Mobile VoIP: Managing, scheduling and refining voice packets to and from mobile phones

MOHAMMAD ABDUS SALAM
Student ID: 01201023

TAPAN BISWAS
Student ID: 01201003

\

Department of Computer Science and Engineering
August 2006

BRAC University, Dhaka, Bangladesh

DECLARATION

I, Mohammad Abdus Salam, ID: - 01201023 have completed our proposed Thesis, Mobile VoIP: Managing, scheduling and refining voice packets to and from mobile phones, Under CSE 400 course based on the result found by me.

I therefore declare that this project has not been published by anyone else previously for any degree.

Signature of   Signature of

Supervisor   Author

ACKNOWLEDGMENTS

Abstract

Voice over IP (VoIP) is rapidly gaining acceptance. VoIP has already become a multi billion dollar industry in the IT world. On the other hand the use of mobile phones in the world has exceeded all estimations. Given these two technologies, we are unfortunate that there is no VoIP support available for mobile phones. The bridge between VoIP and mobile phones is a long way, but an initiative like this paper would surely lay down the roadmap.

Mobile phones have several annoying limitations. Limited data transfer rate, processor speed, lack of protocol implementation etc. are the relevant issues for VoIP. This paper addresses the challenges of VoIP and tries to explore ways to overcome those challenges in mobile phones. The language we used for our experiment is JavaME for its simplicity and wide acceptance.

Keywords: Voice Packet, Bluetooth, VoIP.

Table of Contents

1. Introduction

VoIP on mobile phones can open the door to numerous applications with highest usability, accessibility and mobility. Just to imagine that one can make cheap VoIP calls from anywhere using their mobile phone, is exciting enough. Mobile VoIP can also allow audio/video conferencing between remote participants equipped with mobile phones only. Examples of such applications include audio/video conferencing between doctors and patients in remote area, distance learning and many more.

In this paper we suggest ways to overcome the challenges of VoIP implementations on mobile phones.

2. Challenges of VoIP on Mobile Phone

VoIP poses many challenges. In the following sub-sections we present these challenges supported by our suggestions.

2.1 Accumulation Delay

This delay is caused by the need to collect a frame of voice samples to be processed by the voice coder. It is related to the type of voice coder used and varies from a single sample time to many milliseconds [1]. The following table shows comparison between different algorithms [2]:

Algorithm	Data Rate	Program Memory (Kbytes)	Data Memory (Kbytes)	Processor Loading (MCPS)
G.711	64kbps	0.5	1	0.07
G.722	48/56/64 kbps	9.36	Tables: 1.61 Variables: 0.1	4.1
G.723.1	6.3 (high rate) 5.3 kbps (low rate)	132	Tables: 19.4 Variables: 2.3	8.5
G.729 A	8 kbps	127	Tables: 2.71 Variables: 7.41	7.3
GSM-FR	13 kbps	32	2	2.8

Table 1 Comparison between different algorithms.

2.1.1 Suggestions:

Accumulation Delay is a fixed delay. We've nothing to do about it. But we can choose a specific codec, which leads to lower accumulation delay and at the

same time accommodates a feasible implementation on mobile phone. GSM-FR can be a good choice since it is already being used on mobile phones.

2.2 Processing Delay and Network Delay

The actual process of encoding and collecting the encoded samples into a packet for transmission over the packet network causes processing delay. The encoding delay is a function of both the processor execution time and the type of algorithm used. Often, multiple voice coder frames will be collected in a single packet to reduce the packet network overhead. For example, three frames of G.729 codewords, equaling 30 milliseconds of speech, may be collected and packed into a single packet [1].

On the other hand network delay is caused by the physical medium and protocols used to transmit the voice data, and by the buffers used to remove packet jitter on the receive side. Network delay is a function of the capacity of the links in the network and the processing that occurs as the packets transit the network. The jitter buffers add delay which is used to remove the packet delay variation that each packet is subjected to as it transits the packet network. This delay can be a significant part of the overall delay since packet delay variations can be as high as 70-100 msec in IP networks [1].

