A Cost-efficient Approach: Putting SIP, NAT, and Mobile IP All Together

Shan-yung Yang
b90095@csie.ntu.edu.tw

Hung-wei Tseng
r92022@csie.ntu.edu.tw

Shi-wu Luo
d89015@csie.ntu.edu.tw

Ai-chun Pang
acpang@csie.ntu.edu.tw

Chia-lin Yang
yangc@csie.ntu.edu.tw

Tei-wei Kuo
ktw@csie.ntu.edu.tw

Department of Computer Science and Information Engineering,
National Taiwan University,106, Taipei, Taiwan

November 18, 2003

1. Abstract

Voice over IP (VoIP) is a attractive alternation of traditional circuit-switched telephony for lower cost on operating and equipments, and advantages provided by IP. Internet Protocol version 4 (IPv4) is the most popular network layer protocol in current Internet. However, the 32-bits IPv4 address space is insufficient, and IP mobility support for IPv4 is poor. Internet Protocol version 6 (IPv6) fixes both the problems of IPv4, but the cost of new equipments is high. Network Address Translator (NAT) is a temporary way to solve the problem of insufficient address space before the popularity of IPv6, though the problem of IP mobility is still not concerned. Session Initiation Protocol (SIP) is widely used in internet telephony, but SIP does not consider the use of NAT. In this work, we propose a Voice over IPv4 solution by using NAT, and IP mobility is also concerned. An Application Layer Gateway is included to solve additional problems in the proposed architecture. The feasibility of our solution is proved by phyical implementation, and the delay of voice is also insignificant.

Keywords: Internet Telephony, Session Initiation Protocol, Network Address Translator, Application Layer Gateway, IP Mobility.

2. Introduction

As the growing of Internet, Voice over IP (VoIP) becomes a charming alternation of traditional circuit-switched telephony for serveral advantages such as lower cost on operating and equipments, the possibility of integration of voice and application, and the universal of IP. By using VoIP system, even Personal Digital Assistants(PDAs) and notebook computers can provide functions just as traditional circuit switched telephone, and may provide additional interest functions integrated by applications run on these platforms.

Internet Protocol version 4 (IPv4) is the most popular network layer protocol in current Internet. IPv4 does not considered the scenario of IP mobility, user may change his physical attachment of Internet. The change of attachment of Internet will break all the connections based on IP without mobility support. As the popularity of mobile devices such as notebook computers and PDA, the demand of IP mobility support is growing. RFC 2002 defines mobility support of IP. By the use of Home Agent (HA), Foreign Agent (FA), and care-of address, IP mobility can be achieved. Therefore, the node can change its physical attachment without breaking of current connection.

However, the mobility support for IPv4 is still poor, and the 32 bits IPv4 address space is insufficent for increasing demand of IP addresses. Internet Protocol version 6 (IPv6) fixes both the problems of IPv4. IPv6 has 128 bits address space, and mobility is concerned when IPv6 draws up. But the cost of new equipments those supports IPv6 and the overhead of tunnelling while communicating with IPv4 devices would be nightmares. Network Address Translator (NAT) is a temporary way to solve the problem of insufficient address space before the popularity of IPv6, though the problem of IP mobility is still not concerned.

Session Initiation Protocol (SIP) is a popular protocol for internet telephony signaling system

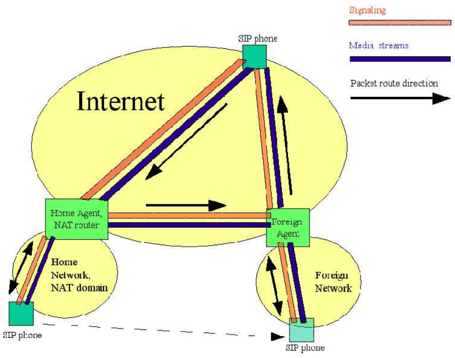

Figure 1: Basic Concept of Our System Architecture

The basic network scenario: the home agent performs the function of NAT router also. While SIP phone mobile to a foreign network, the packets can be tunneled by HA to FA.

that handles the call setup procedure. SIP can be combined with Session Description Protocol (SDP), so that the media stream connections can also be setup by SIP signaling system. However, the design of SIP does not consider the scenario when SIP phones are inside NAT domain.

