[1]Design of SIP Transformation Server for Efficient Media Negotiation

Sangheon Pack, EunKyoung Paik, and Yanghee Choi

Multimedia & Communication Laboratory, Seoul National University, Korea

ABSTRACT

Voice over IP (VoIP) is one of the advanced services supported by the next generation mobile communication. VoIP should support various media formats and terminals existing together. This heterogeneous environment may prevent diverse users from establishing VoIP sessions among them. To solve the problem an efficient media negotiation mechanism is required.

In this paper, we propose the efficient media negotiation architecture using the transformation server and the Intelligent Location Server (ILS). The transformation server is an extended Session Initiation Protocol (SIP) proxy server. It can modify an unacceptable session INVITE message into an acceptable one using the ILS. The ILS is a directory server based on the Lightweight Directory Access Protocol (LDAP) that keeps user's location information and available media information.

The proposed architecture can eliminate an unnecessary response and re-INVITE messages of the standard SIP architecture. It takes only 1.5 round trip times to negotiate two different media types while the standard media negotiation mechanism takes 2.5 round trip times. The extra processing time in message handling is negligible in comparison to the reduced round trip time.

The experimental results show that the session setup time in the proposed architecture is less than the setup time in the standard SIP. These results verify that the proposed media negotiation mechanism is more efficient in solving diversity problems.

Keywords: VoIP, SIP, Media Negotiation, Intelligent Location server, Round Trip Time, Media Gateway, LDAP

1. INTRODUCTION

The advent of IP network technologies and media coding schemes brings a new service called Voice over IP (VoIP). VoIP is a technology that provides a voice call service over IP network. VoIP service can provide customers with a cost effective call service and various integrated multimedia services [1].

To develop more scalable and flexible VoIP services, many factors must be considered. Among many factors, media encoding schemes and network bandwidths are the core factors to support suitable Quality of Service (QoS) to customers. So many media encoding schemes and networks with different bandwidths exist in current VoIP systems. For example, there are International Telecommunication Union-Telecommunications (ITU-T) G.7xx series for voice formats[2] and ITU-T H.261[3], H.263[4] and MPEG-4[5] for video format. And there are Asymmetric Digital Subscriber Line (ADSL), WaveLAN -II and ISDN as available networks in the VoIP system.

Due to these various technologies, we have faced some diversity problems. There are two dominant diversity problems in mobile communication supporting VoIP service: terminal diversity problem and access network diversity problem. The former problem occurs when the users use various VoIP terminals with different encoding schmes for more suitable communication. And the latter problem means that users can use VoIP services through various access networks with different bandwidths. For example, one user with a Personal Digital Assistant (PDA) supporting only the H.263 format wants to use VoIP services through ISDN line. And the other user with a Desktop PC supporting only the MPEG-4 format uses VoIP services through WaveLAN-II. In this case, since the available encoding schemes and network bandwidths are different, a VoIP session between them cannot be established right now and they cannot communicate using VoIP services. To establish a VoIP session in this case, diversity problems must be resolved by media negotiation mechanism.

[1] This work was supported in part by the Brain Korea 21 project of Ministry of Education and in part by the National Research Laboratory project of Ministry of Science and Technology, 2001, Korea.

The diversity problems will become more serious in the 4th generation mobile communication based on IP technology. Because the more various encoding schemes and access technologies will be supported in the 4th generation mobile communication environment for the advanced VoIP services. Therefore efficient media negotiation mechanism is a critical issue in the future VoIP system.

We adopted the Session Initiation Protocol (SIP)[6] as a signaling protocol in our VoIP architecture. It is designed as a signaling protocol for establishing, modification, and termination of multimedia sessions in Internet Engineering Task Force (IETF)[7]. It is a simple, flexible, and extensible protocol and can be used for the various multimedia services so that we have designed the VoIP architecture with SIP[8].

According to the media negotiation mechanism in the standard SIP, only the callee who receives a request for VoIP session initiation can detect diversity problems of encoding schemes or available bandwidths. Therefore, if any diversity problem occurs, the session cannot be established. Then the caller modifies the session initiation message and re-send the modified message to the callee. Because of this retransmission, it takes 2.5 round trip times to negotiate media information between the caller and the callee. It requires unnecessary round trip time for media negotiation.

In this paper, we propose the efficient media negotiation architecture using the SIP-based transformation server. To reduce the media negotiation time, the transformation server uses the Intelligent Location Server (ILS) that keeps the user's location information, the available encoding schemes and so on. Using information in the ILS, the transformation server can determine whether the diversity problems occur and the media negotiation is needed or not.

