CMP-6009B Networks:
Design, Implementation & Evaluation of a VoIP Communication System

100022010 & 100017267

February 29, 2016

Contents

1 Introduction

For the coursework we have developed a VoIP system for four network channel conditions, which have different characteristics in terms of packet loss, burst length and packet delivery order. Datagram Socket 1 has ideal network conditions with no packet losses and it guarantees in order delivery. As for Datagram Sockets 2, 3 and 4, they simulate non-ideal channel conditions, i.e. packets are sometimes lost or corrupted. Therefore they have to be analysed first to identify the channels characteristics and compensation methods, that should be applied to increase the quality of speech.

2 Network Analysis

When analysing each Datagram Socket, we characterised each into five measurable qualities.

- Packet efficiency, the rate of packet retrieval and packet loss.
- Delay, the rate of delay between packets.
- Burst lengths, the length of each packet loss.
- Bit rate, the number of bits per second that can be transmitted along a digital network.
- Quality of speech, this is evaluated using the Perceptual Evaluation of Speech Quality (PESQ) measurement scope and using an opinion poll from 10 people measuring intelligibility scored 0 to 4.5.

In order to identify problems with each datagram socket a sequence number and timestamp was added to the header for each packet and displayed on the receiver side. The timestamps registered the time the packet was created, and the time the packet was played.

When displaying the returning sequences numbers and timestamps it was concluded that Datagram Socket 2 delivers packets in the order they were sent, whereas Datagram Socket 3 does not guarantee in-order delivery/delivers packets out of order.

Also it was provided that Datagram Socket 2 loses 24.9% of packets, whereas Datagram Socket 3 does not deliver 16.4% of packets. Delay in receiving Datagram Sockets 1, 2 & 4 were unnoticeable, however the delay for Datagram Socket 3 was high with an average delay time of 38.565ms per packet.

For Datagram Socket 4 packets were not lost but were corrupted during receiving. An interleaver could resolve both out of order delivery and shorten long bursts of packet losses for Datagram Sockets 2 & 3, but then it is important to estimate the optimal dimension of the interleaver (spread) to avoid long delay.

The following table summarises the characteristics of each channel.

Summary of Network Analysis				
	Datagram Socket 1	Datagram Socket 2	Datagram Socket 3	Datagram Socket 4
Packets Received (2000)	**2000 (100%)**	**1502 (75.1%)**	**1672 (83.6%)**	**2000 (100%)**
Packets Lost	**0 (0%)**	**498 (24.9%)**	**328 (16.4%)**	**0 (0%)**
Average Burst Length	**0**	**5**	**3**	**0**
Total Delay Time	**686ms**	**551ms**	**64480ms**	**798ms**
Average Delay Time	**0.343ms**	**0.367ms**	**38.565ms**	**0.399ms**
Bit Rate	**134kBits/s**	**134kBits/s**	**134kBits/s**	**134kBits/s**
PESQ Score	**4.5**	**1.5923**	**1.2113**	**1.262**
Opinion Score	**4.5**	**2.5**	**1**	**3.5**

Table 1: Network Analysis for Datagram Sockets

It was found that using live audio recordings gave inconsistent PESQ scores due to the background noise or inconsistent speech patterns. Instead of recording in an uncontrolled environment, a sample of spoken word from a podcast that was consistent and had no background noise or audio clips had been recorded in ideal recording settings so there was consistency throughout conducting each different compensation technique for each Datagram Socket.

2.1 Datagram Socket 1

Packet Efficiency

Datagram Socket 1 was a perfect audio quality scenario where 100% of the packets were received. The result of that was an unnoticeable amount of delay at 0.343ms and no packet loss.

Quality of Speech

We measured the quality of speech both through Perceptual Evaluation of Speech Quality (PESQ) and through opinion polls. Both the opinion score and the PESQ score resulted in 4.5.

