3G

A SIP Signaling Retransmission Scheme for Efficient Mobile IPTV Service over 3G Wireless Networks
A SIP Signaling Retransmission Scheme for Efficient Mobile IPTV Service over 3G Wireless Networks

Byungjun Lee, Kyung Tae Kim, Hyung Do Lee, Hee Yong Youn
School of Information and Communication Engineering
Sungkyunkwan University, 440-746, Suwon, Korea
{bjlee, ktkim, krtitan, youn}@ece.skku.ac.kr

Abstract—The rapid advancement of mobile communication technologies recently has led to the advent of mobile IPTV service. For a variety of multimedia services, a process of session setup using SIP (session initiation protocol) is required. However, high frame error rate (FER) of typical wireless network causes failure and delay in session setup. In this paper we propose a new scheme minimizing the session setup delay. The proposed scheme estimates the congestion probability by analyzing the FERs of successive transmissions. Through this, the retransmission timer interval (RTI) is set to minimize the session setup delay time. Computer simulation confirmed that the proposed scheme significantly reduces the setup time by properly adjusting the RTI in the rapidly changing wireless network in comparison to the previously proposed retransmission scheme.

Keywords - Mobile IPTV; SIP; session setup; retransmission time interval; congestion;

I. INTRODUCTION

The recent advancement of super high-speed broadband communication technology has realized large-capacity multimedia services in wireless network. In addition, diverse digital environments as well as development and distribution of mobile terminals have led to the increase in the demand for mobile IPTV service. Mobile IPTV enables the users to send and receive multimedia data through the IP-based wired and wireless networks. Also, with QoS/QoE, security, mobility and interactive functions, it provides the IPTV service anytime, anywhere. To receive various services through mobile IPTV, connection setting between media server providing the IPTV service and mobile terminal is necessary. This connection setup process is facilitated by the SIP (session initiation protocol) [2] of application layer.

SIP is a signaling protocol for call setup between terminals developed at the IETF (Internet engineering task force) and is independent of the lower protocols. Unlike H.323 [3] of ITU-T, etc., expansion of SIP is simple. Also, it is highly compatible with the existing Internet application protocols such as HTTP (hyper text transfer protocol) [4] and SMTP (simple mail transfer protocol) [5]. Therefore, it is applied in various field. In addition, with the recent employment of SIP as the IMS (IP multimedia subsystem) [6] session control protocol of the next-generation mobile communication system, its application range is further expanded in both wired and wireless Internet. With such advantages, SIP is being used in a variety of multimedia applications including Internet phone, Internet TV, and multimedia conference, etc.

Mobile IPTV also uses SIP in the service control function for session connection between media server and mobile terminal. In order to provide seamless service to the users, mobile IPTV must consider frame loss unlike wired IPTV. In general, frame error has not been seriously taken into consideration for multimedia services in the wired environment. However, wireless network environments such as 3GPP and WLAN are subject to high frame error rate due to lowered signal intensity, fading, narrow bandwidth, and congestion, etc. Frame error rate directly affects the users’ service time and quality of mobile IPTV service provided through the wireless network. High frame error rate in wireless environment causes frequent message losses in the process of session setup which causes delay in providing mobile IPTV service. Therefore, to provide seamless mobile IPTV service, a scheme is necessary which reduces the delay time caused by retransmission of messages lost during session setup process [7-8].

So far, the studies investigated this issue proposed the fixed timer scheme and adaptive retransmission timer scheme [1]. With the fixed timer scheme, when message loss is detected in the process of session setup, the session setup message is retransmitted after a prescribed period of time during which the network condition becomes stable enough for message transmission. The adaptive retransmission scheme measures the time taken for frame transmission during the session setup phase and provides an RTI (retransmission timer interval) value suitable to the current situation. With this, unnecessary waiting time of the conventional fixed timer scheme can be reduced. This scheme doubles the waiting time for retransmission compared to the previous session setup phase. Retransmission occurs after the waiting time expires when the congestion is expected to be resolved. However, only network congestion is considered as the cause of retransmission in this scheme. As a result, even in the situations other than congestion, it retransmits the message after waiting transmission time. This may cause unnecessary long waiting. In this paper, thus, we propose a new scheme minimizing the session setup delay in order to provide efficient mobile IPTV service. The proposed scheme analyzes the source of frame error in wireless network environments such as 3G and WLAN, and thereby estimate the congestion probability of the current network. Based on this, RTI minimizing the session setup delay is proposed, which is important for seamless mobile IPTV service. Through comprehensive computer simulation, it was confirmed that the proposed scheme significantly decreases session setup time in the hostile network environment in comparison to the conventional fixed and adaptive retransmission timer scheme.

