public land mobile network (PLMN).
The IMS-MGW interfaces the media plane of the CS networks (PSTN, PLMN) and the IMS. It terminates the bearer channels from the CS network and media streams from the backbone network (e.g. RTP streams in an IP network), executes the conversion between these terminations and performs transcoding and signal processing for the user plane when needed. On one side, the MGW is able to send and receive IMS media over Real-Time Transport Protocol (RTP) as defined in [RFC 3550] [12]. On the other side, the MGW uses one or more TDM (Time Division Multiplexing) time slots to connect to the CS network. The IMS-MGW is controlled by the MGCF using the Mn reference point to control the user plane resources. The Mn interface controls the user plane between IP access and IMS-MGW. It also controls the user plane between CS access and IMS-MGW. It’s based on H.248.
The MGCF is the central node of the PSTN/CS gateway that enables communication between IMS network and CS network. It implements a state machine that does protocol conversion, maps SIP (the call control on the IMS side) to either ISDN User Part (ISUP) over IP or Bearer Independent Call Control (BICC) over IP (both BICC and ISUP are call control protocols in circuit-switched networks) and forwards the session to IMS. In

addition to call control protocol conversion, the MGCF controls the resources in the MGW (Media Gateway). The protocol used between the MGCF and MGW is based on H.248 (ITU-T Recommendation H.248) [2], [14].

The SGW interfaces the signaling plane of different networks such as Stream Control Transmission Protocol (SCTP) /IP-based signaling networks and SS7 signaling networks. The SGW performs lower-layer protocol conversion (both ways) at the transport level between the Signaling System No. 7 (SS7)-based transport of signaling and the IP-based transport of signaling. SGW is responsible for replacing the lower MTP (ITU-T Recommendation Q.701) transport with SCTP [as defined in RFC 2960] [28] over IP [25].

A Border Gateway Control Function (BGCF) is essentially a SIP server that involves routing based on telephone numbers. It identifies if a session terminates on the PSTN and determines which MGCF should handle it. It is responsible for choosing where a breakout to the CS domain shall occur. The outcome of the selection process can be either a breakout in the same network in which the BGCF is located or a breakout in another network. If the breakout happens in the same network in which the BGCF is located, then the BGCF selects a MGCF to perform the breakout. If the breakout is to take place in another network, then the BGCF forwards the SIP signaling via an IBCF (Interconnect Border Control Function) to another BGCF in a selected network [3GPP TS 23.228] [18], [17], [25]. IMS does not specify how the BGCF obtains the information on which to base the decision to route to a particular gateway (it might be static configuration, based on a routing protocol etc).

3.7 Identification in the IMS

The IMS employs Public User Identities and Private User identities (PUIs) in order to uniquely identify its users [2]. The private user identity represents the identity that is authenticated by the network. The public user identity is the one employed by other users to request communication with the IMS subscriber. These two identity formats are explained below:

3.7.1. Public User Identity

Users' identities in IMS networks are called public user identities. They are the identities used for requesting communication with other users. With these identities, users will be able to initiate sessions and receive sessions from any different networks, such as GSM networks and the Internet. To be reachable from the CS side, the public user identity must conform to telecom numbering (e.g., tel: +358501234567). In similar manner, requesting communication with Internet clients, the public user identity must conform to Internet naming (e.g., sip:bob.smith@ims.telia.se).

The IMS architecture specifies the following requirements for public user identity [3GPP TS 23.228, TS 23.003] [18], [29]:

- The public user identity/identities will take the form of either a SIP Uniform Resource Identifier (URI) or a telephone Uniform Resource Locator (tel URL) format.

- At least one public user identity will be securely stored in an IMS Identity Module (ISIM) application.
- It shall not be possible for the UE to modify the public user identity stored in an ISIM application.
- A public user identity shall be registered either explicitly or implicitly before the identity can be used to originate IMS sessions and to initiate IMS session-unrelated procedures (e.g., MESSAGE, SUBSCRIBE, NOTIFY).
- A public user identity shall be registered before terminating IMS sessions, and terminating IMS session-unrelated procedures can be delivered to the UE of the user that the public user identity belongs to. Subscriber-specific services for unregistered users may nevertheless be executed.
- It will be possible to register multiple public user identities through a single registration request via a mechanism within the IP multimedia CN subsystem (e.g. by using an Implicit Registration Set).
- Public User Identities may be used to identify the user's information within the HSS.

The tel URL scheme is used to express traditional E.164 numbers in URL syntax. The tel URL is described in [RFC 3966] [30], and the SIP URI is described in [RFC 3261 and RFC 2396] [9], [31].

3.7.2. Private User Identity

The private user identity is a unique global identity defined by the home network operator, which may be used within the home network to uniquely identify the user from a network perspective [3GPP TS 23.228] [18]. It is valid for the complete duration of the user's subscription with the home network. It does not identify the user herself; on the contrary, it identifies the user's subscription and device. Therefore, it is mainly used for authentication purposes during the registration phase [39]. It is possible to utilize private user identities for accounting and administrative purposes as well. The IMS architecture specifies the following requirements for private user identity [3GPP TS 23.228, TS 23.003] [18], [29], [7]:

- The private user identity will take the form of a Network Address Identifier (NAI) defined in [RFC 2486] [32]. It is possible for a representation of the IMSI to be contained within the NAI for the private identity.
- The private user identity will be contained in all registration requests passed from the UE to the home network although some SIP devices do not include the private user identity.
- The private user identity will be authenticated only during registration of the user (including re-registration and de-registration).
- The S-CSCF will need to obtain and store the private user identity on registration.
- The private user identity will not be used for routing of SIP messages.
- The private user identity will be permanently allocated to a user and securely stored in the HSS and in the IMS Identity Module (ISIM) application, if an

ISIM is used. The private user identity will be valid for the duration of the user's subscription within the home network.

- It will not be possible for the UE to modify the private user identity stored in an ISIM application.
- The private user identity will optionally be present in charging records based on operator policies.

3.8 SIM, USIM and ISIM in 3GPP

Together with the 3GPP terminals is the presence of a Universal Integrated Circuit Card (UICC). The UICC is a removable smart card used to store subscription information, authentication keys, phonebook and messages, among other things that can be inserted and removed from UE.

The operation of GSM and UMTS terminals depend on the presence of a UICC in the terminal. The interface between the UICC and the terminal is standardized as defined in [ETSI TS 102.221 Release 7 and 3GPP TS 31.101] [33], [34].

A UICC may contain several logical applications such as a Subscriber Identity Module (SIM- which is widely used in the 2G networks such as GSM networks), a Universal Subscriber Identity Module (USIM- used to access the UMTS networks, the 3G evolution of GSM) and an IP Multimedia Services Identity Module (ISIM- collection of parameters that allows a terminal to operate in the IMS) [7].

Access to a 3GPP IMS network relies on the presence of either an ISIM or a USIM application in the UICC with preference for ISIM because it is designed for the IMS although access with USIM is also possible.

The ISIM itself stores IMS-specific subscriber data mainly provisioned by an IMS operator. This data is mainly used when a user registers a device to the IMS. The following data can be stored in ISIM e.g. when a user obtains an IMS subscription from an operator [25]:

- Private user identity of the user – it is used in a registration request to identify the user's subscription.
- One or more public user identities of the user – it is used in a registration request to identify an identity to be registered and is used to request communication with other users.
- The name of the entry point of the home network (home network domain name) – it is used in a registration request to route the request to the user's home network.
- Address of P-CSCF – it can be used when the access technology does not support dynamic P-CSCF discovery capabilities.
- Security parameters related to Generic Bootstrapping Architecture. Security parameters enable authentication to IMS.

4. Video Telephony

This chapter focuses on the basic requirement and concept for video transmission over the Internet, and video compression standards. Furthermore, the main protocols, RTP/RTCP, are explained with their various important heard fields

4.1 Introduction to Video telephony

Video telephony is one of the most appealing possibilities promised by broadband Internet. It is full-duplex, real-time audio-visual communication between or among end users. The initial primary challenge facing video telephony has been bandwidth because of the fact that high-resolution, full-motion video data requires far more bandwidth than audio data. With broadband Internet solutions such as DSL (Digital Subscriber Line), cable and land-based wireless, it is now possible to transmit and receive video data at high resolutions and higher refresh rates than with ordinary telephone system, such as ISDN.

Video telephony can be categorized according to its intended purpose, functionality or its method of transmission. In this chapter, emphasis is laid on video telephony over IP (Internet Protocol) for peer-to-peer video telephony and for video conferencing.

The Internet operates as a packet-switched network that interconnects end nodes implementing the TCP/IP protocol stack. The Internet Protocol (IP) resides in the network layer protocol in the TCP/IP protocol stack. Under IP, each host (network device) that communicates directly with the Internet has an address assigned to it that is unique within the network. This is known as the IP address of the host. Further, each IP address is subdivided into smaller parts: a network identifier part and a host identifier part. The former uniquely identifies the access network to which the host is attached [47]. The latter indicates the host (device) within a particular network. Routers periodically exchange routing information about address identifiers of the concerned access networks. As a result of this periodic information exchange, routers are able to build and maintain routing tables that guide packet forwarding among different networks. These tables are used at each intermediate router along a path within the network to indicate the forwarding interface an IP packet (datagram) should take in order to get to its destination.

In the IP layer, fragmentation of packets into smaller parts occurs when the packet size exceeds a maximum frame size of the network (MTU), which is done by the network layer. These fragments are then forwarded by the network in separate packets toward the destination. Reassembly of these fragments into the original packet is done at the destination end. Note that if a fragment is lost within the network, the destination is unable to reassemble the packet that has been fragmented and the surviving fragments are then discarded. The forwarding decision for each packet is individually taken at each intermediate IP signaling node. As incoming IP packets arrive at an intermediate router, they are placed in an input buffer waiting for the router to process a routing decision for each one of them to indicate their appropriate output interface. Persistent packet buffering at routers is known as network congestion. Furthermore, if a packet arrives at a router and the buffer is full, the packet is simply discarded by the router. Therefore, severe network congestion causes packet losses as buffers fill up. As a consequence of these characteristics, IP service offers no guarantees of bounded delay and limited packet loss rate. It is also not guaranteed that packets of a single flow will follow the same path within the network or will arrive at the destination in the same order they were originally transmitted by the source [45].

IP provides a connectionless best-effort service to the transport layer protocols. The main transport protocols of the TCP/IP protocol stack are the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP). TCP offers a connection-oriented and reliable service. In turn, UDP provides a connectionless best-effort service. The choice of transport protocol depends on the requirements of the application. From the viewpoint of an application, UDP simply provides an extension of the service provided by IP with an additional UDP header checksum. Hence, applications using UDP see the IP service as it is. In contrast with UDP, TCP seeks to mask the network service provided by IP to the application protocols by applying connection management, congestion control, error control, packet split up and re-assembly and flow control mechanisms.

The Real-time Transport Protocol (RTP) [12] supports applications streaming audio and/or video with real-time constraints over the Internet. RTP is composed by a data part and a control part. The data part of RTP provides, besides transporting the actual multimedia, functionality suited for carrying real-time content, e.g. the timing information required by the receiver to correctly output the received packet stream. The control part, called Real-time Transport Control Protocol (RTCP) offers control mechanisms for synchronizing different streams with timing properties prior to decoding them. RTCP also provides reports to a real-time stream source on the network Quality of Service (QoS) offered to receivers in terms of delay, jitter, and packet loss rate experienced by the received packets. These functionalities are also available to the multicast distribution of real-time streams.

Video conferencing is an interactive method of communication that combines the use of video, audio and computing technologies to allow people in different locations to *virtually* meet face-to-face to conduct a conference in real-time. If both audio and video media are used in a conference, they are transmitted as separate RTP sessions. RTCP packets are transmitted for each medium using two different UDP port pairs and/or multicast addresses. There is no direct coupling at the RTP level between the audio and video sessions, except that a user participating in both sessions should use the same distinguished (canonical) name in the RTCP packets for both so that the sessions can be associated [12]. One motivation for this separation is to allow some participants in the conference to receive only one medium if they choose. Despite the separation, synchronized playback of a source's audio and video at the receiving end can be achieved using timing information carried in the RTCP packets for both sessions.

Despite the rapid expansion and improvement of the Internet underlying infrastructure, quality-of-service (QoS) is still one of the major challenges of real-time communication over IP networks. The unreliable and stateless nature of today’s Internet protocol results in a best-effort service, that is, packets may be delivered with an arbitrary delay or may even be lost. Transmitted over the best-effort network and suffering from variable throughput, delay, and loss, data packets have to be delivered by a 'deadline' to remain useful. Excessive delay severely impairs communication interactivity; packets loss results in poor audio and picture quality and in may result in frozen or non-synchronized frames in video. The heterogeneity of today's Internet also poses a major challenge for media delivery to users with various connection speeds and also different user equipment capabilities.

4.2 Concept of transmitting video stream over the Internet

To give a basic view of video communication system, a generic functional diagram for a video transmission process is shown in figure 4.1. The first step in the process is to analyze the supplied analog/digital video signal. The analysis can include such operations as filtering, analog to digital conversion (if received video data is analog) and usually, no compression is done with the analysis [56]. Data is only transformed to a format that is more compressible than the original signal format.

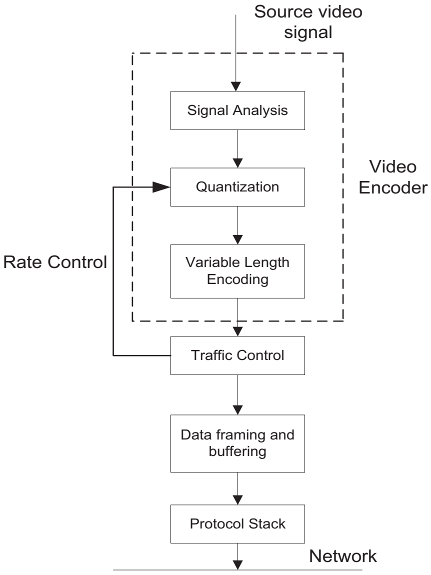


Figure 4.1: Functional diagram of a video transmission process [47]

The second step performs quantization of the signal, either lossless or lossy way. In a lossy system the quantizer reduces signal quality in a way that is as acceptable as possible to the eye. In the variable length coding block each signal event will have a code with different number of bits. That is why it is also called entropy coding. To get compression, short codes are assigned to frequently occurring events and long codes to infrequent events.

The traffic control block follows data flow status in a communication channel, adjusts encoder parameters (Rate control in the above picture) according to the data flow status in order to adapt generated video data to the communication channel. The next block forms data packets according to a used protocol. It also buffers the packets in order to serve a continuous and smooth data stream to the communication channel.

Such system parameters as buffer size and packet length are very essential for the system performance, and therefore they must be considered very carefully during design.

With reception of a video data stream, a receiver must know the received data format. SIP/SDP compliant applications [9], [35] negotiate at the beginning of a communication session to find a suitable format for both sides. When data transmission starts, a goal is that the receiver should serve video frames to its decoder at a rate equal to that at which frames were generated by the coder at the sending side. Due to packet delay variations in the transfer path, received packets are buffered to get required tolerance/resilience for delay variations. In order to have more tolerance for the delay variations, more video data can be buffered so that longer delay 'offset', can be achieved accordingly. This causes an optimization problem between communication delay and the QoS. In addition to the buffering task, the receiver must also check order of packets and rearrange them if necessary, and decide what to do in the case of erroneous or lost packets. RTP protocol [12] helps greatly to implement these functions by offering required parameters such as inclusion of time stamps and sequence numbers in the header of the packets.

4.2.1. Video requirements

The nature of video content directly influences the achieved compression rate of a video encoder and the resulting traffic to be transmitted. For example, on the one hand, a news video sequence usually shows a person just narrating events and, as a consequence, most of the scene is still, thus favoring video compression techniques based on motion estimation. On the other hand, an action movie is less susceptible to compression because of frequent camera movements and object displacements at the movie scenes. Furthermore, scene changes produce disruptions that result in larger coded frames. Therefore, video sequences with frequent scene changes, like music video clips, generate highly bursty network traffic.

The transmission of video sequences over the Internet imposes different requirements on video quality delivery. Processing, transmission, propagation, and queuing delays compose the total delay that takes a packet to be fully transmitted from the video source to its destination. The processing delay consists of the coding and packetization at the source, the packet treatment at intermediate routers, and the depacketization and decoding at the receiver. The transmission delay is a function of, among others, the packet size and transmission capacity at the links [47]. The propagation delay is a characteristic of each communication medium. The queuing delay is unpredictable because it depends on the concurrent traffic video (or other) packets encountered at each intermediate router. Maximum delay is an important metric for interactive applications with real-time requirements, such as videoconferencing and distance collaboration. Streaming video applications also depend on packets arriving within a bounded delay for timely content reproduction. If a packet arrives too late at the receiver, the packet is considered as lost since it is useless for video playback. A bounded jitter is also desirable because it reduces the buffer capacity needed at receivers to compensate for these variations in delay. IP is a connectionless protocol and hence packets may follow different paths through the network unless MPLS[1] is used. As a consequence, they may arrive at the receiver out-of-order. All these issues affect the video quality perceived at the receiver by an end user.

4.2.2. Video compression standards

The need for video and image compression in order for conservation of network bandwidth is of high importance and this has resulted in the development of a number of compression standards over the years. As it is well established, the high heterogeneity of the Internet is characterized by a large number of low-capacity terminal links and core infrastructures. Algorithms for compressing information, and in particular video data, are highly demanded and have been the main focus of many research efforts in the latest decades. There could be a lot of redundancies in an image and we can predict most of the rest of the features by examining part of it. Also there could be a lot of correlations between several consecutive images and we can make a fairly good prediction of one from the others. In general, the main goal of coding is the bit-rate reduction for storage and transmission of the video source while retaining video quality as good as possible. The basic idea behind video compression is to remove spatial redundancy within a video frame and temporal redundancy between adjacent video frames [56]. Two entities have contributed with the most used algorithms and standards for video compression: MPEG (Moving Picture Expert Group) and ITU (International Telecommunications Union).

The process of compression involves applying an algorithm to the source video data stream to create a compressed format which is ready for transmission or storage. To play the compressed file, an inverse algorithm is applied to produce a video that shows (virtually) the same content as the original video. The time it takes to compress, send, decompress and display a file is called latency. Generally, the more advanced the compression algorithm, the higher the latency.

There are two types of compression: lossless and lossy. Lossless data compression is used when the data must be restored exactly as it was before compression. Lossy compression works on the assumption that the data doesn’t have to be restored perfectly. A good deal of redundant information can simply be thrown away from the image, video and audio data and such data will still be of acceptable quality [47].

A pair of algorithms that works together is called a video codec (encoder/decoder). Video codecs of different standards are normally not compatible with each other; that is, video content that is compressed using one standard cannot be decompressed with a different standard. For instance, an MPEG-4 decoder will not work with an H.264 encoder. This is simply because one algorithm cannot decode the output from another algorithm but it is possible to implement many different algorithms in the same software or hardware, which would then enable multiple formats to coexist.

The Motion Picture Expert Group (MPEG), launched in 1998, is an ISO/IEC working group that works towards standards for both audio and video digital formats and for multimedia description frameworks. The main standards for video coding released by this group include MPEG-1, MPEG-2, and MPEG-4. The MPEG-1 standard defines a series of encoding techniques for both audio and video (video is part 2 of the standard), designed for generating flows of up to 1.5 Mbps. The main goal of MPEG-1 was to address the problem of storing video in CD-ROMs, and it has become a successful format for video exchange over the Internet. However, higher rates than the 1.5 Mbps of MPEG-1 became rapidly a need. This led to the definition of MPEG-2, which defines rates from 1.5 to tens of Mbps. MPEG-2 is based on MPEG-1, but proposes a number of new techniques to address a much larger number of potential applications, including digital video storage and transmission, high-definition television (HDTV), and digital video disks (DVDs) [45]. MPEG-1 and 2 achieve good compression ratios by implementing causal and non-causal prediction. Briefly, a video flow is defined as a sequence of group of pictures (GOP), composed of three types of frames (or pictures): I, P, and B. I-frames, also called reference frames, are basically low-compressed pictures that serve as reference for the computation of P and B frames. P-frames are obtained from a past I-frame by using motion prediction, and can then be encoded with higher compression ratios. B-frames, or bidirectional frames, are based on both previous and future I and P frames within a GOP, which leads to high compression ratios [45].

MPEG-4 part 2 (see also H.264 below) covers a gap in the objectives of the previous two standards: the need for a flexible framework that would be adaptable to the wide range of applications and the high heterogeneity found in the Internet. MPEG-4 targets the scaling from a few Kbps to moderately high bit rates (about 4 Mbps). The main innovation brought by MPEG-4 was the use of the concept of video objects, in which a scene is decomposed in a number of objects that can be treated differently one from another during the video transmission (e.g. prioritization among objects) [45].

The International Telecommunication Unit (ITU), in its H series (e.g. H.310, H.320, and H.324), has also addressed the problem of transmitting compressed video over the Internet. The first standard proposed by ITU is the H.261 video codec. In order to deal with a wide range of communication patterns, the H.261 standard defines a number of methods for coding and decoding video at rates of p×64 Kbps, where p varies in the range of 1 to 30. Later, three other standards have appeared: H.262, H.263, and H.264. The H.262 standard targets higher bit rates. H.261 and H.263 are based on the same principles, although H.263 introduces a number of improvements that lead to equivalent video quality for even half of the bandwidth. To solve the incompatibility between different TV standards (PAL, SECAM and NTSC), the CIF and QCIF picture structures were introduced. The CIF (Common Intermediate Format) is a format used to standardize the horizontal and vertical resolutions in pixels of YCbCr sequences in video signals, commonly used in video teleconferencing systems. It was first proposed in the H.261 standard. The QCIF (Quarter CIF) format, which employs half the CIF spatial resolution, is also supported by all H.261 compatible codecs. To have one fourth of the area as "quarter" implies the height and width of the frame are halved [47], [48]. H.261 supports both QCIF (176×144 pixels) and CIF (352×288) resolutions, which are also supported by H.263, although H.263 introduces a number of improvements that lead to equivalent video quality for even half of the bandwidth. H.263 also supports SQCIF (128×96 pixels), 4CIF (704×576 pixels), and 16CIF (1408×1152 pixels). The H.264

standard, also known as MPEG-4 part 10 or H.264/AVC, is the result of a joint work between ITU and MPEG (the partnership is called Joint Video Team -JVT), with the objective of defining a codec capable of generating, without increased complexity, good video quality at lower bit rates than previous formats (H.262, MPEG-2, MPEG-4 part 2). H.264/AVC makes use of advanced coding techniques in order to generate video for a wide range of applications, from low to high resolution and at varying bit rates (for example, DVD, broadcast, 3G mobile content) [45].

4.3 Transport protocols

The Internet network provides no guarantees for traffic. The throughput and delay along a path can vary from time to time and samples can arrive in different order than they were sent (jitter and longer-than-expected latency). When the network is heavily loaded, packets may get lost leaving gaps in the data stream.

Multimedia applications need a transport protocol to handle a common set of services which does not have to be complex like TCP. The goal of the transport protocol (RTP) is to provide end-to-end delivery services that are specific to multimedia applications. The services can be distinguished clearly from conventional data services:

- A basic framing service is needed, defining the unit of transfer, typically common with the unit of synchronization.
- Multiplexing is needed to identify separate media in streams.
- Timely delivery is needed.
- Synchronization is needed between different media and it is also a common service to networked multimedia applications.

The Internet standard for conveying media streams is the Real-time Transport Protocol (RTP) [12]. This application (in OSI layer terms) was developed specifically for streaming data across IP networks. RTP is the most important streaming standard. All media streams, regardless of their format and content, are encapsulated in RTP packets. RTP provides several data fields that are not present in TCP, in particular a Timestamp and a Sequence Number which are formatted specifically for each media type. These are used by the playout and synchronization algorithms. RTP runs on UDP/TCP and uses its multiplexing and checksum functionalities, since it carries source identifiers. It allows control of the buffer so that the video stream is served at the correct speed. The media player is then able to reassemble the received RTP packets into the correct order and play them out at an appropriate speed. RTP transmits packets in real time. Lost or damaged packets are not retransmitted. There are strategies for the client software on how best to cope with the missing packets (e.g. error concealment, packet replication, etc). If the connection speed is lower than the data rate of the media, the transmission breaks up and the media plays poorly (or does not play at all). If the connection is fast, the extra bandwidth remains untouched and the user load only depends on the number of streams and the bandwidth of these streams.

RTP is divided into two parts: the data transmission protocol and the control protocol, RTCP. The RTP Control Protocol (RTCP) is used to convey additional information such as participant information and statistics of packet loss and arrival. RTP is intended to be malleable to provide the information required by a particular application [12]. Real Time Control Protocol (RTCP) provides support for real-time multiparty conferencing within the Internet, including support for gateways and unicast-to-multicast translators. It provides feedback to the service provider on the network reception quality from each participant in an RTP session in real time. The messages include reports on the number of packets sent/lost and the jitter statistics (early or late arrivals). These QoS related information at a receiver end can be used by higher-level applications to control the session and improve the transmission; for example, the bit-rate of a stream could be changed to combat network congestion. This allows a sender to adjust transmission parameters so that a receiver achieves the best possible results. The receiver can get such necessary information as sender identifier, packet sequence numbers, and timestamps of packets, for example. This information is an essential base when QoS is optimized at the receiver end. RTCP message types include *sender report*, *receiver report*, *source description*, *bye* and *application*. RTCP bandwidth uses 5% of the media session bandwidth [12]. The session bandwidth value is calculated by each participant in the session which is a product of the individual sender's bandwidth times the number of participants from which approximately 5% of this value is reserved for RTCP by each participant.

4.4 Establishing video session over IP

The figure below depicts the overall architecture of video call set up over the internet protocol using the IETF signaling protocol, SIP, in order to handle the setup and teardown of sessions between end points. This signaling protocol (SIP) is transported over UDP or TCP and it uses *invitations* to create Session Description Protocol (SDP) messages to carry out capability exchange and to setup call control channel use which allows participants to agree on a set of compatible media types.

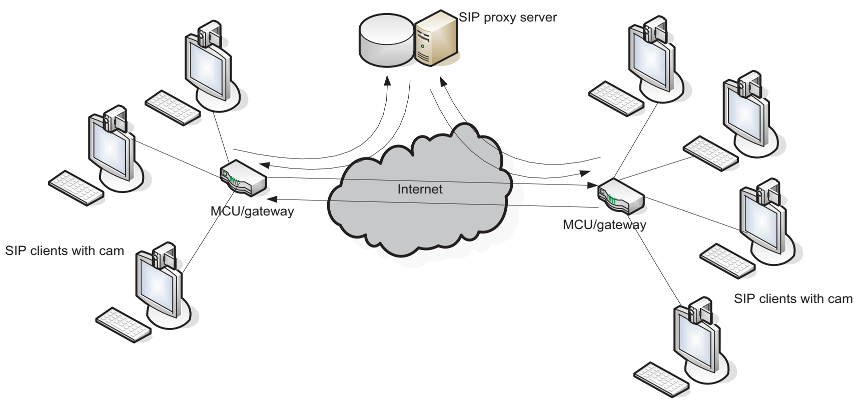


Figure 4.2: Video over IP using SIP proxy architecture [39]

The approach depicted in figure 4.2 can also be used to implement video conferencing. The SIP proxy server can act as a registrar for some or all the participants in the session. The MCU (Multipoint Control Unit) is used for enabling video conferences involving more than two video conferencing systems. The MCU acts as a hub to which all of the videoconferencing systems connect, and distributes the audio and video received from each videoconferencing system to the other videoconferencing systems involved. MCUs are sometimes combined with a terminal, gatekeeper or, most commonly, a gateway between different videoconferencing technologies, for example IP and ISDN. The MCU acts as an aggregation point for all the videoconference participants.

4.5 Encoding data streams into RTP messages

For video telephony over the Internet, where both audio and video media are involved in the session, the media are transmitted as separate RTP sessions. That is, separate RTP and RTCP packets are transmitted for each medium using two different UDP port pairs and/or multicast addresses as depicted in Figure 4.3. There is no direct coupling at the RTP level between the audio and the video sessions, except that a user participating in both sessions should use the same distinguished (canonical) name (this is stated in the SDES parameters or items which include the CNAME) in the RTCP packets for both so that the sessions can be associated. For example, when SIP is used, the two associated media streams are defined in SDP offer/answer negotiated during call set-up and this definition allows for media association by the receiver.