The rests of this paper are organized as follows. Section 2 briefly describes the SIP architecture and the media negotiation mechanism in the standard SIP architecture. Section 3 proposes a new media negotiation mechanism based on the SIP-based transformation server and the ILS. And we will explain that this architecture can integrate easily with the general SIP-based VoIP architecture and it has an advantage to resolve the diversity problems. Section 4 summarizes the experimental results of the proposed mechanism and analyzes the performance of the proposed mechanism in session setup time. Finally, Section 5 concludes the paper.

2. SIP OVERVIEW

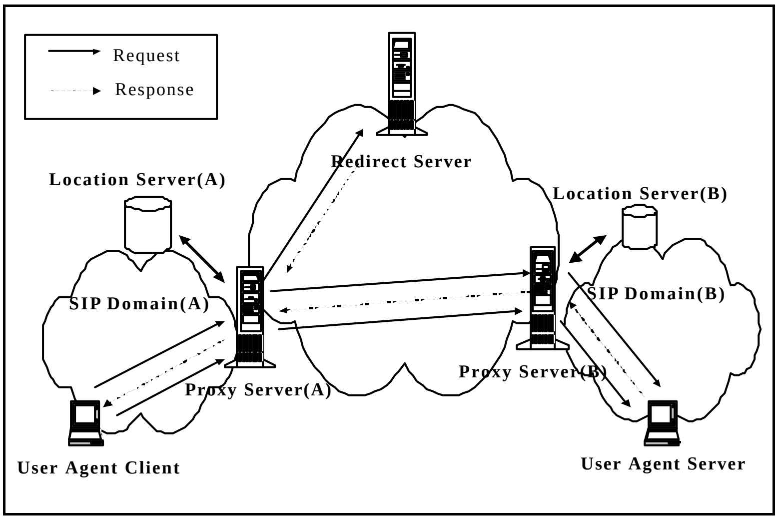
SIP is an application-layer signaling protocol for establishing, modifying, and terminating multimedia sessions. These sessions include VoIP calls, Internet multimedia conferences and so on. There is another signaling protocol, H.323[9] in the VoIP system. Although H.323 is a dominant signaling protocol in the commercial VoIP markets, SIP provides more scalable, extensible and simple features than H.323. Because of these reasons, the VoIP systems using SIP as a signaling protocol are growing more and more. Besides, 3rd Generation Partnership Project (3GPP)[10] and 3rd Generation Partnership Project 2 (3GPP2)[11], organizations for standardization of the next generation mobile communication, have chosen SIP as a control protocol for multimedia sessions including VoIP services.

2.1 The Standard SIP Architecture

The standard SIP architecture is composed of following entities: a user agent client (UAC), a user agent server (UAS), and SIP servers. UAC is a client application that initiates a SIP request. On the other hand, UAS is a server application that contacts the user when a SIP request is received and that returns a response on behalf of the user. The response accepts, rejects or redirects the request. There are a proxy server, a redirect server, and a location server in SIP servers. A proxy server is an intermediary program that acts as both a server and a client for the purpose of making requests on behalf of other clients. A redirect server is a server that accepts a SIP request, maps the address into zero or more new addresses and returns these addresses to the client. A location server is used by a SIP redirect or proxy server to get a user's location information.

SIP is a client-server protocol. This means that requests are generated by a caller entity (UAC) are sent to a receiving entity (SIP server or UAS) which processes them. The UAS generally responds to the requests based on human interaction or some other kind of input. Furthermore, SIP requests may traverse many SIP servers. In SIP proxy server, the server receives a request and forwards it towards a next hop server, which may be another SIP server or the destination UAS. On the other hand, a SIP redirect server informs the client of the address of the next hop server, so that the client can contact it directly.

The basic and simple SIP architecture is composed of UAC, UAS, outbound proxy server, inbound proxy server, redirect server, and location server, as shown in figure 1. The procedures for SIP session establishment are as follows: First, UAC sends INVITE request to initiate a session. A successful SIP invitation consists of two requests, INVITE followed by ACK. The INVITE request asks the callee to join a particular VoIP session. The callee responds to the invitation request by sending response message with status code. In SIP, hierarchical status codes are defined for response messages. In Figure 1, "200 OK" and "302 temporarily moved" responses are used. 200 and 302 are status codes to inform the processing results of request message. More detail status code can be found in RFC 2543. After the callee has agreed to participate in the call with response having "200 OK" response, the caller confirms that it has received that response by sending an ACK request. If the caller no longer wants to participate in the call, it sends a BYE request instead of an ACK.