The rest of the paper is organized as follows. Section 2 describes the basic SIP protocol and the previously proposed...
retransmission timer scheme. Section 3 presents the proposed approach for the congestion probability estimation and retransmission. Section 4 confirms the performance of the proposed scheme through computer simulation. Section 5 gives the conclusion and discusses the future work.

II. RELATED WORKS

This section describes the setup process of an SIP session for mobile IPTV service and briefly examines session setup delay time for which a number of studies have been carried out so far. Then, the conventional session setup timer scheme is discussed.

A. SIP (Session Initiation Protocol)

SIP protocol is a signaling protocol of application layer that prescribes the procedures used to create, delete, or change voice and multimedia communication sessions. SIP protocol has the advantages that both TCP and UDP protocols can be used in the user/server mode and the implementation is simple. SIP also provides flexibility and expandability in voice communication service. Compared to H.323, it can be configured with more convenient protocol operations. SIP session setup and communication process begin with the registration in proxy server by the user. SIP proxy server handles the request for call connection and disconnection sent from the user. In Fig. 1, location server and proxy server operate in a physically same server. In a small-scale LAB environment, proxy server, location server, and register server operate as a single server. Before each user makes a call, the location information is provided to the location server. This information includes the user’s SIP address and IP address, etc. Each user and server receives accurate location information of the other entity through the DNS and location server in the first invitation of the user. This helps the users execute mutual call setup.

Figure 1. The structure of SIP-based system.

SIP message is configured as a text-based message based on the conventional HTTP message type structure. The classification and types of messages are as follows.

- **REGISTER**: SIP clients must provide their own location information to the registration server. In other words, they must register their own SIP address and IP address, etc.
- **INVITE**: This is the message the user (user agent client) sends to the server when starting an SIP session, that is, when making a call. In some cases, the message can be immediately transmitted to the user agent server.
- **ACK**: When the user receives a final response message for INVITE message, it replies with ACK. Regardless of success or failure of response, the user replies with ACK for the final response.
- **BYE**: When a client ends a call, this is used to notify the server of the completion of the call[9-10].

In addition to the four messages described above, SIP communication process is supported with various other messages. A call setup process consists of approximately 10 stages, and the order of the status codes generated during this process is predetermined. The entire stages are grouped into four stage groups.

- **Stage Group 1**: Stage for user registration.
- **Stage Group 2**: Stage to invite another user to communicate and execute call setup through additional ringing, trying, OK, and ACK stage.
- **Stage Group 3**: Stage to execute call setup through RTP/RTCP stages when call setup stages are finished.
- **Stage Group 4**: Stage in which user reenters the waiting status through BYE and OK stage after completing the RTP communication stage[11-12].

B. Retransmission Time Schemes

There may be a number of causes of frame loss when transmitting messages in the wireless network environments. The first one is congestion which is due to large data or frequent message transmissions to the server or terminal. Congestion occurs mainly at the bottleneck of the network. When message loss occurs due to congestion, the user must wait until the congestion status is resolved. This is because continuous attempt of transmission without considering the congestion status would aggravate the situation. Other causes of frame loss are fading of signal intensity and shadows in the buildings or locations far away from the base station. Disconnection also occurs while on the move as observed frequently with mobile terminals receiving IPTV service. If it is
required to retransmit the lost message upon the detection of the loss of session setup message, retransmission can be attempted immediately. Diverse studies have been conducted to reduce delay time due to message loss in the process of session setup. One of the representative studies is on the fixed timer scheme and adaptive retransmission timer scheme [1].

When message loss is detected in the session setup process with the fixed timer scheme, the session setup message is retransmitted after a prescribed period of time after the network is recovered to enable message transmission. However, this scheme uses a fixed RTI value, and therefore incurring unnecessary waiting time. To resolve this, adaptive retransmission timer scheme was proposed. It is similar to the fixed timer scheme in considering network congestion as the cause of message loss. However, it does not use a fixed RTI value. Using the actual message transmission time, the RTI value is changed according to the network status. Note that congestion is not the only cause of message loss in mobile IPTV environments. Unless using RTI suitable for different causes, session setup for mobile IPTV service will spend unnecessary waiting time. As a solution for this, it might be necessary to analyze various causes of frame loss and configure RTI accordingly.

III. THE PROPOSED SCHEME

This section proposes a scheme minimizing the session setup delay time by reducing unnecessary waiting time in the multimedia service for mobile IPTV user. The proposed scheme consists of two phases; network status estimation to identify the cause of session setup delay and decision of retransmission time interval to minimize the session setup delay time.