In this work, we propose a Voice over IPv4 solution by using NAT, and IP mobility is also concerned. We also implement the environment of proposed architecture. An Application Layer Gateway is also included to solve additional problems in the proposed architecture. The feasibility of our solution is proved by phyical implementation, and the delay of voice is also insignificant.

The paper is organized as follows: Section 2 describes the network scenario. Section 3 contains the basic concepts of SIP, NAT, and Mobile IP. Section 4 idicates problems on our basic architecture. Section 5 is our solution, and section 6 concludes the paper.

3. Network Scenario

In this work, we are going to build up an internet telephony system with limited IP addresses. The basic concept of our system architecture is illustrated in figure 1. SIP is the protocol that used by our internet telephony nodes for signaling such as call setup, call forwarding, and etc.

The user terminals in our internet telephony network can be PDAs, notebooks, or normal SIP telephones with wireless network connections. These nodes are possible to run SIP user agent on them. IP mobility support enables the mobile users to roam between wireless access points, just

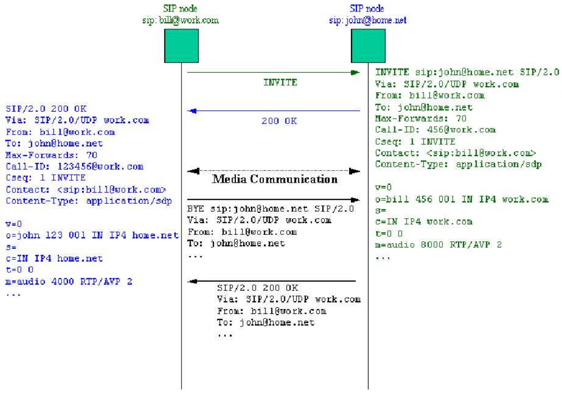

Figure 2: An Simple Example of SIP Session

like they are using cellular phones.

Because of the insufficient number of IP addresses in IPv4, IPv6 is proposed as a solution of this problem. However, IPv6 network is not widely deployed, and the cost of building up a network that support IPv6 protocol is too expensive. Fortunately, NAT that tries to provide sufficient IP addresses over IPv4 seems to be a feasible choice under the current scenario. By using NAT, each node may not have global routable IP addresses except network routers.

Besides, IP mobility is also concerned in our work. Each node may change its physical attachment to the internet, and this will also change the packet route of such node. In order to keep the established telephone connection alive, mobile IP is adopted to enable seamless attachment migration.

4. Background

4.1 Session Initiation Protocol (SIP)

SIP is a signaling protocol that handles the internet telephony call setup, modification, and teardown of multimedia sessions. SIP, mostly combined with session description protocol (SDP), is used to describe the session characteristics to potential session participants. In comparison with triditional telecommunication protocols, SIP is mainly designed for flexibility. For example, SIP uses text based messages, so it's easy to build custom features. SIP also does not care what type of media is to be exchanged during a session, as well as the type of transport to be used for the media.

Figure 2 shows an example of SIP session. SIP nodes communicate to each other by SIP request messages and response messages. The basic SIP message is composed by one start-line and several message-headers. The start-line defines whether this message is a request or response message. For example, the "INVITE" performs a request for call setup, and "200 OK" performs the response for request message such as "INVITE". The message-headers describe information regarding the request or response. For example, "To:" indicates the callee of the request, and "From:" indicates the originator of such request. Besides, "Contact:" and "Via:" are also important headers in our work. "Contact:" provides where the user can be reached for further communication, and "Via:" indicates the path taken by the request so far. In SIP architecture, a SIP proxy may also be included for the use of call forwarding, time-of-day routing, or follow-me services. "Contact:" header is helpful for performing these functions.

In internet telephony, the setup of media connection is also important. SIP protocol uses SDP as its message body for providing the information about media connection setup. SDP is also text based. SDP is composed of descriptions of media streams. These descriptions are like numbers of parameters those indicate the media type, media server address, port number, transport protocol and media format. For example, the "c=" field provides connection type, network type and connection address, and the "m=" field stands for the media type, the transport port and the media format.

SIP and SDP works well in general network. While NAT is used, some problems will appear. We will describe these problems in the following sections.

