VoIP Mobility Issues

Gábor Bányász, Renáta Iváncsy

Department of Automation and Applied Informatics
and HAS-BUTE Control Research Group
Budapest University of Technology and Economics
Goldmann Gy. tér 3, H-1111 Budapest, Hungary
e-mail: {gabor.banyasz, renata.ivancsy}@aut.bme.hu

Abstract: Voice Over IP (VoIP for short) on wired line networks has gained significant attention in the past few years as the bandwidth of the internet accesses of a custom user is increasing continuously. The expansion of using the mobile environment induces an effort for integrating mobile communication with VoIP. However, integrating mobile environment with VoIP does not mean simply applying the techniques used in wired networks to wireless networks. There are many tasks that have to be solved in order to obtain an appropriate quality of such communications. There are two common mobile transmission techniques that can be used for transmitting data across a mobile telephone network, namely, GPRS and WiFi. This paper deals with the problem of integrating VoIP with mobile environment. It describes the various difficulties one faces when using GPRS or WiFi technologies for this purpose.

Keywords: VoIP, Mobile convergence, NGN networks

1 Introduction

Voice Over IP (VoIP for short) on wired line networks has gained significant attention in the past few years as the bandwidth of the internet accesses of a custom user is increasing continuously. The expansion of using the mobile environment induces an effort for integrating mobile communication with VoIP. However, integrating mobile environment with VoIP does not mean simply applying the techniques used in wired networks to wireless networks. There are many tasks that have to be solved in order to obtain an appropriate quality of such communications.

There are two common mobile transmission techniques that can be used for transmitting data across a mobile telephone network, namely, GPRS and WiFi. However, the characteristics of these media differs significantly from the wired media, thus further researches are needed to be able to deliver real-time media, such as voice or video, on these networks.

This paper deals with the problem of mobile VoIP using SIP. The two basic communication schemas, namely, the GPRS and the WiFi is described and the different problems regarding these techniques are discussed from the VoIP point of view.

The organization of the paper is as follows. Section 2 introduces the basic task a VoIP service and briefly introduces the protocols that are used. This section describes the basic requirements of a VoIP communication as well. In Section 3 the different aspects of using GPRS for VoIP purposes is described in detail. Section 4 deals with the problem of using WiFi for VoIP.

2 VoIP

During the last couple of years VoIP technologies have gained much attention from both researchers and industry. Today VoIP can be described as a rapidly rising technology for voice communication that uses IP-based networks to deploy different VoIP devices and services.

Typical Internet applications use TCP/IP, whereas VoIP uses RTP/UDP/IP. While TCP is a reliable transport protocol that uses acknowledgments and retransmission to ensure packet receipt, UDP provides unreliable connectionless delivery service using IP to transport messages between end points in an internet. RTP is a protocol used for transmitting real-time data, such as audio and video. Its fundamental approach is to deliver the data with low latency and low overhead as quick as it possible allowing also packet loss for this reason. [1,2]

There exist different call signaling protocols like H.323 or SIP [4] call using SIP (Session Initiation Protocol). In this paper we focus on SIP. SIP is a control protocol that can set up and tear down a session. It uses SDP to describe the details of the call. After a call is established the voice to be transmitted has to be coded [6], the analogous data has to be converted into digital; furthermore, it has to be compressed as well.

The great challenge of a VoIP service is to ensure real-time communication between the two end points based on the internet. In this case critical requirements from the service are the following:

- Ensuring low latency
- No jitter
- Enabling packet loss, but its amount has to be kept low.

3 GPRS

Wireless wide-area cellular networks such as GPRS are being widely deployed, promising ubiquitous access to IP-based multimedia applications for end-users. The characteristics of the cellular wireless medium, however, pose several new challenges to audio and video multimedia streaming applications. Cellular systems has several features that distinguishes them from the wired networks and from the relatively stable WLAN (802.11) networks. These features are for example the large and mutable round trip time (RTT), the fluctuation of the bandwidth, connection stuttering and burst packet loss. These influence the data transmission significantly [5,7]. These characteristics of the GPRS do not allow simply applying the SIP/VoIP system to the mobile environment.

Two fundamental problems related to VoIP over GPRS is the quality of the communication and the IP address distribution among the mobile users. The following two subsections of this section deals with these kinds of problems.