< Figure 1: SIP architecture >

2.2 SIP Messages and SDP

SIP is similar to the Hypertext Transfer Protocol (HTTP)[12] model of clients, origin and proxy servers. As in HTTP, the user agent client's requests invoke several methods on the server. So SIP messages are divided as request message and response message. Requests and responses are encoded as textual, and contain many header fields, which convey call properties and service information. SIP reuses many of the header fields uses in HTTP, such as the entity headers (e.g., Content-type) and authentication headers. This allows for code reuse, and simplifies integration of SIP server with web servers.

A SIP message is able to have a body part in Session Description Protocol (SDP)[13]. SDP is designed to describe multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation. SDP describes not only the session information but also media information like an available media types. In the standard SIP, this media information is only used for the media negotiation and never changed by the intermediate SIP server. However, in the proposed architecture, it may be changed by the transformation SIP server for the more efficient and faster media negotiation. The proposed media negotiation mechanism will be described in next section 3 in more detail.

2.3 The Media Negotiation Mechanism in the standard SIP

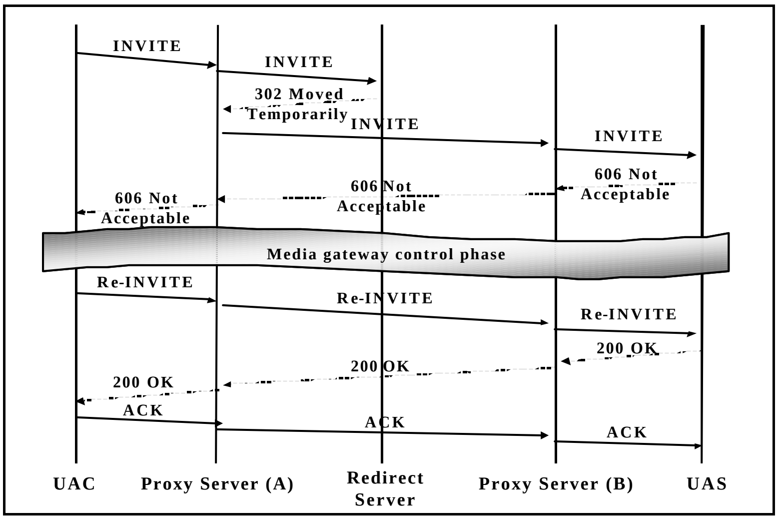
The media negotiation mechanism to resolve diversity problems is divided media conversion phase and session setup phase. The media conversion phase is done by a media gateway[14]. It is an intermediate device located between the caller and the callee. It has a simple ability to convert one media type to another media type. For example, a media gateway is able to convert G.722 format [15] to G.728 format [16]. Many device vendors have developed this media gateway to support various media types in the VoIP system. The media gateway is controlled by some protocols like a Media Gateway Control Protocol (MGCP)[17], MEdia GAteway Control (MEGACO)[18] and so on. The session setup phase is possible only when the media conversion is possible by a media gateway. As mention it, SIP is a signaling protocol for session setup.

In this paper, we assumed that media gateways, having media conversion capabilities, already exist in the VoIP architecture. So we focused on the efficient mechanism for the session setup phase, not for the media conversion phase.

The general media negotiation mechanism is already defined in the standard SIP. This mechanism can be summarized as follows.

(1) A caller sends INVITE request message to a user for initiation of VoIP session. The request contains SDP part describing the caller' s media information and session information.
(2) A UAS received INVITE request determines whether the described media types can be supported by the callee.
(3) If the UAS can support the request' s media types, it responds to the UAC with "200 OK" response. After receiving response, the UAC sends ACK request to the UAS and then a new session is established.
(4) If the UAS cannot support the request' s media types, it responds to the UAC with "606 Not Acceptable" response. Because the media conversion using media gateway is available in our architecture, the UAC modifies SDP part of the previous request and sends modified INVITE message to the UAS again. If the new INVITE message can be accepted by the UAS, the UAS sends "200 OK" response to the UAC and a session is established finally by sending ACK request to the UAS.
(5) Therefore the media negotiation in the standard SIP requires 2.5 round trip times.

Figure 2 shows the message flows in this media negotiation mechanism. This example used the simple SIP architecture in figure 1.



