Voice over IP in PDC Packet Data Network

Pernilla Norlund

June 13, 2004

Abstract

Sending voice as packages over a packet network, voice over IP, has increased with the growth of Internet. This has also affected the mobile cellular market and more services using this technique are provided to the end users. At TietoEnator Telecom Partners in Ursviken the Japanese mobile cellular system PDC is developed, verified and maintained. The main purpose of this masters thesis was to investigate how voice over IP would work in the PDC Packet Data network and the result of the investigation may lead to operators being able to provide more services to the end user. The investigation was conducted by implementing a prototype application that streamed real time speech over the PDC Packet Data network. The quality of the streaming speech was evaluated and analysed. Among the codecs found and used to convert the analogue signal to digital and back again, the CELP codec with a bit rate of 4.5 kbps gave the best reproduction of the speech.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The widespread growth of the Internet has created a mass market for multimedia and information services. Voice over Internet Protocol (VoIP) [4] is a way of packaging and sending voice data over the Internet or a packet data network like GPRS(General Packet Radio Service) [6] or PDC-P(Personal Digital Cellular Packet data network) [18, 21]. The use of VoIP implementations has increased and enables users to carry voice traffic over a packet data network.

To be able to send speech, audio and video package over a packet data network it has to be converted from an analogous waveform to a digital form. This is done with a codec. It is used to encode and decode speech, audio and video. The codec takes the analogue form of the media and tries to reproduce the analogue waveform to a corresponding digital waveform. The digital audio waveform is broken up into smaller packets and sent over the network to the receiver. At the receiver the package is unpacked and played. When this is done in real time it is called streaming audio. This is used in several services on the Internet, e.g., broadcasting radio or making telephone calls, IP telephony.

The low cost for the user is one reason for the growth of the voice over IP market. The user only pays for the data transmitted and not for the amount of time connected to the network. For the operator there are no big additional cost. They already have the data network and only some minor services have to be added.

The use of sending voice packet over the data packet network has grown among the mobile cellular operators. The operators provide a variety of services, e.g, WAP [9] (Wireless Application Protocol) which is a service that makes it possible to access Internet with a mobile phone, and MMS [7] (Multimedia Messenger Service) which is a method to send and receive messages with formatted text, graphics, photographs, audio and video clips. One of the more recent services provided is the push-to-talk [17] technique. It is a walkie-talkie functionality that makes use of the voice over IP procedure.

In this master thesis an investigation is conducted to see how well voice over IP in the PDC Packet data network would work. The study is conducted at TietoEnator Telecom Partners in Ursviken. At this branch of the company the Personal Digital Cellular (PDC)[18, 21] system for the Japanese mobile telephone market is developed, verified and maintained. The PDC mobile telephone standard is a variant of the GSM (Global System for Mobile Communications) standard and is used in the Japanese mobile telephony network. Among other things it includes a service for sending IP packet based traffic over the radio network, a system called PDC Packet data network (PDC-P). It would be beneficial to the operator to examine how audio is handled in the PDC-P network, and specifically how well voice over IP would work, so that the operators could provide more services to the end users.

1.1 Goal and approach

There exist some different algorithms for how the codec encodes and decodes the data stream or signal. An investigation and evaluation of the different speech codec algorithms are made.

The codecs have to fulfil two main criterias, first of all they need to be open source and free to use and secondly it must be developed especially for encoding and decoding speech. The codecs not only convert from analogue sound, speech or video to digital code, they also compress the binary code to the smallest number of bits possible. The file size of the compressed speech also effected the decision, since a smaller file will transmit faster. Based on this evaluation a few codecs are chosen for development of the prototype application.

An investigation is conducted to see what kind of services would be desirable using voice over IP. The investigation is conducted to get answers to the questions about what applications there are on the market and if there are any needs for others. On the basis of these answers a decision is made upon which application would be suited for the project and if it would be possible to implement the application under the amount of time allocated. A prototype application or service that makes use of Voice over IP and with the PDC-P system as IP carrier is implemented. An investigation on how to measure sound quality and IP packet throughput is also a must.

1.2 Thesis outline

In Chapter 2, the Japanese PDC mobile cellular system is described. The history of PDC and the technique used in the PDC is presented. An overview of the mobile cellular system is given. The overview describes the different components in the system and the functions and tasks of the components. It also describes the PDC packet data network and what task the PMSC performs.

An introduction on how to code and compress speech from analogue to digital form is given in Chapter 3. Some of the different codec algorithms are also being explained.

In Chapter 4, the chosen codecs for the project are presented. The criterias for how the codecs are selected are described. The chapter also includes motivations to why they are selected and why they are suited for the project. The advantages and disadvantages are discussed.

Answers to what Quality of Service (QoS) is and how can it be measured in voice over IP is explained in Chapter 5. The chapter will also give explanation to the parameters involved when deciding on good quality. Benchmarks for evaluating voice quality called Mean Opinion Score (MOS) and Perceptual Evaluation of Speech Quality (PESQ) are presented.

Some of the different services and applications using voice over IP are explained to the reader in Chapter 6. The section is ended with a discussion of which service that would be suited to implement in this project.

In Chapter 7, the design and implementation of the application are explained. It describes the architecture structure of the application and what design decisions have been made. The reader will get an overview of what the application can do and how the package is transmitted over the network.

The test result of the voice quality test preformed is discussed in Chapter 8. The criterias and test methods for the test are given.

In Chapter 9, the analyse and evaluation of the result is presented. The short comings and limitations of the application are discussed and what can be improved. Possible further developments are presented.

Chapter 2

PDC overview

2.1 PDC

The Personal Digital Cellular (PDC) [18, 21] system is a second-generation technology that is based on Time Division Multiple Access (TDMA) [20] technique and is used in digital cellular telephone communication in Japan. The PDC system was developed since there was such a high load on the analog cellular network. In April 1989 the Ministry of Posts and Telecommunications in Japan took the first initiative on the development of the PDC system. In 1991, the air interface specification of the system was presented by the Research and Development Center for Radio Systems (RCR) [21].

The TDMA technique splits each channel into different timeslots (see Figure 1), which allows different users to use the same frequency and hence the amount of data that can be carried [20]. For each channel it is possible to support three users under normal cir-

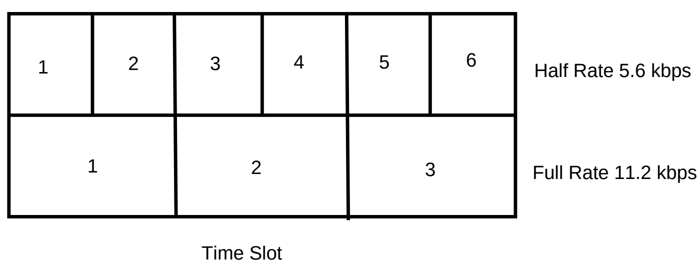

Figure 1: Timeslots and the different bit rates in PDC.

cumstances. When the traffic levels and the load on the network is high, it is possible to use half-rate, which means that the timeslots are divided into two and enable six users to be supported by each channel [18]. Half-rate channels reduce speech quality and data transmission rates, but it allows more users to occupy the same bandwidth. PDC has three full-rate (or six half-rate) channels in a 25 kHz frequency space. PDC offers two alternative rates, 11.2 kbps in full-rate channels and 5.6 kbps in half-rate channels [21].