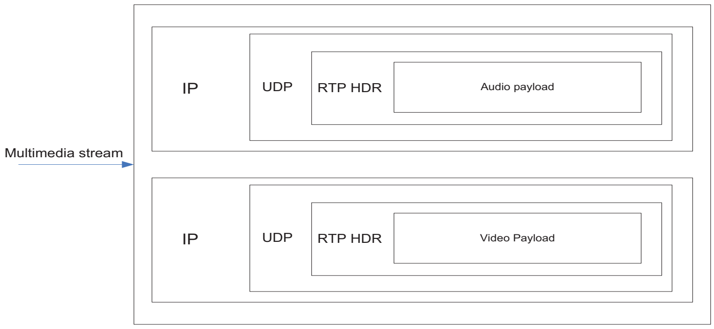


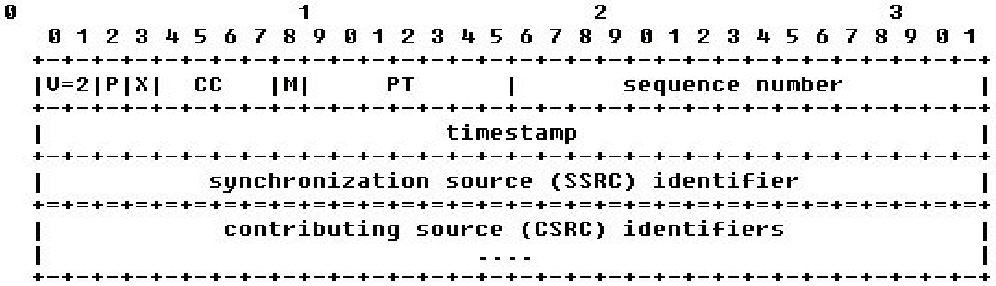
Figure 4.3: A multimedia stream made of two different payloads

Each participant sends audio and video data in small chunks of, say 20ms duration (for video stream, the packetization time is not consistent since it depends on the video source). Each chunk of the audio and video data is preceded by an RTP header; RTP header and data are in turn contained in a UDP packet. This forms the UDP payload which is the data transported by RTP in a packet (for example, the compressed video data). The RTP header indicates the type of audio or video encoding (codec can be any of PCM, GSM, LPC, H.263, H.263+ etc) that is contained in each packet so that receivers can decode the messages by applying the required decoding algorithm on the received packets, hence smooth communication [42].

The Internet, like other packet networks, occasionally loses and reorders packets and delays them by variable amounts of time. To cope with these impairments, the RTP header contains timing information and a sequence number that allow the receivers to reconstruct the timing produced by the source, so that in this example, chunks of audio are continuously played out by the speaker every 20ms. This timing reconstruction is performed separately for each source of RTP packets in the conference. The sequence number data contained in the RTCP sender and receiver reports can also be used by the receiver to estimate how many packets are being lost.

4.6 RTP Header fields

Figure 4.4 shows the format of an RTP header. Typically in one-to-one telephony applications, the size of the RTP header is 12 bytes (no CSRC). After the header, optional header extensions may be present. The header is followed by the RTP payload, the format of which is determined by the particular class of application. The fields in the header are as follows:



```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |V=2|P|X|  CC   |M|     PT      |       sequence number         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                           timestamp                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |           synchronization source (SSRC) identifier            |
 +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
 |            contributing source (CSRC) identifiers             |
 |                             ....                              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4.4: RTP header format [12]

The first twelve octets are present in every RTP packet, while the list of CSRC identifiers is present only when inserted by a mixer.

Payload type (PT): 7 bits

This field identifies the format of the RTP payload and determines its interpretation by the application. A set of default mappings for audio and video is specified in RFC 3551 [42].

Sequence Number: 16 bits

The value of the Sequence Number increments by one for each successive packet. It is used by the receiver to detect packet loss, to re-order out-of-order packets and to delete duplicates. The initial number for a stream session is taken at random. Among the first set of RTP packets received, the first one will contain a marker field bit (true).

Timestamp: 32 bits

This is a sampling instance derived from the transmitter’s 90 kHz clock reference and is synchronized to the system’s program clock reference. It allows for synchronization and jitter calculations. It is monotonic and linear in time.

Synchronization Source Identifiers: 32 bit

This is a unique identifier for the synchronization of the RTP stream. The SSRC is the source identifier of the RTP stream. The CSRC exists when multiple media are involved, for example more than two participants. For a peer-to-peer communication, SSRC and CSRC are the same. In fact, the CSRC is excluded in RTP stream of peer-to-peer communication because it’s the same as the SSRC.

One or more contributing sources (CSRCs) exist if the RTP stream carries multiple media streams such as, for example, a mix of several video and audio sources (for example in a conference video call)

4.7 RTCP Packet format

The RTCP packet types include [12]:

SR: sender report, for transmission and reception statistics from the participants that are active senders.

RR: receiver report, for reception statistics from participants that are not active senders and in combination with SR for active senders reporting on more than 31 sources.

SDES: source description items, including CNAME.

BYE: indicates end of participation.

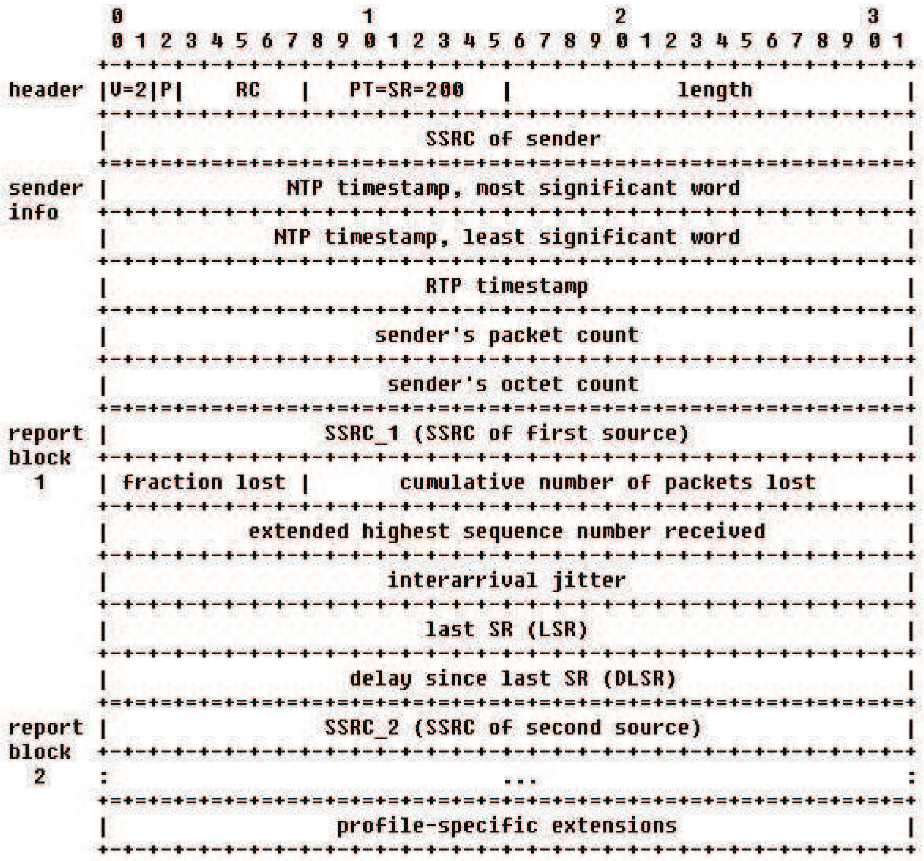
APP: application-specific functions.

4.7.1. Sender and Receiver reports

The sender and receiver reports are of main importance among the RTCP message types because they actually give an insight into the quality of reception of media exchanged between participants in a video/audio session.

RTP receivers provide reception quality feedback using RTCP report packets which may take one of two forms depending upon whether or not the receiver is also a sender. The only difference between the sender report (SR) and receiver report (RR) forms, besides the packet type code, is that the sender report includes a 20-byte sender information section for use by active senders. The SR is issued if a site has sent any data packets during the interval since issuing the last report or the previous one, otherwise the RR is issued.

Figures 4.5 and 4.6 define the format of the two reports.



```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
header |V=2|P|    RC   |   PT=SR=200   |             length            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                         SSRC of sender                        |
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
sender |              NTP timestamp, most significant word             |
info   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |             NTP timestamp, least significant word             |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                         RTP timestamp                         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                     sender's packet count                     |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                      sender's octet count                     |
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
report |                 SSRC_1 (SSRC of first source)                 |
block  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  1    | fraction lost |       cumulative number of packets lost       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |           extended highest sequence number received           |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                      interarrival jitter                      |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                         last SR (LSR)                         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                   delay since last SR (DLSR)                  |
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
report |                 SSRC_2 (SSRC of second source)                |
block  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  2    :                               ...                             :
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
       |                  profile-specific extensions                  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4.5: Sender Report: RTCP packet [12]

Reception report count (RC): 5 bits

The number of reception report blocks contained in this RTCP message.

Packet type (PT): 8 bits

Contains the constant 200 to identify this as an RTCP SR message.

Length: 16 bits

The length of this RTCP packet in 32-bit (4-byte) words minus one, including the header and any padding. In this case, the length is 12, meaning the RTCP packet size is (12+1) x 4 = 52 bytes.

SSRC: 32 bits

The synchronization source identifier for the originator of this SR packet.

NTP timestamp: 64 bits

This is the Network Time Protocol [12] is a protocol for synchronizing the clocks of computer systems in order to resist the effects of variable latency by using a jitter buffer. It indicates the wall clock time when this report was sent so that it may be used in combination with timestamps returned in reception reports from other receivers to measure round-trip propagation to those receivers.

RTP timestamp: 32 bits

Corresponds to the same time as the NTP timestamp (above), but in the same units and with the same random offset as the RTP timestamps in data packets.

Sender's packet count: 32 bits

The total number of RTP data packets transmitted by the sender since starting transmission up until the time this SR packet was generated.

Sender's octet count: 32 bits

The total number of payload octets (i.e., not including header or padding) transmitted in RTP data packets by the sender since starting transmission up until the time this SR packet was generated. This field can be used to estimate the average payload data rate.

SSRC_n (source identifier): 32 bits

The SSRC identifier of the source to which the information in this reception report block pertains.

Fraction lost: 8 bits

The fraction of RTP data packets from source SSRC_n lost since the previous SR or RR packet was sent, expressed as a fixed point number with the binary point at the left edge of the field.

Cumulative number of packets lost: 24 bits

The total number of RTP data packets from source SSRC_n that have been lost since the beginning of reception.

Extended highest sequence number received: 32 bits

The low 16 bits contain the highest sequence number received in an RTP data packet from source SSRC_n, and the most significant 16 bits extend that sequence number with the corresponding count of sequence number cycles.

Interarrival jitter: 32 bits

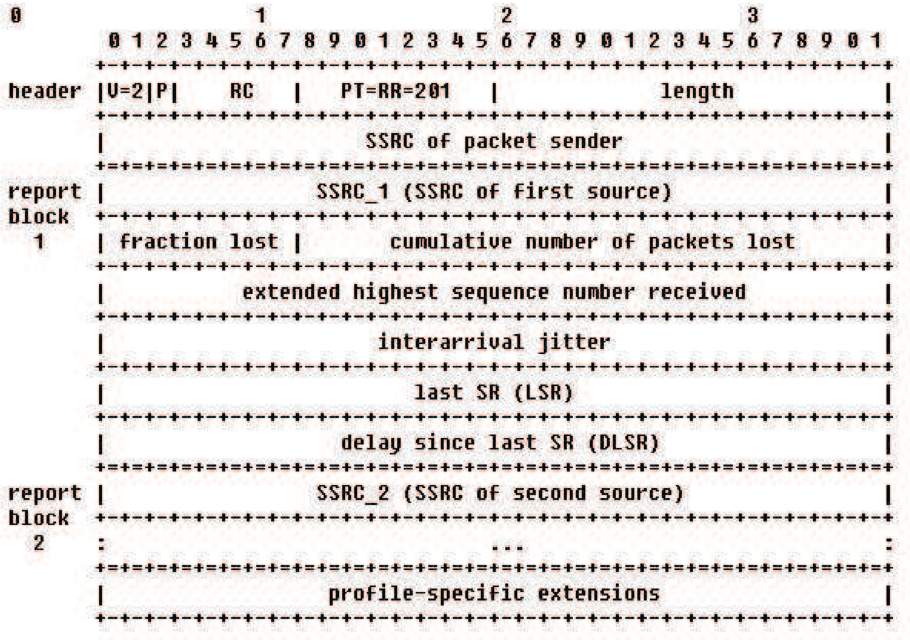
An estimate of the statistical variance of the RTP data packet interarrival time measured in timestamp units and expressed as an unsigned integer.

Last SR timestamp (LSR): 32 bits

The middle 32 bits out of 64 in the NTP timestamp received as part of the most recent RTCP sender report (SR) packet from source SSRC_n.

