Telephony RTP Stream Extractor from Ethernet capture file

BITS ZG628T: Dissertation

by

Chandra Prakash Singh

2014HT12924

Dissertation work carried out at

Plivo Communication Private Ltd, Bangalore

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)**

April 2017

Telephony RTP Stream Extractor from Ethernet capture file

BITS ZG628T: Dissertation

by

Chandra Prakash Singh

ID No. 2014HT12924

Dissertation work carried out at

Plivo Communication Private Ltd ,Bangalore

Submitted in partial fulfillment of M.Tech. Software Systems degree programme

Under the Supervision of
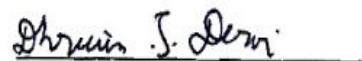Dhruvin J Desai,
Plivo Communication Pvt Ltd, Bangalore

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)**

April, 2017

CERTIFICATE

This is to certify that the Dissertation entitled **Telephony RTP Stream Extractor from Ethernet capture file** and submitted by **Chandra Prakash Singh** having ID-No. **2014HT12924** for the partial fulfillment of the requirements of **M.Tech. Software and System** degree of BITS, embodies the bonafide work done by him/her under my supervision.

Signature of the Supervisor

Place : Bangalore

Date : 10-Jan-2017

Dhruvin J Desai
Senior Voice Core Engineering Manager
Plivo Communication Private Ltd, Bangalore
Name, Designation & Organization &Location

Birla Institute of Technology & Science, Pilani

Work-Integrated Learning Programmes Division

Second Semester 2016-2017

BITS ZG628T: Dissertation

ABSTRACT

BITS ID No.   : 2014HT12924

NAME OF THE STUDENT   : Chandra Prakash Singh

EMAIL ADDRESS   : cps.ucer@gmail.com

STUDENT'S EMPLOYING   : Plivo Communication Private Ltd, Bangalore
ORGANIZATION & LOCATION

SUPERVISOR'S NAME   : Dhruvin J Desai

SUPERVISOR'S EMPLOYING   : Plivo Communication Private Ltd, Bangalore
ORGANIZATION & LOCATION

SUPERVISOR'S EMAIL ADDRESS: dhruvin@plivo.com

DISSERTATION TITLE   : Telephony RTP Extractor from Ethernet Capture File

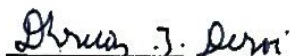_Chandra Prakash_
_____
Signature of the Student

Name: Chandra Prakash Singh

Date: 10-Jan-2017

Place: Bangalore

Dhruvin J. Desai
_____
Signature of the Supervisor

Name: Dhruvin J Desai

Date: 10-Jan-2017

Place: Bangalore

4

ACKNOWLEDGEMENTS

The satisfaction and euphoria that accompanies the successful completion of any task would be incomplete without mentoring the people who made it possible, because success is the epitome of hard work, perseverance, undeterred missionary zeal, steadfast determination and most of all "Encouraging Guidance".

I express my gratitude to my supervisor **Mr. Dhruvin J Desai** for providing me a means of attaining my most cherished goals.

I record my heart full of thanks and gratitude to my additional examiner **Mr. Mohit Tiwari** for providing me an opportunity to carry this project, along with purposeful guidance and moral support extended to me throughout the duration of the project work.

Chandra Prakash Singh

Table of contents

List of Figures

Chapter 1: VoIP Call Quality

1. Background

RTP, the real-time transport protocol, RTP provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services. RTP does not address resource reservation and does not guarantee quality-of-service for real-time services. RTP does not have a well known UDP port (although the IETF recommend ports 6970 to 6999). Instead, the ports are allocated dynamically and then signalled using a different protocol such as SIP(Session Initiation Protocol) . In SIP and other protocols a RTP session is described by SDP(Session Description Protocol), which is not really a protocol itself but rather a formalised way to describe a media session.

Wireshark, a network analysis tool formerly known as Ethereal, captures packets in real time and display them in human-readable format.