The PDC system consist of a network of neighboring radio cells (see Figure 2), which together covers the service area and gives it a complete coverage. Each cell has a Base Station (BS) that operates on as set of radio channels. To avoid interference the BS in the adjacent cells operates on different channels. A group of BS is controlled by the Mobile

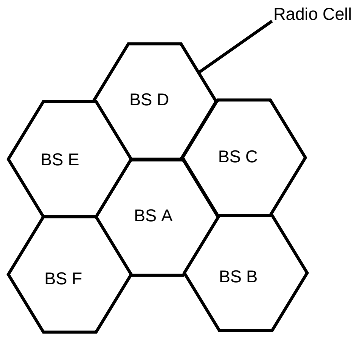

Figure 2: Neighboring radio cells with base stations.

services Switching Center (MSC) which controls calls from and to the Public Switching Telephone Network (PSTN), Integrated Service Digital Network (ISDN) and Public Land Mobile Network (PLMN).

The PSTN is the name of the international telephone system that is based on copper wires. PSTN carries analog voice data. The ISDN system is an international communication standard for sending voice, video and data over digital telephone lines or normal telephone wires. PLMN is a general term that refers to a mobile wireless network. Cellular phone and mobile Internet access are two common uses of a PLMN [8].

In order to make a call to a mobile phone or a Mobile Station (MS) there must be a number of databases in the network to keep track of where to find it (see Figure 3). Home

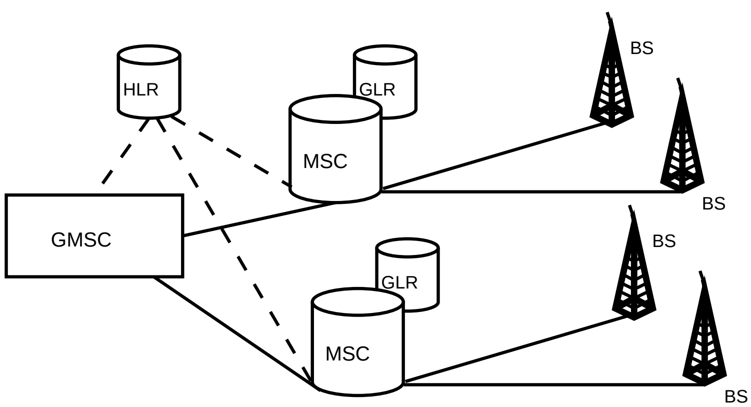

Figure 3: A simple overview of the PDC network.

Location Register (HLR) is one of these. When a subscription is bought from one of the PDC operators, the subscriber is registered at the HLR of that operator. The HLR contains not only the location of the MS, in which MSC area the MS is found, but also some subscriber information, such as supplementary service and authenticating parameters. As the MS move around the information is changed and the MS will send the changed information via MSC/GLR to its HLR. It is then possible for the MS to not only make calls but to receive calls.

Gateway Location Register (GLR) is a database containing information about all the visiting MS currently located in the MSC area. When a MS roams over and into another mobile network, the GLR connected to that MSC requests data from the HLR about that MS. The HLR will at the same time get information about on which MSC area the MS is located on. When the MS wants to make a call, the GLR has all the information needed to set up the call and do not need to ask the HLR each time.

If a caller from the fixed network (PSTN) wants to make a call to a PDC subscriber it will connect to an MSC with a gateway function also called Gateway Mobile services Switching Center (GMSC). It can be any of the MSCs in the PDC network. By interrogating the HLR the GMSC will get the information about where the MS is registered and the address to the current MSC area. When the call reaches the MSC the GLR will have more information about where the MS is located.

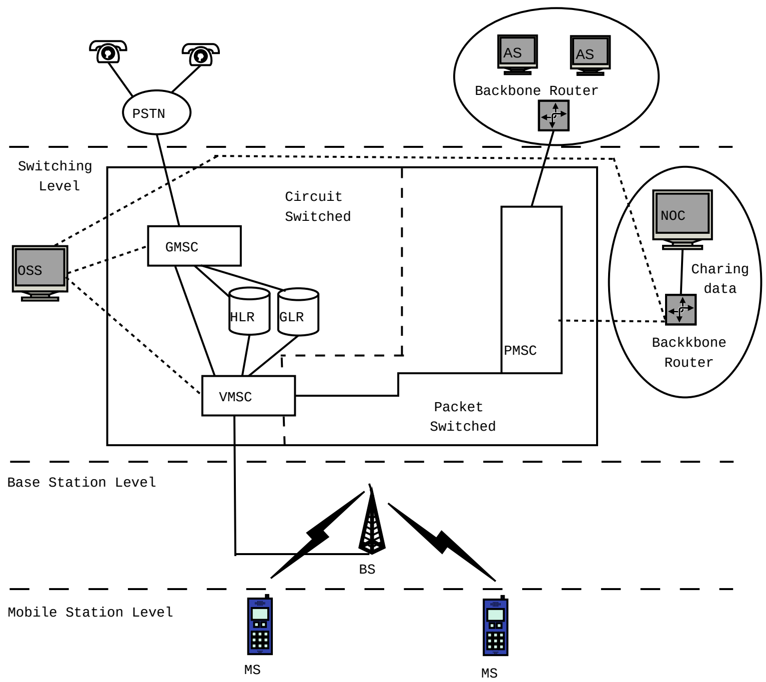

Figure 4: A more detailed overview of the PDC network.

The Operations Support System (OSS) provides the network operator with user-friendly tools for planning, operation and maintaining a cellular network (see Figure 4). The OSS handles the administration of the switching system, such as alarm list presentation, command handling, command log, file transfer etc.

Network Operating Center (NOC) has a similar task as the OSS, it handles alarms and supervision. The base stations are connected to the Visited Mobile services Switching Center (VMSC). The GMSC has the connection to external networks.

2.2 PDC-P and PMSC

PDC-P (PDC mobile Packet Data Communication System) was introduced to facilitate data transmission to a data rate of 28.8 kbps. The PDC-P system let the users use a single channel simultaneously. It makes use of the wasted bandwidth that occurs when only one individual user is being permanently dedicated to a single channel. Here individual packets of data are routed to the receiver. For bursty applications, such as Internet browsing where packets being sent or received with relative long silent periods between each transmission burst, this is very valuable. Users can be permanently online and only pay for the volume of data transmitted. The PDC-P allows a data transfer rate of 9.6 Kbps in uplink and 28,8 Kbps in downlink. Uplink and downlink are terms used when talking about data transfer speed. Uplink describes the amount of data being sent from the mobile phone to the base station. Downlink describes the amount of data being sent to the mobile phone from the base station.

In the PDC-P packet data network the PMSC is connected to the MSC. The node called the PMSC, acts as a bridge and a router between the radio networks and the IP traffic backbone, see Figure 4. It handles the packet data traffic in the PDC network. The IP network can be reached through a packet data enabled mobile phone, providing services for the end users.

Chapter 3

Code and compress speech

A speech signal (input signal) is converted from a various continuously physical value by some electromechanical device into its electrical counterpart. This electrical signal can be converted to a sequence of digital values, which is called a sample. The sample represents the amplitude of the input waveform. The two critical factors that determines the accuracy of the sample, how well the conversion from analogue to digital agree with reality, are the maximum rate of which the sample was made and the number of bits used at each sample.

To achieve a good quality a sample is collected regularly using the Nyquists theorem [4]. The theorem states that the sample rate should be twice the highest frequency on a voice line to achieve good quality, e.g., if a process has a filter that filters out anything above 4000 Hz, then the sample rate should be 8000 samples per second. If you then multiply the eight bit words which stores each sample the result is 64 000 bits per second or 64 kbits [4].

3.1 Sampling from analogue to digital

The speech signal goes through a few stages to be encoded. The process for PCM (pulse code modulation), which is the simplest form of codec, is as follow. First a band-pass filter is applied to eliminate frequencies in the signal that are not in any interest (e.g. frequencies above a certain value, which for telephone speech is above 4 kHz). The signal is then sampled and the analogue signal is converted into a sequence of values that represents the amplitude of the signal over a small time interval, see Figure 6. These values are then quantized, mapped into one of a set of fixed values, see Figure 5 (e.g., telephone quality speech, one of 2^8, 256 possible values). These values are then coded for transmission or storage. The reverse order of this process is applied at the receiver [3].