2.2.1 Suggestions:

Figure 1 is the proposed network diagram where the cell phone will be connected with the server through the WiFi/Bluetooth access point and the call will be forwarded to other cell phone through the Internet.
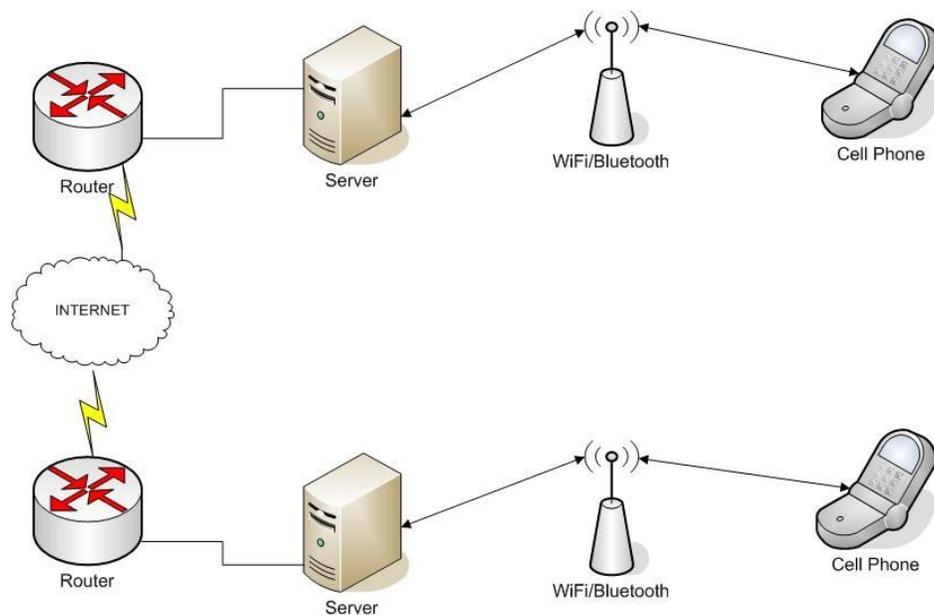

Figure 1 Call forwarding from cell phone to cell phone through the Internet

To minimize the processing delay we can implement the voice codecs, packet-scheduling algorithms in the server as the processor speed is not a big problem here. The con of this design is that it will increase the network delay a bit. On the other hand if we keep the voice coders in the cell phones the network delay will

be lowered but processing delay will be increased. So certainly there is a tradeoff has to be made.

In Figure 2 we proposed a network to forward a call from cell phone to the PSTN network. A Transcoding gateway is attached with the previous network and it will do all the necessary translations needed for a call to be forwarded from the packet switched network to the circuit switched network.
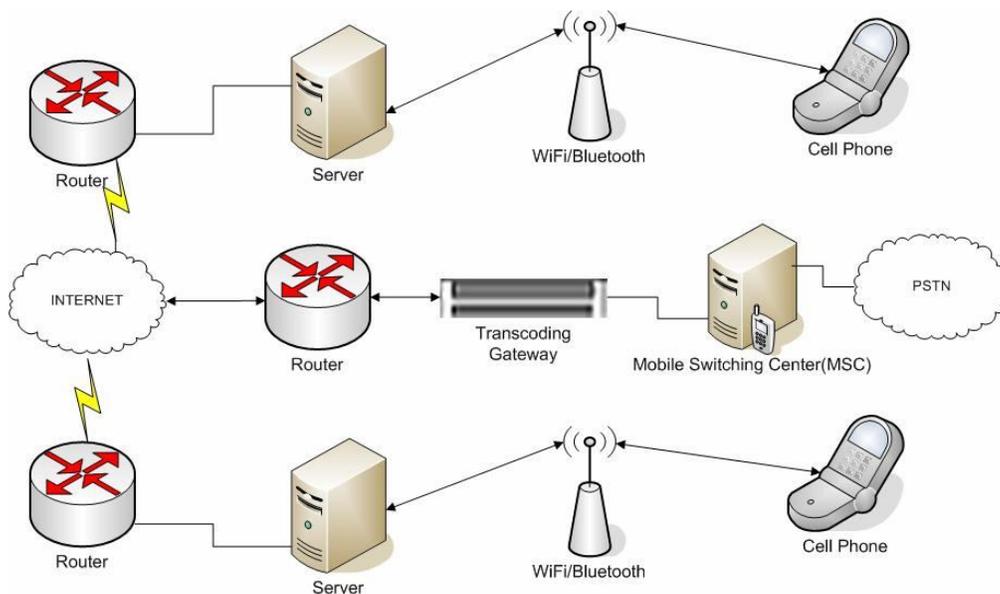

Figure 2 Call forwarding from cell phone to PSTN Network through the Internet.