< Figure 2: Media Negotiation in the standard SIP >

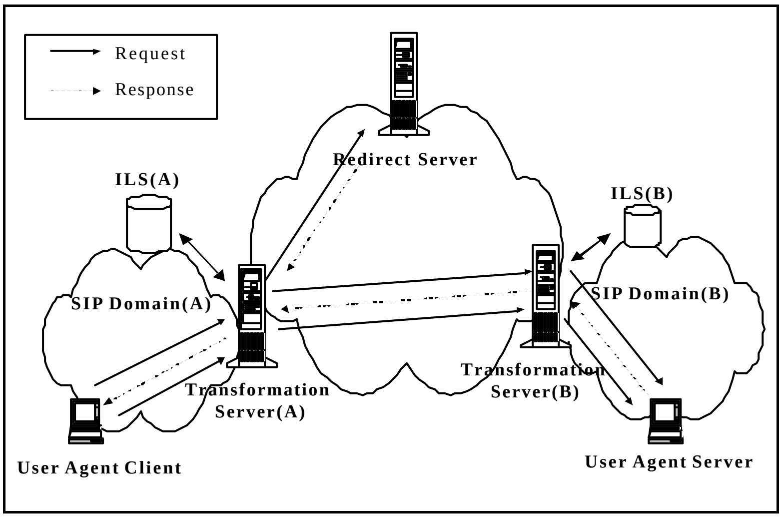
According to above message flows, this mechanism requires inefficient re-INVITE and it takes 2.5 round trip times to negotiate media type between the users with different media types. Because only UAS can determine whether the media types of current INVITE request can be accepted by UAS or not, these long negotiation times is needed.

3. SIP-BASED TRANSFORMATION SERVER AND ILS

The media negotiation in the standard SIP requires 2.5 round trip times as shown in figure 2. However, if the media information of the callee is available before UAS gets INVITE message and transforms the request into the acceptable form, round trip time for media negotiation will become much shorter. Based on this idea, we proposed the new SIP architecture. The proposed SIP architecture improves the media negotiation mechanism and resolves diversity problems. It is composed of Intelligent Location Server and the SIP-based transformation server.

3.1 The Architecture of SIP-based Transformation Server and ILS

Figure 3 shows the new SIP architecture supporting the proposed media negotiation mechanism. In comparison to the figure 1, there are two changes in the new SIP architecture: First, the inbound/outbound proxy servers in a local domain are replaced with SIP-based transformation server. Second, the general SIP location servers are replaced with Intelligent Location Server. The transformation server and the ILS are key components to support the efficient media negotiation mechanism.



< Figure 3: The Proposed New SIP Architecture >

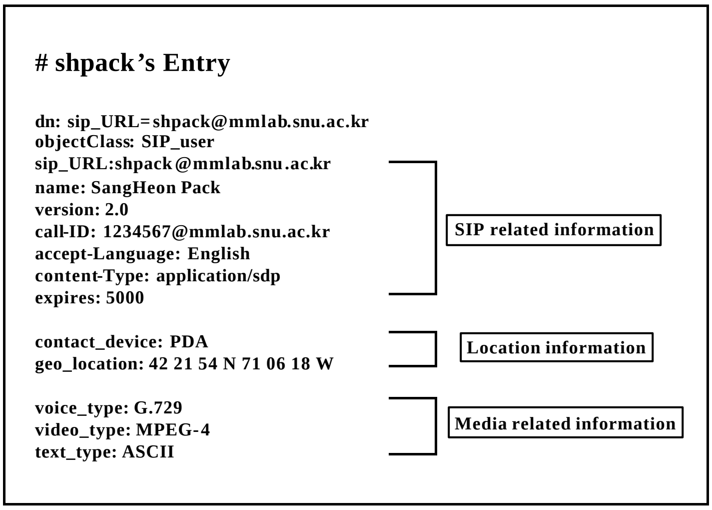
3.1.1 ILS

ILS is the enhanced location server. It stores users' media information as well as location information while a general SIP location server only stores users' location information. Intelligent Location Server is implemented using Lightweight Directory Access Protocol (LDAP)[19]. LDAP is a client-server protocol for accessing a directory service. It was initially used as a front-end to X.500[20], but can also be used with stand-alone and other kinds of directory servers. This protocol is specifically targeted at management applications and browser applications that provide read/write interactive access to directories.