3.2 Speech codec algorithms

There are different techniques to encode/decode speech. The *waveform* codec try to produce a waveform that is as close to the original as possible without any knowledge of how the signal to be coded was generated [24]. Examples of waveform codecs are PCM(pulse code modulation) and ADPCM(adaptive differential pulse code modulation). A *source codec* takes in consideration the characteristics of how speech is generated. The input signal is put into a model that extracts simplified parametric information about the original speech

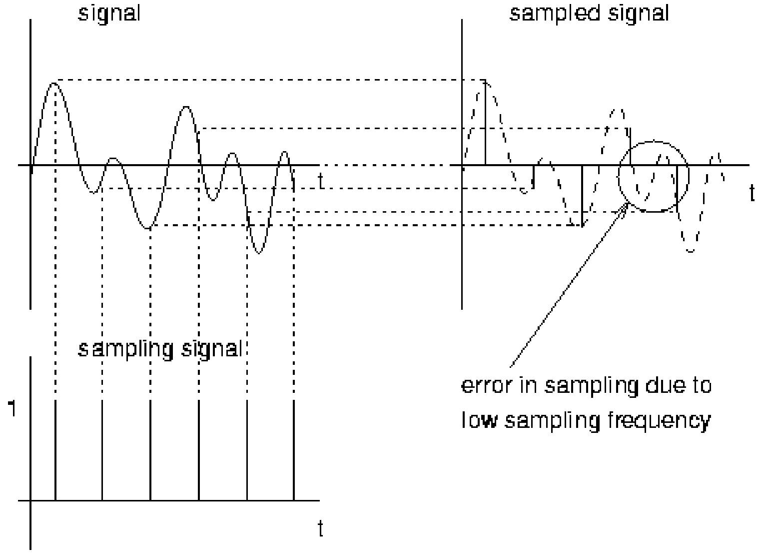

Figure 5: Sample signal from analogue waveform to digital form [3].

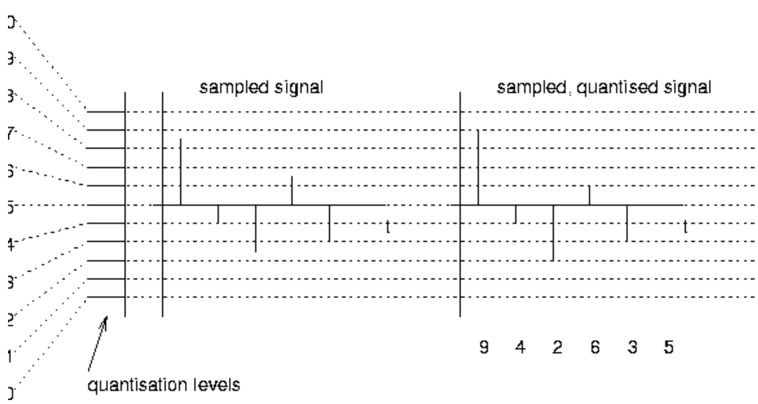

Figure 6: Digital form and quantised form of the signal [3].

Table 1 G series recommendations by the ITU-T.

G Serie	Description
G.711	Describes the 64 Kbps PCM voice coding technique.
G.726	Describes ADPCM coding at 40, 32, 24 and 16 Kbps.
G.728	Describes a 16 Kbps low delay variation of CELP.
G.728	Describes a CELP compression allowing voice to be coded in 8 Kbps.

excitation and vocal tract shaping. The transmission of this information requires less bandwidth. These codecs are also known as vocoders(*vo*ice *coders*) and includes codecs such as LPC and CELP [4].

The ITU-T (International Telecommunication Union Telecommunication Standardization Sector) has standardized the different codecs (e.g., PCM, ADPCM and CELP) into its G series recommendations, see Table 1. There are a few more standards in the G series than mentioned in Table 1.

PCM, pulse code modulation, is the simplest form of waveform codec. There are two common variations of 64 kbits PCM, the μ-law used in North America and the a-law used in Europe. There is not a big difference between the two standards, just minor compression details [4]. PCM takes 8000 samples per second and uses 8 bits per sample and a non-linear quantization that theoretical is supposed to give the best quality, giving a bit rate of 64 kbits [24].

ADPCM, adaptive differential pulse code modulation, differs from PCM. Instead of quantizing the speech signal directly, the ADPCM quantize the difference between a prediction of the speech signal and the real speech signal. The difference between the real and predicted speech is then quantized with fewer bits, 4 bits, than it would be necessary with the original speech sample. At the decoder the difference signal is added to the predicted signal to obtain the reconstructed speech signal. Compared to PCM this type of encoding reduces the number of bits required per sample.

LCD, linear predictive coding, uses a simple, analytic model of the vocal tract to fit speech into. Only the parameters that describe the best-fit are passed to the decoder. The decoder uses the parameters to generate synthetic speech resulting in intelligible and robotic sound. The LCD is used to compress speech at 16 kbps and below [3].

CELP code exited linear predictor is similar to LCD. The idea is to predict the next signal by using a linear combination of the past sample. CELP does the same modeling as LCD but also computes the errors between the original speech and synthetic model. Both the model parameters and the compressed representation of the errors are then included in the transmission. CELP uses long and short-term prediction. CELP is the most common used codec in modern speech coding technology [3].

Chapter 4

Speech codecs chosen for the thesis

There are a variety of different codecs that are used for compressing speech, music and video. The search has been concentrated on finding codecs that are made especially for compressing speech. After searching the Internet for free open source codecs three possible variants are found. The first one is called Speex and is a codec based on the CELP algorithm and implemented in C. The second one is called JSpeex and is a Java port of the Speex speech codec . The last one is the HawkVoiceDI(direct interface) which is a software that includes different speech algorithms and the software is implemented in C.

The Java variant of Speex, JSpeex, is not included in the rest of the development. The Speex and HawkVoice are more suited because they are both developed under the same programming language, C.

4.1 HawkVoiceDI

The HawkVoiceDI [10] software is especially developed for speech coding and includes different codecs that can easily be switched and tested. The software has support for Windows, UNIX and Linux. The different codecs are all free and open sources. The codecs included are shown in Table 2.

The program using HawkVoiceDI can alter between a variety of free and open source codecs. It makes it easy to test and analyze different codecs without having to make changes to the code, just change the argument passed to the application. It also includes the following

Table 2 Codecs included in HawkVoiceDI.

Kbps	Codec name	License	Code type
64	G.711 u-law	LGPL	fixed point
32	Intel/DVI ADPCM	Free	fixed point
13.2	GSM	LGPL	fixed point
4.8	LPC	LGPL	floating point
4.5-2.3	CELP	LGPL	floating point
2.4	LPC10	LGPL	floating point
1.8-1.4	OpenLPC	Free	floating point

features.

Voice Activated Transmission (VOX) process a voice buffer to see if the sample is silent or at least unvoiced. If there is a silent packet the developer can choose to not send the packet to the receiver.

The HawkVoiceID also includes MD5 and Blowfish for *encryption* of the voice stream. The encryption is good to secure a safe communication between the two endpoints.

Automatic Gain Control (AGC) is used to produce a constant volume level. The level is the percent of max volume for the sound buffer.

Not all of the codecs included in HawkVoiceDI are suited to use in the PDC-P network. The bandwidth in uplink is 9.6 kbps, and codec that encodes/decodes in a higher bit rate do not give a satisfying result, and is not included in the project. The ulaw (64 kbps), ADPCM (32 kbps) and GSM (11.3 kbps) are the three codecs that have a bit rate which is too high to be included.

