

# A Caching Scheme for Session Setup in IMS Network

Yufei CAO<sup>1,2</sup>, Jianxin LIAO<sup>1,2</sup>, Qi QI<sup>1,2</sup>, Xiaomin ZHU<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

<sup>2</sup>EBUPT Information Technology Co., Ltd, Beijing, China

E-mail: {caoyufei, liaojianxin, qiqi, zhuxiaomin}@ebupt.com

Received April 29, 2009; revised May 8, 2009; accepted May 10, 2009

## Abstract

In IP Multimedia Subsystem (IMS), the session setup delay is a critical value for Quality of Service (QoS). The existing approaches to improve this metric depend on optimization of Session Initiation Protocol (SIP) message transmitting and signaling flows. Unfortunately, some service features are missing considered although they have been used widely in traditional 2G networks. This paper proposes a novel session setup scheme based on caching, upon the investigation of the performance of IMS session establishment. This mechanism involves cache based local routing policy and an adaptive caching algorithm, which can decrease call setup delay effectively as cached information in the terminating Serving-Call Session Control Function (S-CSCF) hit. The analytical model is deduced, as well as the delay and cost ratio functions are presented based on the model. Moreover, the analytical model is validated through the performance simulation in which the performance of the proposed novel method is evaluated against the basic session setup mechanism in terms of cost and delay.

**Keywords:** SIP Session, Caching, IMS

## 1. Introduction

IMS (IP Multimedia Subsystem) network is introduced in 3GPP R5, which aims to provide mobile user multimedia services such as voice, video and data. It unifies core network as all-IP network architecture and realizes the integration of fixed and mobile communication networks [1–3]. Home service control is selected in IMS, which means the entity that accesses to the subscriber database and interacts directly with service platforms is always located at the user's home network. Thus, location management and session management are pointed to the home network as far as possible. The HSS (Home Subscriber Server) contains all the information related to the users and their services. The S-CSCF (Serving-Call Session Control Function) located in home network provides session control and registration services [2,4,5].

In IMS network, when the caller A wants to establish session with the callee B, SIP INVITE request constructed by UE (User Equipment) is forwarded to the user A's home network via the P-CSCF (Proxy-Call Session Control Function). And then user A's home S-CSCF executes the service control, including interaction with the AS (ap-

plication server), a process of querying DNS to determine the entry of UE B's home network and assigning the S-CSCF through the I-CSCF (Interrogating-Call Session Control Function) which is needed to select the S-CSCF of UE B. This S-CSCF is responsible for dealing with and ending the session, containing the interaction with AS, sending request messages to the P-CSCF which UE B accessed to and forwarding to UE B finally. The response generated by UE B reverses the same path back to UE A. After several forward and back flows, session establishment is completed [6,7].

The excessive signaling of current IMS session setup mechanism results in the long delay from the session initiation of caller and the final response of callee [8,9]. This is unfavorable to those applications which require fast communication handshake. Comparing with traditional 2G network, there are some problems in the flows defined in current specifications: 1) There is no considering that users' session setup takes place within one S-CSCF serving area, so DNS querying and S-CSCF assignment are involved. Thus, this process brings two problems. Firstly, it increases unnecessary interactive signaling and the session setup delay is raised. Secondly, the larger the scale of

users is, the more the traffic load of the entities as DNS, I-CSCF, HSS, etc. is. That not only would be a waste of network bandwidth and resources, but also reduces system reliability. And eventually it leads to long session setup delay for users' experience. 2) A series of interactive inquiries to establish sessions is too complicated. Even if the requested destination server addresses (S-CSCF address) of the current session request is same as the previous one, DNS inquiry process and the S-CSCF selection are still necessary. So it is feasible to optimize signaling traffic load by decreasing the number of signaling interaction.