2. Playing VoIP calls

To access the VoIP calls analysis, use the menu entry "Telephony->VoIP Calls...". The current VoIP supported protocols are:

To play the RTP audio stream of one or multiple calls from the VoIP List, select them from the list and then press the "Player" button:
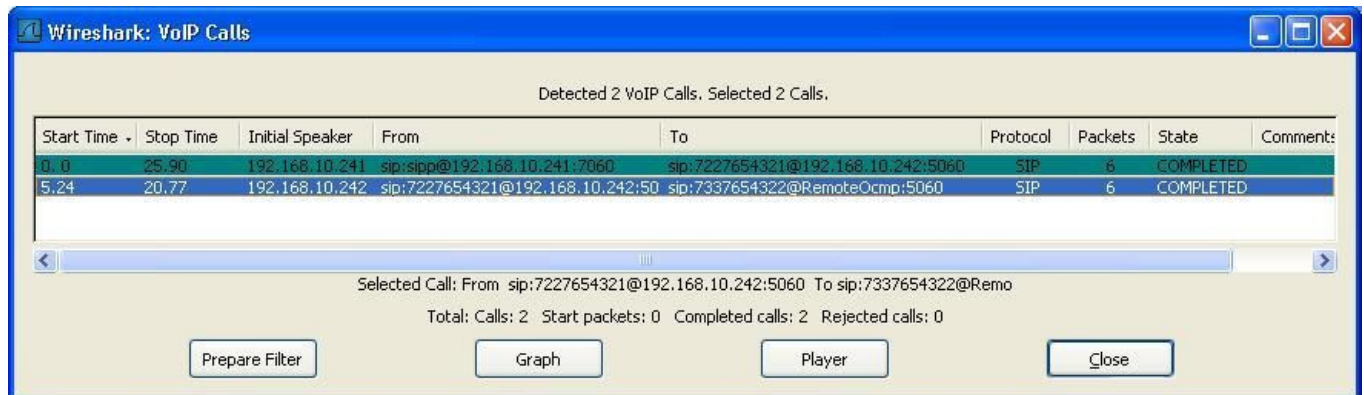
Figure 1: VoIP Calls

You can now see all RTP streams available for the calls that you selected:
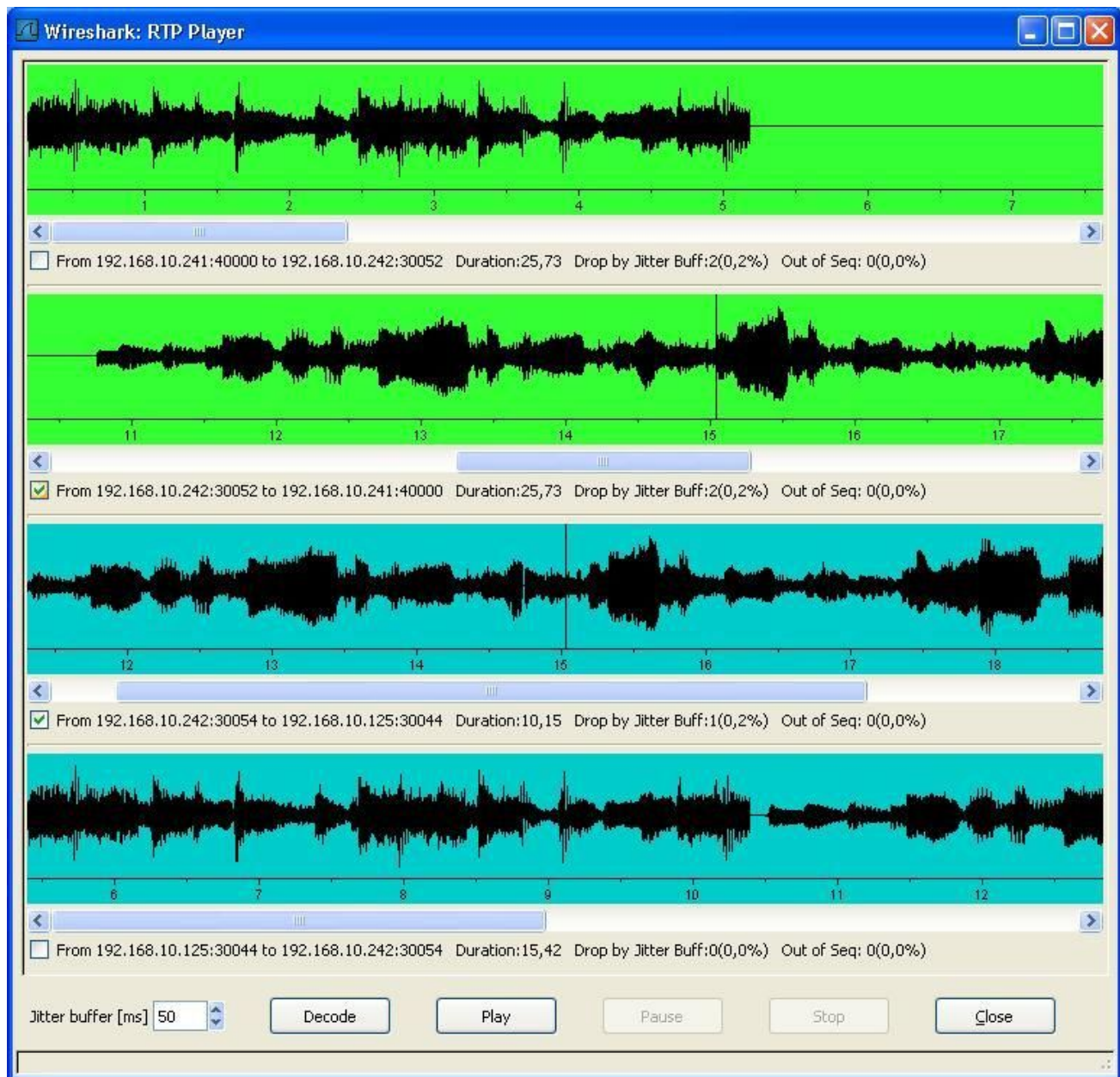
Figure 2: RTP Player