4.2 Speex

The Speex software [22] is also designed for speech coding. The developer of the software wanted to create a codec designed for Voice over IP instead of the cell phone use, meaning more robust to packet lost. Speex is developed for applications like Voice over IP and archiving of speech data e.g., Voice mail. The design goal was to accomplish good quality and low bit rate. These goals led to the choice of CELP(e.g., used in the GSM standards) as the encoding technique in Speex. CELP is designed to compress voice at bit rates ranging from 2 to 44 kbps. The Speex software supports both Windows and POSIX.

Speex is mainly designed for 3 different sampling rates: 8 kHz, 16 kHz, and 32 kHz. These are referred to as narrowband, wideband and ultra-wideband respectively. When encoding with Speex there are different parameters that uses different bit rates and can be used to adjust the quality. Speex also includes some of the following features:

Variable Bit Rate (VBR) allows the coder to adjust the bit rate dynamically depending on the complexity of the audio being encoded. Different sounds, like vowels or for example s and f sounds, requires different bit-rate to achieve good quality. Vowels need a higher bit rate to achieve good quality than s and f sound that can be coded satisfactorily with fewer bits. One problem with VBR is that by only define the quality there is no guaranty of what the final average bit rate will result in.

Average Bit Rate (ABR) solves one of VBR problem by adjusting VBR quality dynamically in order to meet a specific target bit rate.

Voice Activity Detection (VAD) detects if the audio being encoded is speech or silence (background noise). If the encoding is done with VBR the VAD is always implicitly activated. If Speex detects non-speech periods it encodes them with very few bits, just enough to reproduce the background noise.

4.3 Choice of codecs

Both Speex software and HawkVoiceDI is used as codec for speech in different open source project. The codec used in Speex, CELP, is also listed as one of the codecs that is included in the HawkVoiceDI software. The Speex has a better documentation and offers both a manual that provides an introduction to Speex and CELP coding, information about encoding and decoding, API libspeex and also a reference manual that provides information on programming with Libspeex. The HawkVoiceDI does not include the same amount of information as Speex on how to use the software. To HawkVoiceDI advantage it includes different codec types that could easily be changed and vary between the different types. This makes the evaluation and quality testing more efficient.

Chapter 5

Quality of Service

Quality of Service (QoS) [2] is a measurement of the network performance in terms of network transmission quality and service availability. The QoS can take the form of traffic policy, for example if the transmission rate is limited, a certain amount of bandwidth can be guaranteed to the applications.

5.1 Quality of streaming audio

In addition to bandwidth there are also a number of other parameters involved when quality of streaming audio over IP are measured. Three main factors that affect how a person experiences the quality of speech over a network are lost of packets, latency and jitter.

Sending data over IP is not reliable. Packets can be lost or never delivered for several reasons. One reason is when a network gets busy, when the network utilization increases the percentage of dropped packets also increases.

Latency or delay is the amount of time it takes to send a packet from source to destination. Together with bandwidth, latency defines the speed and capacity of a network. The ITU-T has given a recommendation that specifies that for good voice quality no more than 150 ms of one-way end to end delay should occur.

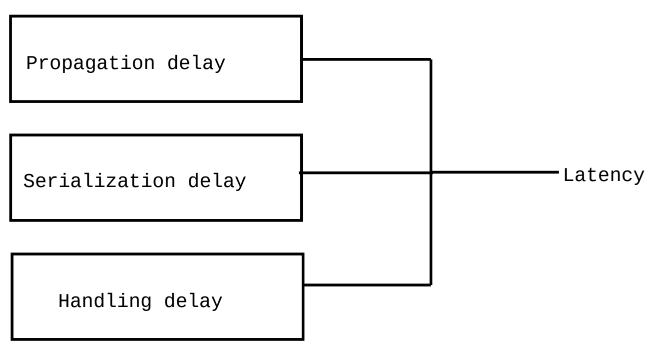

Figure 7: Latency in telephone network.

There are three types of delays in a telephone network, see Figure 7. Propagation delay means that the delay is created by the devices and material used, like fiber or copper wire.

This delay is not by it self imperceptible by the human ear but in conjunction with other delays it can cause noticeable speech degradation. The second delay is the serialization delay which means the time it actually takes to place a bit or byte onto an interface. As for propagation delay serialization delay can not by it self create large noticeable speech degradation and its influence on the overall delay is relatively minimal. The third and last type is the handling delay or processing delay. It includes many different causes of delay. Handling delay can occur when packages are queued because of congestion on an outbound interface. Queuing delays occur when more packages are sent out than the interface can handle with in a given interval. Some other causes of handling delay are the time it takes to package a packet, compression or packet switching. These types of delays are the ones making the biggest impact on the packet data networks [4].

In a voice packet environment it is expected that voice packages are transmitted at a given frequency, e.g., one frame every 10 ms. But a voice package can be delayed or even dropped. This results in packages not arriving at the same frequency to the receiver. The difference between when the packet is expected to arrive and when it actually arrives is called jitter. Jitter occurs when the network is under heavy load and the packet must be buffered. This variation is inconsistent and unreliable [11].

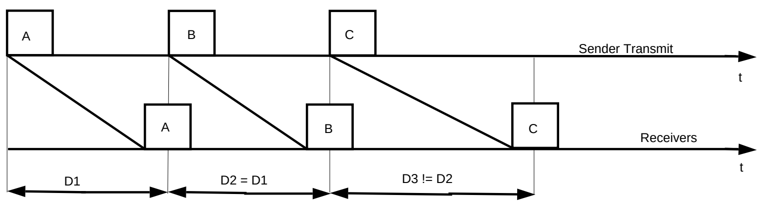

Figure 8: Jitter in packet network.

In Figure 8, the time for packet A and B to be sent and received is equal, but packet C encounters a delay and is received after it is expected. With a jitter buffer the delayed packets can be concealed.

5.2 MOS - Mean Opinion Score

Voice quality can be tested in two ways, subjectively and objectively. The subjective voice testing is preformed by human and the human ear. Computers can perform an objective test of the sound quality and are not fooled by the illusion of the compression schemes that can play a trick on the human ear.

When codecs are developed the developer tries to tune the codec to sound as good as possible on the basis of subjective measurements of the voice quality. If the codecs instead were measured by the objective testing methods it would get a much worse result and it would not correlate with the human ears perception of the voice quality. In the end it is the human perception of the sound quality that determines if the voice compression techniques give a good result.

To be able to compare codecs against each other a subjective benchmark called Mean Opinion Score (MOS) [12] is used. The MOS is one of the parameters that are used to determine the QoS for voice over IP. The MOS is a test that is accepted by the International

Telecommunications Union Telecommunication standardization sector (ITU-T recommendation P.800) and is a subjective score of voice quality perceived by people listening to speech over a communication system. To determine the MOS for a particular codec, a mixed sample of females and males (40 or more people) from different ethnic or language background are given an audio sample that is several seconds long to listen to, each person rates the quality of the audio on a scale 1 to 5, see Table 3. The resulting MOS is the average of all the individual scores, 1 being the worst to 5 being the best. Because voice quality is very subjective to different people it is important to have a wide range of listeners and sample material when conducting a MOS test.

Table 3 MOS score.

Score	Description
5	Excellent, a perfect reception.
4	Good, long distance telephone quality, PSTN.
3	Fair, communications quality such as GSM (requires some hearing effort).
2	Poor, low bit rate vocoder such as LPC (hard to understand the speech).
1	Bad, communications breakdown.

The ITU-T standards state that 5 are a perfect MOS score, and 4 are considered being a high enough standard for toll-quality conversation. It is generally accepted that a minimum MOS rating of 3.6 is required for having a satisfying voice conversations, see Table 4 [23].

Table 4 Desirable MOS rating.