This paper focuses on the study of basic session setup mechanism in current IMS core network, which includes the process of caller's session request arriving at its S-CSCF (the originating S-CSCF) and the connection setup between the originating S-CSCF and the terminating S-CSCF. A cache based session setup mechanism is proposed by improving the originating S-CSCF session setup procedure with taking locality and caching into account. The advantages are: 1) when session participants are in the one S-CSCF, their sessions can be established directly. 2) If the destination server address of current session request is the same as the previous one, session can be established directly. It decreases the times of signaling interaction for DNS query and reduces the load of the HSS by improving the S-CSCF assignment process. 3) There is no change of IMS core network, no adding or varying to the terminal signaling, and no impact on signaling flows. The simulation shows that the cache based method is able to reduce signaling traffic load and session setup delay in IMS. At the same time, as the improvement is mainly about the signaling flows, instead of system hardware or network structure, the cost of its implements is smaller comparatively.

The rest of the paper is organized as follows. Section 2 reviews related work and motivates the cache based session setup in IMS. Section 3 presents the basic session setup flows and analyzes the terminating S-CSCF routing process with explanation of its issues. Section 4 introduces the cache based session setup mechanism with original contribution presented in detail. Section 5 shows the detailed analysis of the cost and the mean delay function of new mechanism. Section 6 shows evaluation of the performance by simulation. Section 7 concludes this paper.

## 2. Related Work

Usually, in Circuit Switch (CS) of both traditional 2G and 3G networks, there are two approaches to improve call setup performance [10].

One is to improve location management policy and management protocols [11–14], which aims at exploring how to manage and query user's location information

efficiently, in order to quickly address entries serving the callee in the core network. For example, there are three-tier location management in [11], and layered management of mobile IP in [12,13] which restricts UE register signaling in its local networks. Also, in [14], a cache scheme of location information is proposed to reduce call setup delay. In [3], two functional entities MMS (mobile management server) and ACS (access control server) are introduced to enhance 3G core network architecture. It separates registration procedure and security service away from CSCFs to achieve the efficiency of session setup by allaying complexity of CSCFs without change to the current IMS session setup procedure.

The other one is to improve call setup process by taking advantage of the users' locality to advance call setup performance [10,15]. The local routing is proposed in [10] to modify the call setup process when the caller and callee are in one VLR (Visitor Location Register), so the cost between the originating MSC (Mobile Switch Center) and the terminating HLR (Home Location Register) can be saved. The work in [15] uses a local routing policy for call setup based on three-tier database architecture in 3G network with the caching in GLR (Gateway Location Register).

Furthermore, the researches in IMS try to improve session setup delay in two aspects. The first is optimizing Session Initiation Protocol (SIP) signaling transmitting. IMS session setup delay is affected by the quality of the wireless link, e.g. frame error rate (FER), which can result in retransmissions of lost packets and can lengthen the session setup time. One way to do is choosing the appropriate retransmission timer and the underlying protocols. The work in [16] focuses on SIP signaling transmitting by optimizing it with an adaptive retransmission timer and evaluates SIP session setup performances with various underlying protocols, such as transport control protocol (TCP), user datagram protocol (UDP), and radio link protocols (RLPs). The work in [17] proposes that choosing an appropriate SIP compression efficiency and transport protocols can improve session setup delay. The work in [18] studies the SIP signaling transmitting, processing and queuing delay in 3G and WiMax networks, and proposes increasing channel rates can reduce IMS session setup delay.

The second is improving SIP signaling flows. The work in [8] investigates the call control procedure in UMTS Packet Switch (PS), and decreases call setup delay through performing Radio Access Network (RAN) resource allocation concurrent with media negotiation. But signaling interactions are reduced at the high cost of air interface resource to achieve fast handshake. The work in [19] is concerned about the in-calling setup delay and enhances the I-CSCF reliability through check point mechanism; also it uses the cache in I-CSCF to accelerate session setup. To the problem of triangular routing for a certain period of time when the user is

moving, Alam, M. T. *et al.* [20] proposes a decision algorithm to select the optimal session setup option.

The features of call service in IMS are similar to those in traditional 2G and 3G network, such as characteristics of localization. However, the previous works seldom take advantage of the locality to improve IMS session setup. Therefore, we propose the cache based session setup mechanism involving the local routing policy along with an adaptive caching algorithm, in order to solve the problem of signaling waste and reduce session setup delay.