3. Problem Definition

For the moment, RTP Player feature works only for G711 A-Law and G711 u-Law RTP streams (other codecs not implemented). To Play other codec, we need to extract the RTP stream from the network capture file to verify the call quality. This step is manual and need a lot of effort in manual testing the audio call quality.

4. VOIP Call Quality solution

- Build, an effective tool for RTP Analysis for mostly used audio codecs like G711, G722 opus, AMR and AMR-WB.
- Provide the complete analysis of Audio RTP for each and every packet like Sequence number, SSRC, Payload, Jitter, Redundancy etc.
- Decoding of RTP payload
- Converting the RTP payload in playable format for manual verification
- Analysis of Audio Quality based on predicted MOS score for automated procedure.

Chapter 2: VoIP RTP Quality

1. About MOS score

RTP quality is important factor to validate the SIP telephony call(audio) quality hence we need to verify the RTP Metrics like Packet Loss, Jitter, latency and come up with a predictive MoS score after **extracting the RTP payload from the network capture file.** MOS measures subjective call quality for a call. MOS scores range from 1 for unacceptable to 5 for excellent.
VOIP calls often are in the 3.5 to 4.2 range.
Mean Opinion Score (MoS) / R-Factor is sometimes used to determine the requirements for QoS.
Quality of Service provides a guarantee of bandwidth and availability for VOIP call quality for
Delay, Jitter, Echo, Congestion, Packet loss and Out of Sequence packets
The MOS is highly subjective. One should not make decisions on a VoIP system based on the MOS alone. Other measurable parameters should be analyzed such as network delay, packet loss, jitter, and so on. As an alternative to the MOS, a different, less subjective rating has been introduced.

R-Factor is an alternative method of assessing call quality. Scaling from 0 to 120 as opposed to the limited scale of 1 to 5 makes R-Factor a somewhat more precise tool for measuring voice quality. R-Factor is calculated by evaluating user perceptions as well as the objective factors that affect the overall quality of a VoIP system, accounting for the Network R-factor and the User R-factor separately.

The following table demonstrates the effect of the MOS and R-Factor on the perceived call quality.

User Satisfaction Level	MOS	R-Factor
Maximum using G.711	4.4	93
Very satisfied	4.3-5.0	90-100
Satisfied	4.0-4.3	80-90
Some users satisfied	3.6-4.0	70-80
Many users dissatisfied	3.1-3.6	60-70
Nearly all users dissatisfied	2.6-3.1	50-60
Not recommended	1.0-2.6	Less than 50

2. What affects your VoIP MOS test score.

First, it is important to understand that MOS - Mean Opinion Score - is a relative scale and is built upon many factors which can affect voice quality. VoIP measurements are collected for after testing the one-way delay or the latency of the connection, packet loss with a metric to include the number of consecutive packets lost, and the amount jitter (difference in time it takes packets to arrive). Calculations then factor anR Factor that can be used to estimate a MOS score.