Score	Description
4.0 - 5.0	Desirable voice quality.
3.6 - 4.0	Generally acceptable.
1.0 - 3.6	Not recommended.

MOS testing is also used to evaluate different codecs under varying circumstances like different background noise levels. The data received is then used in comparison with other codecs. The ITU-T MOS scores of different codecs are listed in the Table 5 [4].

Table 5 MOS score for different codecs.

Compression method	Bit Rate (kbps)	Sampel Size(ms)	MOS Score
G-711 PCM	64	0.125	4.1
G.726 ADPCM	32	0.125	3.85
G.728 Low Delay Code Excited Linear Predictive(LD-CELP)	15	0.6251	3.61
G.729 Cojugate Structure Algebraic Code Exited Linear Predictive(CS-ACELP)	8	10	3.92
G729a CS- ACELP	8	10	3.7
G.723.1 MP-MLQ	6.3	10	3.9
G.723.1 ACELP	5.3	30	3.65

5.3 PESQ - Perceptual Evaluation of Speech Quality

Although the MOS is a subjective method for determining voice quality, it is not the only way. There are several other methods and algorithms developed to evaluate the QoS. Perceptual Evaluation of Speech Quality (PESQ) is an ITU-T standard (recommendation P.862, February 2001), and an objective algorithm that is especially developed to determine the quality of speech. The reason for the development of the PESQ algorithm is that the MOS testing is a very expensive method and takes a lot of time to conduct. The PESQ automatically predict the quality scores that would be given by a typical subjective test. Taking the signal sent through the test environment and a reference signal and then compeering those two performs part of the test. The two signals are put through the PESQ and it results in a PESQ score, see Figure 9 [19].

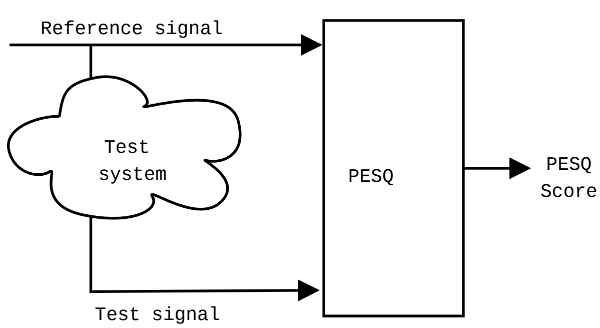

Figure 9: PESQ testing.

As mentioned above one drawback is that the machine cannot perceive some of what a human ear can hear. A person can trick the human ear to hear a higher quality voice, but he can not trick the computer.

Chapter 6

Services and applications using VoIP

An investigation is conducted to get some answers to the questions about what applications there are on the market and if there are any needs for others. On the basis of these answers a decision is made on which application would be suited for the project and if the application is possible to implement. The application implemented is used to test voice over IP in the PDC-P network.

6.1 IP telephone over the Internet

The voice over IP technique makes it possible to conduct telephone calls over a data packet network, e.g., over the Internet. This service is already provided by many different suppliers.

6.2 MMS (Multimedia Messenger Service)

The technique makes it possible for mobile users to send and receive multimedia messages. It is a method to send and receive messages with formatted text, graphics, photographs, audio and video clips [5].

The mobile phone must be equipped with a camera and audio input. Because it is a relatively new technique not all the users have the necessary devices for reading MMS messages. It is then stored at a legacy support application and a Web URL is sent as an SMS to enable the user to view the MMS using a Web browser [15]. Exactly what and how the message is sent to the receiver is different depending on the operator.

6.3 Push-to-talk technique

The technique is based on the IP Multimedia System (IMS), by just pressing a button the user will make his/hers voice heard to a group of mobile phones, a walkie-talkie function. It is a one-way communication, while one person speaks the other listen. This technique demands that the mobile phone have an IP address to get direct connection. The SIP protocol (Session Initiation Protocol) keeps track of where every terminal is located and how every user is available (see Figure 10). Ericsson, Nokia, Siemens and Motorola have all agreed on a push-to-talk standard [13].

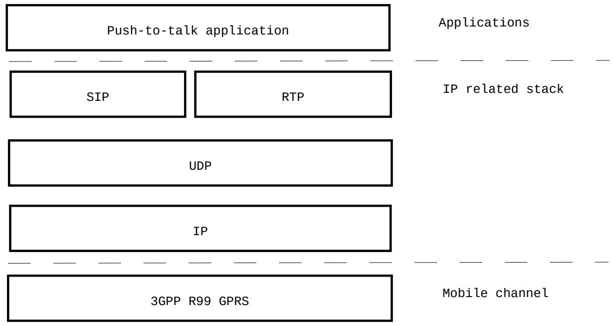

Figure 10: Protocol stack of Push-to-Talk over cellular solution.

Some of the benefits are that push-to-talk works over big distances and the traffic takes little space in the network, because it is only burdened when the voice is sent. The push-to-talk technique also provides a text chat function between the members of the talk group (see Figure 11). The benefit of this feature is that it will make it easier for the users to exchange addresses and textual information within the talk group. The service is based on multi-unicasting, when a packet data is sent to a dedicated push-to-talk server, the server duplicates the packets and sends it to the receivers [16].

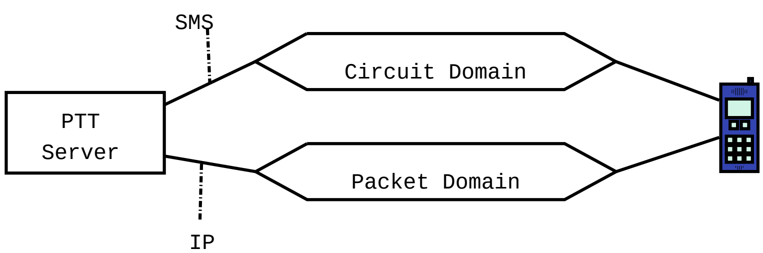

Figure 11: Components needed to be able to Push-to-Talk.

The push-to-talk technique has so far not been introduced by any operators in Europe but they are on the break of launching a mobile phone with push-to-talk technique. The American operators Nextel and Verizon have already an existing version.

How does a push-to-talk call work in GPRS?

1. The sender checks his/hers address book to see who is available to contact and then choose one or several receivers, or a whole group.

2. The sender push the button, wait for two seconds, meanwhile the phone is establishing a connection over the GPRS to the push-to-talk server. With the SIP protocol the server make a connection to the receiving phones.

3. After the connections is made the sender can start to talk, the voice is digitalized and sent as an IP packet over the GPRS to the push-to-talk server and on to the receiving phones. The voice is heard with approximately one second delay at the receiver.

4. When finished talking, the sender release the button and the other members of the group gets a message that it is free to talk/respond.

5. Only one at a time can get a clearance to talk. It is also possible to send a voice message that ends up in the receivers phone. The receiver can listen to the message without having to call the receivers voicebox [14].

6.4 Application suited for the thesis

IP telephone over Internet, MMS and push-to-talk are some of the common applications that use voice over IP technique. The biggest benefit of this kind of application is the cost reductions that the end users of the voice over IP application will obtain. The most interesting application to implement is an application using the push-to-talk technique. For TietoEnator a push-to-talk application could be attractive to implement and the result could in the end give the operator further services to the end users. The implementation demands an investigation on how push-to-talk works and what improvements that has to be done to the packet data network. PDC-P has similar characteristics as the GPRS. By looking on how the technique is used in GPRS some ideas can be given on how to develop push-to-talk in PDC-P.