The basic operation of the proposed scheme is as follows. UAC (user agent client) sends ‘INVITE’ message for session setup to the server, and the server sends 183 messages. Then the terminal transmits PRACK message. In case of PRACK message loss due to high FER of wireless network, the terminal calculates a new RTI value to reflect the network status by using the congestion probability estimation scheme. Then, after waiting for RTI, PRACK message is retransmitted. The subsection below describes the proposed scheme in detail.

A. Congestion Probability Estimation

The proposed scheme tries to set RTI optimized to different situation. Using the proposed scheme, the network status is collected from the lower layers with a fixed interval. When message loss is detected, the cause is identified. The variables used in the proposed scheme are summarized Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size</td>
<td>The window size during which the radio link information from lower layers is collected</td>
</tr>
<tr>
<td>Average FER</td>
<td>The average FER value within WS</td>
</tr>
<tr>
<td>Previous aFER (pFER)</td>
<td>The average FER value of previous window</td>
</tr>
<tr>
<td>Bad Threshold (BT)</td>
<td>Used to decide whether the state of the network is bad</td>
</tr>
<tr>
<td>Good Threshold</td>
<td>Used to decide whether the state of the network is good</td>
</tr>
<tr>
<td>Query Number (QN)</td>
<td>The number of queries for retrieving the lower-layer radio link information</td>
</tr>
<tr>
<td>Cong</td>
<td>The probability of congestion state in the network</td>
</tr>
</tbody>
</table>

- Step 1: Initialization Step. Here each variable is initialized, and WS is set which determines the number of frame error rates to be collected to calculate the average.
- Step 2: Queries are used to receive the network status from the lower layers.
- Step 3: It is checked if frame error rates have been collected as many as the WS set in Step 1.
- Step 4: Average value of the collected frame error rates is calculated.
- Step 5: pFER is checked if is 0. This is applied only once at each setup stage.
- Step 6: The previous average frame error rate and the current average frame error rate are compared to examine the change in the network status. If aFER is larger then pFER, the network condition is getting worse. Otherwise, it is getting better. This step is for identifying the direction of the network condition change.
- Step 7: BT and GT are used which were set empirically to respond to the rapidly changing network status. The frame error rate larger than BT implies that the network status gets worsening. Therefore, the congestion probability must be increased. If frame error rate is smaller than GT, the network status is deemed to recover. Therefore, the probability of congestion must be reduced. Congestion is one the causes of message loss in wireless network, which is expressed by ‘Cong’.
- Step 8: is updated properly according to the situation.
The use of SIP over RLP can reduce the influence of FER (frame error rate) on the session setup time and increase the reliability. RLP in nontransparent mode is a pure Negative Acknowledgment (NAK)-based selective repeat protocol.

B. Retransmission Timer

In case the frame loss is caused by disconnection while the terminal is moved or intermittent cutoff due to other problems in the network environment, unnecessary waiting time incurs if it is too long. To solve these problems, a new retransmission timer scheme is proposed. The components required to calculate the time interval for the proposed retransmission timer are as follows. They are also illustrated in Fig. 4.

- \( K_i \): The size of the \( i \)th message frame required for session setup
- \( D \): The time taken for frame transmission from mobile terminal to proxy server
- \( t \): The time interval between two successive frame transmissions

The RTI based on the components above is as follows.

\[
Tr(i - 1) = D + (K_i - 1) \cdot t + D + (K_{i-1} - 1) \cdot t
\]

(1)

\[
Tr(i) = (1 - Cong) \cdot Tr(i - 1) + Cong \cdot (2 \cdot Tr(i - 1))
\]

(2)

In (1), \( Tr(i-1) \) indicates the time taken for the transmission of the previous message if the message is lost. In other words, when a message transmission fails, RTI is calculated using the previous message transmission time. The term of \( '1 - Cong' \) in (2) indicates the probability of message loss due to disconnection while the terminal is moved or intermittent disconnection due to some reason other than network congestion. The reason why the previous transmission time is doubled in the case of loss due to congestion is that it takes at least this amount of time for the network to escape from the congestion status.

Fig. 5 is an example of the operation of the proposed scheme. During session setup using SIP, assume that loss of SIP PRACK, the third message, is detected. The proposed scheme then calculates the transmission time of SIP INVITE and SIP 183 message using (1), the messages sent before the loss of SIP PRACK message, and sets RTI using (2) by applying the estimated congestion probability. After the timer expires, the SIP PRACK message is retransmitted. By setting the timer according to the cause of message loss as the proposed scheme, unnecessary waiting time will be reduced.

Figure 3. The flow chart of network congestion probability estimation scheme.

Figure 4. The components used in the retransmission timer model.