Propagation delay which is is the time required for a digital signal to travel end-to-end across the entire network. The greater the distance the greater the propagation delay. Additionally the data has to travel through network routers, switches and other devices like firewalls, each adding its own (transport) delay.

Packetization delay which is the time required to digitize the signal for the codec used for sending over the internet and decode it at the far end. A more compressed codec like G.729 has a higher packetization delay than a non-compressed codec like G.711 codec.

Jitter buffer is the delay introduced by the ATA device to hold one or more datagrams, to compensate in variations of arrival times. In VoIP a jitter buffer is an area where voice packets can be collected, stored, and then sent to the processor in more evenly spaced intervals. Variations in packet arrival time (jitter) is usually the result of network congestion or route changes. The jitter buffer, which is located at the receiving end device of the voice connection, intentionally delays the packets so that received voice is presented correctly with less distortion.

The following items can all affect call quality:
- Bandwidth
- Codec in use.
- Hardware
- Jitter
- Latency
- Packet Loss

3. Codecs- how they affect MOS- Mean Opinion Score

A non-compressed codec, G711, will give the best voice quality, as no compression and decompression is required. This in turn makes this G711 less susceptible to packet loss, which even in small amounts will quickly degrade voice quality. Other codecs which use compression techniques consume less bandwidth, a good thing, but this compression process in itself lowers the voice clarity and introduces a delay. Additionally, compressed codecs are even more susceptible to packet loss. In addition G711 will maintain better quality with more latent connections, whereas compressed codecs' voice quality can quickly degrade as latency increase, as the chart below reflects.

Figure3:Comparison of voice quality for G711 & G729 codecs as latency increases.

Chapter 4: Analysis of SIP and RTP details

For any voip call , we need to analyze the for SIP and RTP including RTPEvent (DTMF) to find out call details record. I am using tshark and lua programming for extracting the SIP and RTP details from the capture file.

Figure 4: SIP Summary

Figure 5: RTP Summary

SIP packet:-

```
<packet>
<timestamp>1487248464.977806000</timestamp>
<proto>UDP</proto>
<src_ip>52.205.63.210</src_ip>
<src_port>5080</src_port>
<dst_ip>172.31.18.204</dst_ip>
<dst_port>5060</dst_port>
<sip>
   <sip_call_id>1beb391e-5795-4a9d-b0e8-ccf636a7c842</sip_call_id>
   <sip_cseq_method>ACK</sip_cseq_method>
   <sip_status_code>nil</sip_status_code>
   <sip_request>ACK sip:35.154.101.16:5060;transport=UDP SIP/2.0</sip_request>
   <sip_reply>nil</sip_reply>
</sip>
</packet>
```

RTP packet:-

```
<packet>
<timestamp>1487248464.998570000</timestamp>
<proto>UDP</proto>
<src_ip>52.205.63.210</src_ip>
<src_port>24016</src_port>
<dst_ip>172.31.18.204</dst_ip>
<dst_port>6000</dst_port>
<rtp>
   <seq_no>63563</seq_no>
   <ssrc>0x76349b6b</ssrc>
   <codec>g711a</codec>
   <rtp_timestamp>160</rtp_timestamp>
</rtp>
</packet>
```

RTPEvent packet:-

```
<packet>
<timestamp>1487248470.898571000</timestamp>
<proto>UDP</proto>
<src_ip>52.205.63.210</src_ip>
<src_port>24016</src_port>
<dst_ip>172.31.18.204</dst_ip>
<dst_port>6000</dst_port>
<rtpevent>
   <rtpevent_id>1</rtpevent_id>
   <rtpevent_duration>160</rtpevent_duration>
   <rtpevent_end_of_event>0</rtpevent_end_of_event>
</rtpevent>
</packet>
```

Chapter 4: Reading and Decoding RTP Payload

Pyshark allows parsing from a capture file or a live capture, using all wireshark dissectors you have installed. Using pyshark module filter the UDP packet and decode them as RTP. Save the RTP payload in the raw audio file.

```python
rtp_list = []
cap = pyshark.FileCapture(pcapfile,display_filter='udp',decode_as={'udp.port==10000-65536': 'rtp'})
self.get_rtp_streams_info() // Retreive all the RTP stream in the pcap file)
for rtp in self.rtp_stream_report:
  raw_audio = open(rtp+'.raw', 'wb')
  for i in cap:
    try:
      rtp = i[3] // should be RTP payload or message
      if rtp.payload:
        rtp_list.append(rtp.payload.split(":"))
    except:
     pass
  for rtp_packet in rtp_list:
    packet = " ".join(rtp_packet)
    audio = bytearray.fromhex(packet)
    raw_audio.write(audio)
```