The other way we can minimize Network Delay is:

1. By reducing the number of router hops.

2. By using Real Time Protocol (RTP).

2.3 Jitter Buffering

The delay problem is compounded by the need to remove jitter, a variable inter-packet timing caused by the network a packet traverses. Removing jitter requires collecting packets and holding them long enough to allow the slowest packets to arrive in time to be played in the correct sequence. This causes additional delay. Two approaches to adapting the jitter buffer size are detailed below. The approach selected will depend on the type of network the packets are traversing [1].

2.3.1 Suggestions:

1. The first approach is to measure the variation of packet level in the jitter buffer over a period of time, and incrementally adapt the buffer size to match the calculated jitter. This approach works best with networks that provide a consistent jitter performance over time, such as ATM networks [1].

2. The second approach is to count the number of packets that arrive late and create a ratio of these packets to the number of packets that are successfully processed. This ratio is then used to adjust the jitter buffer to target a predetermined allowable late packet ratio. This approach works best with the networks with highly variable packet inter-arrival intervals, such as IP networks. In addition to the techniques described above, the network must be configured and managed to provide minimal delay and jitter, enabling a consistent quality of service [1].

2.4 Lost Packet Compensation

Lost packets can be an even more severe problem, depending on the type of packet network that is being used. Because IP networks do not guarantee service, they will usually exhibit a much higher incidence of lost voice packets than ATM networks. In current IP networks, all voice frames are treated like data. Under peak loads and congestion, voice frames will be dropped equally with data frames. The data frames, however, are not time sensitive and dropped packets can be appropriately corrected through the process of retransmission [1].

2.4.1 Suggestions:

Lost voice packets cannot be retransmitted because retransmission wastes time, which contradicts with real-time conversation. We have to minimize the number of lost packets since we cannot wait for a retransmission.

2.5 Echo Filtering

Echo in a telephone network is caused by signal reflections generated by the hybrid circuit that converts between a 4-wire circuit (a separate transmit and receive pair) and a 2-wire circuit (a single transmit and receive pair). These reflections of the speaker's voice are heard in the speaker's ear. Echo is present even in a conventional circuit switched telephone network. However, it is acceptable because the round trip delays through the network are smaller than 50 msec. and the echo is masked by the normal side tone every telephone generates.

Echo becomes a problem in Voice over Packet networks because the round trip delay through the network is almost always greater than 50 msec. Thus, echo cancellation techniques are always used.

Echo is generated toward the packet network from the telephone network. The echo canceller compares the voice data received from the packet network with voice data being transmitted to the packet network. The echo from the telephone network hybrid is removed by a digital filter on the transmit path into the packet network [1].

3. VoIP Protocols

As in one of our network diagrams we have kept a server, we need some VoIP protocols implemented in the server for call management and call signaling to other cell phones or PSTN network.

3.1 RTP

RTP stands for Real time transport protocol. The Internet standard protocol for the transport of real-time data, including audio and video. RTP is used in virtually all voice-over-IP architectures, for videoconferencing, media-on-demand, and other applications. A thin protocol, it supports content identification, timing reconstruction, and detection of lost packets.

VoIP data packets live in RTP (Real-Time Transport Protocol) packets which are inside UDP-IP packets.

Firstly, VoIP doesn't use TCP because it is too heavy for real time applications, so instead a UDP (datagram) is used.

Secondly, UDP has no control over the order in which packets arrive at the destination or how long it takes them to get there (datagram concept). Both of these are very important to overall voice quality (how well you can understand what the other person is saying) and conversation quality (how easy it is to carry out a conversation). RTP solves the problem enabling the receiver to put the packets back into the correct order and not wait too long for packets that have either lost their way or are taking too long to arrive (we don't need every single voice packet, but we need a continuous flow of many of them and ordered) [3].

3.1.1 RTCP

RTCP stands for *Real-time Transport Control Protocol*, provides out-of-band control information for an RTP flow. It partners RTP in the delivery and packaging of multimedia data, but does not transport any data itself. It is used periodically to transmit control packets to participants in a streaming multimedia session. The primary function of RTCP is to provide feedback on the quality of service being provided by RTP.

It gathers statistics on a media connection and information such as bytes sent, packets sent, lost packets, jitter, feedback and round trip delay. An application

may use this information to increase the quality of service perhaps by limiting flow, or maybe using a low compression codec instead of a high compression codec. RTCP is used for QoS reporting.