3.1 Different Approaches for Improving Quality

When using VoIP on GPRS different problems can be arisen because of the characteristics of the GPRS. These problems are for example the high latency, low bandwidth, the fluctuation of the size of the bandwidth, the burst-like packet loss. These features cause that we could call somebody very slowly and the voice would suffer delay.

In the following three subsections different approaches are discussed that can be used for making the quality of VoIP over GPRS more convenient for the users.

3.1.1 UDP-Lite

The protocol of the internet was designed assuming the following circumstances:

- High bandwidth
- Low latency
- Low error rate

In case of GPRS connection we can take the following features into account:

- Low bandwidth
- High latency
- High error rate

In case of real time voice transfer the latency tolerance is small, while most of the decoding algorithms can handle also the damaged data. By reducing the reliability the latency can be reduced as well. This could be a good solution in such cases like using VoIP.

Such solution is provided by the UDP-Lite [12]. This is a protocol even lighter than the UDP, that uses only partially checksums. It protects the important information like the header, but it does not care about the content of the message [12].

One of the disadvantages of the protocol is that it should be supported by both sides of the connection. The UDP-Lite has its own IP protocol identifier, thus when the destination host is unable to receive such messages, then it sends back an ICMP “Protocol Unreachable” message.

3.1.2 SIP Compression

When establishing the connection then we have to ensure, that the whole packet is received by the receiver. For this reason in case of SIP packet sending UDP or TCP protocol should be used. The reliability at GPRS means sometimes great answering time. Furthermore, the SIP protocol is textual, thus a great amount of data have to be transmitted through the radio connection. This further increases the time of establishing the connection. While in case of a typical GSM system this time is about 3.6 s, in case of SIP protocol this is about 7.9 s. This can be enhanced when using compression algorithms for reducing the data to be transferred [1,5]. Using this solution the time for establishing a connection can be reduced by 1.5 s.

3.1.3 RTP Header Compression

In the real time communication the compressed voice is transferred in small chunks. In case of good quality compression and having small frame sizes the size of the voice information can be 15-20 byte. The size of the headers is significant compared to the size of the useful information. In case of IPv4, the size of the IP header is 20 byte, the UDP is 8 byte and the RTP is 12 byte that is 40 byte at all. When using IPv6, the size of the IP equals to 40 byte, thus the total size of the header is 60 byte.

When investigating these data we can draw the conclusion that compressing the header is a key issue. However, the features of the cellular networks have to be taken into consideration as well, thus algorithms have to be developed that can also deal with the packet loss. The primary goals are to handle the great compression and to have a robust solution. For this reason was developed the Robust Header Compression (ROHC) [8].

The ROHC is a header compression framework that contains different profiles and compression strategies for the different protocol sets. However to satisfy the requirements of the cellular industry for an efficient way to transport voice ROHC is mainly focusing on IP/UDP/RTP header compression. The ROHC RTP became an efficient, robust compression scheme and is able to compress in great percentage. It can compress the header down to one octet. [13]

Although UDP-Lite is similarly to the UDP, but there exist some differences in the behavior of the fields, and uses different protocol identifier, thus UDP-Lite requires a distinct ROHC profile.

3.2 Handling IP Addresses

The NAT (Network Address Translation) is a technique that can be used in order to modify the source or destination address of an IP packet that is passes over a router or a firewall. Most systems use the NAT to enable hosts on the private network to access the Internet using one or some public IP addresses. Thus more end points can access the internet than the number of the public IP addresses is that have a certain user. Using NAT become widespread as the IP addresses of the IPv4 domain was proven to be small, and reserving new domains was expensive.

The number of the GSM users exceeds the number of the IPv4 addresses, thus the provider cannot distribute public IP addresses for its customers, thus it has to use NAT for solve this problem. This arises new problems when using VoIP through GPRS connection.

Both the SIP and the RTP is based on UDP that because of its nature cannot be used through NAT. While the packets can be sent easily from the private addresses to the internet, sending them in reverse direction is a more complicate task. There are many solutions for solving this king of problem, for example the Application Layer Gateways (ALG), the Simple Traversal of UDP through NAT (STUN) [10], the Traversal Using Relay NAT (TURN) [11], and the Interactive Connectivity Establishment (ICE) [9] that is based on the STUN and TURN protocols.