2.2 Datagram Socket 2

Packet Efficiency

Datagram Socket 2 was an example scenario where some packets would get lost. The delivered packets were in the order they were sent, but the packet loss rate was at 24.9%. The delay in receiving packets from the sender was a small amount of 0.367ms that was unnoticeable.

Burst Lengths

Due to the high rate of packet loss for Datagram Socket 2 there was an average burst length of 5.

Quality of Speech

Because of the packet loss the audio had missing packets of speech therefore the intelligibility was not very high on the opinion scores. This resulted in an average opinion score of 2.5 and a PESQ score of 1.5923.

2.3 Datagram Socket 3

Packet Efficiency
Datagram Socket 3 was an example scenario where the delivery of packets was not guaranteed to be in-order. Additionally, there was a packet loss rate of 16.4% to consider.

Delay
Due to the scenario where packets were returned out of order, the delay time in receiving packets increased substantially. The average delay time being 38.565ms. This created a noticeable delay when listening to audio.

Burst Lengths
Because of the rate of packet loss for Datagram Socket 3 there was an average burst length of 3.

Quality of Speech
The quality of speech for Datagram Socket 3 was the worst overall. The audio would be noticeably out of order, delayed and sometimes audio was lost due to packet loss. The opinion score of 1 and PESQ score of 1.2113 reflects that.

2.4 Datagram Socket 4

Packet Efficiency
Datagram Socket 4 was an example of a scenario where packets would get corrupted during receiving. The audio would play loud interrupting click sounds when the resulting packets got corrupting. The rate of packet corruption was 2.55%. Despite this the result meant an unnoticeable amount of delay at 0.399ms and no packet loss.

Quality of Speech
Although there were loud interrupting click sounds, the intelligibility of the audio was high due to 0% packet loss. Therefore the average opinion score was 3.5. However, due to the corruption of packets the PESQ score was low at 1.262.

3 VoIP System Design & Implementation

All of the systems implemented were done using threading and a full duplex connection across two computers. This was because it was essential that the system had the ability to send and receive/play packets concurrently.

Upon analysis, receiver-based and sender-based compensation methods were implemented to conceal/compensate packet loss and to increase sound quality. Originally, splicing was considered as a method of compensation, however when investigating the effect it would have on error correction the results showed that the quality of audio produced was un-intelligible and to make the quality of audio high, significant delay would have been needed to return the audio to normal speed and at an intelligible quantity. Due to the behavioural characteristics of the Datagram Sockets it was not decided as a method of compensation.

Table 2 shows all the compensation methods implemented and to which Datagram Sockets they were applied to. The implementation of Datagram Socket 1 was straight forward as this was a scenario where no error correction was required.

3.1 Implementation of Compensation Methods

3.1.1 Fill-In with Silence & Repetition

Fill-in compensation works by placing an empty sound frame where there has been a packet lost. Throughout transmission packet loss and burst lengths were kept track of. Whenever there was a packet loss, a 32ms frame of silence was played. Due to bursty packet loss, repeating silence for long periods caused delay. Therefore a maximum of 4 silence frames were played if the bursts were over 4 consecutive packets.

The implementation of repetition followed the exact same structure as the previous compensation method. On playing a packets audio, it was saved in case of packet loss on the next delivery of packets. Where previously silence was played, the previous sound frame was played instead. Putting a limit to the number of repeats that could occur works well for this method, as it stops speech becoming distorted due to excessive repetition.

Both these compensation methods where good for sockets 2 and 3 because both had long bursts of lost packets. Although the sound was not perfect, these methods are still more effective given the conditions than for example using splicing.

3.1.2 Block Interleaving

It was determined from analysis that Datagram Socket 3 delivered packets out of order. Block interleaving was used to solve this issue. In addition to solving long bursts of lost packets, it also helps to deliver packets in the correct order. At the sender, packets were grouped into a 2 dimensional ($d * d$) array, where d is the dimension of the interleaver. Before sending the packets, the 2D array was rotated by 90 degrees anti-clockwise. As a result packets were sent in a non-consecutive order.The receiver sorted the packets into their correct order by looking at their sequence numbers.