Delay since last SR (DLSR): 32 bits

The delay, expressed in units of 1/65536 seconds, between receiving the last SR packet from source SSRC_n and sending this reception report block.



```
0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
header |V=2|P|    RC   |   PT=RR=201   |             length            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                     SSRC of packet sender                     |
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
report |                 SSRC_1 (SSRC of first source)                 |
block  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  1    | fraction lost |       cumulative number of packets lost       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |           extended highest sequence number received           |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                      interarrival jitter                      |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                         last SR (LSR)                         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                   delay since last SR (DLSR)                  |
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
report |                 SSRC_2 (SSRC of second source)                |
block  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  2    :                               ...                             :
       +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
       |                  profile-specific extensions                  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4.6: Receiver Report: RTCP packet [12]

The format of the Receiver Report (RR) packet is the same as that of the Sender Report (SR) packet except that the packet type field contains the constant 201 and the five 32-bit words of sender information are omitted (these are the NTP and RTP timestamps and sender's packet and octet counts). The remaining fields have the same meaning as for the SR packet.

4.8 Overview of existing codecs

An overview of some of the existing codecs (according to the RTP A/V profile) [42] is given here together with their payload types as shown in Table 4.1 [42].

All the video encodings use an RTP timestamp (sampling) frequency of 90,000 Hz, the same as the MPEG presentation time stamp frequency. For most of these video encodings, the RTP timestamp encodes the sampling instant of the video image contained in the RTP data packet. If a video image occupies more than one packet, the timestamp is the same on all of those packets. Packets from different video images are distinguished by their different timestamps [42].

Most of these video encodings also specify that the marker bit of the RTP header should be set to one in the last packet of a video frame and otherwise set to zero. Thus, it is not necessary to wait for a following packet with a different timestamp to detect that a new frame should be displayed.

CelB

This image compression algorithm is a variable bit rate video coding scheme which allows applications to transport CellB video flows over protocols used by RTP. The CELL-B encoding is a proprietary encoding proposed and implemented by Sun Microsystems. This scheme has been specially adapted for network-based video applications, especially robustness to packet loss [49].

JPEG

In this encoding scheme, the RTP packet format is optimized for real-time video streams where codec rarely change from frame to frame. The encoding is specified in ISO Standards 10918-1 and 10918-2. The JPEG scheme was initially designed to compress still images. It is applied to video by compressing each frame of video as an independent still image and transmitting them in series [50].

H261

H261 uses the temporal redundancy of video to perform compression. This scheme is sensitive to packet loss. In order to combat this, the INTRA-frame encoding refreshment rate can be adjusted according to the packet loss observed by the receivers [51].

The encoding is specified in ITU-T Recommendation H.261, "Video codec for audiovisual services at px64 Kbit/s".

H263

The scheme is designed for video coding at very low data rate. Three modes are available in this compression scheme depending on the desired network packet size to be used in the video streams. H.263 supports four negotiable coding options to improve performance which can be used in any combination.

The encoding is specified in the 1996 version of ITU-T Recommendation H.263, "Video coding for low bit rate communication"[52].

H263-1998

Numerous coding options were added to this scheme to improve codec performance over the H263. These added options provide enhanced error resilience capability, prevention of propagation of errors from one segment of the picture to the others. The encoding is specified in the 1998 version of ITU-T Recommendation H.263, "Video coding for low bit rate communication" [53].

MPV

MPV designates the use of MPEG-1 and MPEG-2 video encoding elementary streams as specified in ISO Standards ISO/IEC 11172 and 13818-2, respectively. The scheme is designed to support maximum interoperability with MPEG system environments and to provide maximum compatibility with other RTP-encapsulated media streams [54].

MP2T

MP2T designates the use of MPEG-2 transport streams, for either audio or video. In this scheme, the RTP timestamp is primarily used to estimate and reduce any network-induced jitter and to synchronize relative time drift between the transmitter and receiver [54].

PT	Encoding name	Media type	Clock rate (Hz)
24	unassigned	V	
25	celB	V	90,000
26	JPEG	V	90,000
27	unassigned	V	
28	nv	V	90,000
29	unassigned	V	
30	unassigned	V	
31	H261	V	90,000
32	MPV	V	90,000
33	MP2T	AV	90,000
34	H263	V	90,000
35-71	unassigned	-	
72-76	reserved	N/A	N/A
77-95	unassigned	-	
96-127	dynamic	dynamic	
dyn	H263-1998	V	90,000

Table 4.1: Payload types (PT) for video and combined encodings [42]

4.9 Session setup

An important issue in multimedia communications is how sessions are set up. Session establishment applies for multimedia communications in general, and can be set up in two different ways: by announcement or by invitation. A TV program is an example of session established by announcement, whereas a videophone call is an example of session established by invitation. The Internet Engineering Task Force (IETF) has proposed a set of protocols for session setup, covering description, announcement, and invitation phases. The ITU-T has also defined a standard, H.323 [46], but in this report the focus is on the IETF's solutions.

The common protocol for media session description is the Session Description Protocol (SDP) [35], which is used to describe the characteristics of the session to be established. In SDP, sessions are characterized by a well-defined set of descriptors in textual form, including: the name of the session, the objective, associated protocols, information about codecs, timing, among others. SDP operates in a complementary way with the protocols defined in the following. The Session Announcement Protocol (SAP) is a very simple protocol for announcing future multimedia sessions [55]. It basically sends over a multicast session, in a periodic fashion, the description of the session defined by SDP. A bit more complicated is the Session Initiation Protocol (SIP), standardized to control multimedia sessions by invitation. One of the main contributions of SIP is the way it addresses corresponding nodes. In the classical telephone network (PSTN), when making a call, the initiator knows exactly where the destination phone is physically located, but is neither sure that the person that will respond is the one she/he searches nor that this latter will be at the other side of the line. In SIP, the idea is to call a person and not the phone this person may be near. This principle also applies to GSM. In a situation where SIP is used for residential wireline telephony, the same situation of calling a home-bound phone and not a person also arises.

4.10 Quality of Service (QoS)

The best-effort approach of the conventional Internet has become inadequate to deal with the very diverse requirements on network QoS of video streams and other video applications. The hierarchical structure of video encoding with possible error propagation through its frames imposes a great difficulty on sending video streams over lossy networks because small packet loss rates may translate into much higher frame error rates [45]. Besides being lost, some packets may also suffer unpredictable amounts of delay or jitter due to network congestion at intermediate routers, compromising their accomplishment of real-time constraints. All these issues related to the best-effort service of IP may seriously contribute to degrade the perceived quality of an end user at the video reproduction.

An alternative way of recovering from transmission errors, which can turn a frame undecodable, is to apply error concealment techniques. In the absence of a frame due to errors, error concealment may replace the lost frame by a previous one or roughly estimate it from adjacent well-received frames. Further, error concealment techniques differ on the roles the encoder and decoder play in recovering from errors. Different QoS schemes may also be applied to the video stream in order to adapt it for transmission given the network conditions. These adaptive strategies involve applying redundancy; either by using Forward Error Correction (FEC) to tolerate some losses or by using a different compression factor in the video encoding that may be achieved in changing the adopted GOP pattern (changing the number of I, B or P frames in the pattern e.g. IBBPBBBPBBI). FEC schemes protect video streams against packet losses up to a certain level at the expense of data redundancy [47] (i.e. less efficient data transmission). Adopting different frame patterns allows a video stream to better adapt itself to the available transmission conditions. Within the network, unequal protection based on frame type may avoid quality degradation due to the loss of one particularly important frame and the possible propagation effect of this loss throughout the hierarchical structure of compressed video streams. The joint adoption of these QoS strategies may as well contribute to a better delivery quality of video streams.

5. Definition of test cases

5.1 Methodology

The aim of this thesis research work is to study video over IP calls and analyze the calls in order to represent the quality of experience of users using SIP, RTP/RTCP protocols within the framework of the Internet Protocol Multimedia Subsystem (IMS). It is also aimed at showing the effect of these important protocols when they are in synchronization with each other and otherwise. In order to achieve this, different video call-case scenarios have been established and the respective data are collected.

There are various possible test case scenarios that can be used for our measurements hence the choice of certain scenarios which are explained in section 5.5 in order to have as near as possible what happens in real life situations when video calls are being established between peers and in some cases, third party video call (video conferencing) using the IMS network. The analysis is done on the results from these scenarios which are based on the following criteria:

- Network specifications and standards (3GPP, IETF, ITU)
- Test results and measurements

These criteria provide the following information to analyze the different functionalities of the SIP, RTP/RTCP as they pertain to video call over IP using the IMS as a platform:

1. Functionality. Analyzes the different functions and services provided by these protocols
2. Architecture specific. Looks at the different configuration details and the network setup necessary for the protocols to work, and also into the associated protocols and transport.

5.2 Evaluation

The final result of the analysis is provided by the evaluation criteria. The degree of conformity of these various protocols associated with quality of call measurements with standards as defined by the standardization bodies (3GPP, IETF, ITU) are carefully evaluated. The aspects of these protocols that are used for the study are:

1. Flexibility. It denotes the degree of adaptability of the protocols for different configurations and scenarios.
2. Reliability. This is a measure of how dependent the information provided by these protocols is to evaluate quality.
3. Complexity. This evaluates the level of complexity these protocols can assume under different call case scenarios.

5.3 Requirements

According to IETF RFC 3261 [9] and IETF RFC 3550 [12], as well as a number of accompanying IETF standards, the following functionalities of SIP, RTP/RTCP have been defined (not exhaustive):

SIP:

- Allowing users to register as VOIP subscriber and place their contact binding in a designated registrar
- Keeping track of user's availability
- Exchanging user capabilities during session negotiation and establishment
- Establishing media session parameters for both ends of the communication
- Media session management

RTP:

- Transfer of real-time multimedia
- Re-sequencing, by receiver, of data fragments
- Loss detection for quality estimation, recovery
- Supports intra-media synchronization: remove delay jitter through playout buffer
- Intra-media synchronization: drifting sampling clocks
- Inter-media synchronization (lip sync between audio and video) alongside RTCP
- Quality-of-service feedback and rate adaption
- Source identification (CNAME)

RTCP:

- Primary function is to provide feedback on the quality of the data distribution
- Carries a persistent transport-level identifier for an RTP source called the canonical name (CNAME) which is also used in association data streams/sessions
- Used to control rate of sending report packets by participants in a session or associated sessions

5.4 Test setup

In order for this assignment to be carried out, different Ethernet capture setups can be adopted. Based on available equipment and in order for simplicity and ease of measurements, the test setup adopts the method of shared media or hub to connect the setup entities together. A hub is connected to the Ethernet LAN, meaning that all IP messages ('IP packets') are received by all nodes connected to the hub on the network. The Ethernet adapter on the PC is put into *promiscuous*[1] (open) mode so that all packets that are transferred on the Ethernet connection that hub forms part of can be seen by the adapter and thus captured with that adapter. The figure below shows the test setup used for this work.