### 3. Session Setup in IMS

#### 3.1. Basic Session Setup Flow

Figure 1 shows a classic session setup procedure in IMS network. With no loss of general we could suppose that as follows:

1) UE A and UE B are IMS terminals with the same type of properties.

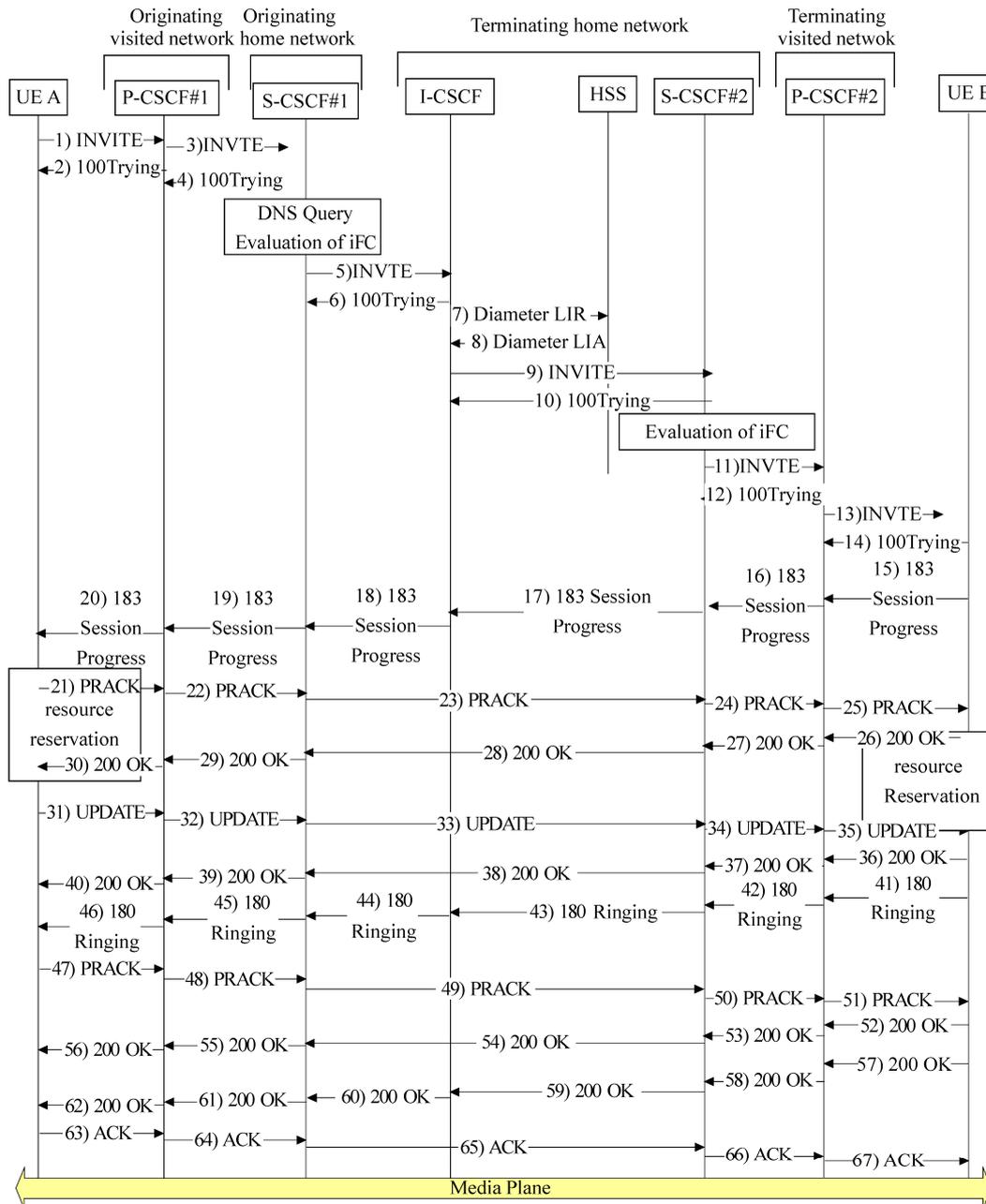


Figure 1. Session establishment procedure in IMS.