Chapter 5: Converting into wav/mp3 format

Using SoX(Sound Exchange) , I am converting the raw audio file into wav file format and mp3 format.SoX reads and writes audio files in most popular formats and can optionally apply effects to them. Even It can combine multiple input sources, synthesise audio, and, on many systems, act as a general purpose audio player or a multi-track audio recorder. It also has limited ability to split the input into multiple output files.

```python
for raw_file in raw_files:
    audio_file = '%s'%(raw_file.split('/')[-1][:-3] + 'wav')
    mp3_audio_file='%s'%(raw_file.split('/')[-1][:-3] + 'mp3')
    subprocess.call('sox -r 8k -e a-law  %s -e signed %s'%(raw_file, audio_file), shell = True)
    subprocess.call('sox -t wav %s -t wav -e signed-integer -c 1 -r 8000 - rate | lame -b 16 - %s ' %
(audio_file,mp3_audio_file), shell=True)
```

Example:-
sox -r8000 -c1 -t ul out/rtp.0.0.raw -t wav out/0.wav > /dev/null 2>&1
sox -r8000 -c1 -v 5 -t ul out/rtp.0.1.raw -t wav out/1.wav > /dev/null 2>&1
sox -m out/0.wav out/1.wav out/call.wav > /dev/null 2>&1
sox -t wav out/call.wav -t wav -e signed-integer -c 1 -r 8000 - rate | lame -b 16 - out/call.mp3 > /dev/null 2>&1

Chapter 6:Quality Checking Using PESQ MOS

Quality Checking Using Mean Opinion Score (MOS) ,we need to check the quality of audio RTP flow. There are many alternatives available to check the quality of RTP flow, i.e., we can measure throughput, average latency, jitter, etc. One standard measurement is MOS (Mean Opinion Score). MOS score value ranges between 1 to 5, with 5 being excellent quality and 1 being the worst quality. MOS score depends on the various parameters, i.e., link speed, codec used, jitter, round trip time (RTT), packet loss percentage etc.

PESQ is a full-reference algorithm and analyzes the speech signal sample-by-sample after a temporal alignment of corresponding excerpts of reference and test signal. PESQ can be applied to provide an end-to-end (E2E) quality assessment for a network, or characterize individual network components.

PESQ Output:-

Reading reference file 0x76349B6B_52.205.63.210_172.31.18.204.wav...done.
Reading degraded file 0x7698771D_52.205.63.210_172.31.18.204.wav...done.
Level normalization...
IRS filtering...
Variable delay compensation...
Acoustic model processing...

P.862 Prediction (Raw MOS, MOS-LQO): = 4.500   4.549

Chapter 7: RTP Analyser Server

A RTP Analyser Server provides Web GUI for your scripts and remote execution facility.We will be able to access your scripts via web-browser and execute them. Everything will run on your machine, so users shouldn't care about setting up an environment or working via ssh.
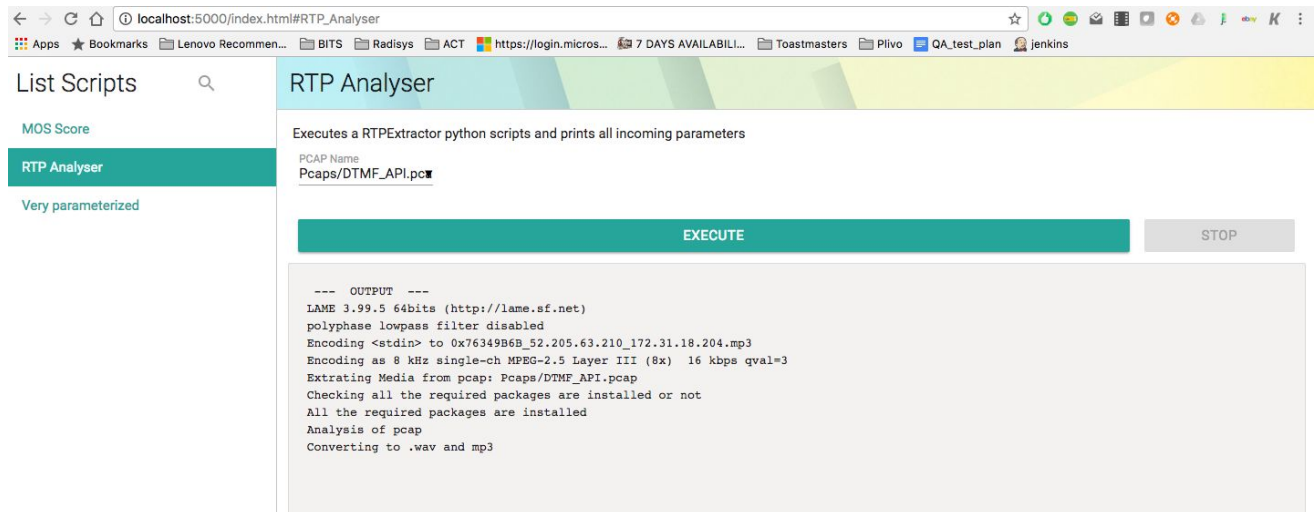
Chapter 8: MOS Server