[1] Promiscuous mode in this context refers to a configuration of a network card that makes the card pass all traffic it receives to the central processing unit rather than just frames addressed to it. It is used in packet sniffing

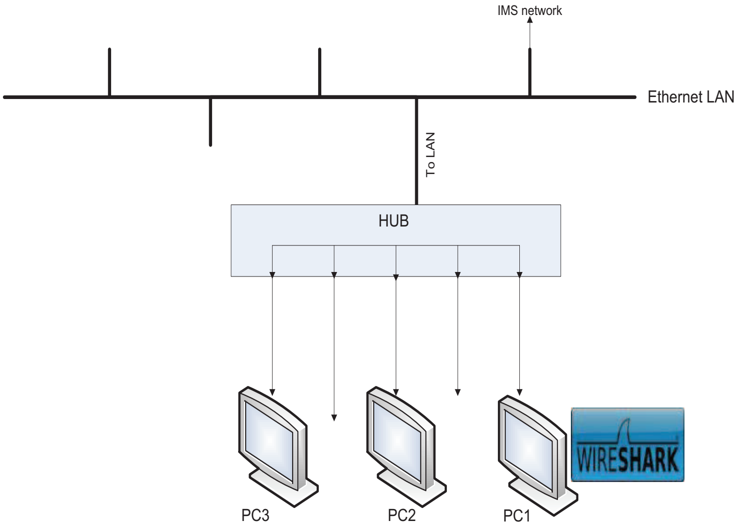
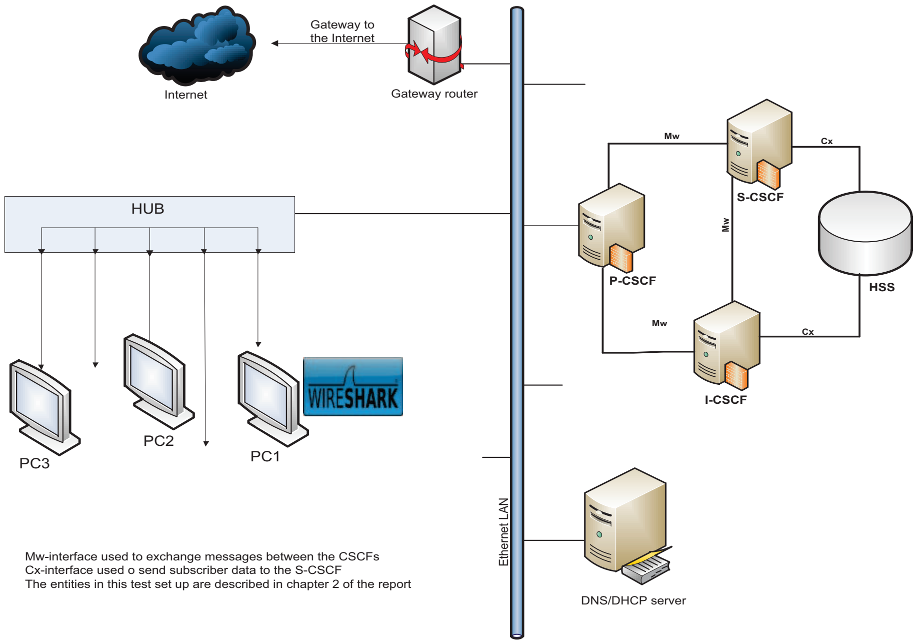


Figure 5.1: Test bed set up

For the test setup, the following hardware and network protocol analyzers were used:

- 3 PCs (2.2GHz, 1GB RAM) each with webcam. PC1 represents the A-party which is the session initiator for all test cases. PC2 and PC3 represent B-party and C-party which are the remote user parties.
- PC softphone (Counterpath X-lite 4.0)
- Wireshark version 1.2.2 (free and open source packet analyzer used for network troubleshooting, analysis, software and communications protocol development and education)
- SIP workbench v1.0.0.3970 (a graphical SIP, RTP, STUN, and TURN protocol analyzer and viewer designed to help illustrate and correlate VoIP and IM network interactions)
- Replace pioneer (a professional text, binary, replace and conversion utility used to split and parse file)
- IMS-in-a-box (represents the IMS network)

IMS-in-a-box (which is an Ericsson solution) is a single entity/system deployed on a single PC. This is a small-scale representation of an IMS network, running on a single Linux-based PC. It is on this entity that the IMS service runs. The figure shows the detailed structure of the IMS network together with the hub connection.



Mw-interface used to exchange messages between the CSCFs
Cx-interface used o send subscriber data to the S-CSCF
The entities in this test set up are described in chapter 2 of the report

Figure 5.2: Test bed network as it connects with the LAN and the Internet

Counterpath X-lite Beta 4.0 was installed on all the PCs used in the experiment. Wireshark (network protocol analyzer) was also installed on PC1, the PC used for the packet captures during the experiment. The DHCP/DNS server does the IP address allocation to the SIP phones and the PCs connected to the hub.

5.5 Video Test Case Scenarios

For this experiment, four different video call-case scenarios have been chosen and designed in order to have an insight into the behavior of the IMS network as it supports video over IP call sessions. For each of the scenarios, the call lasted two minutes. In all the scenarios, SIP and RTP/RTCP protocols are the protocols of interest as they give an insight into the quality of service behavior of the sessions. The results of the thesis work and the analysis based on each of the highlighted test case scenarios are included in chapter 6 of this report.

In each of the test cases, a specific set of codecs have been chosen so that negotiation on the particular one to be used among this set has been left to X-lite. The remote party accepts one of these codecs during negotiation. Port selection for both RTP and RTCP were also left to the PC operating systems. For all the test cases, media runs end-to-end without traversing media proxies (e.g. Access Border Gateway or NAT).

One of the main factors that have led to the choice of the simple test case scenarios used in this experiment is the availability of equipment. Complex call forwarding scenarios would require certain functionalities on the SIP phones and designated services in the IMS. Some desk top SIP phones do not support RTCP and for that reason, their use in this experiment was ruled out. X-lite soft client phones have been used, which are freely available for download on the Internet.

Measurements have been done on the PC where the call is established which gives a certain view on the results, such as media arriving earlier than 200 Ok. Measurement on the PC from the B-party (remote party) gives a different view. The impact of this has been considered in the further analysis of our results.

The following are the different call case scenarios tested in the experiment:

5.5.1. Peer-to-peer video session

This scenario has been chosen because of the fact that it represents the normal everyday video call sessions between two end users. Here, an end user sets up a video session with another user whose phone possesses the video call capability.

This scenario involves establishing a video call between two peers (PCs). Figure 5.3 depicts the signaling and media flows involved.

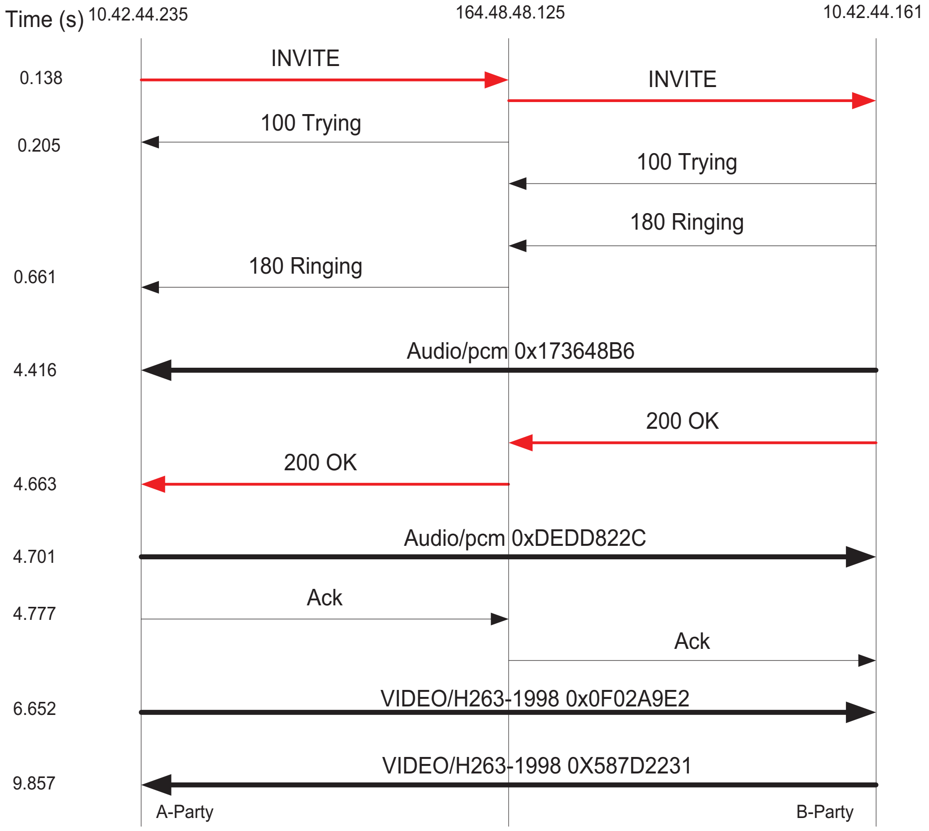


Figure 5.3: Peer-to-peer video call establishment and session

From figure 5.3, it can be noticed that media starts arriving at the calling party before the 200 Ok is sent by the called party. This is due to media arriving earlier than the SIP signaling. Applying source filtering by the A-party would result to it dropping this early media because the A-party needs to have the remote party IP address in order to accept remote media. This is stated in the SIP INVITE. Also due to the fact that media transmission is end-to-end between parties while control message (200 Ok) has to traverse intermediate proxies, media gets to the called party earlier. As a result of the modulation and compression techniques coupled with differences in packet sizes, the video data transmission starts later than the voice media component of the call.

5.5.2. Peer-to-peer voice call

Although this research work focuses mainly on the analysis of video calls over IP, this scenario has been chosen in order to be able to compare the behavior of an end-to-end video session with that of an end-to-end voice call, showing the network resource usage, delay experienced and also jitter.

Similar to the way the call session is established in the first scenario (Peer-to-peer video session) described above, this second scenario call session is set up with only audio data stream involved between the peers. Figure 5.4 gives the details of the signaling and media flow between the peers.

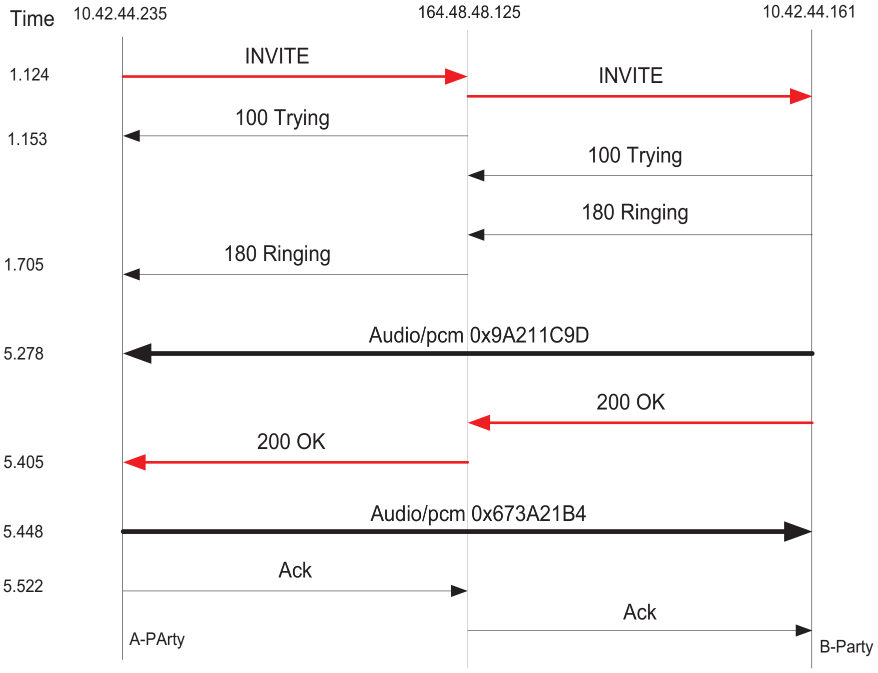


Figure 5.4: An end-to-end voice call signaling and media exchange

As shown in the figure, media starts arriving at the A-party earlier than the 200 Ok as perceived from the PC1.

5.5.3. Peer-to-peer video call with a video call to a third party while the former is put on hold

This scenario has been chosen in order to present the situation whereby during an ongoing video call session, another video call is initiated towards a third party by one of the peers involved in the ongoing video session. This situation may happen frequently in real life situations so therefore, this section depicts the network situation during the scenario showing the signaling involved and the order of steps involved the signaling and media exchange between all the parties involved in the scenario. The synchronization of the main protocols of focus is shown and the effect of them being out of sync is also displayed in the result analysis.

This third scenario is partly similar to the first scenario (Peer-to-peer video session). A video call is established between the two peers, A-party and B-party, with another call established towards a third party (C-party) by the A-party while placing the B-party on hold.

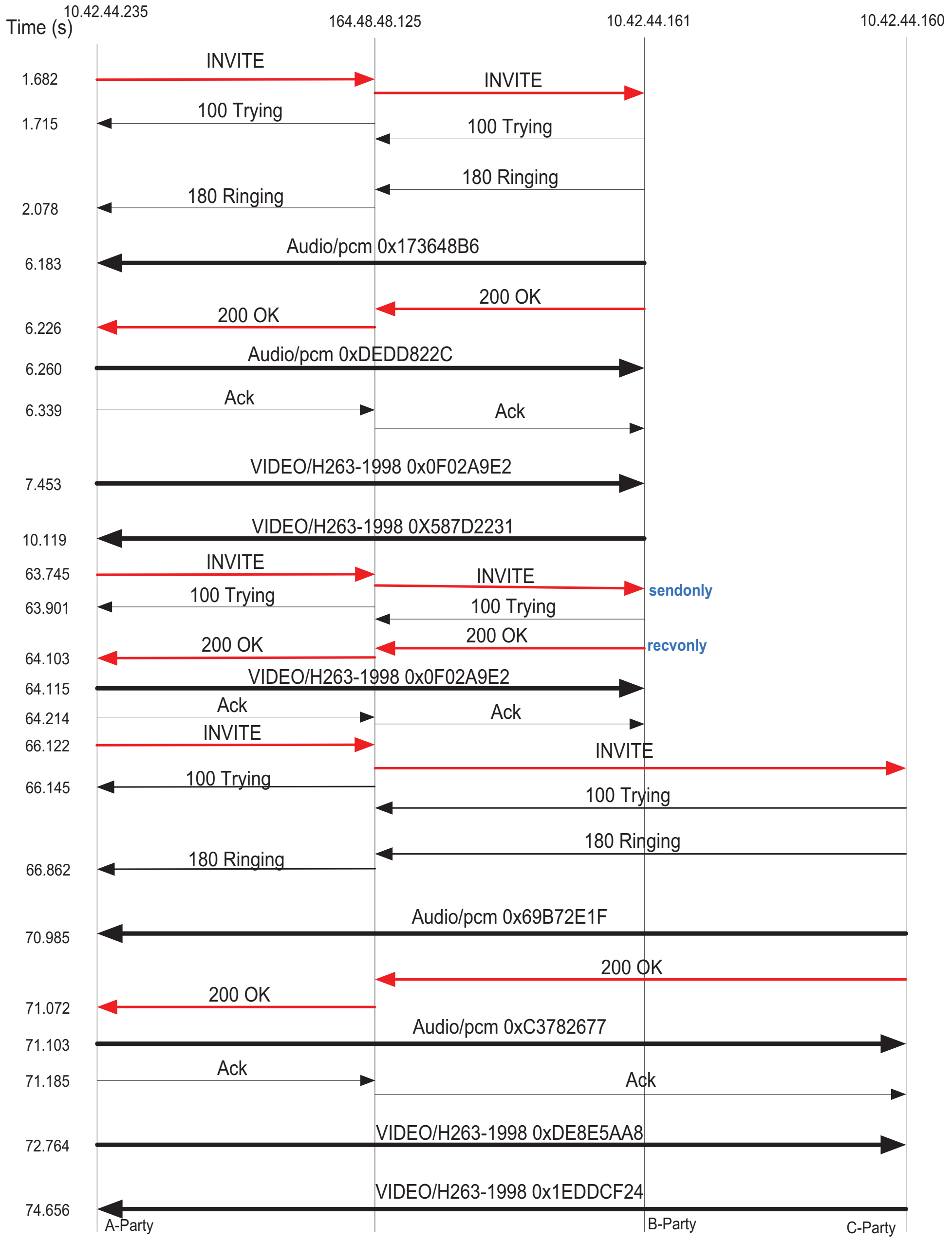


Figure 5.5: Peer-to-peer video call session with a video call established towards a third party while the former session is put on hold

When remote party is put on hold, the stream attribute in the updated SDP offer from the A-party (who is applying the call hold) is set to a=sendonly and not to a=inactive.

This is to avoid RTP timeout on its receiving RTP port. RTP timeout results into the closure of RTP port thereby ending the call.

The updated SDP is used when previously agreed upon call parameters are to be changed during an on-going session like, codecs, IP addresses, ports numbers, type of media, bandwidth, packetization time etc. this SDP update is used to change the status of the RTP ports during the test case depicted in this scenario. The a=sendonly and a=recvonly on the A and B-parties RTP ports are updated using this SDP updates.

5.5.4. Video conference between three users

The fact that video conferencing holds a potential for significant reduction in travel expenses as well as conducting more productive meetings leading to more effective and efficient decision making for the enterprise environment and for individuals, this scenario has been chosen in order to analyze the network situation during a conference call involving three end users. More users could have been involved but, the fact that the soft client used only supports three users (one user and two remote users) at a time (during a session) hence, the use of three users. The session initiator is denoted as the A-party, while two other remote users are denoted by B-party and C-party respectively.

This scenario actually shows the beauty of setting up many sessions between a number of users. This involves quite a large amount of signaling between the peers in order for proper coordination of the whole session (audio and video).

A-party sets up and commences a video session with B-party. While the session is on-going, A-party initiates another session towards C-party while placing the session with the B-party temporarily on hold. Once the session with C-party is established, the A-party re-invites the B-party into the session therefore creating a platform for the three parties to start exchanging media with each other.

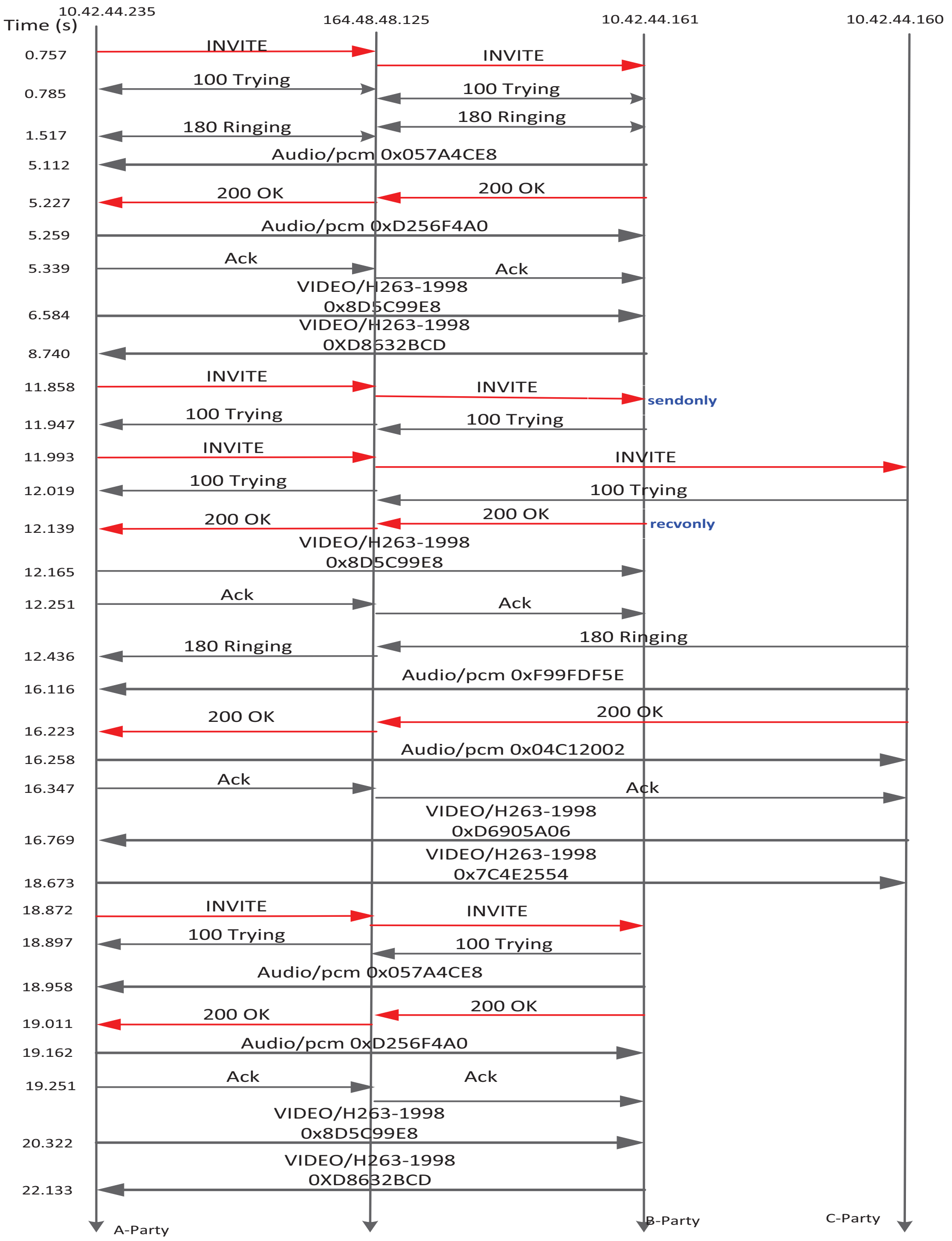


Figure 5.6: Video conference between three users

Figure 5.6 shows the signaling flow during the call establishment phase of the session, placing a session on hold, the re-invite and the media exchange flows between all the parties involved.

6 Results and analysis

In order to represent the results of the measurements carried out in each of the video call case scenario as described in chapter 5, the Wireshark capture (SIP, RTP/RTCP) of each of the scenario was filtered (removing unwanted headers) using Python scripting (a general-purpose high level programming language) and parsing the resulting data compiled in an excel spreadsheet. Graphical representations are then made from the compiled data.

In each of the test case scenarios adopted in order to carry out this work, the measurement has been done in the following way and steps:

1. The three parties involved in the sessions are provisioned in the IMS network, i.e. subscriber data in HSS. A call is initiated from party-A to party-B and the whole call session lasts for two minutes. Wireshark is started on the machine on which the SIP phone that initiates the call resides, i.e. PC1, before the start of each call.
2. The Wireshark programme has been configured to only display and save our main protocols of focus which are SIP, RTP/RTCP. The IMS network has been configured to use SIP over UDP. RTP/RTCP also in principle runs over UDP which is also the situation in each of our test cases.
3. At the end of the two minutes call session period, the Wireshark programme is terminated with captured data saved.
4. The captured data from Wireshark is saved by default in .pcap format (captured packet format used by Wireshark). In order for our filtering procedure, the format of the data is converted into plain text (.txt) format.
5. The fact that the captured data is quite large and contains a lot of headers that we don’t need for our analysis, filtering and parsing of unwanted headers (reception report count, length, SSRC, packets lost etc.) had to be done on each of the captured data in each of the cases.
6. To do the parsing and filtering, Replacement Pioneer software using Python scripting (a professional text, binary, replace and conversion utility used to split and parse file) was used in order to filter out all unwanted headers in the protocols leaving behind only the information needed (interarrival jitter, sender octet count, extended highest sequence number, timestamp information) to give a report about the network situation during each of the case cases sessions.
7. These filtered headers of interest are now imported into Microsoft excel spreadsheet in order to make plots and give a graphical representations of our results.
8. For the analysis, each of the test case sessions is divided into the audio stream and the video media stream thereby showing the usage of network resource by each of these components of the call sessions.
9. All the measurements have been done on the PC1 initiating the calls. The effect of this is that, media arrives earlier than the SIP signaling.

Also in the analysis of the results, the RTCP and RTP components have been considered and analyzed separately. X-lite is configured to offer G.711, GSM, iLBC, Broadvoice-32

and other audio codecs and for video codecs, H.263, H.264. The result of the negotiation is in any case G.711 for audio and H.263 for video. The following IP addresses were allocated to each of the participating parties by the DHCP server:
10.42.44.161 to A-party, 10.42.44.235 to the B-party and 10.42.44.160 to the C-party. All the port numbers indicated are related to the RTP. For each of the data streams, different SSRC are used in order to distinguish between the streams.

6.1 Test Case 1

For this scenario, a video call was initiated, established end-to-end between party-A and party-B with the call session lasting for two minutes. Captures of all signaling flows in the control plane and media exchange in the user plane were done with Wireshark on PC1. Below are the plots of the parameters that depict the situation of the call session. In this scenario, it is expected that as soon as call is established, exchange of audio and video stream should commence at the same time.
Forward direction involves the flow of signaling and media streams from A-Party to B-Party while the reverse direction is from the B-Party to A-Party.

Audio stream from 10.42.44.161:31878 to 10.42.44.235:1476

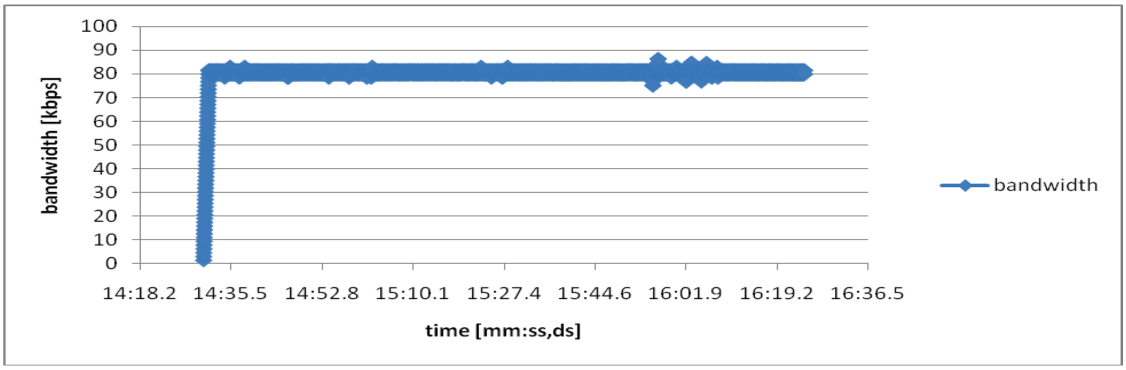


Figure 6.1: Bandwidth usage by the audio stream in the forward direction

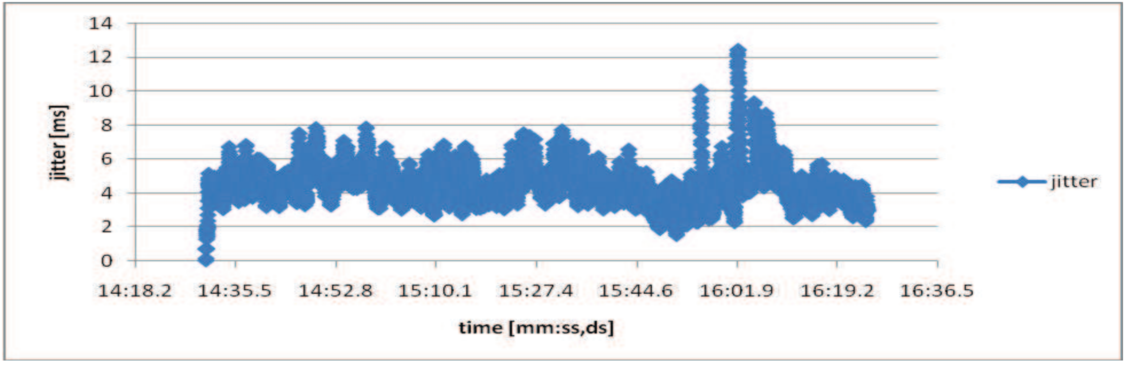