Figure 5. The components used in the retransmission timer model.
Average SIP session setup delay increases exponentially with the FER (p in the model). The number and the size of the messages exchanged also affect the average session setup delay. The reduction of these factors leads to a shorter session setup delay. The number of frames is considered in the study, and we consider two types of channels, 9.6Kbps and 19.2Kbps. The duration of each radio frame is assumed to be 20ms, which corresponds to 24 bytes in a 9.6Kbps channel and 48 bytes in a 19.2Kbps channel, respectively. The values of the delay, D, and the inter frame time, t, are set as in [17], respectively, 100ms and 20ms. For SIP over TCP and UDP, the maximum number of transmissions is set to seven.

A. SIP over UDP

<table>
<thead>
<tr>
<th>Message</th>
<th>Payload size (bytes)</th>
<th>Message size (Bytes)</th>
<th># of frames (9.6Kbps)</th>
<th># of frames (19.2Kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIP INVITE</td>
<td>700</td>
<td>728</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>SIP 183</td>
<td>835</td>
<td>863</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td>SIP PRACK</td>
<td>558</td>
<td>586</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>SIP 200OK</td>
<td>545</td>
<td>573</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>SIP 180</td>
<td>349</td>
<td>377</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>SIP ACK</td>
<td>300</td>
<td>328</td>
<td>17</td>
<td>9</td>
</tr>
</tbody>
</table>

Table II shows the size of the UDP datagram and the number of frames per datagram for both the channels. We assume that each UDP datagram is carried over one IP packet. The overall header is assumed to be 28 bytes (20 bytes for IP header and 8 bytes for UDP header). The average session setup delay is evaluated for various FER between 0 ~ 35%.

As illustrated in the figures, the session setup time of the conventional retransmission scheme and the proposed scheme are similar for low FER. However, as FER increases, there is a significant difference between the session setup times. The time reduction by the proposed scheme is larger than 6 seconds for an FER higher than 30% or more of the 9.6Kbps channel. This is about 20% improvement. This is because the proposed scheme decides the retransmission timer interval suitable to the environment. We observe similarly result with higher speed channel of 19.2 Kbps.

<table>
<thead>
<tr>
<th>Message</th>
<th>Payload size (bytes)</th>
<th>Message size (Bytes)</th>
<th># of frames (9.6Kbps)</th>
<th># of frames (19.2Kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP SYN</td>
<td>74</td>
<td>114</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>TCP SYN-ACK</td>
<td>74</td>
<td>114</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>TCP ACK</td>
<td>66</td>
<td>106</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>SIP INVITE</td>
<td>700</td>
<td>740</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>SIP 183</td>
<td>835</td>
<td>875</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td>SIP PRACK</td>
<td>558</td>
<td>598</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>SIP 200OK</td>
<td>545</td>
<td>585</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>SIP 180</td>
<td>349</td>
<td>389</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>SIP ACK</td>
<td>400</td>
<td>440</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>TCP ACK</td>
<td>66</td>
<td>106</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Table III shows the size of the TCP segments and the number of frames per segment for both the channels. We assume that each TCP segment is carried over one IP packet. The overall header is assumed to be 40 bytes (20 bytes for the
IP header and 20 bytes for the TCP header).

Fig. 8 and 9 also show the average session setup time with 9.6Kbps and 19.2Kbps channel of TCP environment. While there is not much differences from the UDP case, the proposed scheme allows significant improvement over the existing scheme.

![Figure 8](image8.png)

**Figure 8.** Average session setup delay in 9.6Kbps channels for SIP over TCP.

![Figure 9](image9.png)

**Figure 9.** Average session setup delay in 19.2Kbps channels for SIP over TCP.

V. CONCLUSION AND FUTURE WORK

In this paper, as a way to overcome the problem of long session setup process due to high frame error rate of wireless network, we have proposed a scheme deciding message retransmission time interval suitable to the cause of message loss. The proposed scheme assists in effectively handling the harsh wireless network environments of high frame error rates for IPTV multimedia service. It significantly reduces the delay time in session setup process for multimedia service by estimating the congestion probability which is used in deciding the RTI. The users will thus be relieved of inconvenience in using multimedia service.

As future course of study, we plan to optimize the session setup time in wireless environment by applying the hybrid ARQ [16]. In addition, a study will be conducted coping with for changing patterns of FER.

ACKNOWLEDGMENT

This research was supported by the IT R&D program of MKE/KEIT. [10033396, Development of the USN based Convergence Service], and a grant (CR070019) from Seoul R&BD Program funded by the Seoul Development Institute of Korean government. Corresponding author: Hee Yong Youn.

REFERENCES