Chapter 7

Design and implementation

7.1 Architectural structure of the prototype application

The application is implemented in C under the UNIX platform and is a real time audio streaming application. It consists of two parts (see Figure 12), one is the encoding part that takes the input signal of a microphone and encodes and compress the speech into a packet. The packet is sent via UDP(user datagram protocol) to a receiving application server. At the application server the other part of the application is running the decoding part. It takes the packet and decodes and decompresses it. The decode packet is then put to a speaker and a reproduction of the input signal is heard.

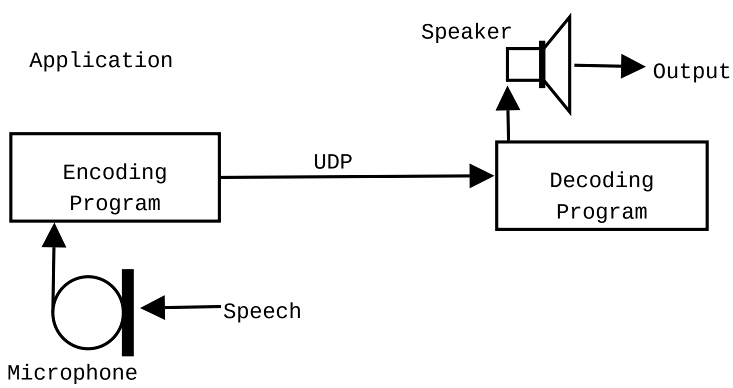

Figure 12: Overview of how the application work.

For packet transmission either TCP(transmission control protocol) or UDP can be used. TCP guarantees the delivery of data. All packages sent over the network are also delivered in the same order which they were sent to the end user. There is no need for a reliable transmission of the package when sending voice data. If a packet is not delivered to the end user, it only results in a reduction of the quality at that moment of the transmission. It would be more disturbing if there would be a delay because the package had to be retransmitted. With UDP the packet is sent with no guarantee given that the packet is delivered to the receiver. No retransmission of undelivered packets is done. UDP does not care what happens with the packet when it has been sent. Therefore UDP is used for the transmission of the voice packet [1].

The equipment used are a microphone for collecting the input signal and loudspeakers to play the output signal. It is also necessary to use the NetDevil equipment to strangle the bandwidth to 9.6 kbps, one timeslot, which represents the actual uplink bandwidth.

7.2 How packages are sent through the network

The PMSC is the node that handles the packet data traffic in PDC-P. It acts as a bridge and a router between the radio network and the IP traffic backbone. To this network an application server (AS) can be connected which can be reached through a packet data enabled mobile phone (MS), providing services for the end users.

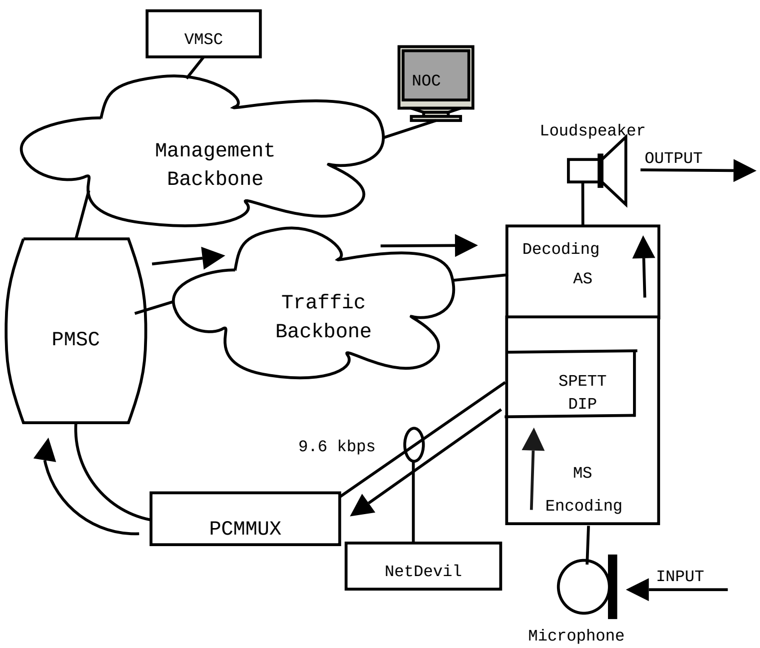

Figure 13: Overview of how the packages are sent through the network.

When running the application in the test plant a few more components are used (see Figure 13). A test tool called SPETT is used in combination with DIP (demultiplexing IP). SPETT simulates a mobile phone and DIP creates a virtual interface which opens a stream into SPETT that any application can use to send data into the PDC-P network. SPETT/DIP has one disadvantage, it uses all available bandwidth and it is not possible in SPETT/DIP to limit the bandwidth to the wanted bandwidth.

To ensure that the uplink is operating at 9.6 kbps, as in the PDC-P, a test tool called NetDevil is used to control the bandwidth. NetDevil is a test tool for disturbing IP traffic in a predetermined and controlled way. Two of its features are the capability to strangle the bandwidth and the ability to simulate packet loss.

To run the application correctly a VMSC simulator has to be started on a management machine. The SPETT test tool and DIP server must be started on a traffic machine. The application encoding and decoding programs should be started on the traffic machine. First the decoding program is started and then the encoding program. The PCMMUX also has to be started, the PCMMUX is used to convert between the radio network and Ethernet.

7.3 Problems encountered

At first the idea was to use prototypes of real mobile phones, used on the Japanese market, to send and receive voice packets. While using the mobile phones as sender and receiver a problem was encountered. When sending to fast the mobile phone crashed. It could only handle one packet per second and that is not a good thing when handling real time audio, the quality of the speech would not be satisfying. So a decision was made to change direction and use a test environment called SPETT/DIP to simulate the mobile phone. It required that the application had to work under a UNIX platform instead of Windows. This resulted in some problems when converting the code from Windows to UNIX.

The problem was the usage of big and little endian. Big or little endian refers to what bytes are most significant, it describes the order in which a sequence of bytes is stored in a computers memory. In a big endian system the most significant value in the sequence is stored at the lowest storage address (i.e., first). In a little endian system the least significant value in the sequence is stored first. Windows is a little endian system and UNIX is a big endian system. The big and little endian problem resulted in excluding the Speex software from the project.

Chapter 8

Evaluation of voice quality for VoIP in PDC-P

The test of the encoding and decoding application was made at the test plant at TietoEnator Telecom Partners, Ursviken. The different codecs were investigated and evaluated. The investigation was conducted to examine if it was possible to obtain a good quality sending voice packages through the PDC-P network, and how well it would work.

8.1 Testing criterias

The investigation was intended to get some answers to how the quality of the streaming voice was received at the end user. It was also interesting to examine how the quality was affected when packets were lost and never delivered. When performing the test, some criterias were more closely looked at. The criterias were compiled in collaboration with TietoEnator. How did the delay affect the perception of the voice? How good was the packet throughput when the bandwidth was strangled to 9.6 kbps and did this affect the quality? As mention before packet loss was simulated with the test tool NetDevil.

8.2 Testing methods

The voice quality was tested with a subjective method. By listening to the different codecs at work a subjective evaluation was made upon which codec sounded most natural and gave the best reproduction of the speech for the human ear. The MOS test was not implemented in full.

An objective testing of the perception of the speech could also have been done. An investigation for a free to use software was conducted, but no product of that kind was found. The test was performed by speaking into the microphone, and listening and evaluating the resulting output from the loudspeaker. The codecs used in the investigation are shown in the Table 6.

Table 6 Codecs used in the test.

Codec	Bit rate (kbps)
CELP	4.5
CELP	3.0
CELP	2.3
LPC	4.8
LPC10	2.4
openLPC	1.8
openLPC	1.4

8.3 Test result for CELP

The different CELP codecs gave a natural reproduction of the original speech. The best of the different CELP codecs was the one having the highest bit rate of 4.5 kbps. The CELP codecs with lower bit rates (CELP 3.0 kbps and CELP 2.3 kbps) did also provide an acceptable result. But at the lower bit rate it required some hearing effort, it sounded like some words were missing and gave a stuttered result.