Figure 6.2: Jitter experienced by the audio stream in the forward direction

Figures 6.1and 6.2 show respectively, the bandwidth consumption and jitter experienced by the audio stream component of the call session in the forward direction. This includes the RTP data (PCM samples + RTP header), UDP header data and IP header data. As soon as the call is established after the initial SDP offer/answer procedure (call set up), a very sharp rise in bandwidth usage is noticed due to the fact that media streams are immediately being exchanged between the peers due to the use of network resource. The fluctuation in the jitter is fairly high between the values of 2ms and 8ms for the audio component with a peak of 12ms. The reason for this being partly because of congestion of IP packets on the hub interface on which it is measured.

Video stream from 10.42.44.161:53942 to 10.42.44.235:9008

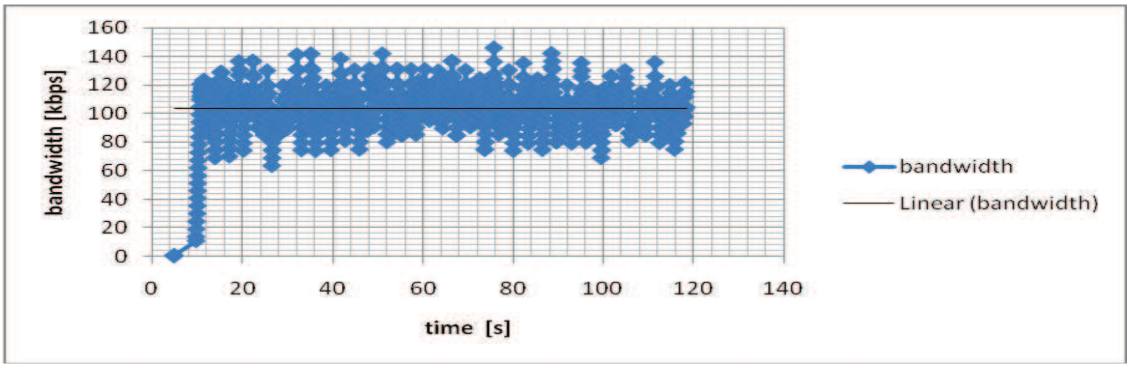


Figure 6.3: Bandwidth usage by the video stream in the forward direction

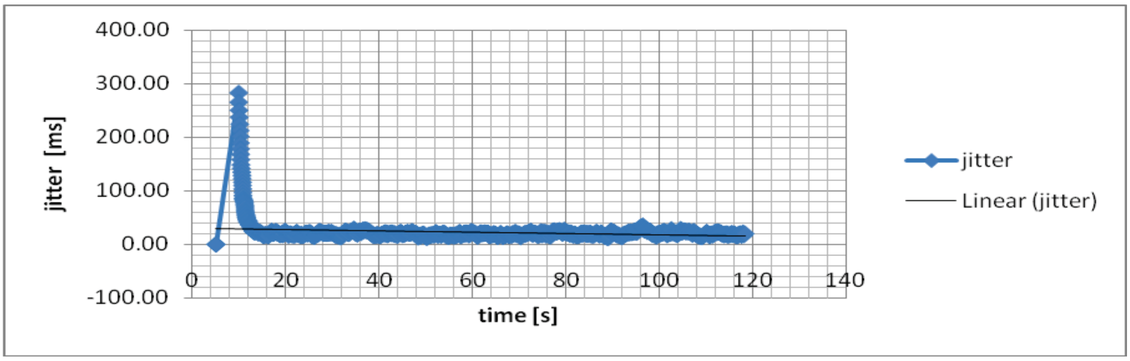


Figure 6.4: Jitter experienced by the video stream in the forward direction

Figures 6.3 and 6.4 are representations of the video stream component also in the forward direction. As expected, there is an increase in the bandwidth usage with a fairly stable jitter due to the fact that the RTP packets making up the video component are transmitted at a fairly steady rate. For the bandwidth usage, it can be noticed that there is fluctuation. This is due to the video compression technique that has been adopted (the choice of pattern is adopted by X-lite). In the video encoding, there is a fairly large number of I-frames present and these I-frames are fairly large in size, hence the rise in bandwidth usage when an I-frame is transmitted. The higher the number of I-frames transmitted per time period, the better the picture quality. The high peak in jitter is due to reason that at the beginning of the media transfer, initial RTP packets transferred contains audio data with timestamp zero. After a few seconds, streaming of audio packets begins and at this point, jitter becomes highly stable.

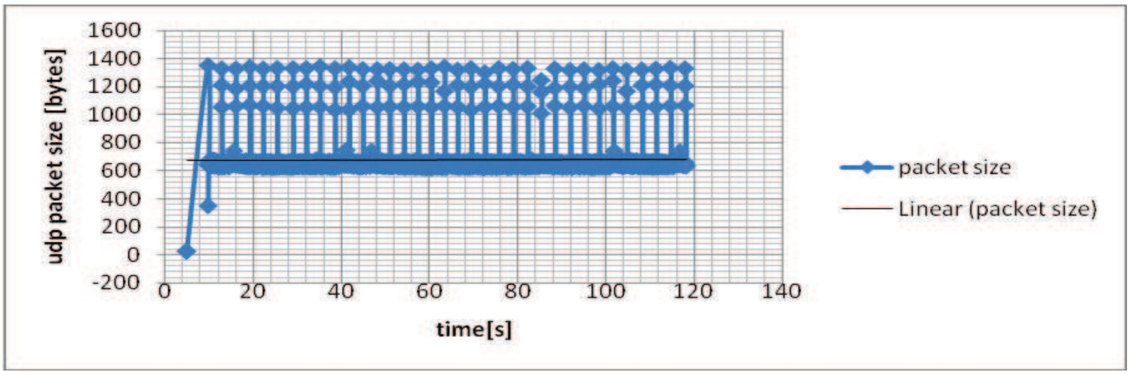


Figure 6.5: Payload over the duration of the call session

Figure 6.5 is a plot of the packet size distribution. This includes the UDP and RTP headers as they are seen by the network at IP level. The packet size is fairly stable at 600ms. The occasional peaks in the plot show the start of a new block of picture frame (I-picture frame), hence the large size of these frames.

Audio stream from 10.42.44.235:1476 to 10.42.44.161:31878

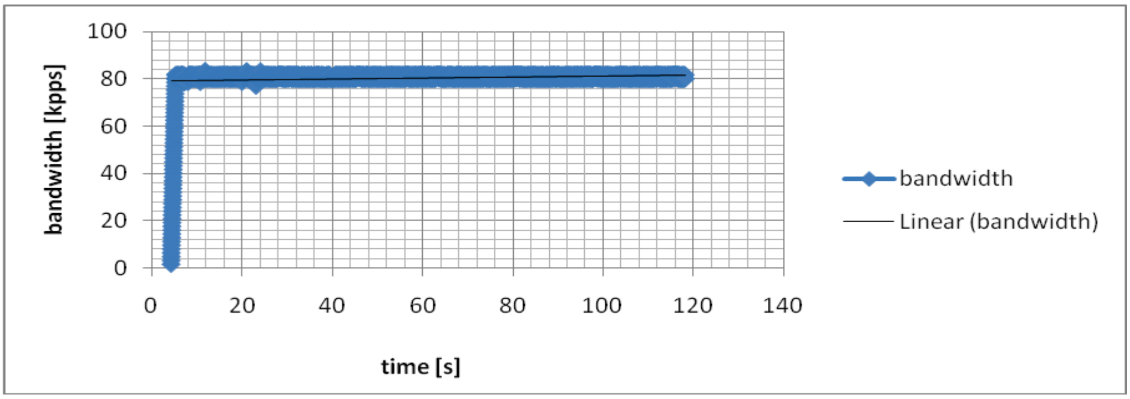


Figure 6.6: Bandwidth usage by the audio stream in the reverse direction

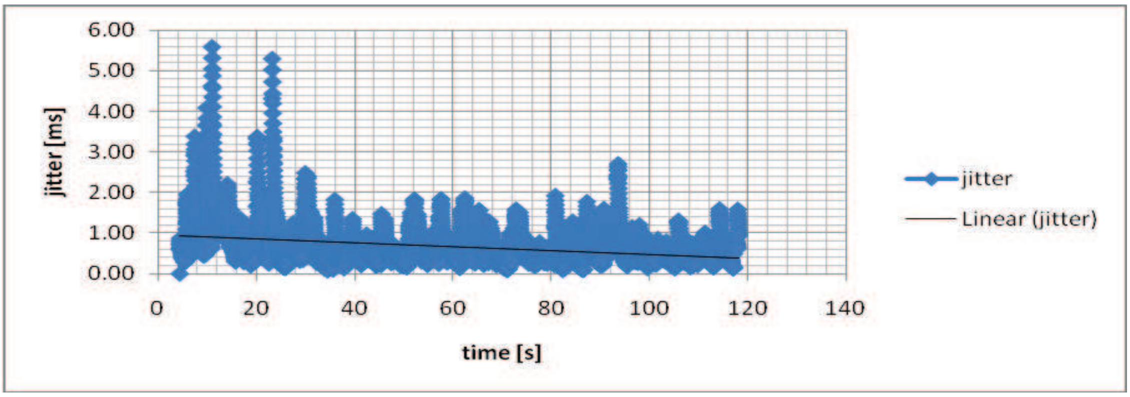


Figure 6.7: Jitter experienced by the audio stream in the reverse direction

Figures 6.6 and 6.7 are similar to figures 6.1 and 6.2. This is in the reverse direction for the audio stream. Fluctuation in jitter is higher in figure 6.7 than in figure 6.4 due to congestion at the hub port of the user with a lower average jitter value. The congestion is due to the fact that each port on the hub broadcast all d traffic traversing the network.

Video stream from 10.42.44.235:9008 to 10.42.44.161:53942

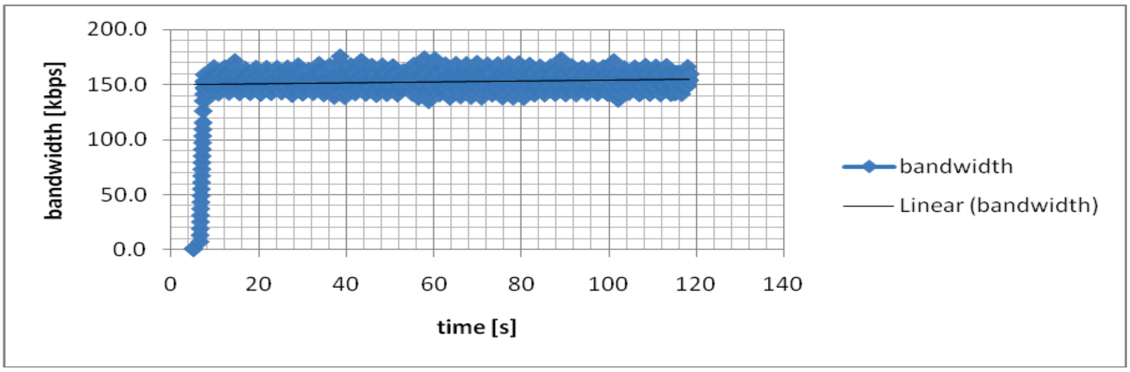


Figure 6.8: Bandwidth usage by the video stream in the reverse direction

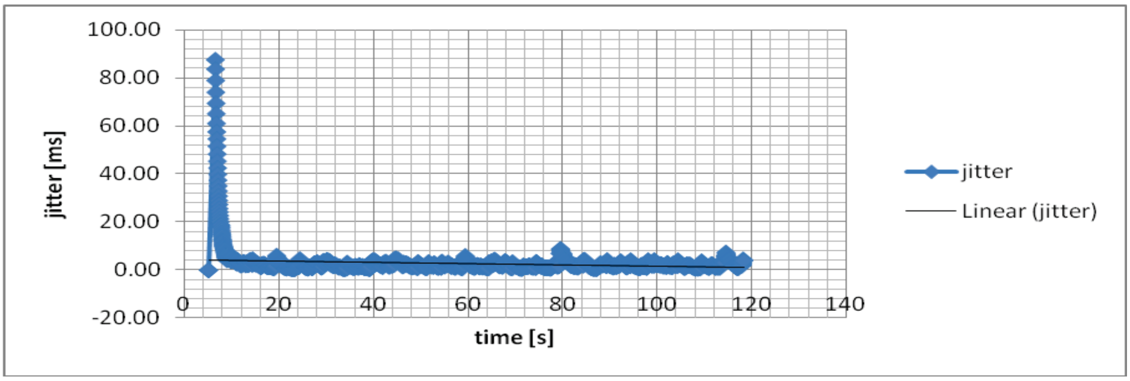


Figure 6.9: Jitter experienced by the video stream in the reverse direction

Comparing figure 6.8 with figure 6.3, it is noticed that in the former, fluctuation in bandwidth consumption is fairly stable. This is due to the video encoding pattern adopted by the SIP soft client (X-lite) which invariably affects the picture quality and user experience.

Figure 6.9 shows a fairly stable jitter variation.

The RTCP reports plots analyses of this test case scenario are done in Appendix A.

Summary

In summary, based on the various information obtained from all the plots covering this test case, there were times where quality of the video stream dropped due to high jitter and also due to video encoding type adopted by an end user. But in overall, communication over the period of the call session has been quite acceptable with occasional clipping sounds and freezing of video frames. The fact that at each interface on the hub, all data traffic traversing the network is broadcast, this has also contributed to the effect of high jitter in the data transmission.

6.2 Test Case 2

The analysis of the simple end-to-end voice call depicted in this test case scenario is comparable to the voice component in test case 1. The G.711 was agreed during codec negotiation and the default packetization time as indicated in RFC 3551 is used.

Party-A initiated, set up and established an end-to-end voice call with party-B. The whole call session lasted for two minutes. A capture of all the signaling flows between the two end users together with the media exchange was done with Wireshark.

Audio stream from 10.42.44.161:26324 to 10.42.44.235:11090

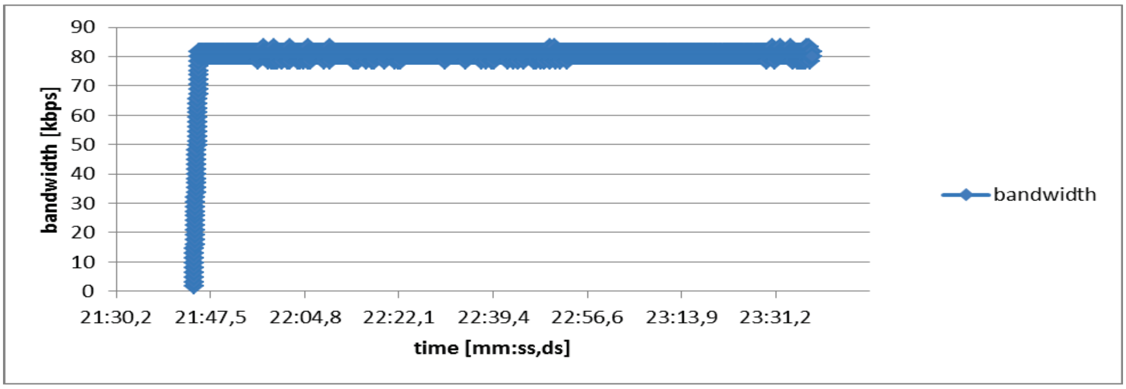


Figure 6.10: Bandwidth usage in the forward direction

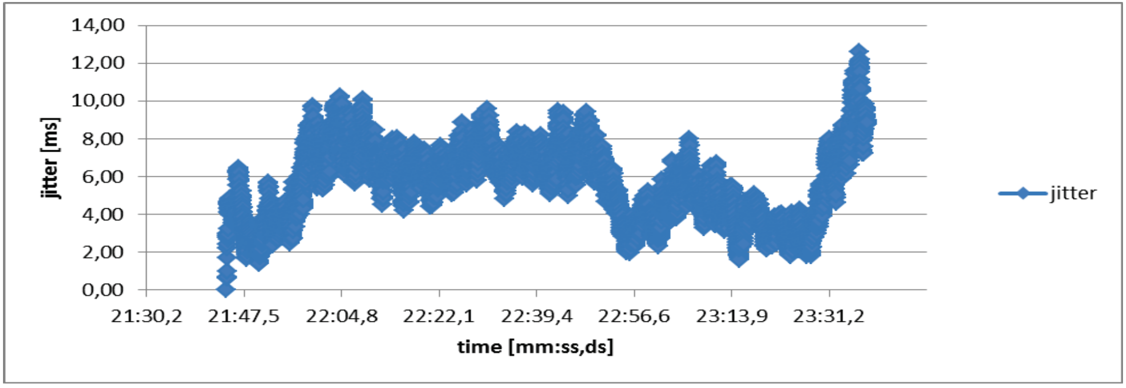


Figure 6.11: Jitter experienced by the audio stream in the forward direction

Figure 6.10 shows a fairly stable bandwidth consumption of 80kbps in the forward direction. This is fairly normal for the audio stream when compared with the audio stream in test case 1 in the forward direction.

Figure 6.11 shows a rather high jitter fluctuation between 2ms and 10ms (and higher). This is probably due travel paths of each RTP packet, queuing and congestion at the interface of connection on the hub.

RTCP audio: 10.42.44.161 to 10.42.44.235 in the forward direction

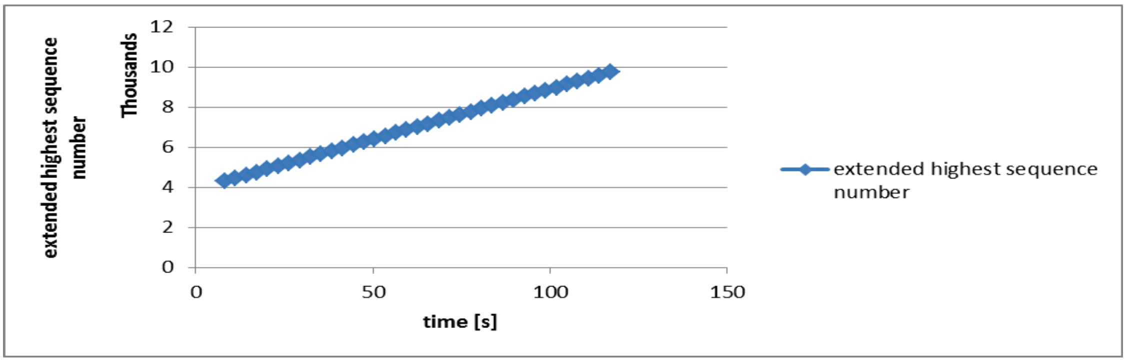


Figure 6.12: RTCP report on the Extended Highest Sequence Number

Figure 6.12 shows a fairly steady rise in the plot of the extended highest sequence number. This indicates that little or no number of packets were lost during the transmission for the duration of the call session.

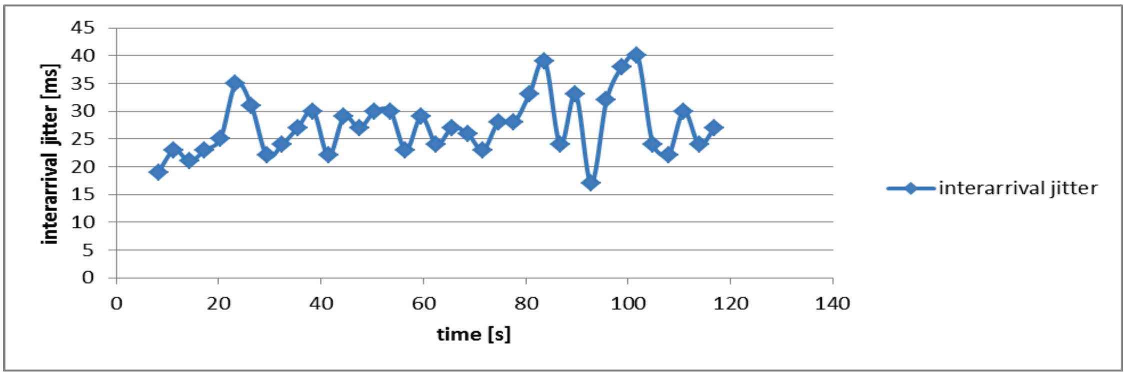


Figure 6.13: RTCP report on the Interarrival jitter

From this plot in figure 6.13, we observed a number of high peaks in the jitter value. At those points/periods where these high peaks were observed, quality of the voice drops due to these packets arriving late at the receiving end.

Audio stream from 10.42.44.235:11090 to 10.42.44.161:26324

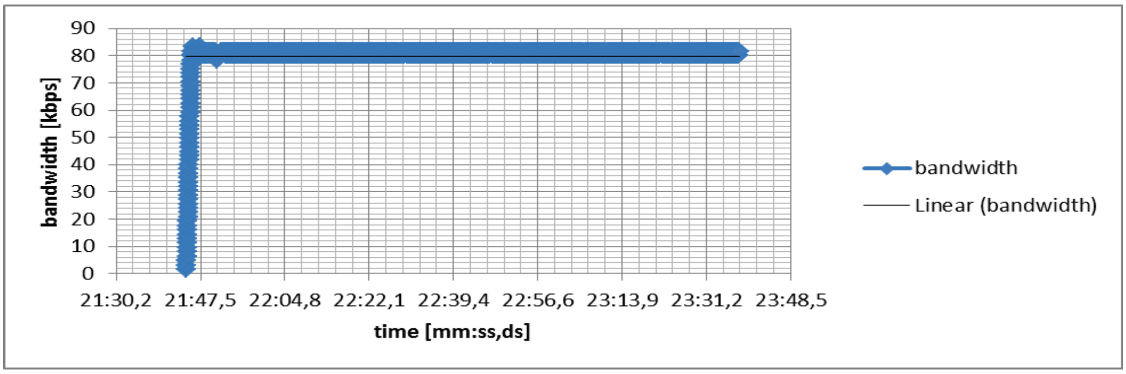


Figure 6.14: Bandwidth usage in the reverse direction

This plot can be compared to figure 6.10. Figure 6.14 shows stable bandwidth consumption.

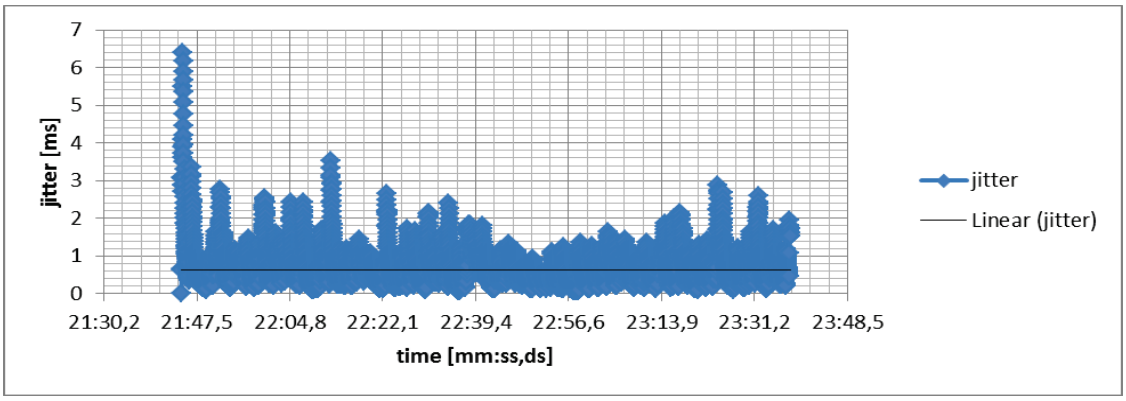


Figure 6.15: Jitter experience by the audio stream in the reverse direction

Figure 6.15 shows a fair variation in jitter between 0 and 3ms (and with a peak of above 6ms). Due to the fact that more audio packets are sent during transmission and possible congestion at the port the user is connected on the hub.

RTCP audio 10.42.44.235 to 10.42.44.161

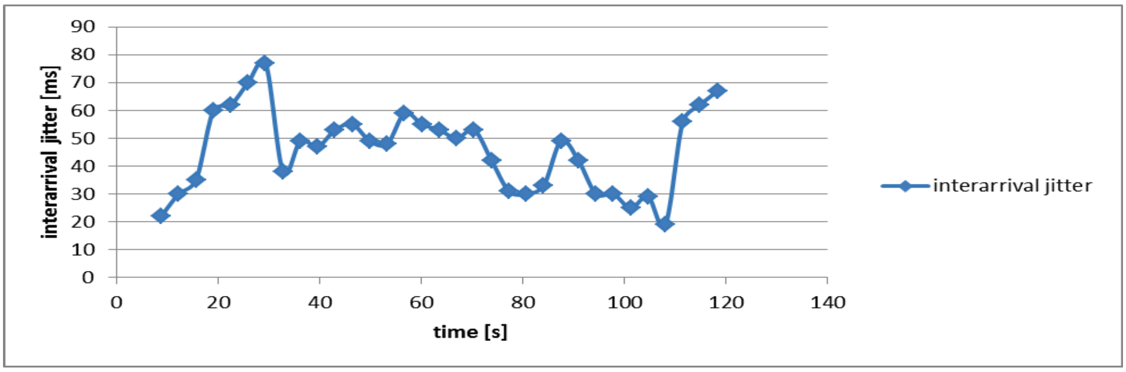


Figure 6.16: RTCP report on Interarrival jitter

In figure 6.16, high variation in the interarrival jitter can be seen with high peaks at a couple of places. At these peaks, some audio packets arrive late (but still within the de-jitter buffer value) at the receiving end.

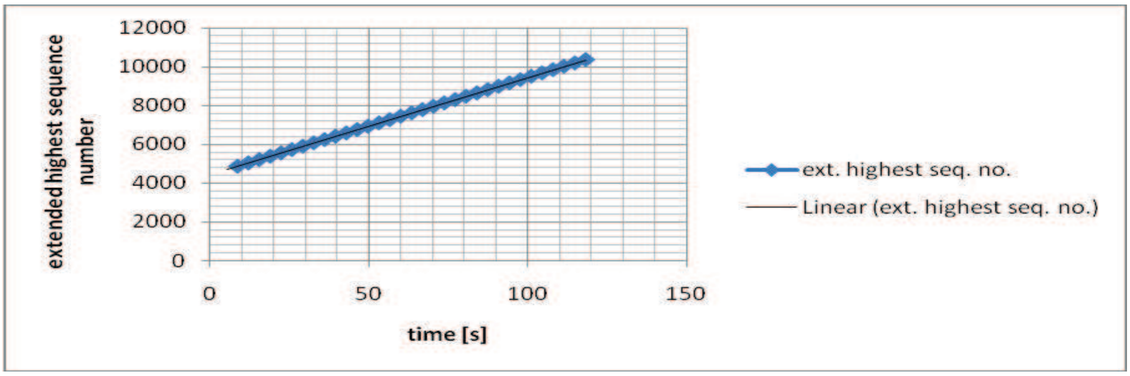


Figure 6.17: RTCP report on the Extended Highest Sequence Number

The plot in figure 6.17 shows a steady rise of the extended highest sequence number. This translates to little or no packet loss during the interval which these reports are transmitted hence a fairly smooth communication between the two calling parties participating in call session.

Summary

In summary, the plots in this test case scenario indicate fairly stable bandwidth consumption with occasional clipping of audio signals due to congestion at the port of connection to the hub as well as high interarrival jitter as reported in the RTCP reports.

The de-jitter buffer time is an important parameter in the phone. Packets arriving outside this value constitutes a high packet lost value at that instant. The hub congestion situation would be crucial especially in a network where hundreds of SIP phones are connected leading to higher SIP and data traffic on the hub ports. The average data rate of the network on which the hub is connected has to be determined. Also the transmission and queuing rate of the hub should be known before it is deployed on the network. This is to make sure that its transmission rate matched matches that of the network it s serving.

6.3 Test Case 3

The scenario involved in test case 3 is of interest because it depicts what happens in the control and user planes respectively when calls are initiated, set up and established. An already established call session is subsequently placed on hold when the need arises for one of the users to be involved in another call session.

In order to set up this scenario, a simple end-to-end video call was established between party-A and party-B which includes the codec negotiation (G.711 and H.263 are negotiated as audio and video codecs respectively), and media exchange. After approximately 60s, party-B was put on hold by party-A in order for it to establish a new session with a third party-C. In this test case scenario, it is expected that once a remote party is put on hold, exchange of RTP packets towards that user should be suspended (both audio and video) with exchange only towards the new remote that has been invited to the call session. The RTCP reports should be sent and received by each participant involved in the session throughput its duration.

The RTCP report plots analyses of this test case scenario are placed in Appendix B.

Audio stream from 10.42.44.161:6348 to 10.42.44.235:46506

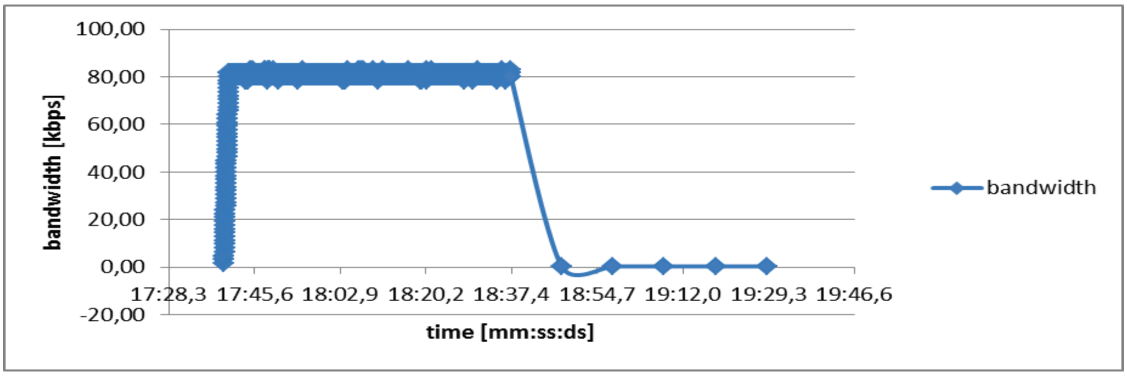


Figure 6.18: Bandwidth usage by the audio stream in the forward direction

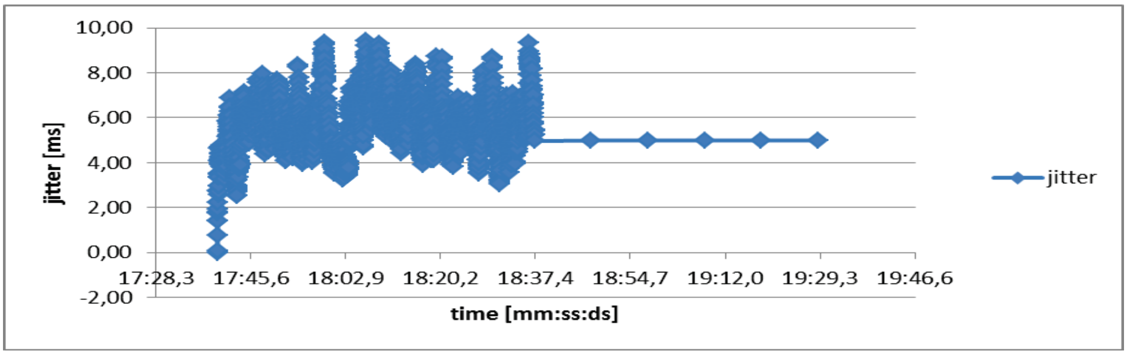


Figure 6.19: Jitter experience in the forward direction

Figures 6.18 and 6.19 depict bandwidth and jitter distribution in the first 60s. Bandwidth usage and jitter values are plotted with the two parameters dropping to a minimum after party-B was put on hold by party-A thereby releasing bandwidth used by this audio media stream to the barest minimum enough to transmit synchronization information (occasional RTP messages) and RTCP reports. Figure 6.19 shows a fairly heavy fluctuation in jitter during the first 60 seconds due to congestion at the port of connection to the hub.

Video stream from 10.42.44.161:45248 to 10.42.44.235:4114

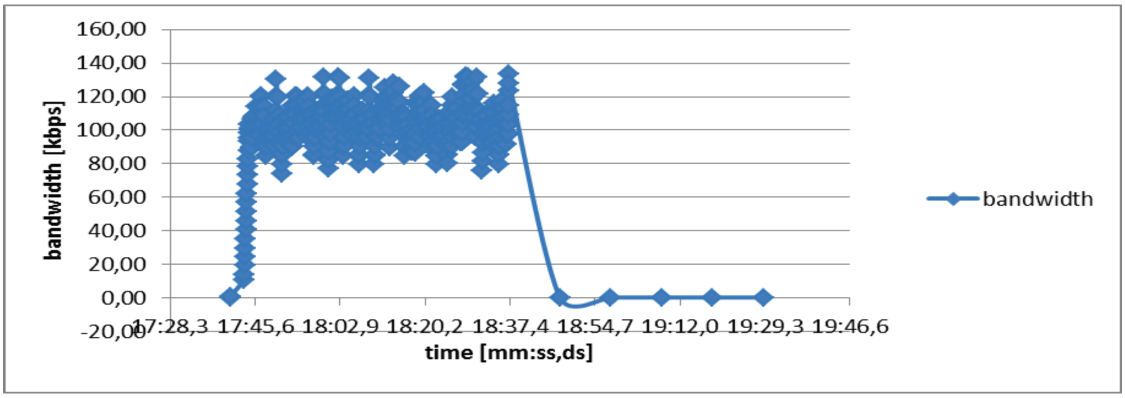


Figure 6.20: Bandwidth usage by the video stream in the forward direction

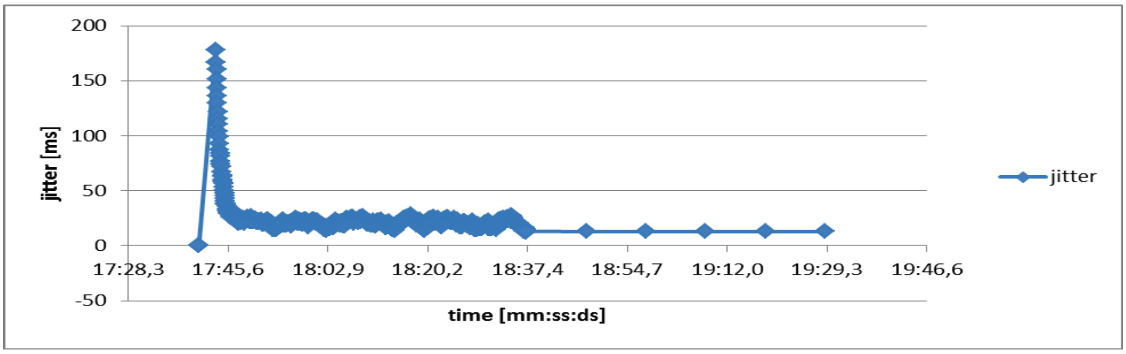


Figure 6.21: Jitter experience by the video stream in the forward direction

Just like the audio component of the call, figures 6.20 and 6.21 show the network resource usage by the video stream with the jitter experienced. In figure 6.20, heavy fluctuation is noticed in the bandwidth usage during the first 60 seconds. This is due to the video encoding technique adopted by the user. The transmission of I-frames which are quite large packets leads to these high peaks in the bandwidth consumption.

RTCP audio: 10.42.44.161 to 10.42.44.235 in the forward direction

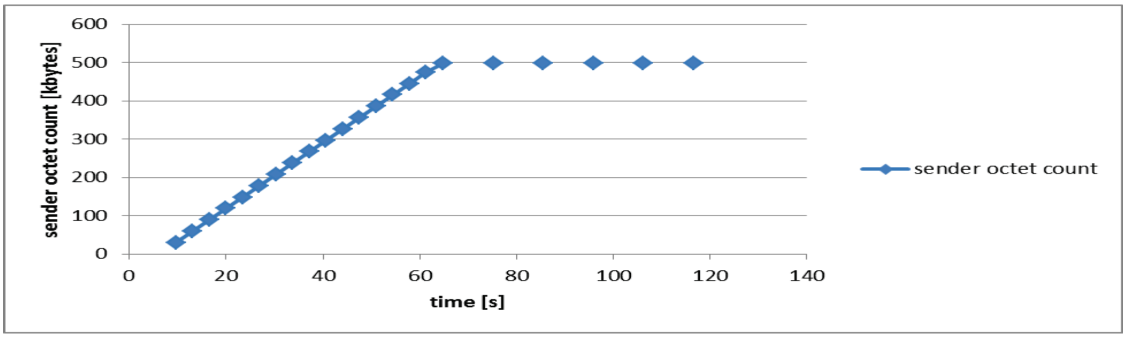


Figure 6.22: RTCP report on the Sender Octet Count

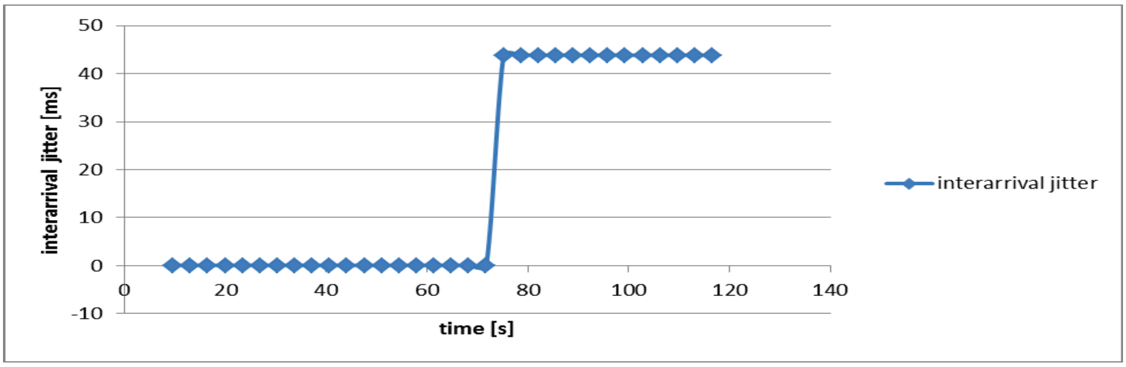


Figure 6.23: RTCP report on the Interarrival jitter

The sender octet count plot in figure 6.22 shows a steady climb in the number of bytes transmitted since the inception of the call and at around 60s, the climb levels off with occasional RTP exchange showing that for this audio component, transmission is put on hold for the remaining period of the call. Also, as depicted by figure 6.23, after 60s, the interarrival jitter rose so high (4s) stating that occasional RTP data were being sent. It is noticed that at around 70s the jitter shoots up so high because of the reduction in RTP packets transmission between the two caller parties and for the next 60s, it remains high. The reason for this rise in jitter value is due to the fact that at the point when the call was placed on hold, RTP packets containing media was suspended only for RTP packets containing only synchronization information being transmitted occasionally and at a later time.

RTCP video: 10.42.44.161 to 10.42.44.235 in the forward direction

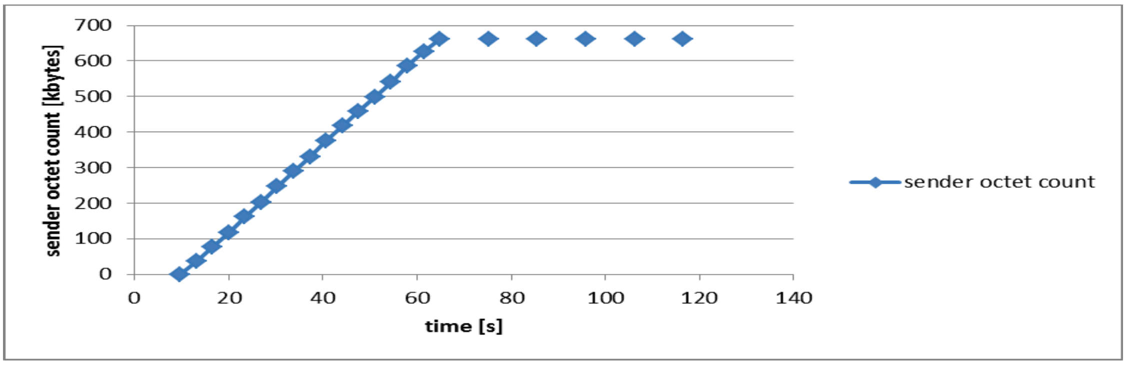


Figure 6.24: RTCP report on Sender Octet Count

Just like the audio component, figure 6.24 shows that once party-B end is put on hold, transmission of video stream becomes very minimal with only synchronization (in the RTP) information being transmitted. This is to keep the RTP ports open. The steady rise in the sender octet count indicates minimal to no packet loss in the first 60 seconds

Audio stream from 10.42.44.235:46506 to 10.42.44.161:6348

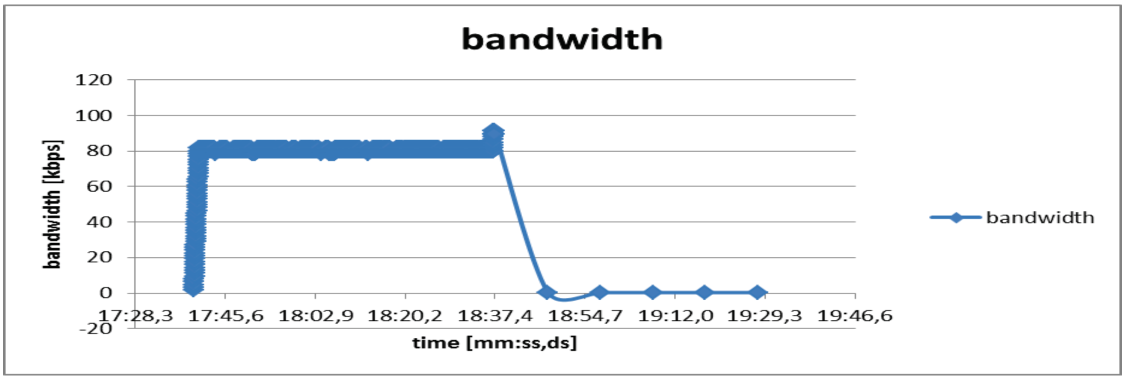


Figure 6.25: Bandwidth usage by the audio stream in the reverse direction

Figure 6.25 is the reverse direction of figure 6.18. Since the call is end-to-end, this kind of similarity is expected with bandwidth usage dropping to the barest minimum after party-B is put on hold by party-A thereby releasing network resource (bandwidth) allocated to this call session.

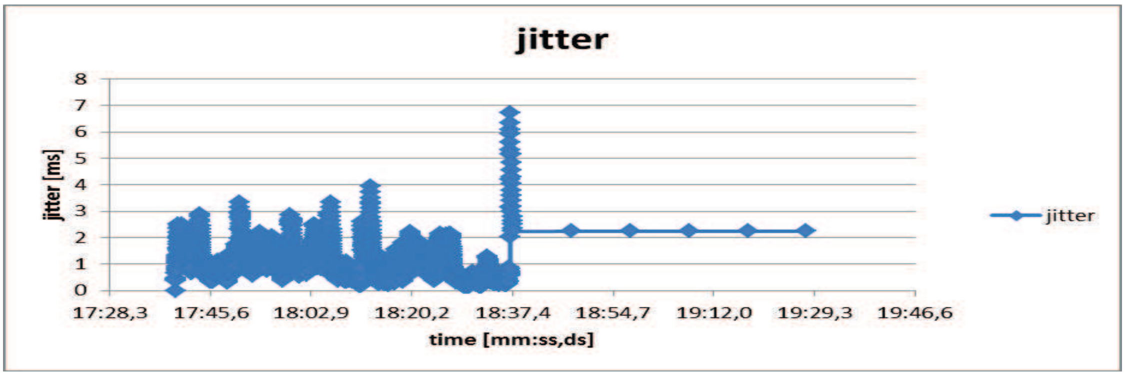


Figure 6.26: Jitter experience by the audio stream

Figure 6.26, similar to its counterpart figure 6.19, shows the variation in jitter values in the reverse direction. The average jitter value is lower here compared to that of figure 6.19.

Video stream from 10.42.44.235:4114 to 10.42.44.161:45248

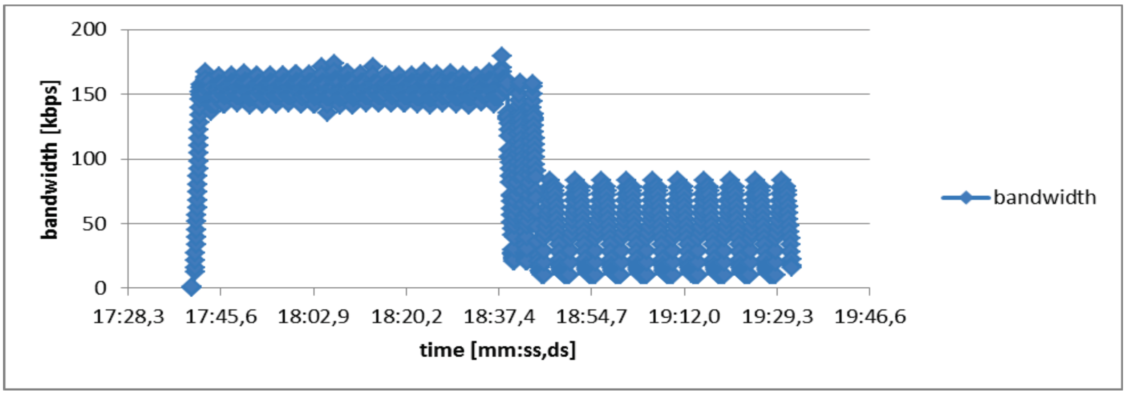


Figure 6.27: Bandwidth usage by the video stream in the reverse direction

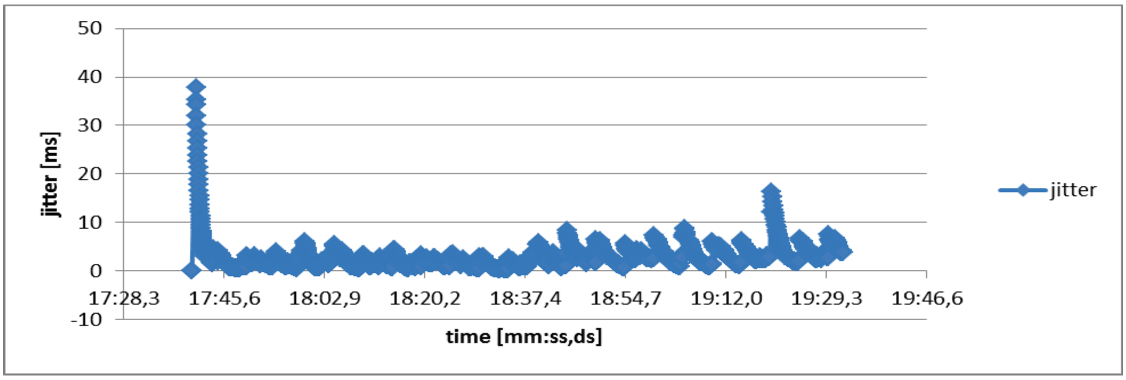


Figure 6.28: Jitter experience by the video stream in the reverse direction

Figures 6.27 and 6.28 show that for the video component of this call session in the reverse direction, some network resources are still being utilized after party-B has been placed on hold. During the latter 60s, bandwidth usage drops but not completely to zero because some video packets (RTP and RTCP) are still being sent to the party placed on hold for synchronization purposes, should in case the call is to be retrieved. For the jitter, the fluctuation went on the increase after party-B was placed on hold due to the fact that data packets containing synchronization information were transmitted occasionally.