Interleaving is a form of sender-based compensation making receiver based compensation much more effective. However this method introduces some delay due to the time that is required to jumble the packets at the sender, and sort at the receiver before playing audio. The size of the block also affects delay times. For this reason different dimensions were tested to determine which was the most appropriate size for the socket. After investigating we decided the optimal dimension was *3 by 3* as it produced better quality audio as well as reasonable delay times. This was the case for both Datagram Socket 2 and Datagram Socket 3. *See 4. Evaluation Section for details.*

3.1.2.1 Interleaving with Fill-In with Silence or Repetition The interleaver was tested with two received based compensation methods, fill-in with silence and repetition. As before, blocks of silence or previous packet audio are played where packets were lost. However this time bursts are much shorter as a result of interleaving making these methods more effective than using them alone.

3.1.3 Cyclic Redundancy Check (CRC)

Cyclic redundancy check was implemented to detect corruption of packets. The sender applies a checksum value to the header of the packet, which is the remainder of a polynomial division of the audio data. At the receiver, the CRC calculation is done on the audio received and compared to the checksum value attached to the header. If these values do not match, a corruption has occurred.

3.1.3.1 CRC with Repetition & Fill-In with Silence CRC was very successful at identifying corrupt received packets. With this information receiver-based compensation was applied. Both replacing corrupt packets with repetition or silence worked well because the bursts of corruption were very low. This meant that the corruption became almost unnoticeable using the mentioned concealment methods.

4 Evaluation

Each socket was tested using all appropriate compensation methods for that channel and evaluated using the same parameters as for the basic network analysis. A table for each socket has been produced showing the effects of each compensation method on the sockets. A combination of parameters were used to identify the most suitable compensation method, quality of service being the most important.

4.1 Datagram Socket 2

Socket 2	Packets Received	Packets Lost	Average Burst Length	Total Delay Time	Average Delay Time	Bit Rate	PESQ Score	Opinion Score
Basic	1502	498	5	551ms	0.367ms	134kBits/s	1.5923	2.5
Fill-in w/ Silence	1502	498	5	620ms	0.413ms	134kBits/s	1.726	2.5
Repetition	1502	498	5	672ms	0.447ms	134kBits/s	2.8109	3.5
Interleaving	1506	494	1	752517ms	560.743ms	71kBits/s	2.6716	1
Interleaving w/ Repetition	1506	494	1	751228ms	559.782ms	71kBits/s	2.184	2
Interleaving & Fill-in w/ Silence	1506	494	2	749916ms	558.805ms	71kBits/s	1.6212	1.5

The best compensation method to use for a VoIP system over Datagram Socket 2 was repetition. This method did not introduce much more delay than the original system but significantly improved the quality of sound. Datagram Socket 2 had high packet loss which was quite bursty. This was improved by using interleaving, interleaving with repetition and by filling in with empty packets.

Although the quality of sound scored high on the PESQ scoring system, it was the lowest amongst the opinion scores. Interleaving also introduces delay which depreciated the quality of sound. To further improve the quality of sound, a combination of interleaving and repetition would have needed slight adjustments to reduce the delay and adjustments to make sure packets were transmitted in sequence.