A MOS Server provides Web GUI for your MOS score calculation facility.We will be see MOS score in web-browser after execute it. Everything will run on your remote machine, so users shouldn't care about setting up an environment or working via ssh.
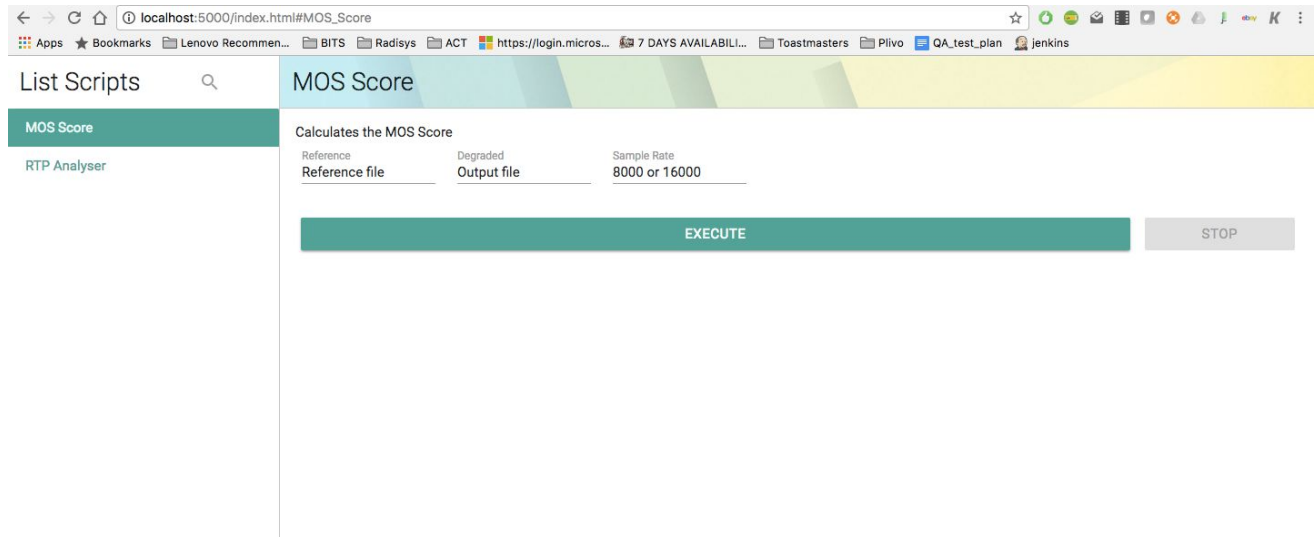
Chapter 8: Audio Server

A Audio Server provides Web GUI to listen the extracted audio server.You will be hear the audio file from the web-browser. Everything will run on your remote machine, so users shouldn't care about setting up an environment or working via ssh.