4.2 Network Address Translator (NAT)

Because of the limitation of traditional IPv4 addresses, NAT is proposed as a short-term solution. In the most common scenario, the router of a stub domain acts as an address translator. (A stub domain is a domain, such as a corporation network, that only handles traffic originated or destined to hosts in the domain). The translator keeps a table consisting of pairs of local IP addresses and globally unique IP addresses. The local IP addresses inside the stub domain are not globally unique, and can be reused in other stub domains.

In general, there are four kinds of NAT implementation, including full cone, restricted cone, port restricted cone and symmetric. They differs from each other in how the IP addresses and port numbers been mapped. For example, a full cone NAT is one where all requests from the same internal IP address and port are mapped to the same external IP address and port. So any external host can send a packet to the internal host in a full cone NAT domain, by sending a packet to the mapped external address. But a symmetric NAT is one where all request from the same internal IP address and port, to a spiecific destination IP address and port, are mapped to the same external IP address and port. So in symmetric NAT, only the external host that recieves a packet can send a packet back to the internal host.

4.3 IP Mobility

When IP mobility is concerned, we regard the node that may change their physical attachment as a mobile node (MN), and the node to which the mobile node is connecting as a corresponding node (CN). We also regard the original subnet of the mobile node as the home network, and the newly attached network as the foreign network. In order to keep the established connection alive, the router of the home network acts as a home agent (HA) that forwards packets to the foreign network by IP tunneling. The mobile node in the foreign network then extracts the IP-in-IP packets so that the original IP address is still available.

When the foreign network is a stub domain using reusable IP addresses, the router of the foreign network, which is called foreign agent (FA), is required. In such a scenario, the home agent cannot establish IP tunnel directly with the mobile node, and the foreign agent is used as the tunnel peer. The foreign agent then extracts the IP-in-IP packets and send them to the mobile node.

If the mobile node is in some foreign network, triangle routing may be adopted. Packets destined to mobile node are sent through the home agent, but those destined to corresponding node are sent directly through the internet, as illustrated in figure 1. Triangle routing reduces the cost of sending packets from the mobile node to its corresponding node, but may leads to some routing problems. Since the source address of packets sent by the mobile node are still the address in the home network, these packets may be regard as IP spoofing packets and dropped by some routers.

5. Problems and Related Works

Unfortunately, SIP does not consider interworking with NAT. In original design of SIP, all SIP nodes will have a global routable address, so they can transmit their media stream to each other directly. When the node assigned with reusable IP in a stub domain tries to establish media connection to some node outside, it cannot recieve signal and media from the outside node because it leaves its IP address and port number used in NAT domain. The address this node used to communicate with other nodes is translated by the NAT box.

We also want to make the nodes inside the stub domain be able to recieve calls. Since they are not using global routable addresses, they cannot recieve any requests sent by outside nodes in general case.

There are indeed solutions, which extends the SIP protocol, to make the nodes in NAT domains being able to communicate with outside nodes. However, these extensions are still in drafts and have not become standards yet. Furthermore, these extensions do not solve the problem that outside nodes cannot send requests to the nodes inside NAT domains.

Other solutions, like STUN protocol, allow applications to discover the presence and types of NATs and provide the ability for applications to determine their public IP address allocated by the NAT. But this solution does not fit in all types of NAT. In a symmetric NAT domain, the address/port mapping may be different because the node is communicating with other outside nodes, not the STUN server.

To sum up, there are two problems: under the SIP protocol, the nodes inside NAT domains cannot setup call sessions with outside nodes, neither can they recieve calls from outside nodes.

6. Our Approach

6.1 Solution

The basic approach that enables the nodes inside NAT domains to communicate with outside nodes is to adopt application layer gateway (ALG) on the routers of NAT domains. In our work, the ALG can be considered as a SIP proxy server with some additional functions for solving the problems in predefined network scenario. The new architecture that adopts ALG is illustrated in figure 3. This solution does not extend the SIP protocol, and only requires the nodes inside the NAT domain to make use of the SIP proxy.