3.2.1 Application Layer Gateways

It could be one of the solutions for address translating when the mobile service provider would install a SIP proxy that can control an RTP proxy on the NAT router. However, this requires the cooperation with the service provider which makes it difficult to be executed in the practice.

3.2.2 STUN

STUN does not provide a general solution for all NAT problems, but it makes possible receiving UDP packets though some NAT implementations.

The four type of NAT observed in the implementations are:

- Full Cone: The full cone NAT maps all requests from the same internal IP address and port to the same external IP address and port. Furthermore any external host can send a packet to the external address which will be mapped and send to the internal host.

- Restricted Cone: The restricted cone NAT works like the full cone NAT, but the external host can send in a packet only if the internal host sent a packet previously to the same external host.
- Port Restricted Cone: The port restricted cone NAT works like the restricted cone NAT, but the restriction includes port numbers too.
- Symmetric: The symmetric NAT is that type of NAT where all requests from the same internal IP address and port to a specific destination IP address and port, are mapped to the same external IP address and port. But if the same internal host sends a packet with the same source port to a different destination IP, then a different mapping is used. Furthermore, only the addressed external host can send back an UDP packet.

STUN does not make possible to receive UDP packets through symmetric NAT but in other cases it can be useful. The main function of STUN is helping the network entity to discover its environment and to identify the type of the NAT in the communication path. Then the network entity can do the needed actions based on the discovered information to control the NAT.

The STUN has a client-server architecture where the client is placed on the private network in the network entity. In case of VoIP applications it is integrated in the VoIP application. The STUN server is placed in general on the public network. The client knows the address of the server in the given domain using the DNS SVR record.

The VoIP device or application firstly discovers its environment. When it starts a call it determines the IP address and the port number that can be sent in the SIP request with the help of the STUN. Using this IP address and port number it can accept the media stream connection. However, it can be happen that pauses are present during the communication meanwhile UDP connection can be lost in the NAT. For this reason keep-alive packets have to be sent periodically to the other application.

In case when the two communication partners are on the same network and the end points are determined using STUN, then the connection is built up using the NAT is unnecessary and in some cases it does not work. For this reason using the STUN should be combined with a direct solution.

3.2.3 TURN

STUN does not provide solution for every firewall or NAT implementations, TURN can be used in these cases. TURN is a protocol still under development. It is also client-server based where the client is placed on the end application, on the private network, and the server is places on the public network. It is like the STUN protocol, however, in case of TURN the server works as a packet transmitter. The packets are received on the TURN server, and they are sent through a new connection to the client that is placed behind a NAT. The answering packets are sent on the same way.

Because TURN inserts a further item into the communication path, furthermore, it modifies the path of the packets with a solution placed in the application layer; it increases the latency and the possibility of packet loss. In the worst case between two endpoints placed on the same communication network the path of the packets could be lead through a long way. Furthermore, operating a TURN server is not cheap, thus using this protocol is advisable only in case when there is no other solution possible.

3.2.4 ICE

ICE (Interactive Connectivity Establishment) is not a new protocol, but a method that discovers the environment, and chooses the most appropriate NAT technology that has to be used for the certain conditions. During its process it builds on the solutions presented before (STUN, TURN) and provides a tool that can be used generally.

3.2.5 Symmetric RTP

The RTP does not define the ports used when sending or receiving data during the communication. Because neither the RTP nor the UDP was developed for two-directional communication, thus they do not deal with this kind of problem. However, most of the firewalls and NAT implementations can only send the packets in both directions properly when the ports used for sending and for receiving are the same. This is called symmetric RTP [14].

3.3 GPRS Solution Summary

It can be seen well that there exist many effort for solving the various problems regarding VoIP over GPRS. However, to guarantee appropriate qualtity of VoIP communication over GPRS is still a great challenge. The most problematic aspect is to keep the latency under a certain limit that can be accepted by the users. Its reason is that until now the “Push to Talk” is the only service that is popular which does not need any real time transfer.

Another technique to connect 2.5G mobile networks with wired VoIP networks is to use GSM VoIP gateways. This solution is used, however, only small companies in order to reduce their costs [15].

4 WLAN

WLAN (Wireless Local Area Network) (also called WiFi) [2,3] is another technology for supporting mobile data transfer. This technology is in general supported by the novel mobile phones. When using WLAN for mobile VoIP communication not the size of the bandwidth raises a problem but switching the different Acces Points (AP). While we are moving with our mobile phones we leave the zone of an AP and get into another AP's zone. The key issue is how to manage the VoIP communication during these events.