RTCP audio 10.42.44.235 to 10.42.44.161 in the reverse direction

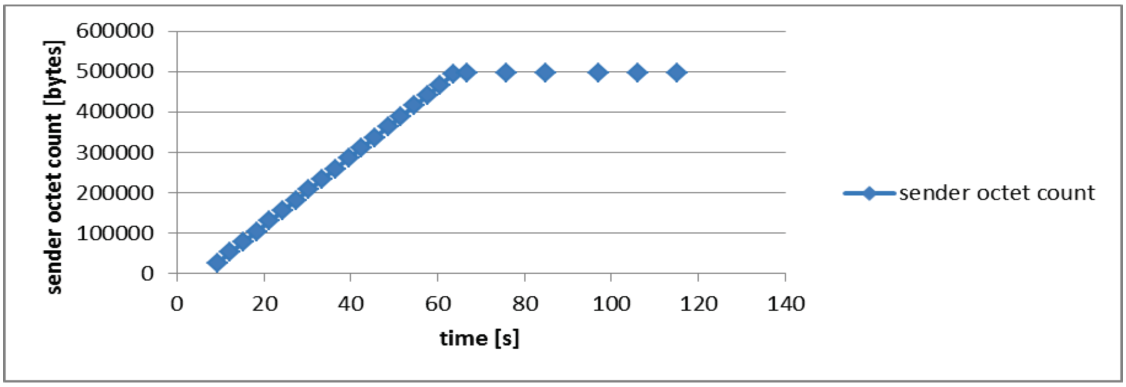


Figure 6.29: RTCP report on Sender Octet Count

The plot in figure 6.29 is similar to what happens in figure 6.22. The plot shows a steady climb in the number of bytes transmitted since the inception of the call and at around 60s, the climb levels off with occasional RTP and RTCP exchange showing that for this audio component, transmission is put on hold for the remaining period of the call. Transmission of video stream becomes very minimal with only synchronization information being transmitted.

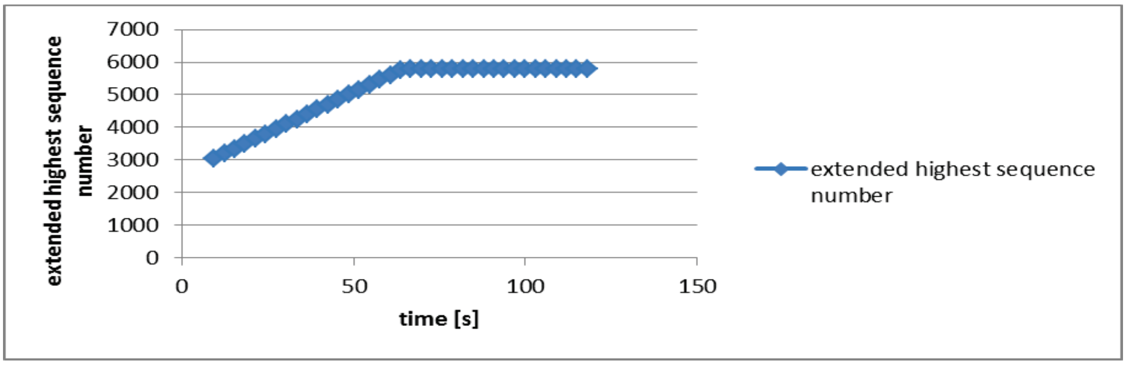


Figure 6.30: RTCP report on Extended Highest Sequence Number

The plot in figure 6.30 gives the same insight as the sender octet count plotted in figure 6.29. As soon as the call session between party-A and party-B is placed on hold, media packets exchange between these two parties is suspended. It is at this point that party-A adapts its own RTP port to a=sendonly towards party-B and a=reconly at party-B RTP port. Hence, party-A will not expect any further RTP packets from party-B as long as the call session is placed on hold.

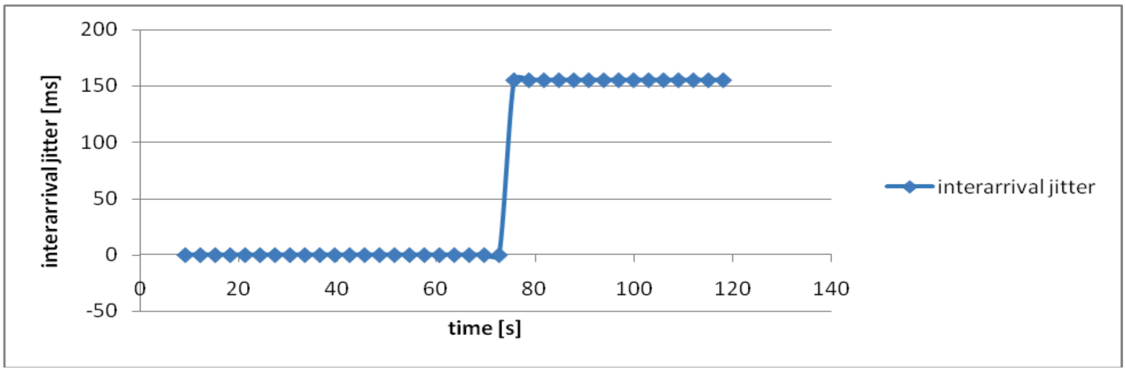


Figure 6.31: RTCP report on Interarrival jitter

Similar to figure 6.23, the plot in figure 6.31 shows an abrupt rise in jitter as soon as call session towards party-B is placed on hold and the jitter remains high for the remaining period of the call.

RTCP video 10.42.44.235 to 10.42.44.161 in the reverse direction

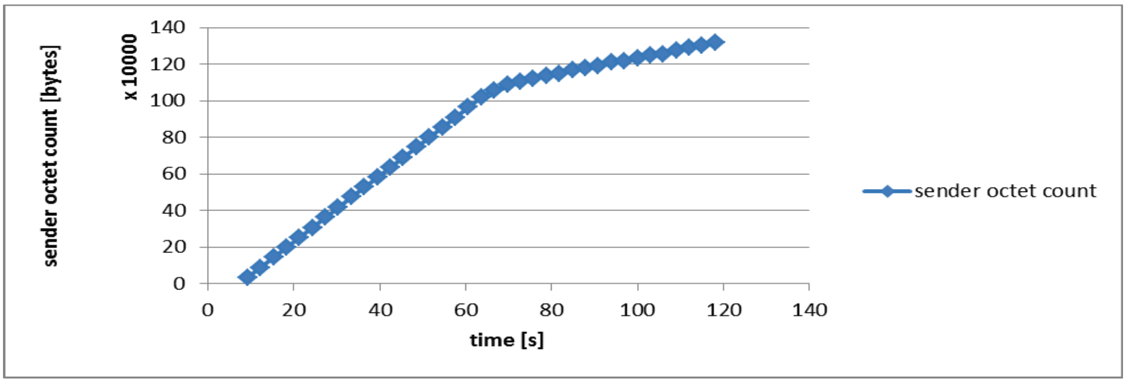


Figure 6.32: RTCP report on Sender Octet Count

Compared to the plot in the forward direction, figure 6.32 shows that in the reverse direction, the RTP octet count continues to rise until the call is put on hold after which, there is a drop in the number of RTP video octet being transmitted. This drop indicates that the rate at which packets are being transmitted has been reduced, thereby releasing network resource to a certain level with only synchronization information being transmitted.

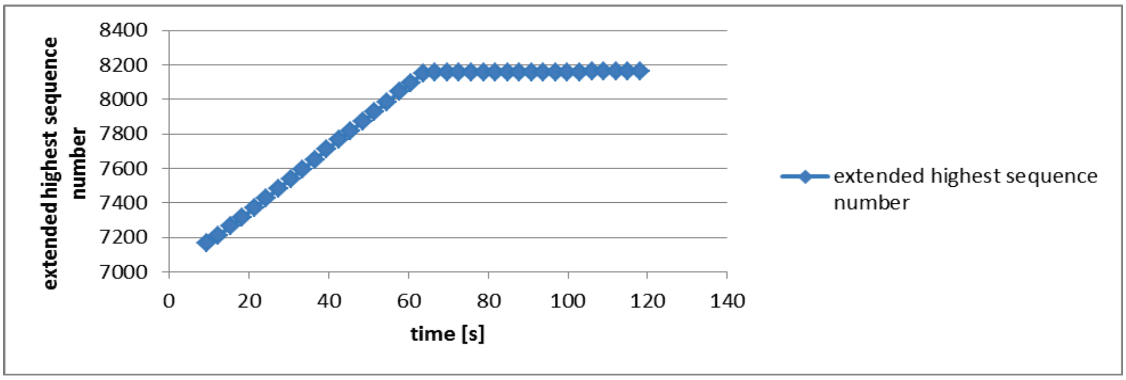


Figure 6.33: RTCP report on Extended Highest Sequence Number

Figure 6.33 corroborates figure 6.32. As soon as the call session is put on hold, the media packets exchange between A-party and B-party is suspended.

The analysis of the call session from the A-party to the C-party (third user) is included in Appendix B.

Summary

In summary, the plots obtained and shown in this test case depict that during the phase of placing a call session on hold and initiating a new one, quite a number of signaling flows are involved in order for a smooth transition from the former session to the new one. Also during the second phase of new call session, the old session is still being kept in synchronization by intermittently exchanging sync information between party-A and party-B. Based on the result obtained, it was noticed that media clipping occurred during the transition phase from old session to the new one due to signaling being out of sync with the media exchange. During this period, cracks and blurry effect were experienced in the picture quality. This is due to the effect of RTP messages being transmitted before the reception of the 200 Ok by the party-A. But as soon as the 200 Ok was received, picture quality became clear.

6.4 Test Case 4

With the so many advantages associated with video conferencing, this scenario has been chosen in order to look into and analyze the volume of signaling flows and media exchange steps and procedures involved in establishing call sessions involving at least three or more users. This scenario is also of interest because it shows the call set up, negotiations, signaling and media exchange between parties involved in conference call. A video call session is negotiated and established between a caller party-A and party-B after which a third party-C is invited to join in the on-going video call session making the session a conference consisting of the three parties.

Once a session is established between party-A and party-B, a new session is established towards party-C. While the new session is being established towards party-C by party-A, party-B is placed on hold. As soon as the new call session with party-C is established, the session with party-B is retrieved and joins the new session enabling the three parties to exchange signaling and media among themselves.

For the analysis, the forward and backward directions of the signaling flows and media exchange by each of the participants have been considered. In this test case scenario, it is expected that during the establishment phase of the video conferencing (set up phase), does a lot of signaling and media transfer to both remote parties, consuming a considerable amount of bandwidth.

Audio stream from 10.42.44.161:1600 to 10.42.44.235:50504

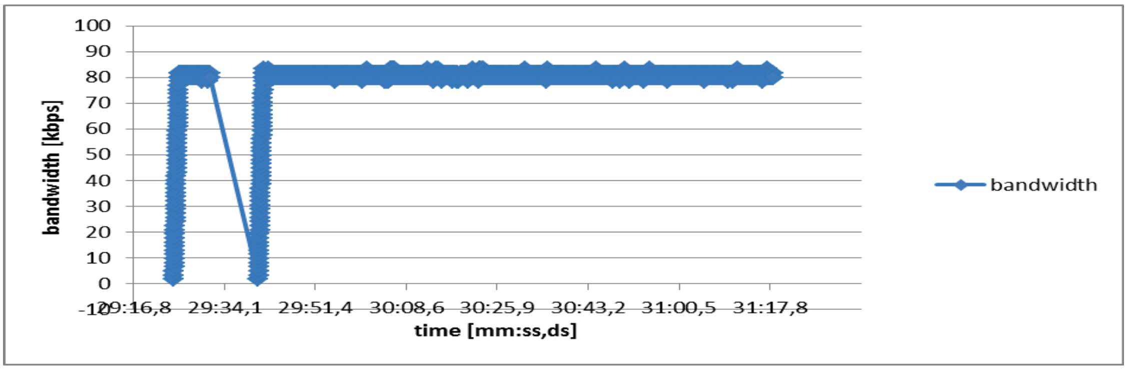


Figure 6.34: Bandwidth usage by the audio stream in the forward direction

Just like it is stated in the test case definition, the plot above depicts the bandwidth usage during the call session between parties A and B. Figure 6.34 shows an abrupt sharp rise in the bandwidth usage only to decline at after a few seconds when the third party is invited to join the on-going call. Once the third party joins the conference, usage of the bandwidth rises again, for the remainder of the session.

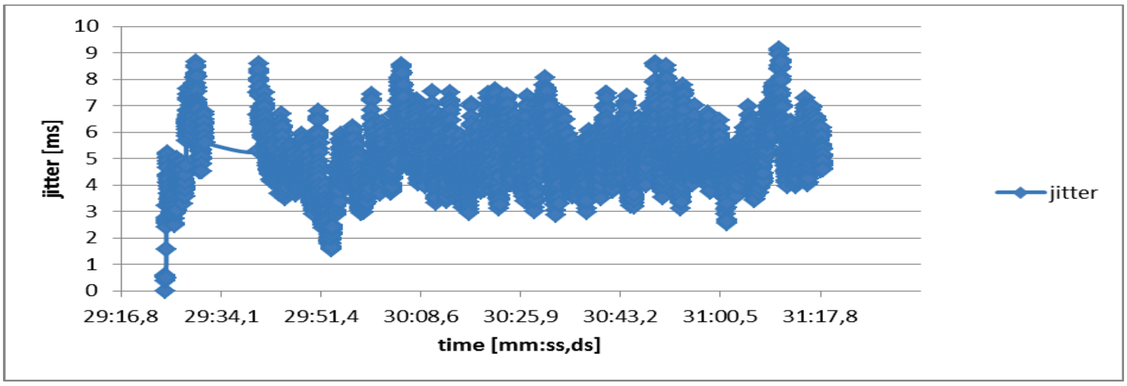


Figure 6.35: Jitter experience by the audio stream in the forward direction

It can be seen from figure 6.35 that after a few seconds, the interarrival jitter leveled off due to he fact that at this period, the remote party was put on hold in order for the local party to initiate a call to another third party after which the jitter resumes its fluctuation. This is the period during which the call to the remote end that was put on hold was retrieved.

Video stream from 10.42.44.161: 33894 to 10.42.44.235:34404

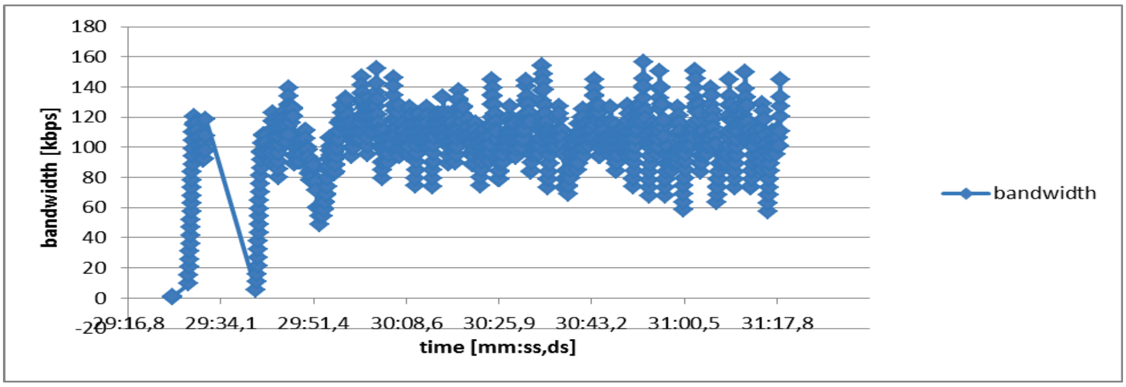


Figure 6.36: Bandwidth usage by the video stream in the forward direction

Similar to figure 6.34, the video stream is depicted in figure 6.36. After a few seconds, the bandwidth usage drops before rising again. This drop is the period in which the remote end is put on hold and another call was being established to a third party after which the call to the remote end is retrieved.

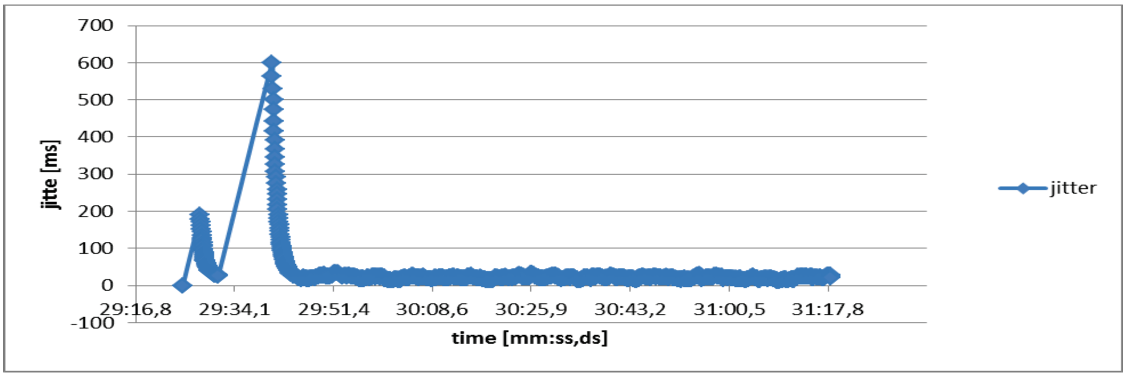


Figure 6.37: Jitter experienced by video stream in the forward direction

As depicted in the plot of figure 6.37, the jitter trend is represented during the initial call establishment between parties A and B, with an initial rise which after a while comes down and become stable. After the third party-C, was invited and has joined the call, there was another sharp increase in jitter which becomes stable for the remaining duration of the call.

RTCP audio report: 10.42.44.161 to 10.42.44.235 in the forward direction

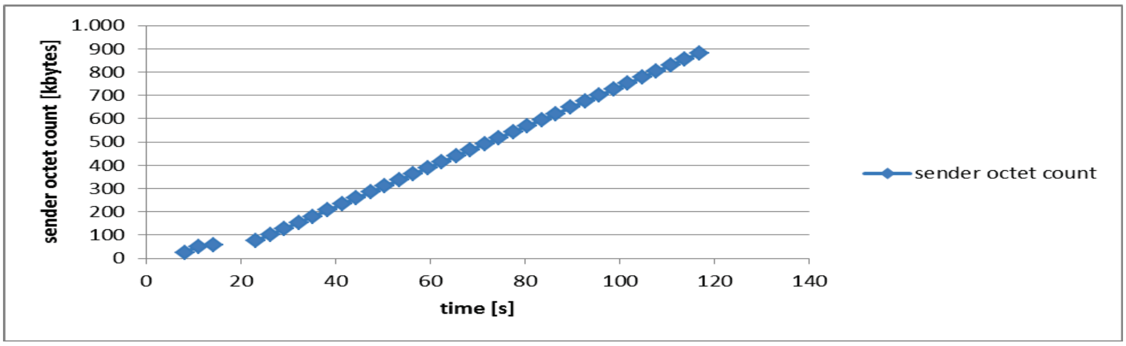


Figure 6.38: RTCP report on Sender Octet Count

The steady rise and sudden break in the sender octet count of the plot of figure 6.38 shows the first established call between the two parties, the period the third party was being invited and the session that involves the three parties.

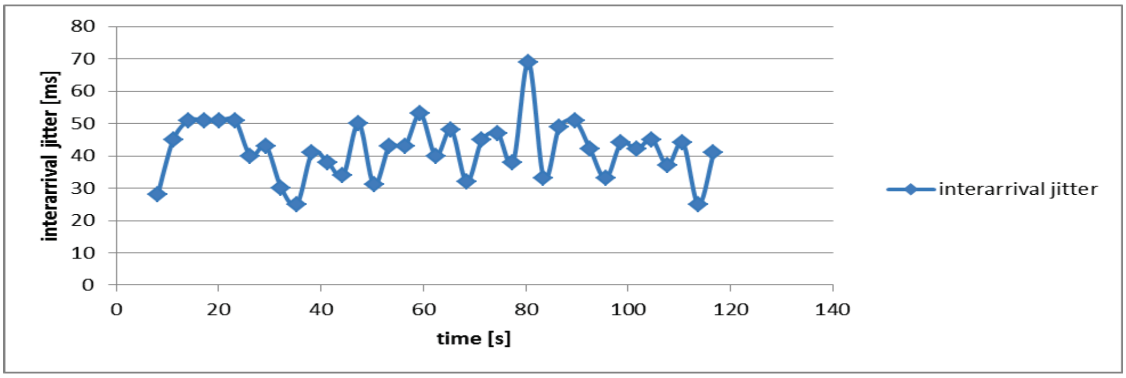


Figure 6.39: RTCP report on the Interarrival jitter

Figure 6.39 shows the variation in the interarrival jitter. It can be inferred from the plot that during the few seconds (approximately 8s), the variation was suspended due to the fact that the remote end was put on hold while establishing call towards a third party. Once the third party joins the conversation and the remote call retrieved, the fluctuation in the jitter resumes.

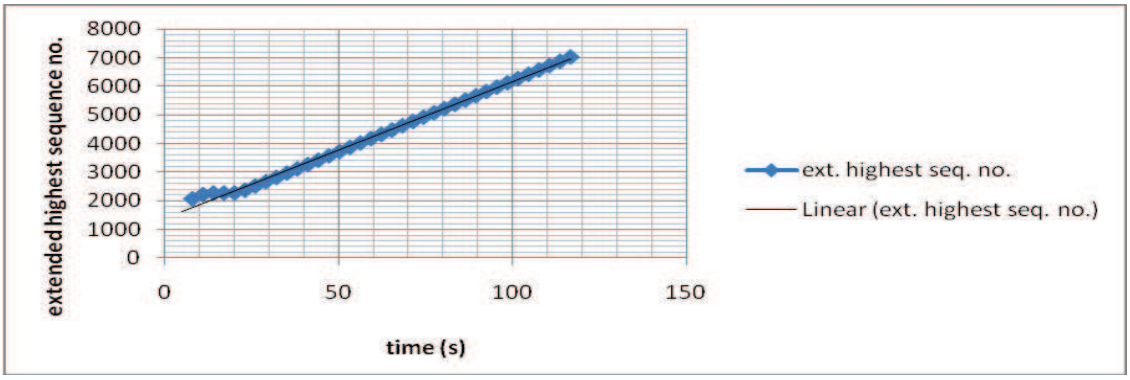


Figure 6.40: RTCP report on the Extended Highest Sequence Number

Figure 6.40 corroborates figure 6.38. The portion of the plot that shows no rise is that period that the remote end was put on hold after which the call was retrieved and RTP packets transmission resumed.

RTCP video 10.42.44.161 to 10.42.44.235 in the forward direction

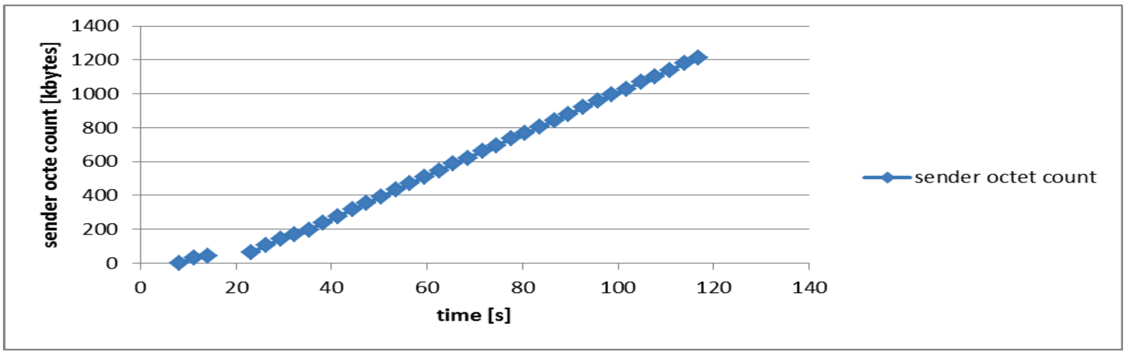


Figure 6.41: RTCP report on the Sender Octet Count

Figure 6.41 shows the video stream of the conference call between the three parties. It is similar to the audio stream in figure 6.38. The break in sender octet count depicts the period call was established to the third party-C.

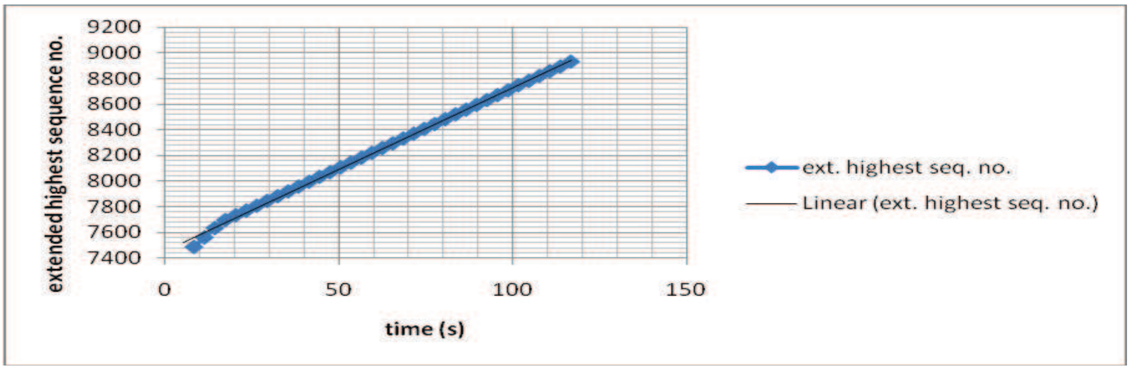


Figure 6.42: RTCP report on the Extended Highest Sequence Number

Figure 6.42 corroborates figure 6.41.

Audio stream from 10.42.44.235:50504 to 1042.44.161:1600

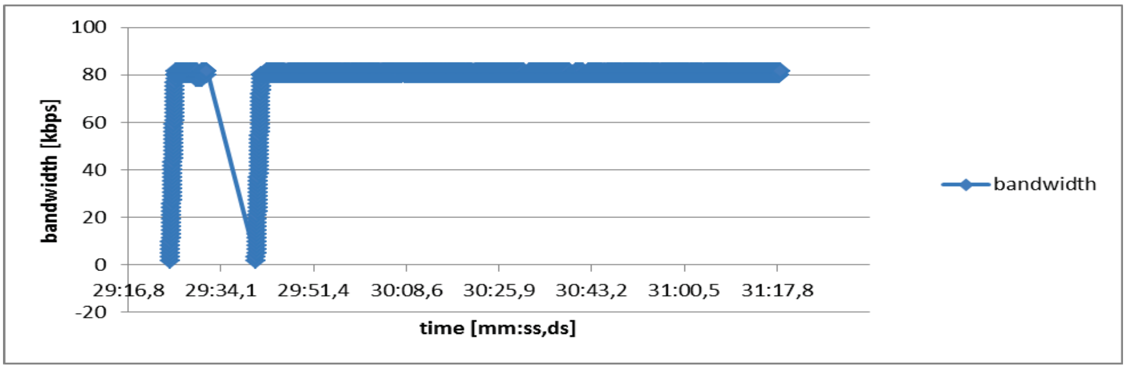


Figure 6.43: Bandwidth usage by the audio stream in the reverse direction

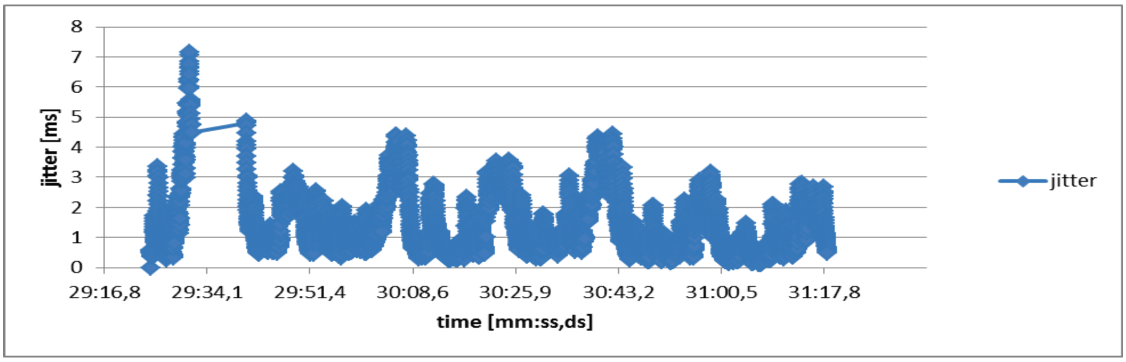


Figure 6.44: Jitter experienced by the audio stream in the reverse direction

Just like in the forward direction for the audio stream, figures 6.43 and 6.44 show the audio stream similar to figures 6.35 and 6.36 respectively.

Video stream from 10.42.44.235:34404 to 10.42.44.161:33894

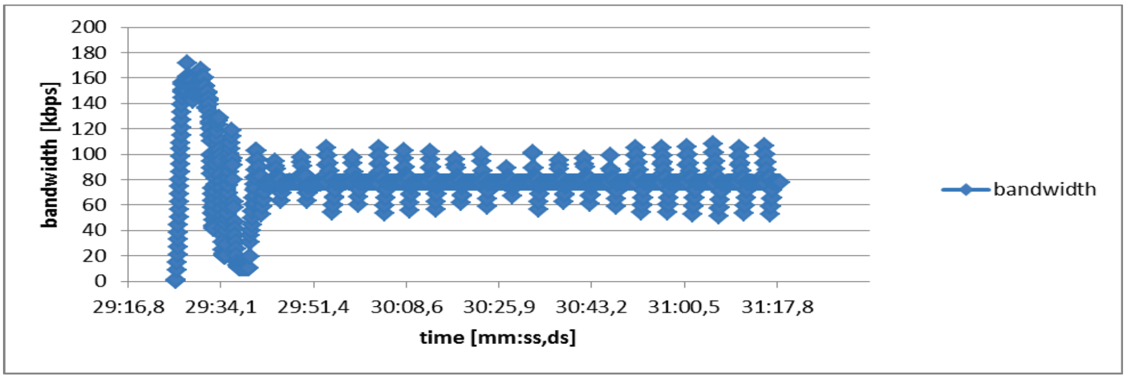


Figure 6.45: Bandwidth usage by the video stream in the reverse direction

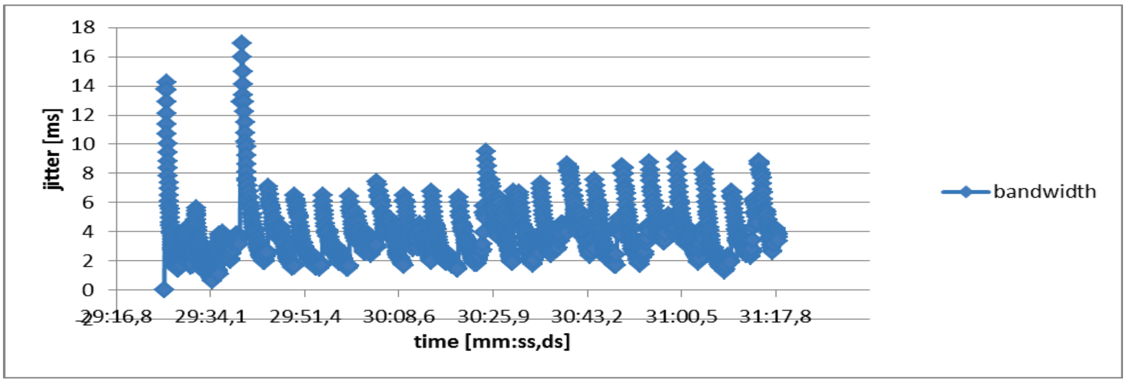


Figure 6.46: Jitter experienced by video stream in the reverse direction

Figures 6.45 and 6.46 show the network usage by the video stream in the reverse direction. This is similar to figures 6.36 and 6.37 in the forward direction. The peaks noticed in 6.45 are due to the large I-frames being transported.

RTCP audio report: 10.42.44.235 to 10.42.44.161 in the reverse direction

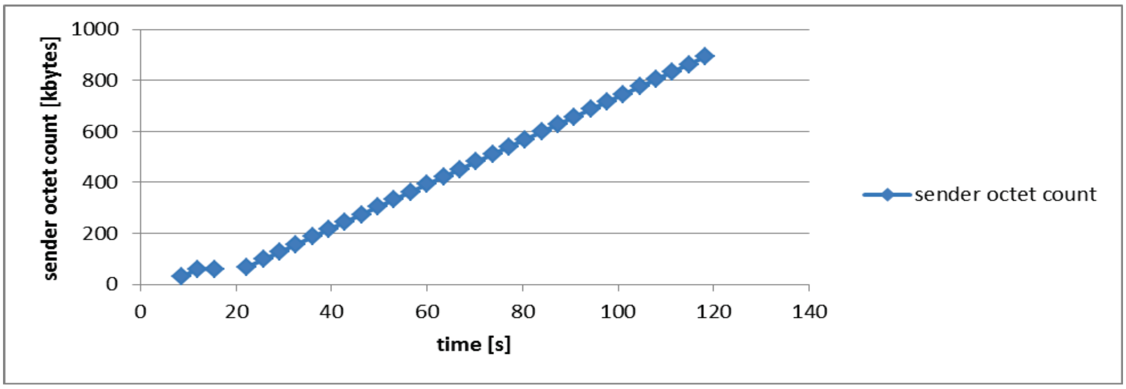


Figure 6.47: RTCP report on the Sender Octet Count

Figure 6.47 is similar to its counterpart in the forward direction (figure 6.38). The steady rise and sudden break in the sender octet count of the plot of figure 6.47 shows the first established call between the two parties, the period the third party was being invited and the session that involves the three parties.

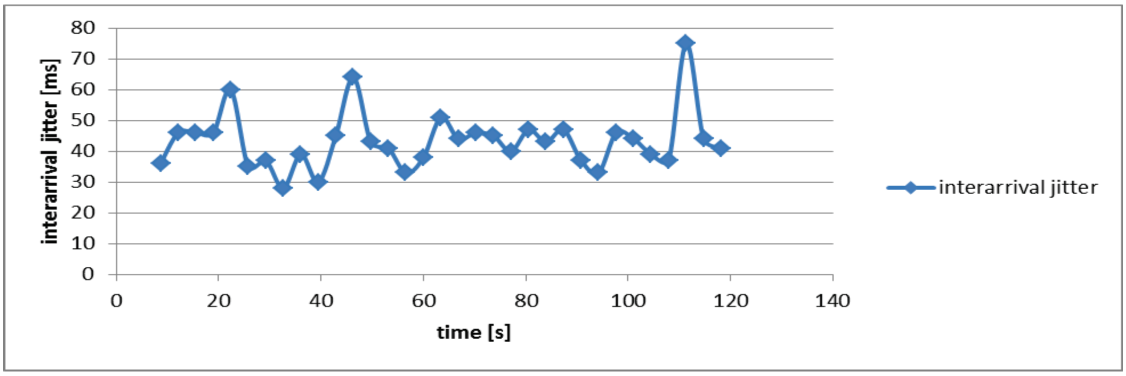


Figure 6.48: RTCP report on the Interarrival jitter

Figure 6.48 is similar to its counterpart in the forward direction (figure 6.39). Figure 6.48 shows the variation in the interarrival jitter. It can be inferred from the plot that during the few seconds, the variation was suspended due to the fact that the remote end was put on hold while establishing call towards a third party. Once the third party joins the conversation and the remote call retrieved, the fluctuation in the jitter resumes.

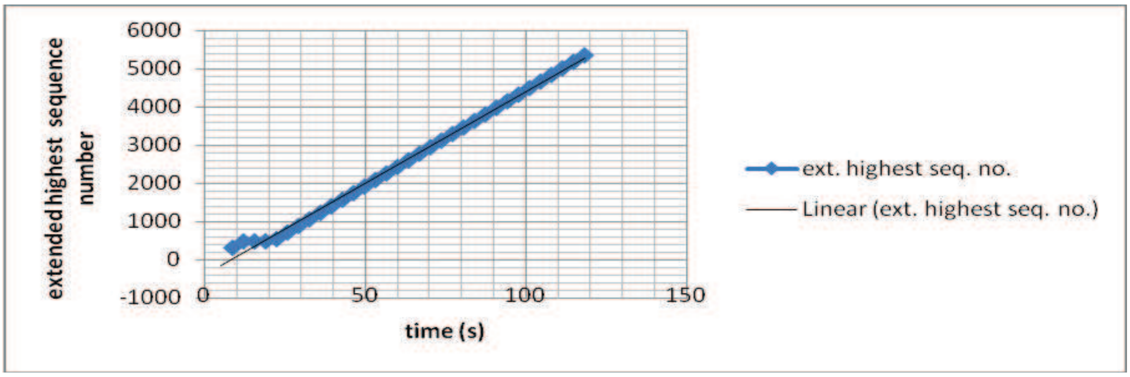


Figure 6.49: RTCP report on the Extended Highest Sequence Number

Figure 6.49 is similar to its counterpart in the forward direction (figure 6.40). The portion of the plot that shows no rise is that period that the remote end was put on hold after which the call was retrieved and RTP packets transmission resumed.

RTCP video report : 10.42.44.235 to 10.42.44.161 in the reverse direction

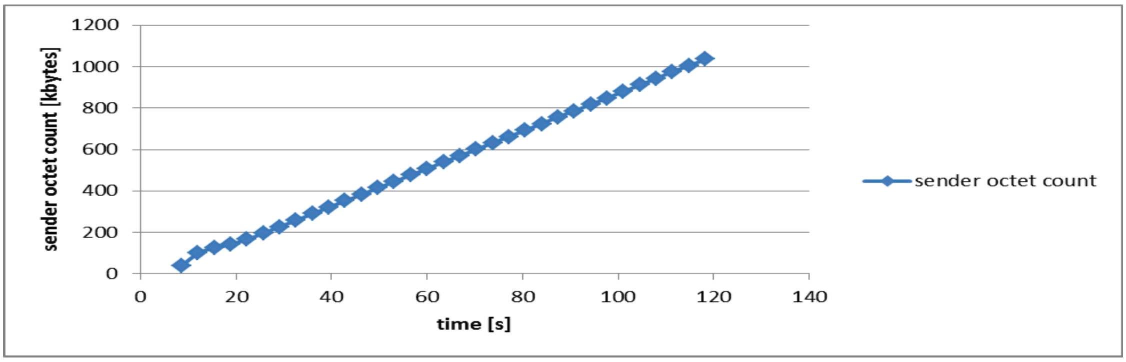


Figure 6.50: RTCP report on the Sender Octet Count

Figure 6.50 is similar to its counterpart in the forward direction (figure 6.41).

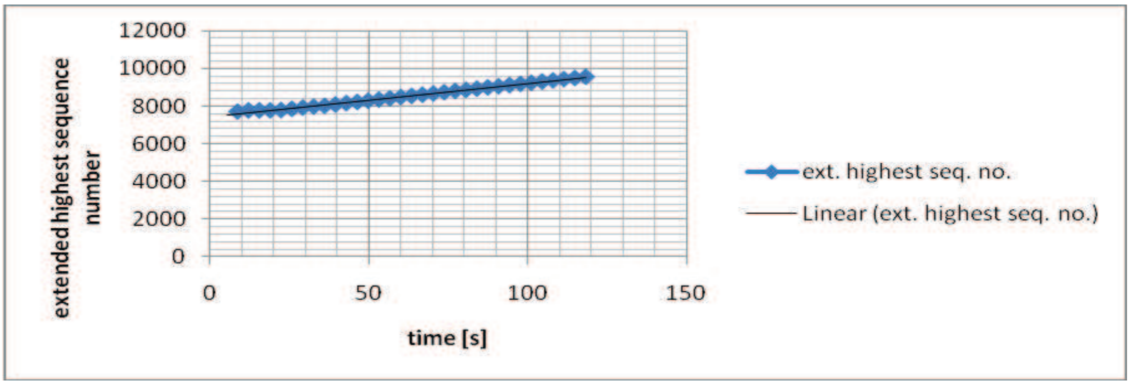


Figure 6.51: RTCP report on the Extended Highest Sequence Number

Figure 6.51 is similar to its counterpart in the forward direction (figure 6.42). As RTP packets are being transferred, the value of the expected sequence number increases linearly over time.

The analysis of the session from A-party to C-party is in Appendix C which includes the RTP and RTCP reports plots.

Summary

In summary, the plots shown in this test case confirm that demand on network resource usage is on the increase during conference call between three or more parties due to the fact that lots of signaling is involved with also a fairly large amount of data traffic between exchanged between the participating parties. It is also discovered that there is heavy fluctuation in the jitter due to paths traversed by each RTP data packet, queuing and high volume of traffic (RTP media packets) passing through the interfaces on the hub associated with each of the users involved in the session.

7. Conclusions and topics for further study

7.1 Conclusions

As a result of the work carried out on this thesis project, the following conclusions could be highlighted:

- With reference to IETF RFC 3550, which states that the interval between RTCP messages transmitted between caller parties should be 5 seconds, it was discovered in the course running some of the test cases that there were deviations from this stated interval time. RTCP transmission intervals of 3 – 3.5 seconds were recorded due to the session bandwidth available during the call session. Algorithms could be developed to fix the minimum transmission interval of RTCP messages irrespective of the amount of available network resource (bandwidth).
- During the SIP INVITE and SDP Offer/Answer phase of call initiation and establishment, some information pertaining to codecs, ports, and other parameters necessary for the media transmission are exchanged between users. Information about packetization time for the video media is not included in this exchange and this has made it impossible to be able to calculate the jitter experienced in the transfer of video packets from end-to-end. This has actually made it impossible to have a clearer picture about the delay variation (jitter) experienced by the video component of the media exchange.
- Accurate synchronization between the media stream (RTP packets) and the RTCP report messages is of very high importance in order for real time analysis of the quality of a call session. In other words, just enough provision must be made available for the RTCP report messages especially in situations whereby online monitoring is to be implemented so as to be able to know the quality of call sessions at any instant in time.