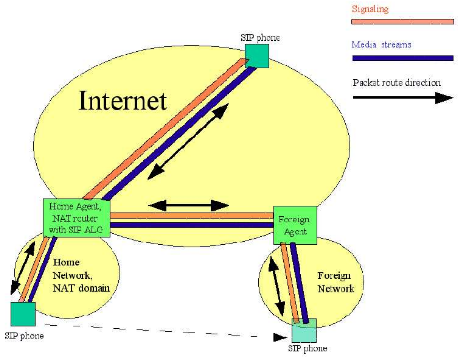

Figure 3: New Architecture Adopting ALG
The architecture modified: an ALG is added, and also notice that triangle routing is not allowed.

When a node inside the NAT domain wants to send request signals, it first send the request packet to the SIP proxy, i.e. the ALG on the NAT router. The ALG on the NAT router, as a SIP proxy, have some functions similar to the NAT box. When the ALG recieves a request, it first allocates two socket pairs, one for SIP signaling and one for media communication. Each socket pair, behaving as socket pairs in the NAT scenario, is used to translate local IP addresses to global unique addresses. The socket pair for signaling pass the message modified by ALG to the nodes inside or outside NAT, and the socket pair for media communication simply recieve media packets and forward them to the nodes inside NAT.

The ALG forwards the outgoing packets by the socket allocated on the NAT router, which has a global unique IP address. Since the outside node regard the ALG as the session initiator, the incoming packets ,including the response signal and media, are sent back to the ALG. Finally these packets are forwarded to the nodes inside the NAT domain.

When the ALG forwards the SIP signals, some of the SIP headers and option fields of SDP description are modified. Since SIP uses the "Contact:" header to inform the callee where the callee should send the response, it should be replaced by the signaling socket address/port allocated by the ALG. Besides, SDP uses "c=" field to provide media connection information which includes connection address, and "m=" field to indicate the media type and transport port. These two option fields should also be replaced by the other socket address/port allocated by the ALG.

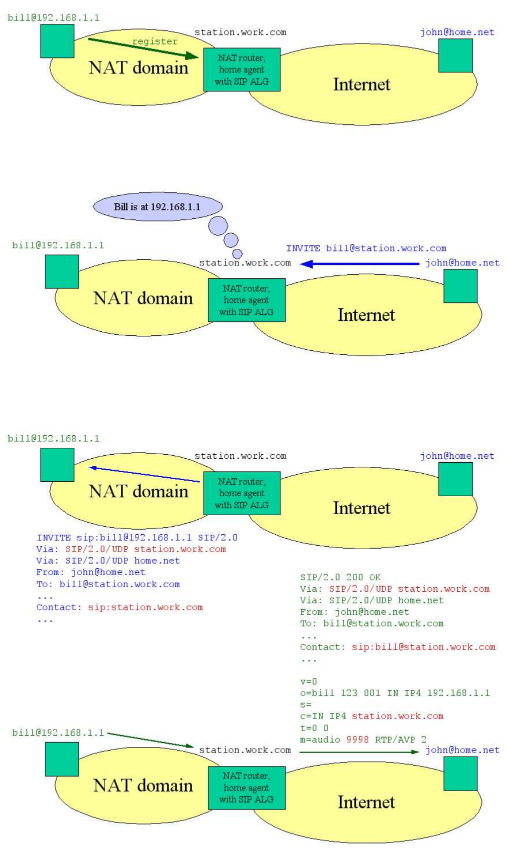

Figure 4: Session Establishment for Nodes Inside NAT Domains.

The mechanism which enables the outside nodes to communicate with the nodes inside the NAT domain is similar. After a node inside the NAT domain registers with the ALG by a unique username, it is ready to be called. The session establishment process is illustrated in figure 4. When the ALG recieves requests of some outside node, it first finds out which node inside the NAT domain corresponds to the username specified in the request line. Therefore, as same as the above method, the ALG allocates two socket pairs, one for SIP signaling and one for media communication. The SIP contact header, as well as "c=" and "m=" fields in SDP description, are also replaced by the socket address/port allocated by the ALG.

IP mobility is well supported in IPv6, and the architecture of mobile IPv6 is simple. The 128 bits address space of IPv6 also provides sufficient global routable IP addresses for each mobile nodes. However, IPv6 is not popular in current network. IPv4 still dominates the internet, and it's impossible to upgrade all devices to IPv6 in a short period. Therefore, we still use IPv4 with mobility support. Since SIP ALG solves the problems of SIP under NAT, using NAT over mobile IPv4 becomes a feasible solution for constructing SIP internet telephony network with IP mobility support.