In LDAP, the stored data are managed by directory structure. Directory structure stores a user's information as an entity type. An entity type has several attributes, describing the features of the entity. The entity type in the ILS is defined to present SIP related information, user's location information, and media related information. The figure 4 shows the defined entity type and attributes. In the ILS, we defined SIP_user entity type to register a user's information. The SIP_user entity has several attributes. These attributes are divided three classes by their properties. The first attribute class presents the SIP-related information. SIP URL, user name, and SIP version belongs to this class. The second attribute class contains location-dependent information, e.g., the current available contact point and geographical location. The third attribute class presents available media format of a user. There are voice_type attribute for voice call service and video_type attribute for advanced multimedia service in this class. The transformation server can detect diversity problems in media types by utilizing information in this attribute class.

After designing an entry and its attributes, we have to add entries to the ILS. The information required for registration can get by REGISTER request. A user agent program sends REGISTER request to default proxy server in its own domain when it starts up or a user moved into another SIP domain. The REGISTER request consists of SIP part and SDP part like another request message. SIP part contains the information such as SIP URL, user name, originator/destination of the message and so on. SDP part describes session information such as session name, connection information, and supportable media formats. The default proxy server receiving REGISTER request parses the message and stores some information in location server using LDAP to support user mobility[21] and the efficient media negotiation mechanism.

Figure 4 shows the example of SIP-user entry. It has some attributes such as sip_URL, name, version, contact_device, voice_type, and video type. But these attributes can be changed according to designing of entry and directory structure.



< Figure 4: Example for SIP_user entry >

3.1.2 The SIP-based Transformation Server

The SIP-based transformation server is the extended SIP proxy server. So it includes basic functions of the general SIP proxy server. In addition to these functions, it has also a few special functions to support efficient media negotiation. These are ILS searching function, message modification function, and media gateway control function. The detail specifications of these functions are described below.

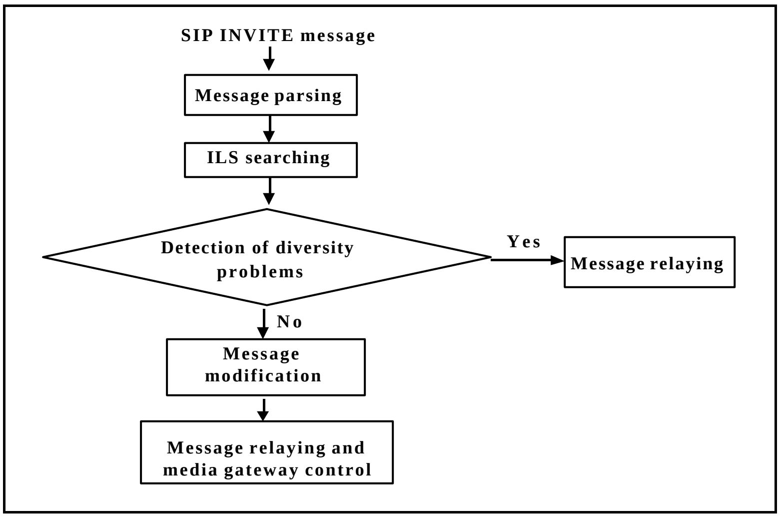
(1) ***ILS searching function***: When the transformation server received the INVITE request, it searches the ILS to get a user's stored information. The results of the ILS searching include a user's supportable media information and this information is used for examination of diversity problems. SIP URL is used as an searching key in the ILS. SIP URL is a unique address to identify a user. SIP URL can get by parsing received request.

(2) ***Message modification function***: If the media types of the received request cannot be supported by the UAS, media negotiation is needed. So the transformation server modifies that SDP part in request message and informs media gateway of necessity of the media type conversion phase.

(3) ***Media gateway control function***: If the ILS searching function and modification function is done, media gateway control function has to run for media type conversion. The controls of media gateway and the media conversion schemes are not main issue in this paper, so we will not explain this function in more detail.

Figure 5 shows message flows in the transformation server based on the three functions above. Since SIP and SDP are text-based protocols, the transformation server must parse the received message to obtain information in it. Then the server searches user's information in the ILS using SIP URL as a searching key. In this way, the ILS can contain users' available media types so that the transformation server can determine whether the UAS supports the requested media types or not.

If the media types between two users are interoperable, no media negotiation is required and the server just forwards received request to another SIP server or UAS without any modifications. If the media types are not interoperable, the server modifies SDP part of the request to interoperable media types according to capabilities of media gateway. At last, the transformation server configures the media gateway to support the suitable media conversion and send the modified request to UAS.