We differentiate micro and macro mobility based on the fact whether two APs are in the same domain or not. In the former case we are speaking about micro mobility, while in the latter about macro mobility. In case of micro mobility the mobile phones moves within APs that are operated by the same provider and that APs are in the same domain. Thus there is no need for external coordination when handling the movement. The latency caused by the AP switching depends on the structure of the network. In case of macro mobility the mobile phone moves between two APs that are operated by different providers. In this case the cooperation of the two providers is needed in order to keep on the communication.

Currently, there are two basic approaches to support mobility in VoIP services. The first one seeks to solve mobility in the network layer by using Mobile IP. The other approach is to solve the mobility problem in the application layer by extending existing VoIP protocols such as Session Initiation Protocol (SIP) [].

4.1 Mobility with Mobile IP

The IPv4 assumes that the endpoints are placed on fixed places on the internet. The unique address is bind with the physical location, and it plays an important role in the transmitting of the packets. If the location of the endpoint is changing, and it is connected to the internet through another sub network then it makes the process of the routing mechanism impossible. For reaching the packets their destination, it is necessary that their IP addresses are changed to a valid IP address of the given subnet. However, changing the IP address means disconnecting the current session and establishing a newer one. This is inadmissible in case of a VoIP session. In order to solve this problem Mobile IP [16] is used.

The Mobile IP introduces the following concepts:

- **Mobile node** (MN): the device that is moving between the networks and sub networks, thus it changes the location of the connection point to the internet.
- **Home Agent** (HA): a router in the home network of the mobile node, which forwards the packets to the mobile node when the mobile node resides in foreign network. Furthermore the HA handles the information regarding the location of the mobile node.

- **Foreign Agent** (FA): A router on the network visited by the mobile node, which provides routing services for the mobile node registered on the network.
- **Home Address or Permanent Address**: the permanent IP address of the mobile node, which is a valid IP address in its home network. On the home network directly this home address is used. In this case there is no need for using Mobile IP. When the mobile node is on a foreign network, the home address solves as the source address for the IP datagrams.
- **Care-of Address** (CoA): When the mobile node uses foreign networks, then the CoA is the address inside the domain of the given network. This is the endpoint of the tunnel that is used for forwarding the packets. The address can be an FA address or an address that is assigned to the mobile node.

The MN sends an Agent Advertisement message in order to discover its environment. From the answer it can determine to which network it is connected, to its home network or to a foreign one. When it is connected to its home network then there is no need for using Mobile IP, thus it communicates on the traditional way. When the MN is connected to a foreign network then it requests a care-of-address. The address obtained is then registered in the HA using Registration Request-Registration Replay exchange of message. When the FA supports the communication, then the registration is fulfilled through it.

Further on when a packet is receiving to the home address of the MN then the HA catches it and forwards it through a tunnel to the CoA, from where the packet reaches the MN. The CoA can be assigned to the FA, but it can happen that the MN gets directly a temporary address. In the latter case there is a possibility to use Mobile IP on networks that do not have any FAs and provides only DHCP service.

The packets to the other direction are sent using the traditional routing mechanism to the destination that means that the packet transmission is asymmetric on Mobile IP.

However, using Mobile IP the problem of changing the AP is solved in such a way, that the connection remains, there exist several problems when using it in VoIP environment. Firstly the packets are sent through the HA in the one direction that causes longer path and higher latency. The other problem that increases the latency is the tunneling mechanism that attaches a further header to the packet which enlarges the size of the packet. In real time voice data transmission the size of the packets are in general small compared to the size of the header, thus an additional part to the header can raise problems. The most important drawback of using mobile IP is that each MN requires a fixed IP address that is problematic because of the limited number of IP addresses in IPv4.

4.2 Mobility with SIP

Another issue to handling macro mobility is to solve the problem in the application layer using the SIP protocol. Originally SIP can handle offline roaming, that is the user can choose between the location and the device on which he is accessible. This can be announced by sending a REGISTER message to the registrar server. It means that when the user has not a voice conversation the application registers its address where the user can be accessed.

In order to get a valid address a DHCP service is needed to be operated on the network. The MN receives temporary IP addresses from the DHCP servers when changing the network. By registering this address the user became accessible. When the user changes the network during a voice conversation then it is necessary that also the partner knows about this fact. In this case the MN sends a directly an INVITE message to the partner.