In our solution, there's no need to modify FA except that reverse tunnelling is forced. Since all signaling messages need to be rewritten, they must be sent through the ALG which is installed on the home agent. This also prevents the routing problem we mentioned in section 4.3.

6.2 Implementation and Experiment

We implement SIP ALG with Perl, a programming language that well support for text processing and network functions. A well functioned SIP user agent with the ability to encode voice with size less than 8Kbps is installed on notebooks so that these notebooks function like SIP phones. In the aspect of IP mobility support, we use Dynamics 0.8.1 on HA, FA, and mobile nodes (the notebooks). Both HA and FA are Linux based. ALG and NAT are run on HA.

After such devices are installed, we can successfully setup SIP session between nodes. The delay of voice is insignificant even when nodes roam to foreign networks and IP-in-IP tunnels are constructed.

6.3 Comparing with STUN

In comparision with the solution using STUN protocol, SIP ALG has both advantages and drawbacks. ALG solution needs to embed additional functions on routers of NAT domains, and this does not fit if the NAT function is embeded into wireless access points or ADSL routers.

However, STUN solution needs to add NAT discovry function to all SIP applications. When the application is run on a simple telephone with wired network connected, it may be impossible to perform software upgrade. Besides, STUN solution does not fit in all types of NATs. It is not a reliable way to ensure the ability that the nodes inside all NAT domains communicate with outside nodes. Finally, STUN protocol does not solve the problem that the nodes inside NAT domains cannot recieve requests from outside nodes.

7. Conclusion

IPv4 has been used for several decades. New applications enhance the convenience of internet, but also bring new challenges to IPv4. Internet telephony is one of these applications. For example, the addresses needed to build up a telephone system are much more than IPv4 provided. IPv6 solved some problems in current IPv4. However, the popularity of IPv6 takes a long time, and the cost of purchasing IPv6 devices is still high. NAT is used in many places where the number of IP addresses is limited to provide sufficient IP addresses before the popularity of IPv6. The IP addresses provided by NAT are not globally routable, that is, some applications those need two way communication like internet telephonys will have problems. We use application layer gateway, a simple and small program on NAT router, to solve these problems without any modification of internet telephony user agent and internet phone architecture. As the growing of portable devices, the demand of IP mobility grows, too. With the help of mobile IPv4 drivers, home agents and foreign agents, IP mobility can be achieved. Triangle routing should be disabled under this approach, just like GPRS.

The delay of voice is acceptable by users though it still takes a little time break while handoff. The small break of handoff might be eliminated by modifying OS kernel to have better mobility support, or multicasting packets to foreign agents. Our approach is a short-term solution over IPv4, a very old protocol. There are still many other limitations for applications like internet teleohone in the IPv4 world, and much research still has to be conducted to find out optimized solution under such networks.

References

[1] K. Egevang, P. Francis, *The IP Network Address Translator (NAT)*, IETF RFC 1631, May 1994

[2] D. Forsberg, J.T. Malinen, J.K. Malinen, T. Weckström, M. Tiusanen, *Distributing Mobility Agents Hierarchically under Frequent Location Updates*, Sixth IEEE International Workshop on Mobile Multimedia Communications, San Diego 1999.

[3] M. Handley, V. Jacobson, *SDP: Session Description Protocol*, IETF RFC 2327, April 1998

[4] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg, *SIP: Session Initiation Protocol*, IETF RFC 2543, March 1999

[5] C. Perkins, *IP Mobility Support*, IETF RFC 2002, October 1996

[6] C. Perkins, *IP Mobility Support for IPv4*, IETF RFC 3344, August 2002

[7] J. Rosenberg, H. Schulzrinne, *An Extension to the Session Initiation Protocol (SIP) for SymmetricResponse Routing*, <draft-ietf-sip-symmetric-response-01.txt>, work in progress, June 2003.

[8] J. Rosenberg, J. Weinberger, C. Huitema, R. Mahy, *STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)*, IETF RFC 3489, March 2003