- In some of the test case scenarios set up in the course of this project, it was discovered that media starts arriving at A-party who is initiating a call even before receiving the 200 Ok message. The reason for this is that media are exchanged end-to-end between users once the media path is established (B-party starts media transfer to A-party once it has sent the 200 Ok). Media transfer is end-to-end between parties unlike the signaling path that has to be traveled by the 200 Ok which involves intermediate proxies. This situation could result into media clipping especially in a situation whereby the resource reservation phase is yet to be completed by the A-party in the direction from local party to remote party. It was also discovered that during video calls, video streams starts quite a while after the audio stream after the call is answered. This effect could be resolved by implementing an SDP offer/answer model negotiating the packetization time of the video stream to match that of the negotiated audio stream.

- When a remote party is put on hold, the stream attribute is set to a=sendonly and not a=inactive. This attribute setting is to avoid RTP timeout which would definitely lead to the termination of the call to the remote party. While the call is on hold, RTP packets containing synchronization information are still being exchanged in order to keep the remote party RTP port open and available should the call to the remote party be retrieved in order to continue the call session.
- The measurements have all been done on the calling-party PC so that the result is from the view of this PC. Taking measurements on the called-party PCs would give a slightly different result especially in arrival times of signaling and RTP data packets. This difference in arrival times as measured from both ends could be improved by taking measurements at both ends while a session is on-going and afterwards, synchronize both information by using time values at each end. The IMS charging Identifier could be used correlate SIP messages captured at different points using the sessions' different SSRC identifiers.
- Evaluation of throughput (at low load and high load) is of high importance at the hub because if there are hundreds of SIP phone connected to the IMS network, congestion at the hub then becomes crucial due to the heavy traffic traversing the network. The data transmission rate and queuing value of the hub should match that of the network traffic rate in order for optimum performance.

7.2 Suggested topics for further study

With the increase in demand for video telephony, there is the need for service providers (VoIP and Video-over-IP providers) to monitor their network real time and determine the quality of service of the video services they are offering their customers. The field of study of this report has been into how the SIP, RTP and RTCP protocols are synchronized and used in analyzing and stating the quality of reception in video call sessions. The report has researched into the control and user planes of call establishment and media transfer; media exchange and network resource consumption (bandwidth), exchange of RTCP message reports between call parties and the frequencies of these RTCP messages. The recommendation provided by this report focuses mainly on these three protocols analyzed:

- The user capability information exchanged during SIP INVITE, SDP Offer/Answer phase of call establishment should be researched into so that parameters like packetization time, minimum required bandwidth are also included.
- In the report of this project, the H.263 has been the video codec used in the test cases analyzed. An interesting topic would be to explore other video codecs like MPEG 1, 2, 4 etc and analyze their effects on the RTCP report parameters like interarrival jitter, sender octet count, delay since last SR timestamp.
- The report has analyzed test cases using a simulated IMS environment (IMS-in-a-box). A topic for further study could be a practical IMS network where many SIP signaling and voice/audio streams are being handled. This will enable the possibility of taking measurements even between network entities in order to have a better insight into more accurate propagation time of media and signaling

flows. Measuring at different points in the network will give a total and clearer view of the signaling and media exchange processes.

- RTP and RTCP messages are transported using UDP transport protocol which is quite unreliable. Since guaranteed and timely delivery of RTCP messages is required in order for accurate interpretation of the network situation, use of TCP to transport RTCP reports would be an interesting area of study. The mix of UDP (for RTP) and TCP (for RTCP) could be applied and the effect analyzed.
- RTCP report size determines the transmission interval. Study into reducing the RTCP report just to the size that would give an insight into the quality of reception of call sessions would be an interesting area of research.

References

[1] <http://3gpp.org//article/ims>

[2] Camarillo, Gonzalo and Miguel Angel Garcia-Martin: *The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds.* John Wiley & Sons, 3rd edition, 2008, ISBN: 9780470516621

[3] Dan Leih and Dave Halliday: Introduction to IMS: Standards, protocols, architecture and functions of the IP Multimedia Subsystem, IMS White paper series: part 1, Motorola Inc. Embedded Communications Computing. <http://www.motorola.com/mot/doc/6/6403_MotDoc.pdf>

[4] Boris IV. Kalaglarski & Emilio Di Geronimo: IMS Interworking: Master Thesis report, KTH Information and Communication Technology, Stockholm Sweden 2007.

[5] Symbian Developer Network: An Introduction to IMS, version 1.1, January 2008, <http://developer.symbian.com/main/downloads/papers/IMS_Introduction_Part1.pdf>

[6] Travis Russel: *Session Initiation Protocol (SIP): controlling convergent networks*. McGraw-Hill Communications 2008. ISBN 978-0-07-148852-5

[7] Travis Russel: *The IP Multimedia Subsystem (IMS): Session control and other network operations*. McGraw-Hill Communications 2008. ISBN 978-0-07-148853-2

[8] M. Handley, C. Perkins and E. Whelan, *SAP: Session Announcement Protocol*, IETF, RFC 2974, October 2000, < http://www.ietf.org/rfc/rfc2974.txt>

[9] J. Rosenburg, H. Schulzrinne, G. Camarillo, A. Johnston, R. Sparks, M. Handley and E. Schooler, *SIP: Session Initiation Protocol*, IETF, RFC 3261, June 2002 < http://www.ietf.org/rfc/rfc3261.txt>

[10] M. Handley, V. Jacobson and C. Perkins, *SDP: Session Description Protocol*, IETF, RFC 4566, July 2006, <http://tools.ietf.org/html/rfc4566>

[11] H. Schulzrinne, A. Rao and R. Lanphier, *RSTP: Real Time Streaming protocol*, IETF, RFC 2326, April 1998, <http://tools.ietf.org/html/rfc2326>

[12] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications*, IETF, RFC 3550, July 2003 < http://tools.ietf.org/html/rfc3550>

[13] P. Calhoun, J. Loughney, E. Guttman, G. Zorn and J. Arkko, *Diameter Base Protocol*, IETF, RFC 3588, September 2003 <http://tools.ietf.org/html/rfc3588>

[14] ITU-T Recommendations H.248.1: http://www.itu.int/dms_pubrec/itu-t/rec/h/T-REC-H.248.1-200509-I

[15] C. Groves, M. Panteleo, T. Anderson and T. Taylor, Gateway Control protocol Version 1, IETF, RFC 3525, June 2003 < http://tools.ietf.org/html/rfc3525>

[16] 3GPP, *IP Multimedia Subsystem (IMS), Stage 2*, TS 23.228 v9.3.0, 2010-03-26 <http://www.3gpp.org/ftp/Specs/html-info/23228.htm>

[17] Miika Poikeselka, Georg Mayer, Hisham Khartabil, Aki Niemi: *The IMS: IP Multimedia Concepts and Services in the Mobile Domain*. John Wiley & Sons Ltd 2004. ISBN: 0-470-87113-X

[18] 3GPP, *IP Multimedia Subsystem (IMS), Stage 2*, TS 23.228 v9.3.0, 2010-03-26 <http://www.3gpp.org/ftp/Specs/html-info/23228.htm>

[19] 3GPP, IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP) Stage 3, TS 24.229 v9.0.0, 2009-06-08

[20] R. Price, C. Bormann, J. Christoffersson, H. Hannu, Z. Liu and J. Rosenberg, *Signaling Compression (SigComp)*, IETF, RFC 3320, January 2003 < http://www.rfc-editor.org/rfc/rfc3320.txt>

[21] M. Garcia-Martin, C. Bormann, J. Ott, R. Price and A. B. Roach, *The Session Initiation Procol (SIP) and Session Description Protocol (SDP) Static Dictionary for Signaling Compression (SigComp)*, IETF RFC 3485, February 2003 < http://www.rfc-editor.org/rfc/rfc3485.txt>

[22] G. Camarillo, *Compressing the Session Initiation Protocol*, IETF, RFC 3486, February 2003 <http://tools.ietf.org/rfc/rfc3486.txt>

[24] J. Peterson, *Enumservice registration for Session Initiation Protocol (SIP) Address-of-Record*, IETF, RFC 3764, April 2004 < http://www.ietf.org/rfc/rfc3764.txt>

[25] Miika Poikeselka, Georg Mayer: The IMS: IP Multimedia Concepts and Services. John Wiley & Sons Ltd 2009. ISBN: 978-0-470-72196-4

[26] 3GPP, *Open Service Access (OSA) Application Programming Interface (API); Part 1: Overview*, TS 29.198 v8.0.0, 2010-12-16

[27] 3GPP, *GSM: Network Architecture*, TS 23.002 v9.2.0, 2009-12-14

[28] R. Stewart, K. Morneault, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang and V. Paxson, *Stream Control Transmission Protocol*, IETF, RFC 2960, October 2000 <http://www.ietf.org/rfc/rfc2960.txt>

[29] 3GPP, *Numbering, addressing and identification*, TS 23.003 v9.1.0, 2009-12-18

[30] H. Schulzrinne, *The tel URI for Telephone Numbers*, IETF, RFC 3966, December 2004 < http://www.rfc-editor.org/rfc/rfc3966.txt>

[32] B. Aboba and M. Beadles, *The Network Access Identifier*, IETF, RFC 2486, January 1999

[33] ETSI TS 102 221 Release 7: "*Smart Cards; UICC-Terminal interface; Physical and logical characteristics ".*

[34] 3GPP TS 31.101 V8.0.0: “UICC-terminal interface; Physical and logical characteristics (Release 8)

[35] M. Handley and V. Jacobson: *SDP: Session Description Protocol*. RFC 2327, April 1998. <http://www.ietf.org/rfc/rfc2327.txt>

[36] P. Faltstrom, M. Mealling: *The E.164 to Uniform Resource Identifiers (URI), Dynamic Delegation Discovery System (DDDS) Application (ENUM)*, IETF RFC 3761, April 2004 <http://www.ietf.org/rfc/rfc3761.txt>

[37] 3GPP TS 23.221: Architectural Requirement for U-TRAN and EU-TRAN based Systems <http://www.3gpp.org/ftp/Specs/html-info/23221.htm>

[38] Next-generation Intelligent Networks: Migrating to IMS Ericsson white paper (An Ericsson white paper)

[39] Rogelio Martinez Perea: *Internet Multimedia Communications using SIP*: Morgan Kaufmann Publishers 2008. ISBN: 978-0-12-374300-8

[40] J. Rosenberg and H. Schulzrinne, *An Offer/Answer Model with the Session Description Protocol (SDP)*, IETF, RFC 3264, June 2002 < http://www.ietf.org/rfc/rfc3264.txt>

[41] J. Rosenberg and H. Schulzrinne, Reliability of Provisional Responses in the Session Initiation Protocol (SIP), IETF, RFC 3262, June 2002 < http://www.ietf.org/rfc/rfc3262.txt>

[42] H. Schulzrinne and S. Casner, RTP Profile for Audio and Video Conferences with Minimal Control, IETF, RFC 3551, July 2003 < http://www.ietf.org/rfc/rfc3551.txt>

[43] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, S. Jarmin: *Resource ReSerVation Protocol (RSVP),* IETF RFC 2205, September 1997<http://www.ietf.org/rfc/rfc2205.txt

[44] T. Berners-Lee, R. Fielding and L. Masinter, *Uniform Resource Identifier (URI): Generic Syntax,* IETF, RFC 3986, January 2005 < http://www.ietf.org/rfc/rfc3986.txt>

[45] <http://encyclopedia.jrank.org/articles/pages/6928/Video-Over-Ip.html#ixzz0uPd8JVMx>

[46] Packet-based multimedia communications systems, ITU-T H.323

[47] Ismo Anttila, Markku Paakkunainen: *Transferring real-time video on the Internet.* < http://www.tml.tkk.fi/Opinnot/Tik-110.551/1997/iwsem.html>

[48] <http://en.wikipedia.org/wiki/Common_Intermediate_Format>

[49] M. Speer, D. Hoffman: *RTP Payload Format of Sun's CellB Video Encoding.* RFC 2029 October 1996. <http://www.ietf.org/rfc/rfc2029.txt>

[50] L. Berc, W. Fenner, R. Frederick, S. McCanne and P. Stewart*: RTP Payload for JPEG Video*. RFC 2435, October 1998. <http://www.ietf.org/rfc/rfc2435.txt>

[51] T. Turletti and C. Huitema: *RTP Payload for H.261 Video Streams*. RFC 2032, October 1996. <http://www.ietf.org/rfc/rfc2032.txt>

[52] C. Zhu: *RTP Payload for H.263 Video Streams*. RFC 2190, September 1997. <http://www.ietf.org/rfc/rfc2190.txt>

[53] C. Bormann, L. Cline, G. Deisher, T. Gardos, C. Maciocco, D. Newell, J. Ott, G. Sullivan, S. Wenger and C. Zhu: *RTP Payload for the 1998 Version of ITU-T Rec. H.263 Video (H.263+)*. RFC 2429, October 1998. <http://www.ietf.org/rfc/rfc2429.txt>

[54] D. Hoffman, G. Fernando, V. Goyal and M. Civanlar: *RTP Payload Format for MPEG1/MPEG2 Video*. RFC 2250, January 1998. <http://www.ietf.org/rfc/rfc2250.txt>

[55] M. Handley, C. Perkins and E. Whelan: *Session Announcement Protocol.* RFC 2974 (Experimental), October 2000. <http://www.ietf.org/rfc/rfc2974.txt>

[56] Richard Schaphorst. Videoconferencing and Video telephony. *Artech House, Inc.*1996

Abbreviations

3G	3rd Generation
3GPP	3rd Generation Partnership project
3GPP2	3rd Generation Partnership project 2
AAA	Authentication, Authorization and Accounting
AKA	Authentication and Key Agreement
AoR	Address of Record
AS	Application Server
B2BUA	Back-to-Back-User-Agent
BGCF	Breakout Gateway Control Function
BICC	Bearer Independent Call Control
CCF	Charging Collector Function
CDMA	Code Division Multiple Access
CIF	Common Intermediate Format
CN	Connectivity Network
CNAME	Canonical Name
CS	Circuit-Switched
CSCF	Call Session Control Function
CSRC	Contributing Source
DiffServ	Differentiated Services
DNS	Domain Name System
DSL	Digital Subscriber Line
E-CSCF	Emergency-CSCF
ENUM	Telephone Number Mapping
FEC	Forward Error Correction
FQDN	Fully Qualified Domain Name
GOP	Group of Pictures
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HDTV	High Definition Television
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
HTTP	HyperText Transfer Protocol
I-CSCF	Interrogating-CSCF
ID	Identification
IETF	Internet Engineering Task Force
IFC	Initial Filter Criteria
IETF	Internet Engineering Task Force
IM	Instant Messaging
IMS	IP Multimedia Subsystem
IMS-MGW	IMS-Media Gateway
IN	Intelligent Network
IntServ	Integrated Services
IP	Internet Protocol
IP-CAN	IP-Connectivity Access Network

IPSec	IP Security
IPv4	IP version 6
IPv6	IP version 4
ISDN	Integrated Services Digital Network
ISIM	IP Multimedia Services Identity Module
ISUP	ISDN User Part
ITU-T	International Telecommunication Union - Telecommunication