< Figure 5: Message Flows in the transformation server >

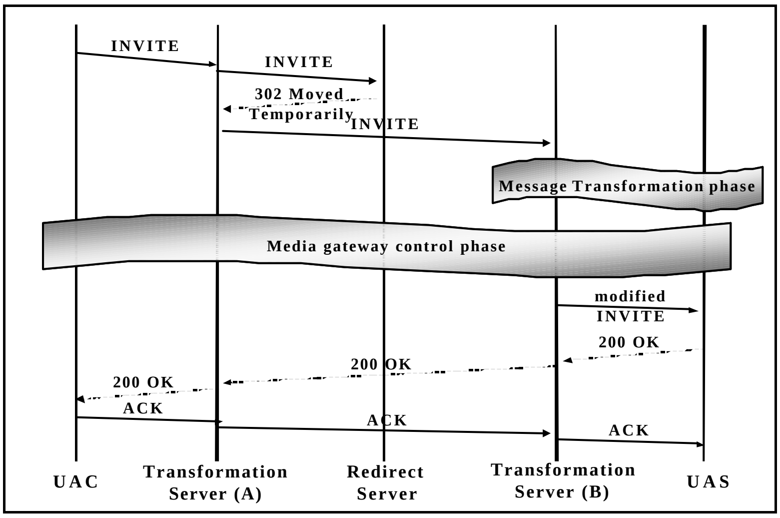
The transformation server handles other request/response messages just like a general SIP proxy server. So it can easily cooperate with the existing SIP components: general SIP proxy server, redirect server, UAC, and UAS. In our proposed architecture, this transformation server will replace only the existing inbound/outbound SIP proxy server in a local domain. So it maintains the flexibility and scalability.

3.2 The Media Negotiation Mechanism using the SIP-based Transformation Server

In this section, we will explain the media negotiation mechanism using the transformation server and the ILS. The proposed media negotiation schemes are summarized as follows.

(1) A caller sends INVITE request message to initiation a VoIP session with a user.
(2) The transformation server receiving this request parses the message and examines whether the media types of the request can be supported by the callee or not. In decision process, the transformation server finds and uses a user's information recorded in the ILS. This information is registered in the ILS by users' registration procedure in advance. A user registers own various information, e.g., name, SIP URL, supportable media types and so on, in the ILS using REGISTER request message. Also the server must consider the capabilities of available media gateway.
(3) If the current request is acceptable to UAS, the transformation server just relays that request to UAS. Then the UAS responds to the UAC with "200 OK" response. After receiving the response, the UAC sends ACK request to the UAS and then a new session is established.
(4) If the current request is not acceptable to UAS, the transformation server modifies the request to the acceptable request. In this phase, the transformation server uses the results of the ILS searching as guideline for modification. If the modified INVITE request can be accepted by the callee, the UAS sends "200 OK" response to the caller and a VoIP is established by sending ACK request to the UAS.
(5) Therefore the media negotiation in this architecture requires only 1.5 round trip times.

Figure 6 shows the media negotiation mechanism based on the proposed SIP architecture.



< Figure 6: Media Negotiation in proposed architecture >

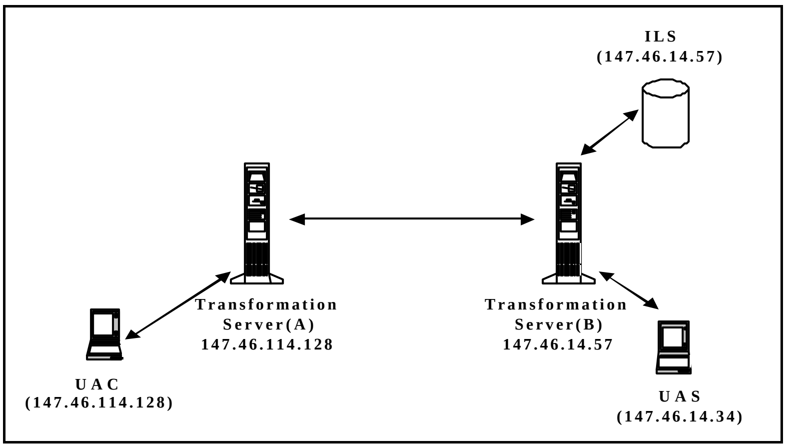
In comparison to the mechanism in the standard SIP, media negotiation in the proposed architecture does not require "606 Not Acceptable" response and re-INVITE messages. Therefore round trip times for session setup can be reduced from 2.5 to 1.5. It is possible due to new SIP architecture with the transformation server and the ILS. Inevitably, this mechanism requires additional processing time for message transformation and LDAP searching. However, these functions can be processed fast by high-speed system. Therefore to reduce message round trip times is more important issue than to remove additional processing times for the efficient media negotiation and the fast session setup.