4.2 Datagram Socket 3

Socket 3								
	Packets Received	Packets Lost	Average Burst Length	Total Delay Time	Average Delay Time	Bit Rate	PESQ Score	Opinion Score
Basic	1672	328	3	64480ms	38.565ms	134kBits/s	1.2113	1
Fill-in w/ Silence	1672	328	3	64656ms	38.669ms	134kBits/s	2.437	2
Repetition	1672	328	3	64735ms	38.711ms	134kBits/s	1.156	3.5
Interleaving	1677	323	2	791216ms	558.374ms	71kBits/s	1.5329	2
Interleaving w/ Repetition	1677	323	2	788882ms	556.727ms	71kBits/s	1.3208	2.5
Interleaving & Fill-in w/ Silence	1677	323	2	787593ms	555.817ms	71kBits/s	1.4634	2

Initially it was believed that the best compensation method for Datagram Socket 3 would be a variation of interleaving. This was because Datagram Socket 3 was quite bursty and had packets arrive out of order. Upon analysis it was found that interleaving did ensure in-order delivery but the delay was reasonably high. This causes the quality of sound to drop considerably, and so these systems were not favoured.

Different interleaver dimensions were tested. The higher the dimension, the more delay was caused. When extra receiver based compensation was added, such as repetition, even more delay was introduced. This caused the quality of audio to be really low and un-intelligible to people.

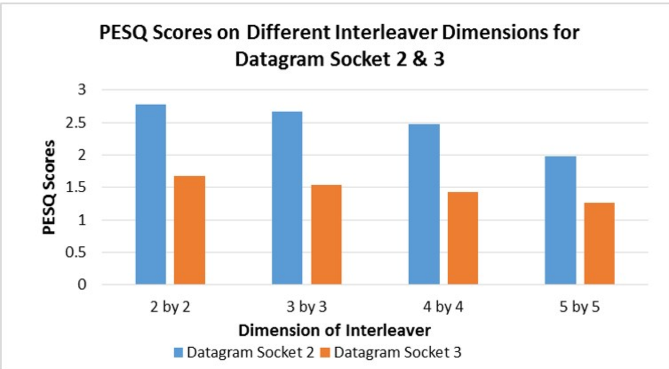
Listening to the sound alone, using repetition was the method which made the audio more intelligible. Even though this method scored low on the PESQ scoring system, opinion scores evaluated this method to be the best one. Reflecting back upon this decision, further improvements could have been made to the interleaving as packets were sometimes transmitted out of sequence. Although the opinion score rated repetition a 3.5, it could have been higher if adjustments to the interleaver were made to work alongside the repetition.

4.3 Datagram Socket 4

Socket 4								
	Packets Received	Packets Lost	Average Burst Length	Total Delay Time	Average Delay Time	Bit Rate	PESQ Score	Opinion Score
Basic	2000	0	0	798ms	0.399ms	134kBits/s	1.262	3.5
Fill-in w/ Silence	2000	0	0	883ms	0.4415ms	134kBits/s	0.9127	3
CRC & Repetition	2000	0	0	605ms	0.336ms	134kBits/s	3.0413	4.5
CRC & Fill-in w/ Silence	2000	0	0	638ms	0.358ms	134kBits/s	1.8504	4

From the analysis it was determined that this socket was occasionally corrupting packets. To solve this problem cyclic redundancy checks were applied to identify corruption of packets. This worked really well and solved this problem. Repetition and filling in with empty packets was used to conceal the corruption. Considering both concealment methods worked well, repetition made the audio almost perfect. This was due to the corrupted packets being spaced out. Repetition scored both high on PESQ and opinion scores.

Even though the quality of the compensated audio was very high, it could be further improved. Measuring the bursts of corrupted packets would allow the analysis of other methods to see which could have been applied. If corruption bursts were high, a small interleaver could have been applied to disperse the corrupted packets.