RTCP itself does not provide any flow encryption or authentication means. SRTCP protocol can be used for that purpose [4].

3.1.2 Suggestions:

To bridge between the cell phone and the RTP and RTCP protocol those are running in the server a middleware application can be implemented in the cell phone. This middleware will do all the necessary works like packaging and delivery of multimedia data to the server.

3.2 SIP and H.323

Session Initiation Protocol (SIP) is a protocol developed by IETF MMUSIC Working Group and proposed standard for initiating, modifying, and terminating an interactive user session that involves multimedia elements such as video, voice, instant messaging, online games, and virtual reality. In November 2000, SIP was accepted as a 3GPP signaling protocol and permanent element of the IMS architecture.

SIP clients traditionally use TCP and UDP port 5060 to connect to SIP servers and other SIP endpoints. SIP is primarily used in setting up and tearing down

voice or video calls. However, it can be used in any application where session initiation is a requirement. These include, Event Subscription and Notification, Terminal mobility and so on. All voice/video communications are done over separate transport protocols, typically RTP.

A motivating goal for SIP was to provide a signaling and call setup protocol for IP-based communications that can support a superset of the call processing functions and features present in the public switched telephone network (PSTN). The SIP Protocol by itself does not define these features, rather, its focus is call-setup and signaling

SIP enabled telephony networks can also implement many of the more advanced call processing features present in Signalling System 7 (SS7), though the two protocols themselves are very different. SS7 is a highly centralized protocol, characterized by highly complex central network architecture and dumb endpoints (traditional telephone handsets). SIP is a peer-to-peer protocol. As such it requires only a very simple (and thus highly scalable) core network with intelligence distributed to the network edge, embedded in endpoints (terminating devices built in either hardware or software). SIP features are implemented in the communicating endpoints (i.e. at the edge of the network) as opposed to traditional SS7 features, which are implemented in the network. [5].

H.323 is an umbrella recommendation from the ITU-T, that defines the protocols to provide audio-visual communication sessions on any packet network. It is currently implemented by various Internet real-time applications such as NetMeeting and Ekiga (the latter using the OpenH323 implementation). It is a part of the H.32x series of protocols which also address communications over ISDN, PSTN or SS7. H.323 is commonly used in Voice over IP (VoIP, Internet Telephony, or IP Telephony) and IP-based videoconferencing

H.323 was originally created to provide a mechanism for transporting multimedia applications over LANs but it has rapidly evolved to address the growing needs of VoIP networks.

One strength of H.323 was the relatively early availability of a set of standards, not only defining the basic call model, but in addition the supplementary services, needed to address business communication expectations. H.323 was the first VoIP standard to adopt the IETF standard RTP to transport audio and video over IP networks.

H.323 is based on the ISDN Q.931 protocol and is suited for interworking scenarios between IP and ISDN, respectively between IP and QSIG. A call model, similar to the ISDN call model, eases the introduction of IP Telephony into existing networks of ISDN based PBX systems. A smooth migration towards IP based PBX systems becomes plannable.

Within the context of H.323, an IP based PBX is, simply speaking, a Gatekeeper plus supplementary services [6].

3.2.1 Suggestions:

SIP or H.323 are used along with RTP for call setup, management and tear down. So the suggested middleware application should also be able to understand SIP and H.323 signaling methods.

4. Limitations

Mobile VoIP is still in its initial stage. So there are lots of limitations.

1. Not much work is done for mobile VoIP.
2. There's no on device testing tools available for testing the applications.
3. The bandwidth of the bluetooth is low (max 3mbps).
4. The processor speed of the cell phones is still not good enough.

5. Conclusion

Voice over IP services on mobile phone offer lucrative advantages to customers and service providers alike. However, as with any new technology, it brings its own sets of network design and optimization issues. By understanding the important parameters, and acquiring the proper tools, we can reap the benefits of voice over packet services on mobile phone.

References:

[1] Texas Instruments, VOIP White Paper Fine - Tuning Voice over Packet Services.

[2] http://www.radisys.com/products/datasheet_page.cfm?ProductDatasheetsID=53

[3] http://en.wikipedia.org/wiki/Real-time_Transport_Protocol

[4] http://en.wikipedia.org/wiki/RTCP

[5] http://en.wikipedia.org/wiki/Session_Initiation_Protocol

[6] http://en.wikipedia.org/wiki/H323