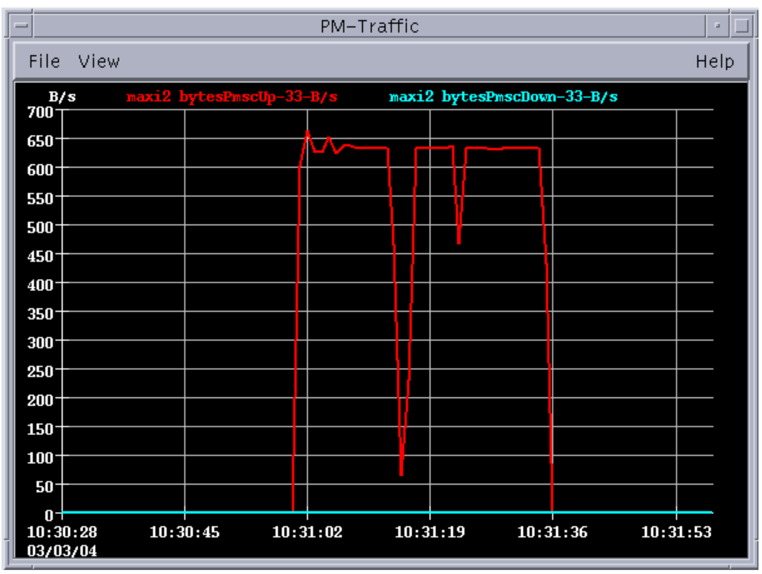

Figure 14: Graph showing CELP encoding in 4.5 kpbs sending packets through the PMSC.

In Figure 14 the graph shows the amount of bits sent per second via PMSC in the PDC-P network when using the CELP 4.5 kbps. The unit of measurement is in bit per second, multiplying this with eight gives approximately a bit rate of 5.0 kbps. This is a higher bit rate than 4.5 kbps and the result of the headers put on by the different protocols like UDP and IP. This resulted in a higher bit rate than the one expected when using a codec with a specific bit rate.

8.4 Test result for LPC

The different LPC codecs gave a more metallic or robotic reproduction of the original speech. The speech recognition was not as good as with the CELP codec. Among the LPC codecs it was the LPC with 4.8 bit rate that produced the best result, see Figure 15 to see the amount of bits sent per second via PMSC in the PDC-P network when using the LPC 4.8. The ones with lower bit rate (LPC10 2.4 kbps, openLPC 1.8 kbps and openLPC 1.4 kbps) did not give a satisfactory quality of speech and demanded very big hearing effort. See Figure 16 to see the amount of bits sent per second using the LPC encoding in 1.4 kbps. It was just the LPC 4.8 kbps that gave a somewhat of acceptable speech quality.

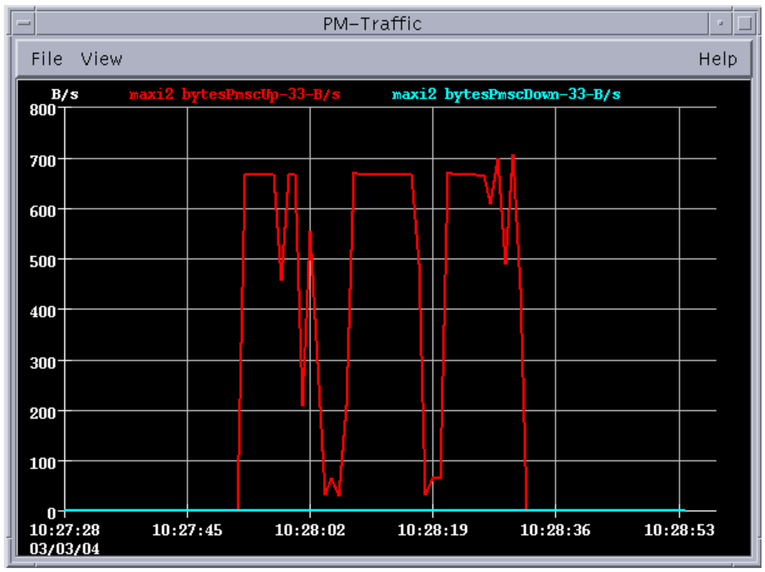

Figure 15: Graph showing LPC encoding in 4.8 kpbs sending packets through the PMSC.

8.5 Conclusion of the test

When comparing the CELP 4.5 against LPC 4.8 the CELP gave a much more realistic reproduction of the speech. But the two different codecs both gave a satisfying quality of speech. It was not a surprising result that the two codecs with the highest bit rate gave the best result. When the bit rate increases less speech is missed than with having a lower bit rate. With a higher bit rate more of the speech can be encode and sent per second.

When packets were dropped, the quality fell drastically. The application did not stand packet loss well. Already with a small packet loss the result was descending and it was very noticeable that the speech quality were much poorer. It required a big hearing effort to understand every sentence. The encoding and decoding application has no packet buffer implemented. Implementing a buffer could result in the application being more tolerant against packet delay and packet lost. If a buffer is set up the absence of a packet or a delay may be less noticeable.

When not using the NetDevil to strangle the bandwidth, the test gave a better result.

There was no visible decrease in the amount and size of sent packets when using NetDevil, but the packets were somehow delayed and some even dropped. This contributed to a decrease of the voice quality.

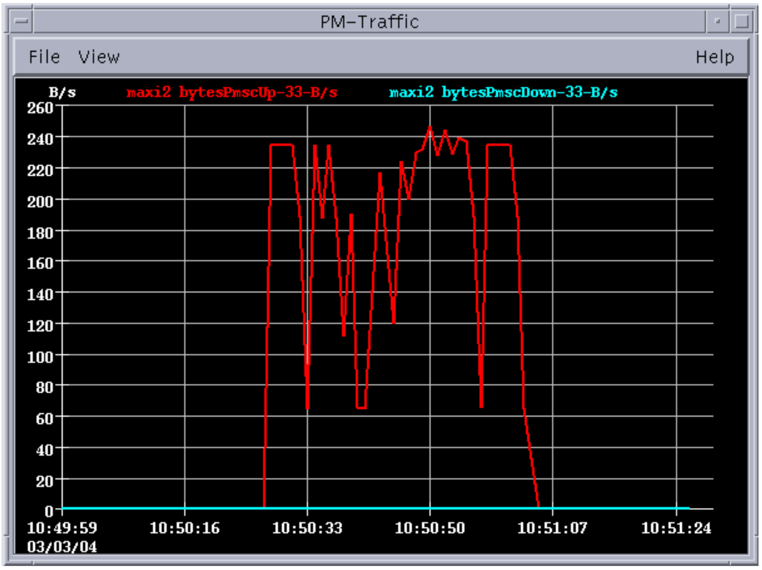

Figure 16: Graph showing LPC encoding in 1.4 kpbs sending packets through the PMSC.

Chapter 9

Discussion and further work

The encoding and decoding application was first implemented for Windows. At first an actual prototype mobile phone was intended to use for sending packets through the PDC packet data network to a receiving application server. The mobile phones were prototypes of mobile phones that are used on the Japanese market. The application was running and the phones were sending and receiving audio packets via the PMSC, but the mobile phones could only manage to send one packet per second. When increasing the transmitting rate and trying to transmit and receive at a higher rate the mobile phone crashed. This made it impossible to implement an application that would handle real time voice streaming, which requires a much higher transmitting rate.