LAN	Local Area Network
LRF	Location Registration Function
LTE	Long Term Evolution
MCU	Multipoint Control Unit
MEGACO	Media Gateway Control Protocol
MG	Media Gateway
MGC	Media Gateway Controller
MGCF	Media Gateway Control Function
MIME	Multipurpose Internet Mail Extensions
MPEG	Moving Picture Expert Group
MRF	Media Resource Function
MRFC	Media Resource Function Controller
MRFP	Media Resource Function Protocol
MTU	Maximum Transmission Unit
NAI	Network Access Identifier
NAT	Network Address Translator
NGN	Next Generation Networks
NTP	Network Time Stamp
OCS	Online Charging System
OMA	Open Mobile Alliance
OSI	Open Systems Interconnection
P-CSCF	Proxy-CSCF
PDA	Personal Digital Assistant
PDSN	Packet Data Serving Node
PLMN	Public Land Mobile Network
PS	Packet-Switched
PT	Packet Type
PSAP	Public Safety Answering Point
PSTN	Public Switched Telephone Network
PUI	Private/Public User Identity
QCIF	Quarter CIF
QoS	Quality of Service
RADIUS	Remote Authentication Dial in User Service
RAN	Radio Access Network
RFC	Request For Comment
RPC	Remote Procedure Call
RR	Receiver Report
RSVP	Resource Reservation Protocol
RTCP	RTP Control Protocol
RTP	Real-Time Transport Protocol
RTSP	Real Time Streaming Protocol

R-UIM	Removable User Identity Module
SCIM	Service Capability Interaction Manager
S-CSCF	Serving-CSCF
SCTP	Stream Control Transport Protocol
SDES	Session Description Protocol Security Descriptions
SDP	Session Description Protocol
SGSN	Serving GPRS Support Node
SGW	Security Gateway
SIM	Subscriber Identification Module
SIP	Session Initiation Protocol
SLF	Subscriber Locator Function
SMTP	Simple Mail Transfer Protocol
SR	Sender Report
SS7	Signaling System 7
SSRC	Synchronization Source
TCP	Transport Control Protocol
TDM	Time Division Multiplexing
THIG	Topology Hiding Internetwork Gateway
UA	User Agent
UAC	User Agent Client
UAS	User Agent Server
UDP	User Datagram Protocol
UE	User Equipment
UICC	Universal Integrated Circuit Card
UMTS	Universal Mobile Telecommunications System
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
USIM	Universal Subscriber Identity Module
VoIP	Voice over IP
WAN	Wide Area Network
WCDMA	Wideband Code Division Multiple Access
Wi-Fi	Wireless Fidelity
WLAN	Wireless LAN

APPENDIX A. Plots of Test Case 1

Analyzed in this section are the RTCP report messages exchanged by the participants in the call session. This test case involves the initiation, set up and establishment of an end-to-end video session between A-party and B-party. The signaling (SIP) and RTP/RTCP measurement of the session is done on the PC initiating the session. Forward direction means the direction of signaling and data messages from the A-party to the B-party while reverse direction is the opposite.

The following IP addresses were allocated to each of the participating parties:
10.42.44.161 to A-party, 10.42.44.235 to the B-party

RTCP audio from 10.42.44.161 to 10.42.44.235 in the forward direction

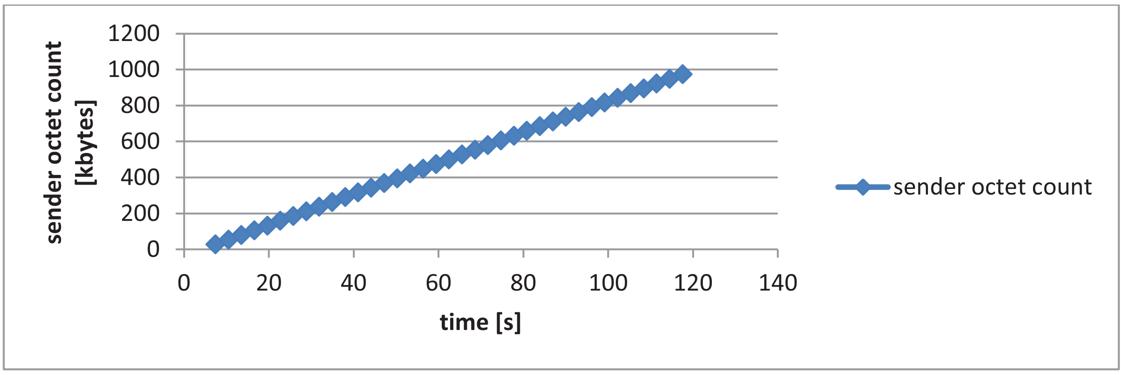


Figure A.1: RTCP report on sender octet count

The plot above relates to figure 6.1 in chapter 6. This plot indicates the RTCP report messages exchanged between the two parties involved in the session. The steady rise in the octet count (bytes) depicted in figure A.1 shows that voice component of the call is being received by the peer at the receiving end. It shows the increase in packet size count from the inception of the call. A fluctuation from this steady rise would indicate that packets are being lost, resulting in drop in reception quality of this component of the media.

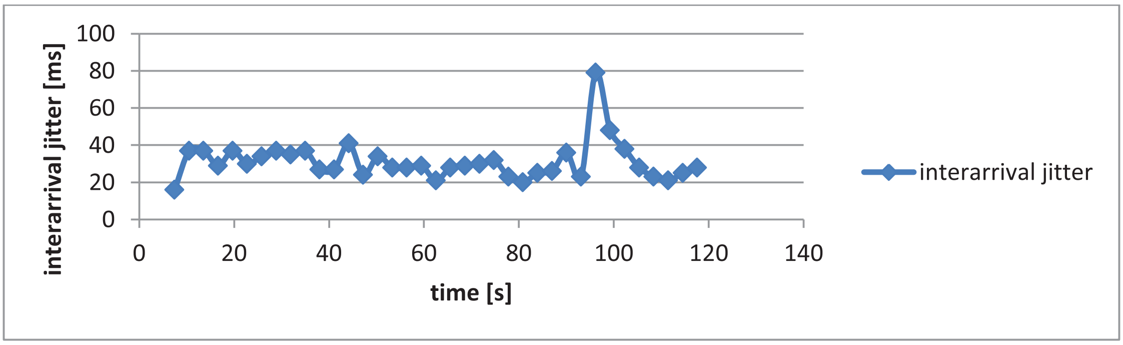


Figure A.2: RTCP report on interarrival jitter (audio stream)

The plot in figure A.2 above shows the interarrival jitter as indicated by the RTCP reports exchanged between the two parties involved in the call session. The interarrival jitter is an estimation of the time interval of received RTP data packet which is measured in timestamp unit and which is in the form of whole number. It is the relative transit time between two data packets. The interarival jitter is calculated for each data packet received by the source SSRC_n (refer to chapter 4.7.1). A relative smooth plot shows a report of a good performance in media exchange between call parties. For the duration of the session, it is noticed that it was only at 98 seconds mark that a sharp rise was experienced. The frequency of this sharp rise gives a feedback on the jitter experienced which translates to quality of user experience.

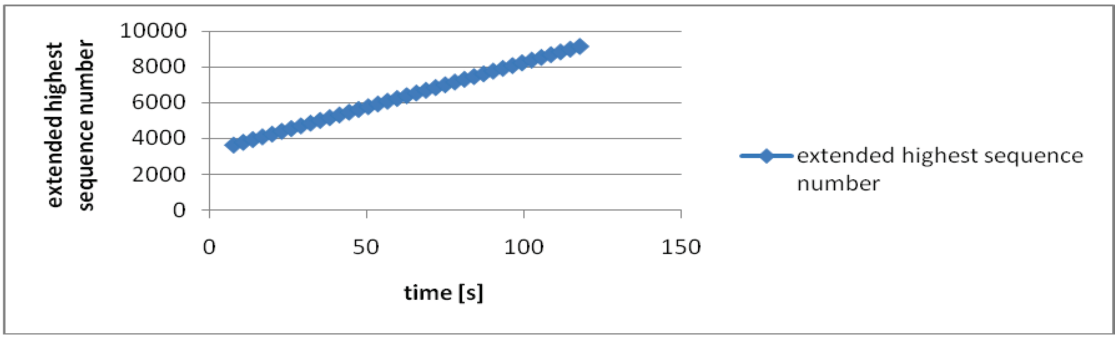


Figure A.3: RTCP report on Extended Highest Sequence Number

Similar to figure A.1, the plot in figure A.3 shows a steady rise in the extended highest sequence number which is an indication that no packets were lost during transmission. The extended highest sequence number gives an indication of the RTP packet sequence number expected at the next RTCP report. This helps to indicate packet loss between the time the last report was sent and the most recent report.

RTCP video 10.42.44.161 to 10.42.44.235 in the forward direction

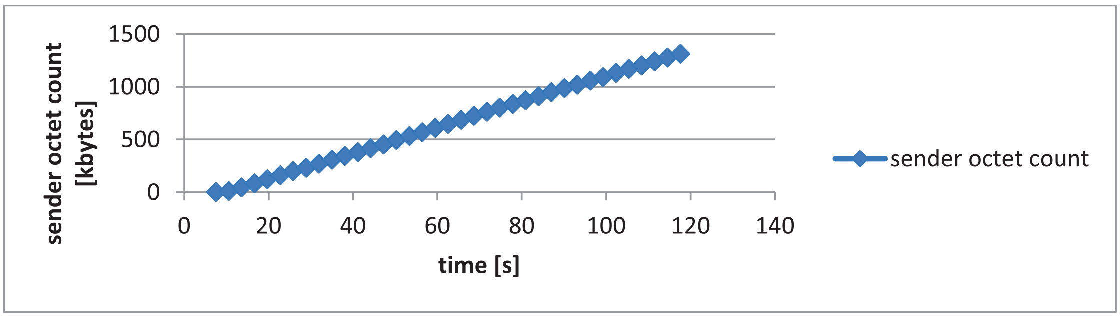


Figure A.4: RTCP report on sender octet count

The plot in figure A.4 is similar to figure A.1 with an increase in octet count (bytes) value due to the video component.

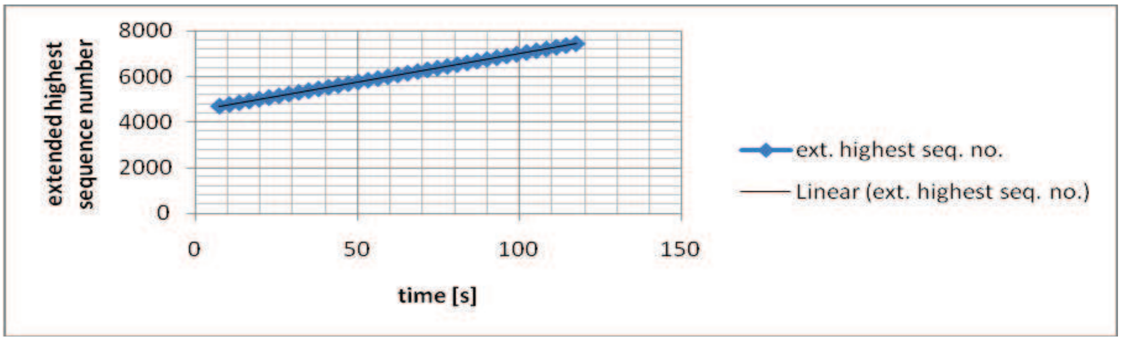


Figure A.5: RTCP report on Extended Highest Sequence Number

Similar to figure A.3, figure A.5 shows no packet loss (good performance). Also for the RTCP report for the video stream, it is observed that the interarrival jitter is reported as zero for the period of the call session. This is so because the packetization time for video is not defined and depends only on the video compression technique adopted.

RTCP audio 10.42.44.235 to 10.42.44.161 in the reverse direction

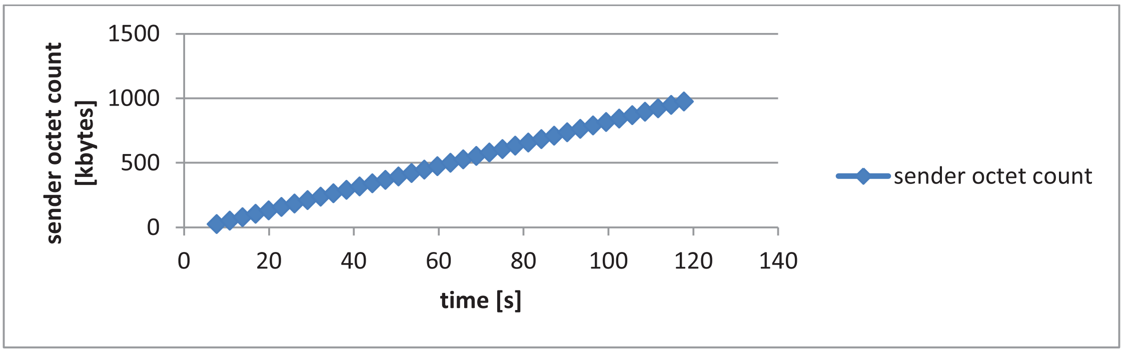


Figure A.6: RTCP report on Sender Octet Count

Similar to figure A.1, figure A.6 shows a steady increase (smooth reception) of the audio stream.

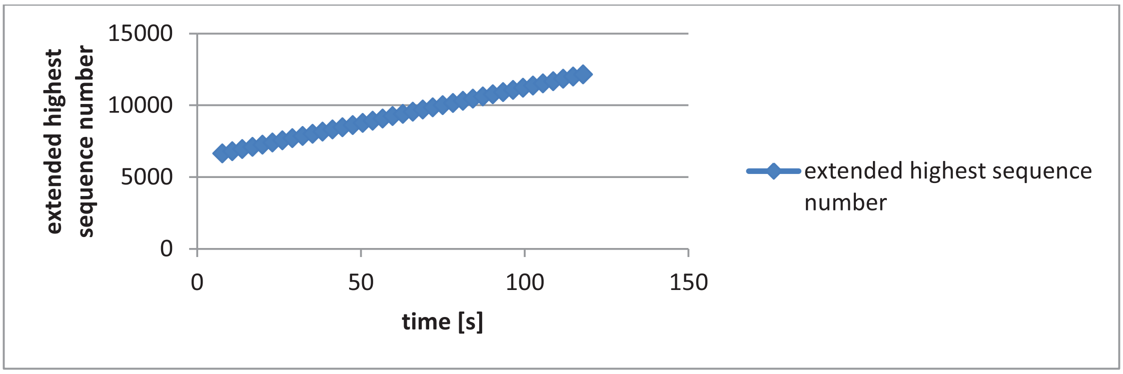


Figure A.7: RTCP report on the Extended Highest Sequence Number

Figure A.7 is similar to figure A.3. It depicts that no packets were lost during the transmission of the audio stream.

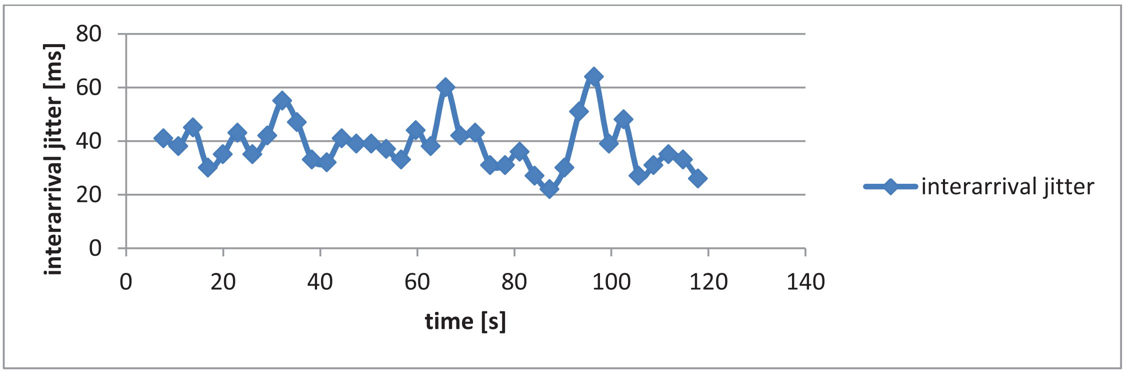


Figure A.8: RTCP report on Interarrival jitter

The plot in figure A.8 above is similar to that of figure A.2 (in the reverse direction). Here, it is noticed that the frequency in sharp rise of the jitter is higher than that of figure A.2. This translates to drop in the quality of user experience which was observed during the call session as cracks and occasional stillness in picture frames.

RTCP video: 10.42.44.235 to 10.42.44.161 (similar to that in the forward direction)

All the plots in the reverse direction for the video stream are not shown but they are all similar to the ones shown in the video stream for the forward direction.

APPENDIX B. Plots of Test Case 3

This test case scenario involves three parties. A video session is established between the A-party and B-party. Once this session is established, after approximately 60 seconds, the A-party (session initiator) initiates a new session towards another third party-C. In order for the A-party to negotiate and exchange media with this third party, the former session it has with B-party had to be put on hold.

The following IP addresses were allocated to each of the participating parties:
10.42.44.161 to A-party, 10.42.44.235 to the B-party and 10.42.44.160 to the C-party.
Forward direction is the direction of signaling and media transfer from A-party to B-party or C-party as indicated. The reverse direction is the opposite.

Audio stream from 10.42.44.160:47748 to 10.42.44.235:59200

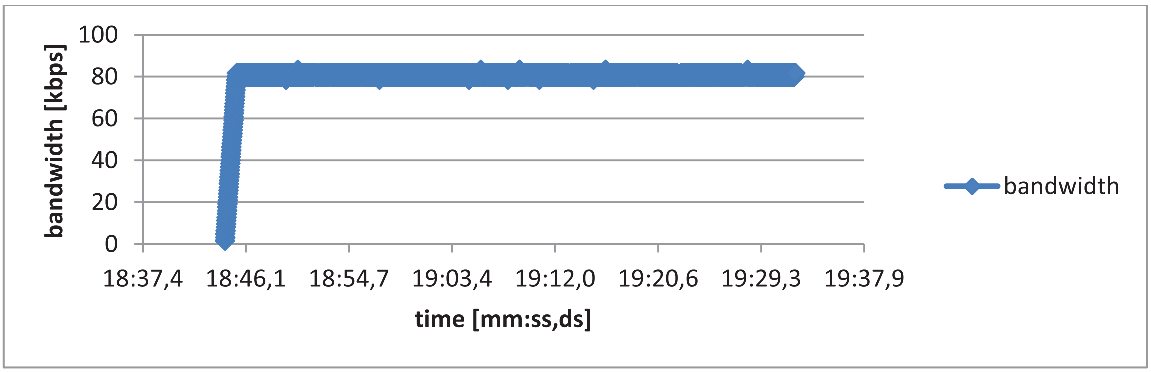


Figure B.1: Bandwidth usage by the audio stream in the forward direction

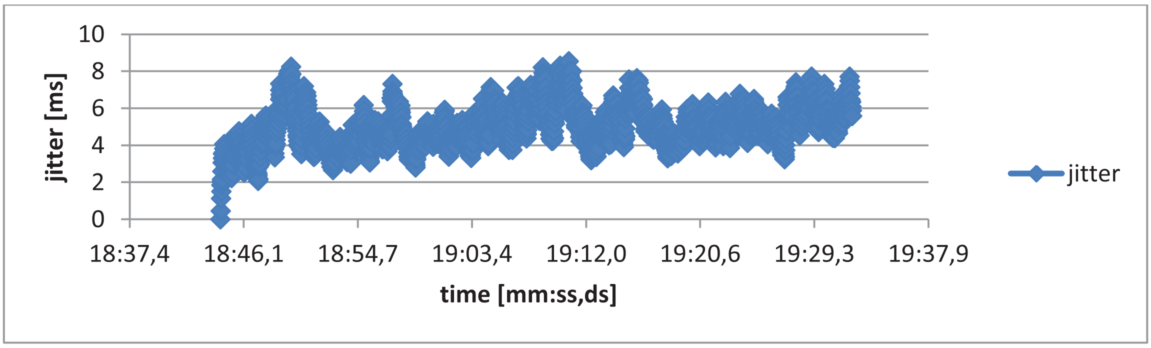


Figure B.2: Jitter experience by the audio stream in the forward direction

Figures B.1 and B.2 show what happens in the latter 60s of the call session. This is the part where a new call session is established with a third party-C while the former session with party-B is placed on hold. The call to this third party is a normal video/voice call established end-to-end between two parties hence similarity in network resource usage and a quite heavy jitter experience due to congestion and queuing at the port on which the user is connected to the hub.

Video stream from 10.42.44.160:34446 to 10.42.44.235:57362

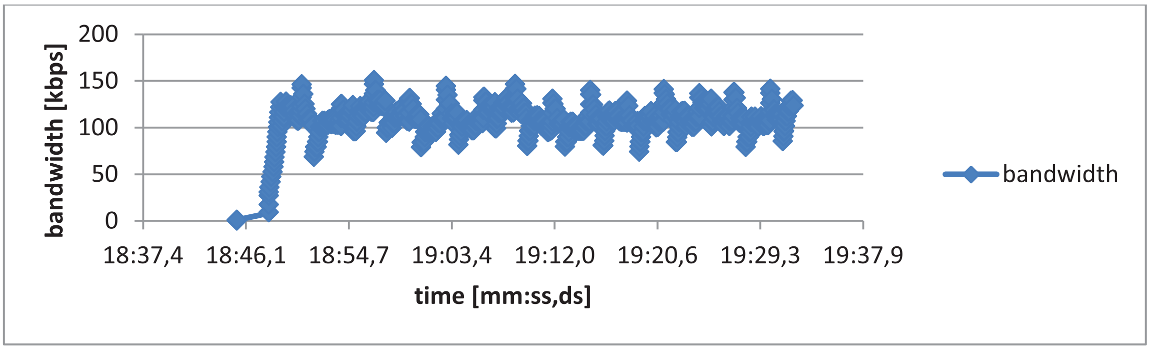


Figure B.3: Bandwidth usage by the video stream in the forward direction

Figure B.3 shows the bandwidth usage for the call session between party-A and third party-C. Heavy fluctuation is noticed with occasional peaks in the bandwidth size used. This is due to the video encoding format used. Those peaks correspond to when I-frame were transmitted.

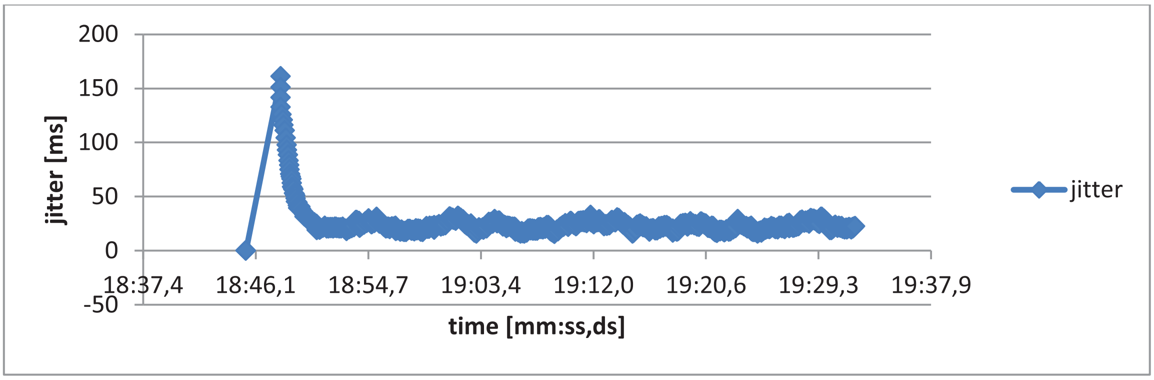


Figure B.4: Jitter experience by the video stream in the forward direction

Figure B.4 depicts the jitter variation during the call session for the video stream. At the beginning the jitter vale goes very high due to the fact that the first few RTP packets transmitter do not contain video data (RTP timestamp of zero).

RTCP audio report: 10.42.44.160 to 10.42.44.235 in the forward direction

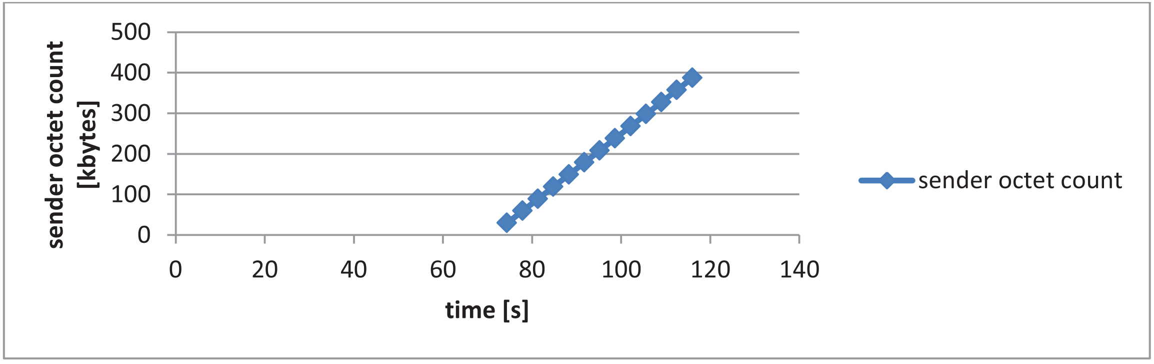


Figure B.5: RTCP report on Sender Octet Count

As clearly shown here, figure B.5 shows a steady rise in transmitted RTP octet count towards the third party-C for the latter 60s.

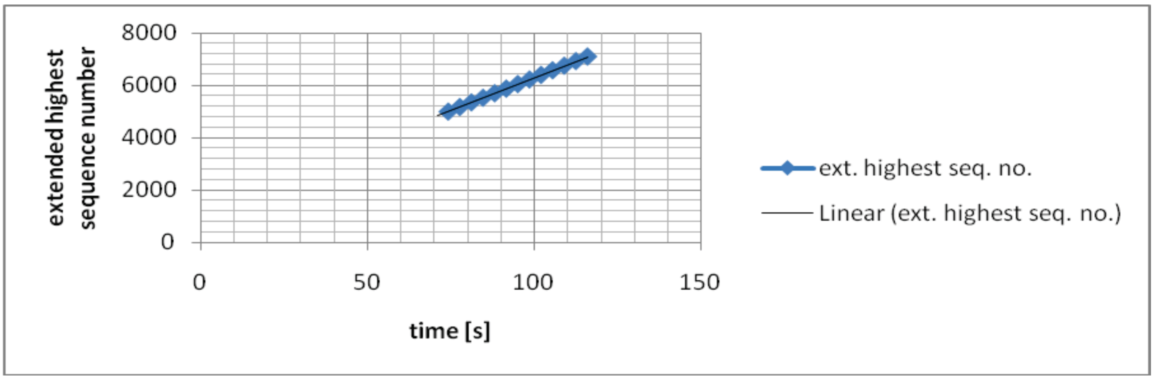


Figure B.6: RTCP report on Extended Highest Sequence Number

Figure B.6 corroborates figure B.5showing the increase in received RTP packets by the third party-C.

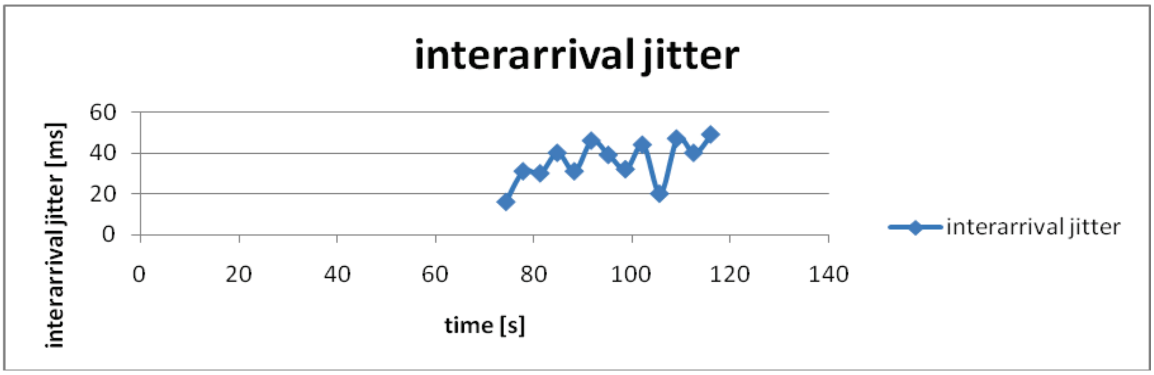


Figure B.7: RTCP report on the Interarrival jitter

Figure B.7 indicates delay variation experienced by the RTP packets as they traverse the network from the C-party to the A-party together with the congestion and queuing at the port interface on the hub to which the A-party is connected.

RTCP video report: 10.42.44.160 to 10.42.44.235 in the forward direction

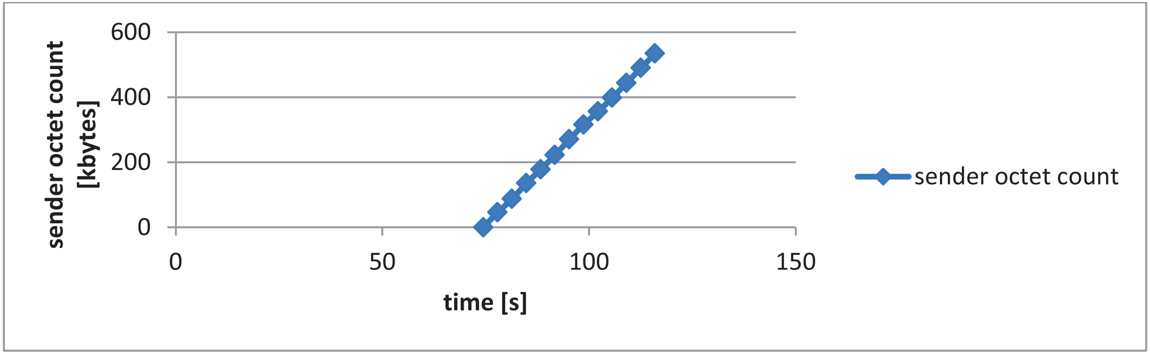


Figure B.8: RTCP report on the Sender Octet Count

Similar to the audio stream in figure B.5, figure B.8 shows a steady rise in RTP video packets.

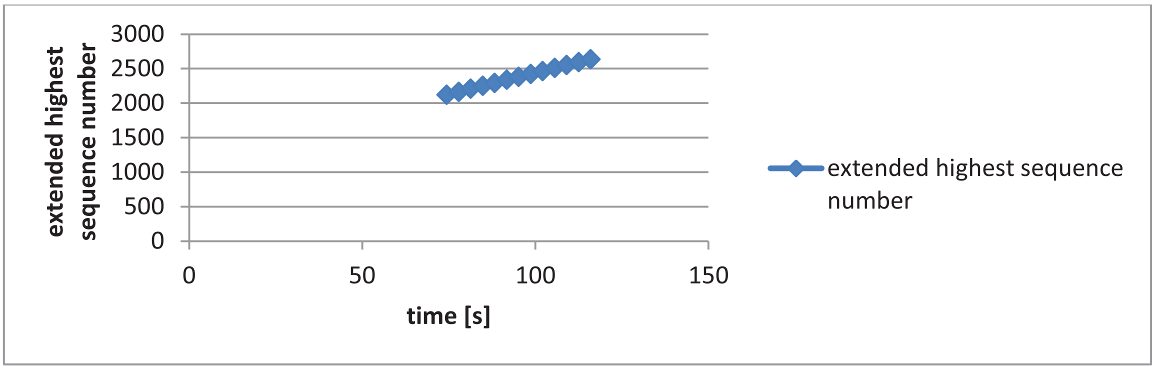


Figure B.9: RTCP report on the Extended Highest Sequence Number

Figure B.9 corroborates figure B.8 showing the increase in received RTP packets by the third party-C.

APPENDIX C. Plots of Test Case 4

Test case 4 involves the scenario involving three parties in a conference call. The A-party initiates and establishes a video session with the B-party. Once established, the call is put on hold in order for a third party, C-party to be invited to another session. Once this new session is established between the A-party and the C-party, the former session (A-B party session) put on hold is retrieved so that the three parties can now participate in the same video call session.

Included in this appendix is the analysis of the call session from A-party towards the third party (C-party) which includes the RTP and RTCP report analysis.

The following IP addresses were allocated to each of the participating parties:
10.42.44.161 to A-party, 10.42.44.235 to the B-party and 10.42.44.160 to the C-party.

Audio stream from 10.42.44.160:20300 to 10.42.44.235:57772

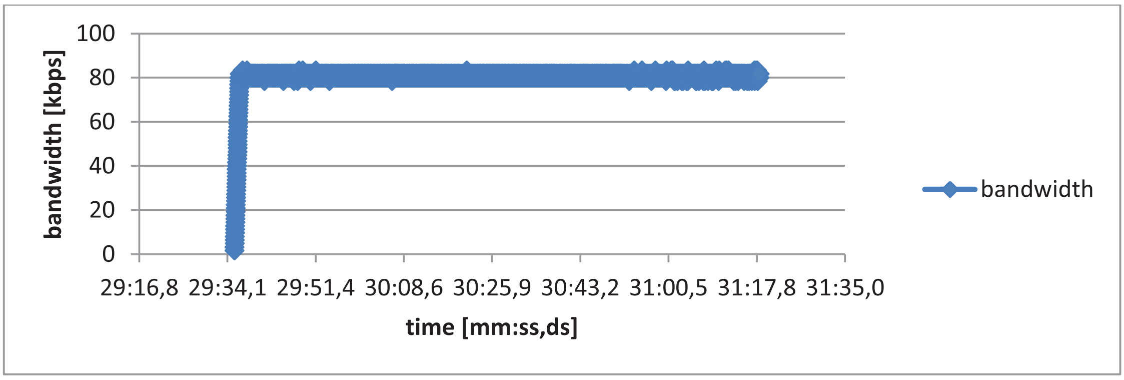


Figure C.1: Bandwidth usage by the audio stream in the forward direction

Just like it was explained in the definition of this call case, the call towards the third party is like a normal established video call, hence the similarity of the plot to that of test case 2. As depicted in the plot of figure C.1, the bandwidth usage remains fairly constant throughout the duration of the participation of the third party in the conference call.

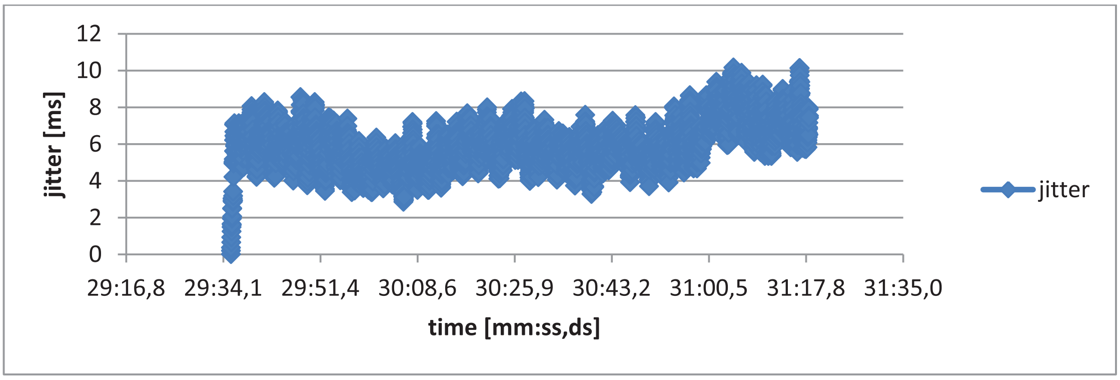


Figure C.2: Jitter experienced by the audio stream in the forward direction

Figure C.2 is similar to a normal video call. The plot shows the variation jitter as audio stream is being transferred between the third party and the A-party (call initiator). The variation in jitter is quite minimal.

Video stream from 10.42.44.160:26042 to 10.42.44.235:14014

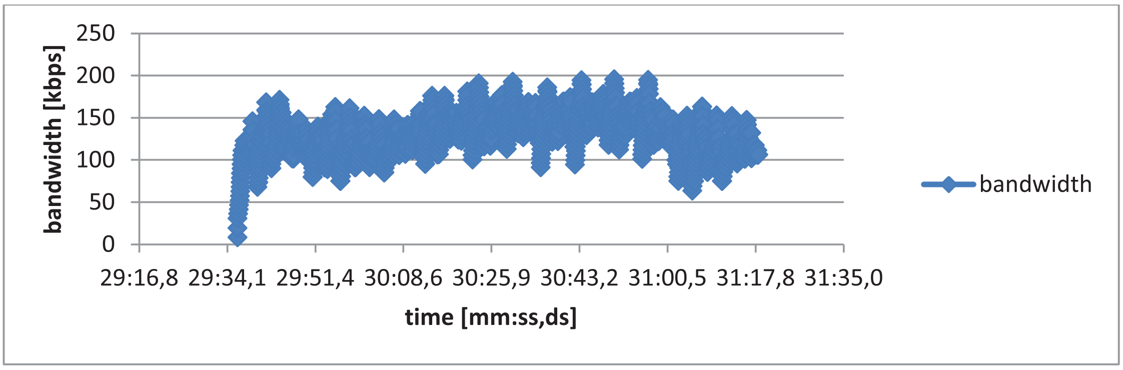


Figure C.3: Bandwidth usage by the video stream in the forward direction

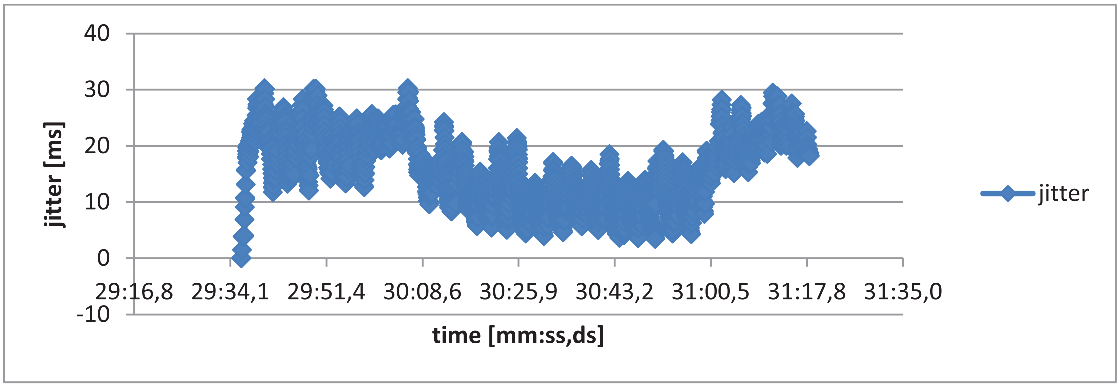


Figure C.4: Jitter experienced by the video stream in the forward direction

Figures C.3 and C.4 depict the bandwidth usage and jitter variation for the video stream in the reverse direction of the third (called) party which shows quite a heavy fluctuation.

RTCP audio report: 10.42.44.160 to 10.42.44.235 in the forward direction

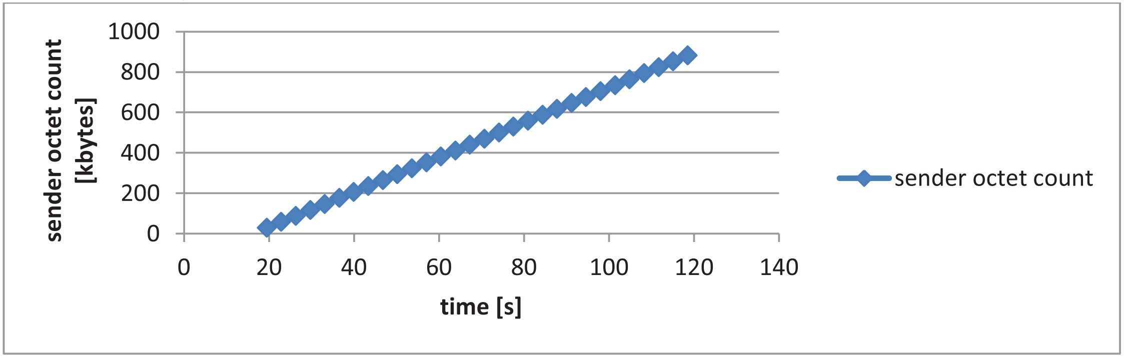


Figure C.5: RTCP report on the Sender Octet Count

Figure C.5 shows a steady rise in sender octet count showing there is no loss of RTP packets during the media exchange.

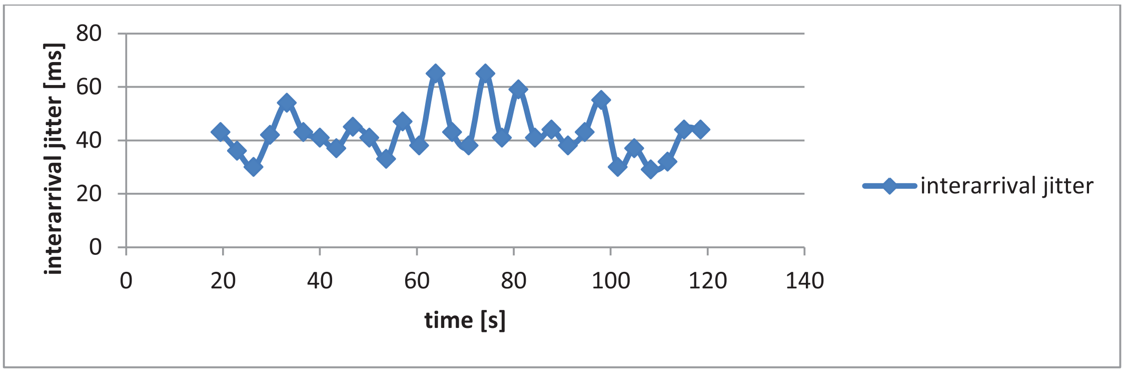


Figure C.6: RTCP report on Interarrival jitter

Figure C.6 shows a not too heavy fluctuation in the interarrival jitter. This jitter variation is similar to those experienced during the audio stream exchange associated with an end-to-end video call.

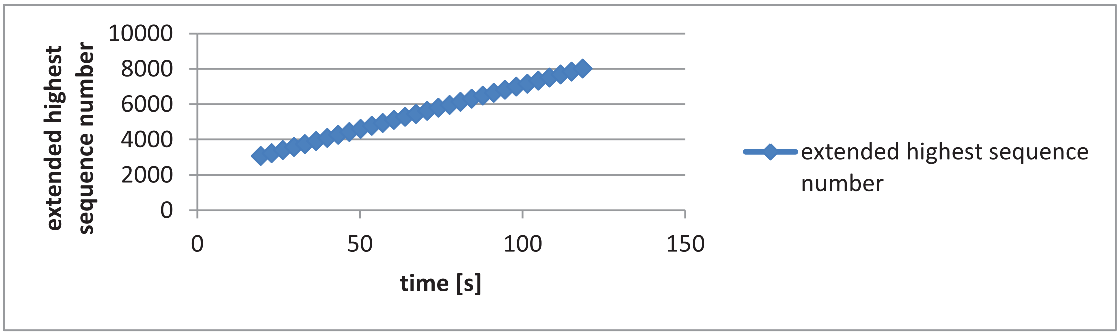


Figure C.7: RTCP report of the Extended Highest Sequence Number

The steady rise in the extended highest sequence number in the plot of figure C.7 depicts no loss in packet transmission between the peers involved in the call.

RTCP video from 10.42.44.160 to 10.42.44.235 in the forward direction

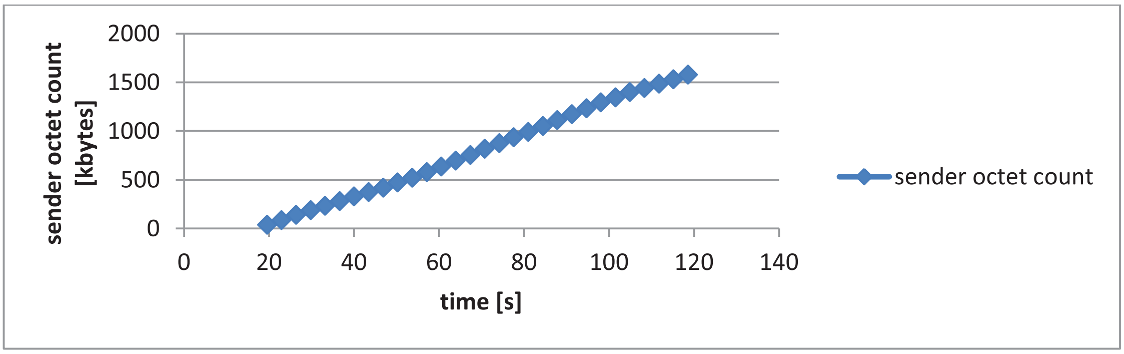


Figure C.8: RTCP report on the Sender Octet Count

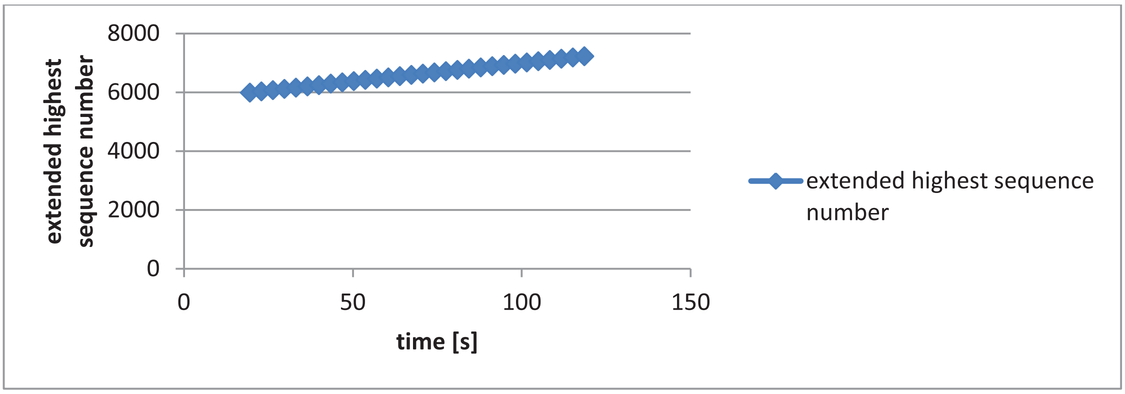


Figure C.9: RTCP report on the Extended Highest Sequence Number

Figures C.8 and C.9 are similar to the plots of a normal end-to-end video call between two parties. The steady increase in the sender octet count and extended highest sequence number respectively shows no packet loss was experienced during video stream exchange.

For the reverse direction of both audio and video streams in this section (between party-A and party-B), which are not shown, all the plots are similar to a normal end-to-end video call session between two end users.