Summary

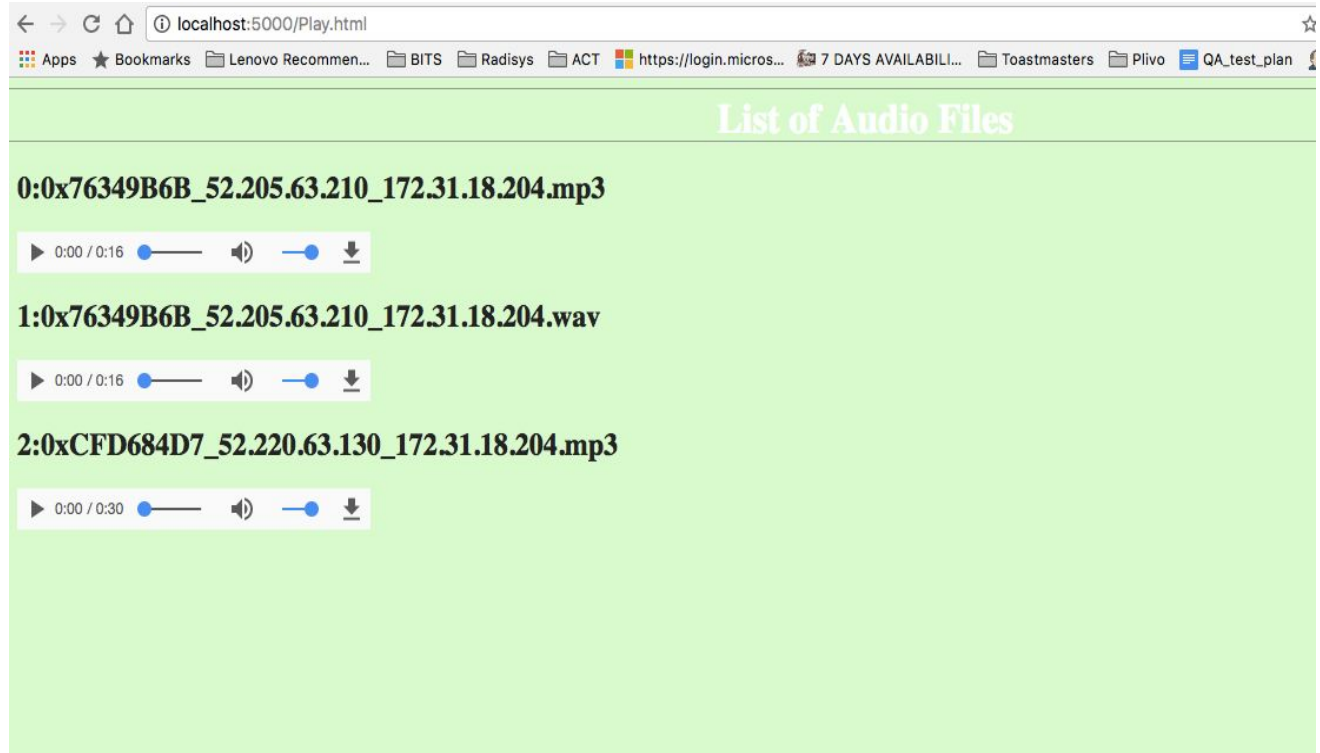To conclude, I have summarize the way to extract the audio from ethernet capture pcap file. As a first step, we have extracted the SIP summary and RTP summary from the pcap file using the tshark. To calculate the MOS score , we need to two reference and degraded wav sample. To get the degraded wave sample below are the steps performed:-

1. Reading the ethernet capture pcap file
2. Decoding of all the RTP session from the pcap file
3. Saving on RTP payload into raw file.
4. Converting into wav and mp3
5. Audio server to listen the extracted audio files.
6. Quality checking using PESQ-MOS

Details of work done till date and Plan of work yet to be done:
The items which are marked √ are completed work till date and items which are marked X are yet to be done.

Table: Details Of Work Done Till Date And Plan Of Work Yet To Be Done

SL No	Description of Work	Proposed date of completion	Current Status
1.	SIP and RTP Analysis	04-02-2017	√
2.	Decoding of all RTP session	12-02-2017	√
3.	Saving RTP session into raw file	19-02-2017	√
4.	Converting into wav and mp3	27-02-2017	√
5.	Quality checking using PESQ-MOS	05-03-2017	√
6.	Testing and audio server to listen the extracted audio files	17-03-2017	√
7.	Testing and final report preparation	26-03-2017	√

Conclusions and Recommendations

The design of Call Quality Framework is discussed in detail in this report. Based on this design the components are implemented, tested and validated. Addition of SIP statistics and RTP Events are requested during the our discussion with additional examiner.