One of the drawbacks of the SIP-based mobility handling is that in this case the amount of the communication is double as high than in case of Mobile IP when requesting for CoA. In addition the time for the ARP communication and for the refreshing increases the latency when changing a network.

4.3 WLAN Solution Summary

Because constructing micro mobility systems is relatively easier than macro mobility system, the expansion of such systems is expected in the near future. Also macro mobility can be handled technically; however, there is a need for serious cooperation between the different providers in handling information about authentication, authorization and accounting (AAA). Thus the expansion of systems like this are expectable only in the far future.

Conclusions

This paper deals with the problem of maintaining VoIP in mobile environment. After introducing the basic steps of a VoIP conversation, the different approaches related to mobile VoIP was described in detail. Two technologies can be used for mobile VoIP, namely, GPRS and WLAN. Both of the technologies have their difficulties that can be solved in different ways. The basic problems detected are the features of these media that does not really fit the requirements of a real-time communication. Another key aspect is the handling of the large number of IP addresses that are raised when using mobile phones as internet end points.

Acknowledgement

This work has been supported by T-Com, the fund of the Hungarian Academy of Sciences for control research and the Hungarian National Research Fund (grant number: T042741).

References

[1] M. Fidrich, V. Bilicki, Z. Sogor and G. Sey, SIP Compression, 2002, http://citeseer.csail.mit.edu/637425.html

[2] T. T. Kwon, M. Gerla and S. Das, Mobility management for VoIP service: Mobile IP vs. SIP, **IEEE Wireless Communications, Vol. 9(5), ISSN 1536-1284**, 2002. pp. 66-75

[3] http://www.iec.org/online/tutorials/ti_voip_wlan/ Including VoIP over WLAN in a Seamless Next-Generation Wireless Environment

[4] J. Rosenberg, H. Schulzrinne, G. Camarillo et al, SIP: Session Initiation Protocol, RFC3261, http://www.ietf.org/rfc/rfc3261 (SIP)

[5] H. Hannu, Signaling Compression (SigComp) Requirements & Assumptions, RFC3322, http://www.ietf.org/rfc/rfc3322

[6] B. Goode, Voice Over Internet Protocol (VoIP), **Proceedings of the IEEE, Vol. 90(9),** 2002, pp.1495-1517

[7] J. Chesterfield, R. Chakravorty, I. Pratt, S. Banerjee and P. Rodriguez, Exploiting Diversity to Enhance Multimedia Streaming Over Cellular Links, *IEEE INFOCOM 2005*

[8] C. Bormann, C. Burmeister, M. Degermark et al, RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed, RFC3095, http://www.ietf.org/rfc/rfc3095

[9] J. Rosenberg, Interactive Connectivity Establishment (ICE): A Methodology for Network Address Translator (NAT) Traversal for Multimedia Session Establishment Protocols, http://www.jdrosen.net/papers/draft-ietf-mmusic-ice-06.txt

[10] J. Rosenberg, J. Weinberger, C. Huitema and R. Mahy, STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs), RFC3489, http://www.ietf.org/rfc/rfc3489

[11] J. Rosenberg, R. Mahy, C. Huitema, Traversal Using Relay NAT (TURN), http://www.jdrosen.net/papers/draft-rosenberg-midcom-turn-08.txt

[12] L-A. Larzon, M. Degermark, S. Pink et al, The Lightweight User Datagram Protocol (UDP-Lite), RFC3828, http://www.ietf.org/rfc/rfc3828

[13] G. Pelletier, RObust Header Compression (ROHC): Profiles for User Datagram Protocol (UDP) Lite, RFC4019, http://www.ietf.org/rfc/rfc4019

[14] D. Wing, Common Local Transmit and Receive Ports (Symmetric RTP), http://www.ietf.org/internet-drafts/draft-wing-behave-symmetric-rtprtcp-01.txt

[15] J. D. Gibson and B. Wei, Tandem Voice Communications: Digital Cellular, VoIP and Voice over Wi-Fi, *Global Telecommunications Conference 2004 (GLOBECOM '04)*, *Vol. 2*, 2004, pp. 617-621

[16] C. Perkins, IP Mobility Support for IPv4, RFC3344, http://www.ietf.org/rfc/rfc3344