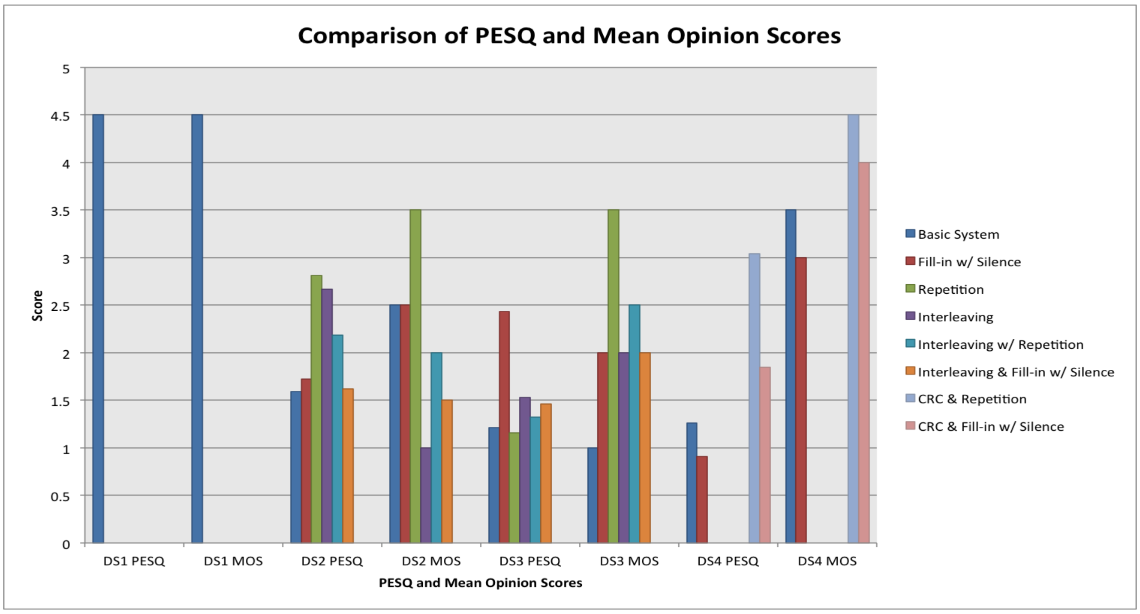
Graph 2: PESQ Scores on Different Interleaver Dimensions for Datagram Socket 2 & 3

4.4 Overview of Results

Testing Different Interleaver Dimensions PESQ Scores

Testing interleaver dimensions was done on Datagram Sockets 2 and 3 since they had different conditions. The smaller the dimensions the higher quality the sound. Having a large interleaver block, even though it reduced bursts introduced more delay to the audio. It can be seen in Graph 2 that the scores drop as the dimensions increase for both of the Datagram Sockets.

Comparison of PESQ & Mean Opinion Scores



Graph 1: Comparison of PESQ and Opinion Scores

The comparison of PESQ and opinion scores showed that in general the improvement experiments conducted improved the overall quality of audio and the intelligibility. It is clear that the PESQ score do not always reflect the opinion scores. Repetition is shown to reflect the overall highest opinion score both for Datagram Socket 2 & 3, which is why it was chosen as our best compensation method for these channels. Implementing a combination of CRC & repetition showed as the highest rated both in PESQ and opinion score, which is why it was chosen as our best compensation method for Datagram Socket 4.

5 Conclusion

It can be concluded that repetition proved to be the best compensation method to implement across all the channels. When compensating for packet loss and corruption, the implementation that filled in with silence removed splicing and jitters, however it meant that large amount of silence occurred which in the end sounded worse. Due to the average burst lengths of packet loss not being too high, it meant that repetition concealed the losses well across all the Datagram Sockets.

PESQ scores were not always consistent when recording live audio. Additionally they did not reflect the quality of the audio in terms of its intelligibility. For this reason opinion scores were taken to get a more formative analysis of each compensation method. Moreover, live audio recordings gave inconsistent PESQ scores because of background noise or inconsistent speech patterns. Instead of recording in an uncontrolled environment, a sample of spoken word from a podcast that was consistent and had no background noise or audio clips had been recorded in ideal recording settings so there was consistency throughout conducting each different compensation technique for each Datagram Socket.

Interleaving proved to solve some of the issues related to the sockets, but it introduced delay which depreciated the quality of the audio. This meant that it sounded less intelligible to people, even though it may have scored higher on the PESQ scoring system.