A new approach was to set up a virtual interface representing a mobile phone and send packets through the interface instead of an actual mobile phone. The SPETT/DIP test tool made this possible. The first implementation, running under Windows, had to be dropped and rewritten. The SPETT/DIP test tool required that the application had to be run in a UNIX environment. The SPETT/DIP test tool also made it possible to send data packets as a stream. That enabled a real time speech stream to be set up, taking input from microphone and sending it to the receiver that pust the output to a loudspeaker.

The overall result of the quality of sending voice packets over the PDC-P network is good. It requires some hearing effort and when the packet loss is high the quality is drastically lowered. Of the codecs found in the investigation the CELP 4.5 kbps gave the best reproduction of the speech.

It would have been interesting to get the Speex program to work under UNIX. The Speex software makes it possible to encode the speech at a varying bit rate. It would have been interesting to examine how good the quality of speech became when having the speech encoding at the highest bit rate that the bandwidth allowed. It would have resulted in a bit rate just below 9.6 kbps to be able to include the headers. But when sending bigger packets across the network there also exist a risk that when packages are lost the disturbance of the quality increases because more information is lost than when sending smaller packages.

The test plant at TietoEnator is very clean, there exist very few elements that can create a congestion or packet loss. The packets were almost always delivered to the receiver. Therefore a test tool called NetDevil was used to simulate packets being dropped. When simulating packet loss the quality drastically decreased.

In the beginning of the thesis the intention was to implement an application that made use of both voice over IP and push-to-talk (PTT) technique. In the end the amount of time allocated for the thesis was not enough and the PTT was changed to a real time streaming audio application.

To be able to do a PTT application an investigation was conducted to see what improvements and changes that had to be made to the network in order for the PTT technique to work. The PTT has already been tested in the GPRS system and the technique will demand new software in the mobile phones. It exists software that can be downloaded to a mobile phone that has not got the PTT technique implemented. This software enables the mobile phone to communicate with a mobile phone that has the PTT technique implemented.

It will also demand new software in the mobile network. The operators will have to conduct an upgrade of the GPRS network with Quality of Service, to guarantee the speed of the data traffic in the network.

The PDC-P is very similar to the GPRS system and all the demands mentioned above have to be applied to the PDC-P system as well. A change in the network is costly and it would perhaps be more efficient to change the PTT use-case to match the bearer capabilities.

Acknowledgments

Thank you to my mentor at TietoEnator, Lars Linder, for the help during the time at TietoEnator. Thank you to my mentor at Umeå University, Pedher Johansson, for the help with the report.

Bibliography

[1] Kenneth L Calvert and Michael J Donahoo. *The pocket guide to TCP/IP sockets C version*. Morgan Kaufmann Publishers, Inc., San Francisco, California, 2001.

[2] Inc Cisco Systems. Quality of service(QoS). Webpage, 13 April 2004. `http://www.cisco.com/univedcd/cc/td/doc/cisintwk/ito_doc/qos.htm`.

[3] Jon Crowcroft, Mark Handley, and Ian Wakeman. *Internetworking multimedia*. Taylor & Francis, London, 1999.

[4] Jonathan Davidson and James Peters. *Voice over IP Fundamentals*. Cisco Press, Indianapolis, 2000.

[5] Ericsson. Multimedia Messaging Service. Webpage, 20 November 2003. `http://www.ericsson.com/technology/tech_articles/MMS.shtml`.

[6] Ericsson. GPRS - General Packet Radio Service. Webpage, 18 March 2004. `http://www.ericsson.com/technology/tech_articles/GPRS.shtml`.

[7] Ericsson. MMS - Mulitmedia Messaging Service. Webpage, 18 March 2004. `http://www.ericsson.com/technology/tech_articles/WAP.shtml`.

[8] Ericsson. Understanding telecommunications. Webpage, 18 April 2004. `http://www.ericsson.com/support/telecom/book-index/index.shtml`.

[9] Ericsson. WAP - Wireless Application Protocol. Webpage, 18 March 2004. `http://www.ericsson.com/technology/tech_articles/WAP.shtml`.

[10] Phil Frisbie. Hawkvoice - free speech compression. Webpage, 10 November 2003. `http://www.hawksoft.com/hawkvoice`.

[11] Rick Gallaher. What ever happend to quality of service(qos)? Webpage, 19 January 2004. `http://www.convergedigest.com/tutorials/qos1/page1.asp`.

[12] ITU-I International Telecommunication Union. MOS terminology. Webpage, 22 March 2004. `http://www.itu.int/rec/recommendation.asp?type=items&lang=E&parent=T-REC-P.800.1-200303-I`.

[13] Lars Anders Karlberg. Ericsson vill göra mobilen till walkie talkie. Webpage, 20 November 2003. `http://www.nyteknik.se/pub/ipsart.asp?art_id=26515`.

[14] Mats Lewan. Global walkie-talkie nästa mobilfluga. Webpage, 13 November 2003. `http://www.nyteknik.se/art/30327`.

[15] Nokia. MMS entering into the next phase. White paper, 2003.

[16] Nokia. Push to talk over cellular - real-time always-on voice service. White paper, 2003.

[17] Nokia. Push to talk. Webpage, 18 March 2004. http://www.nokia.com/nokia/-0,8764,46740,00.html.

[18] Ian Poole. Personal digital cellular (PDC). Webpage, 20 February 2004. http://www.radio-electronics.com/info/cellulartelecomms/pdc/-pdcsummary.htm.

[19] Psytechnics. PESQ: An introduction. Webpage, 11 January 2004. http://www.psytechnics.com/downloads/wp_pesq_intro.pdf.

[20] Jochen Schiller. *Mobile Communications*. Addision-Wesley, London, 2000.

[21] SPG Media Limited. Mobile/Cellular Technology - PDC (Personal Digital Cellular). Webpage, 2 February 2004. http://www.mobilecomms-technology.com/-projects/pdc/index.html.

[22] Jean-Marc Valin. Speech compression. Webpage, 20 November 2003. http://www.speex.org.

[23] Viola Networks. Predict the Voice Quality of VoIP on your network. Webpage, 10 January 2004. http://www.violanetworks.com/moscalculator.asp.

[24] Jason Woodard. Speech coding. Webpage, 20 November 2003. http://-rice.ecs.soton.ac.uk/jason/speech_codecs/standards/.

A complete tutorial about voip configuration in Packet Tracer 7.3 simulation software. Learn how to configure IP phones and Call Manager Express on a Cisco 2811 router.Â This tutorial is designed to help you to configure the voice over ip (voip) features available in Packet Tracer 7.3. It will show you the steps required to : Configure Call Manager ExpressTM on a 2811 router, Use the various telephony devices.Â This configuration will separate voice and data traffic in different vlans on SwitchA. data packets will be carried on the access vlan. SwitchA(config)#interface range fa0/1 â€“ 5 #Configure interface range# SwitchA(config-if-range)#switchport mode access SwitchA(config-if-range)#switchport voice vlan 1 #Define the VLAN on which voice packets will be handled#. Voice over Internet Protocol (VoIP) is the convergence of traditional voice onto an IP data network. VoIP provides better application integration by using a common protocol and can help lower costs by melding separate support staffs. Other real-time traffic, such as uncompressed video and streaming audio, is also converging onto data networks. VoIP is very complex because it involves components of both the data and voice worlds.Â In order for a VoIP solution to function well, the network must be able to give voice packets priority over ordinary data packets or sufficient bandwidth must always be available. Avaya products for VoIPâ€" Communication Manager and Avaya CajunÂ® data switchesâ€"all include several standard strategies to prioritize voice traffic. Remember that an IP network is packet switched and ideal for non-real-time data communication and not designed for real-time VoIP. To overcome this problem UDP is used. UDP is unreliable and connectionless protocol.Â TCP/IP over modern routers and networks is very fast. It is more than capable of handling voice over IP communication. (I've done it myself). My guess is that your implementation has some bugs in it related to buffer sizes.