4. EXPERIMENTS AND RESULTS

To show the effectiveness of the proposed mechanism, we measured the session setup times of the standard SIP architecture and the proposed SIP architecture. The comparison results of two times show performance improvement.

4.1 The Experimental Environment

Figure 7 shows the experimental environment. It is more simplified SIP architecture than the architecture in Figure 3. We calculated the session setup time in two cases: with the transformation server and the ILS and without them.



< Figure 7: Experimental Environment >

As described in section 3, the transformation server has three functions for media negotiation: ILS searching function, message modification function, and media gateway control function. These functions were implemented using Java Development Kit 1.2 (JDK 1.2) in the Linux operating system. It uses User Datagram Protocol (UDP) to communicate with other SIP servers or user agent programs.

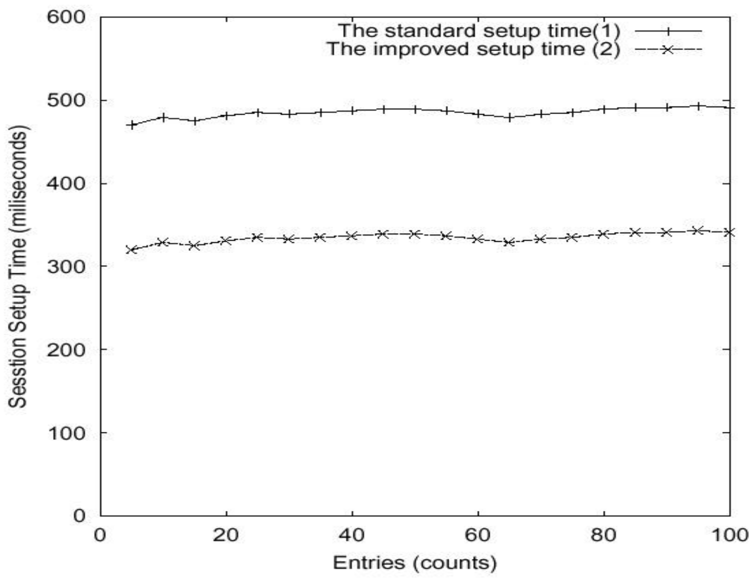
The ILS is implemented using OpenLDAP 2.0.10[22]. OpenLDAP is an open source LDAP program based on LDAP version 3. The ILS stores users' information as SIP-user entity type described in section 3.

4.2 The Experimental Results

In this experiment, the session setup time is calculated as the addition of packet transmission time and message processing time. The loading time of compiled program in Java excluded in calculation setup time.

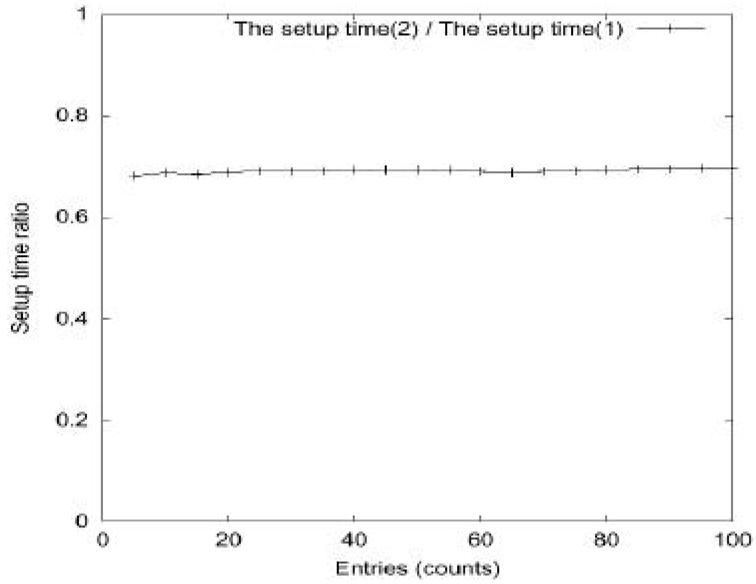
Figure 8 shows the experimental results. In the standard SIP architecture, the session setup time is about 480ms. However, the setup time is at most 340ms in the proposed architecture. Since the proposed architecture removed unnecessary retransmission, the session setup time is less than the setup time in the standard SIP structure. Although the proposed architecture requires additional CPU processing time for media negotiation mechanism, these tasks are processed in internal system. Therefore these additional times hardly affect the total session setup time.