Directions of future work

As of decoding of RTP payload, works for G711 audio codec only. I will try to enhance this framework for various audio codecs like opus, amr, amr-wb evs, g722, g729 etc,

References

1. **Informative Reference**

   a. **PCAP**
      https://en.wikipedia.org/wiki/Pcap
   b. **Tshark**
      https://www.wireshark.org/docs/man-pages/tshark.html
   c. **RTP**
      https://tools.ietf.org/html/rfc3550
   d. **G711**
      https://tools.ietf.org/html/rfc5391
   e. **OPUS**
      https://tools.ietf.org/html/rfc6716

2. **PESQ**

   ESQ results principally model mean opinion scores (MOS) that cover a scale from 1 (bad) to 5 (excellent). A mapping function to MOS-LQO is outlined under P.862.1.
   https://en.wikipedia.org/wiki/PESQ

3. **Pyshark**

   Python wrapper for tshark, allowing python packet parsing using wireshark dissectors
   https://pypi.python.org/pypi/pyshark

4. **SOX(Sound Exchange)**

   SoX is a cross-platform (Windows, Linux, MacOS X, etc.) command line utility that can convert various formats of computer audio files in to other formats. It can also apply various effects to these sound files, and, as an added bonus, SoX can play and record audio files on most platforms.
   http://sox.sourceforge.net/
   http://sox.sourceforge.net/sox.html

5. **Mean Opinion Score (MOS).**

   https://en.wikipedia.org/wiki/Mean opinion score

6. **Mean Opinion Score.**

   http://www.voipmechanic.com/mos-mean-opinion-score.htm.

7. **Mean Opinion Score calculation in Pingplotter.**

   https://www.pingman.com/kb/article/how-is-mos-calculated-inpingplotter-pro-50.html.

8. **lame**

   LAME is a high quality MPEG Audio Layer III (MP3) encoder licensed under the LGPL.
   http://lame.sourceforge.net/about.php

APPENDIX

1. Acronyms

Acronyms	Description
UDP	User Datagram Protocol
TCP	Transmission Control Protocol
SIP	Session Initiation Protocol
RTP	Real Time Protocol
AMR	Adaptive Multi-Rate (AMR or AMR-NB or GSM-AMR) codec
MOS	Mean Opinion score using PESQ
PESQ	Perceptual Evaluation of Speech Quality

Checklist of items for the Final Dissertation Report

Checklist of items for the Final Dissertation Report

This checklist is to be duly completed, verified and signed by the student.

1.	Is the final report neatly formatted with all the elements required for a technical Report?	Yes
2.	Is the Cover page in proper format as given in Annexure A?	Yes
3.	Is the Title page (Inner cover page) in proper format?	Yes
4.	(a) Is the Certificate from the Supervisor in proper format?	Yes
	(b) Has it been signed by the Supervisor?	Yes
5.	Is the Abstract included in the report properly written within one page?	Yes
	Have the technical keywords been specified properly?	Yes
6.	Is the title of your report appropriate? The title should be adequately descriptive, precise and must reflect scope of the actual work done. Uncommon abbreviations / Acronyms should not be used in the title	Yes
7.	Have you included the List of abbreviations / Acronyms?	Yes
8.	Does the Report contain a summary of the literature survey?	Yes
9.	Does the Table of Contents include page numbers?	Yes
	(i). Are the Pages numbered properly? (Ch. 1 should start on Page # 1)	Yes
	(ii). Are the Figures numbered properly? (Figure Numbers and Figure Titles should be at the bottom of the figures)	Yes
	(iii). Are the Tables numbered properly? (Table Numbers and Table Titles should be at the top of the tables)	Yes
	(iv). Are the Captions for the Figures and Tables proper?	Yes
	(v). Are the Appendices numbered properly? Are their titles appropriate	Yes
10.	Is the conclusion of the Report based on discussion of the work?	Yes
11.	Are References or Bibliography given at the end of the Report?	Yes
	Have the References been cited properly inside the text of the Report?	Yes
	Are all the references cited in the body of the report	Yes
12.	Is the report format and content according to the guidelines? The report should not be a mere printout of a Power Point Presentation, or a user manual. Source code of software need not be included in the report.	Yes

Declaration by Student:
I certify that I have properly verified all the items in this checklist and ensure that the report is in proper format as specified in the course handout.

Place: Bangalore

Date: 27-03-2017

Signature of the Student

Name: Chandra Prakash Singh

ID No.: 2014 HT12924