When the number of entries in the ILS increased, the setup time remains almost same. Since the entries in the ILS are managed LDAP program by directory structure with internal index so that the number of entries doesn't affect the ILS searching time.



< Figure 8: Session Setup Time >

Figure 9 shows the improvement ratio of session setup time. The improvement ratio is calculated by dividing the setup time in the proposed architecture with the setup time in the standard SIP architecture. The ideal improvement ratio without CPU processing time is 0.6. Because the round trip time for session setup in the standard architecture is 2.5 round trip times and the round trip time in the proposed architecture is 1.5 round trip time. However, the real improvement ratio is about 0.7. Namely, 10% of the total setup time was used for message modification, LDAP searching and so on. Because this additional time is so small, we can assure that the proposed architecture has better performance than the standard SIP architecture in provision of the media negotiation.



< Figure 9: Performance Improvement Ratio>

5. CONCLUSIONS

In the 3rd and 4th mobile communication environment with various diversity problems, the media negotiation mechanism is essential to support diverse terminals and media types. It is an apparent issue in VoIP service, which provides call services over IP data network and must support many encoding schemes.

In this paper, we proposed the new SIP architecture for the more efficient media negotiation. The new SIP architecture employs the transformation server and the ILS to support the efficient media negotiation.

The ILS is an extended location server in the proposed SIP architecture. It is implemented using LDAP and contains users' media information as well as users' location information. So the transformation server can determine the interoperability between media formats using the ILS. The transformation server has three distinctive functions, the ILS searching function, the message modification function, and the media gateway control function. The round trip times for the session setup become from 2.5 round trip times to 1.5 round trip times.

Though the proposed media negotiation mechanism requires the extra CPU processing time, it hardly affects the session setup time. The packet transmission time in network is more critical factor for the fast session setup and the efficient media negotiation. The experimental results in section 4 show these facts.

The transformation server performs the media negotiation process only when an INVITE message arrived from another entities. When the other requests or responses arrived, it acts just like a general proxy server. Therefore, the transformation server can integrate easily with the existing SIP entities and the proposed SIP architecture is still scalable and flexible.

6. REFERENCES

1. H. Schulzrinne et al., “The Session Initiation Protocol: Internet-Centric Signaling,” IEEE Communication Magazine, October 2000.
2. ITU-T homepage, http://www.itu.int/ITU-T/.
3. ITU-T Recommendation H.261, “Video codec for Audio-Visual Services at p*64kbit/s,” 1993.
4. ITU-T Recommendation H.263, “Video codec for Low Bit Rate Communication,” 1996.

5. ISO/IEC JTC1/SC29/WG11 N4030, "Overview of the MPEG-4 Standard," March 2001
6. M. handley et al., "SIP: Session Initiation Protocol," IETF RFC 2543, March 1999.
7. IETF homepage, http://www.ietf.org.
8. H. Schulzrinne and J. Rosenberg, "A Comparison of SIP and H.323 for Internet Telephony," NOSSDAV, July 1998.
9. ITU-T Recommendation H.323, "Visual Telephone Systems and Terminal Equipment for Local Area Networks which Provide a Non-Guaranteed Quality of Service," 1996.
10. 3GPP homepage, http://www.3gpp.org.
11. 3GPP2 homepage, http://www.3gpp2.org.
12. R.Fielding et al., "Hypertext Transfer Protocol – HTTP/1.1," IETF RFC 2616, June 1999.
13. M. handley et al., "SDP: Session Description Protocol," IETF RFC 2327, April 1998.
14. T. Warabino et al., "Video Transcoding Proxy for 3Gwireless Mobile Internet Access," IEEE Communication Magazine, October 2000.
15. ITU-T Recommendation G.722, "7kHz Audio-Coding Within 64kbit/s," 1988.
16. ITU-T Recommendation G.728, "Coding of Speech at 16kbit/s Using Low-Delay Code Excited Linear Prediction," 1992
17. M.Arango et al., "Media Gateway Control Protocol (MGCP) Version 1.0," IETF RFC 2705, October 1999.
18. F. Cuervo et al., "Megaco Protocol Version 1.0," IETF RFC 3015, November 2000.
19. M.Wahl et al., "Lightweight Directory Access Protocol (v3)," IETF RFC 2251, December 1997.
20. ITU-T Reccomendation X.500, "The Directory: Overview of concepts, models and service," 1993.
21. E. Wedlund and H. Schulzrinne, "Mobility Support using SIP," Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM), August 1999.
22. OpenLDAP homepage, http://www.openldap.org.