2) For simple consideration, both the caller and callee have no service contacts with the session.

3) P-CSCF # 1 and S-CSCF # 1 in the originating network are the entities of P-CSCF and S-CSCF providing services to the caller. Similarly, P-CSCF # 2 and S-CSCF # 2 in terminating network are the entities of P-CSCF and S-CSCF providing services to the callee.

As shown in Figure 1, when originating user A wants to establish session with terminating user B, UE A initiates a call to UE B by sending SIP 'INVITE'. As the 'INVITE' request arrives at S-CSCF # 1 via P-CSCF # 1, S-CSCF # 1 controls the services and the session. This procedure of each flow is given below: verify initial filter criteria (iFC) ; determine the address of network entrance the I-CSCF by querying DNS; forward 'INVITE' request to I-CSCF; the I-CSCF queries HSS to obtain the address of S-CSCF # 2, and then forwards the 'INVITE' request to S-CSCF # 2; finally when S-CSCF # 2 completes verifying iFC, the 'INVITE' request is forwarded to terminating UE B via P-CSCF # 2. Originating and terminating users take Quality of Service (QoS) negotiation by Session Description Protocol (SDP), which is carried in SIP message. After the two sides' QoS negotiation, resource reservation (3-40 steps in Figure 1) is processed. When UE A completes resource reservation, it notifies UE B by 'UPDATE' request, and UE B responses '200 OK' to confirm it. Following UE B's completion of resources reservation, '180 Ringing' message is sent back to the originating UE for ringing. When the terminating user answers, UE B sends the '200 OK' which is the response of 'INVITE' request. As soon as UE A receives this response, 'ACK' is sent to confirm (41-67 steps in Figure 1) that the session between originating and terminating users is established.

### 3.2. Determine the Address of the Terminating S-CSCF

In IMS basic session setup procedure, the originating S-CSCF (S-CSCF#1) is the first node that tries to forward the SIP request based on destination address which is in

the 'Request-URI' field carried by SIP 'INVITE'. The P-CSCF and the I-CSCF are not concerned the destination address, which means they don't inspect 'Request-URI' field in the SIP request. So the originating S-CSCF is the first point parsing the destination address, that is, according to the 'Request-URI' of SIP request it determines the next-hop address in the terminating network. During this procedure, the originating S-CSCF may find two different types of 'Request-URI': 'SIP URI' or 'TEL URI'. If the 'SIP URI' is found, a normal SIP process is adopted and 'INVITE' request is forwarded to the I-CSCF in terminating network through multi-steps DNS querying to determine next SIP server's address, which consists of transport protocol, hostname and port number that the I-CSCF supports [21]. And if the 'TEL URI' is found in 'Request-URI', DNS ENUM (E.164 number and DNS) is needed to decide the right I-CSCF address in terminating network. After receiving the 'INVITE' request, the I-CSCF gets the terminating S-CSCF (S-CSCF#2) address from the HSS by Diameter LIR (Location-Information-Request) and Diameter LIA (Location-Information-Answer) messages and relays 'INVITE' request. Thus, as shown in Figure 2 (A), the route between originating and terminating network has been set up. The detailed signalling flow presented in the specification [6] for the origination S-CSCF towards the terminating S-CSCF is summarized as follows. 1) If the analysis of the destination address determined that it belongs to a subscriber of a different operator, the request is forwarded to a well-known entry point in the destination operator's network, i.e. the I-CSCF. Then the I-CSCF queries the HSS for current location information and forwards the request to the S-CSCF. 2) If the analysis of the destination address determines that it belongs to a subscriber of the same operator, the S-CSCF forwards the request to a local I-CSCF, who queries the HSS for current location information. Then, the I-CSCF forwards the request to the S-CSCF.

Obviously, in originating network so many DNS queries and terminating S-CSCF discovery processes have seriously impact on the session setup delay and the cost of network transport. For IMS network which serves tens of thousands of users, performance will be significantly

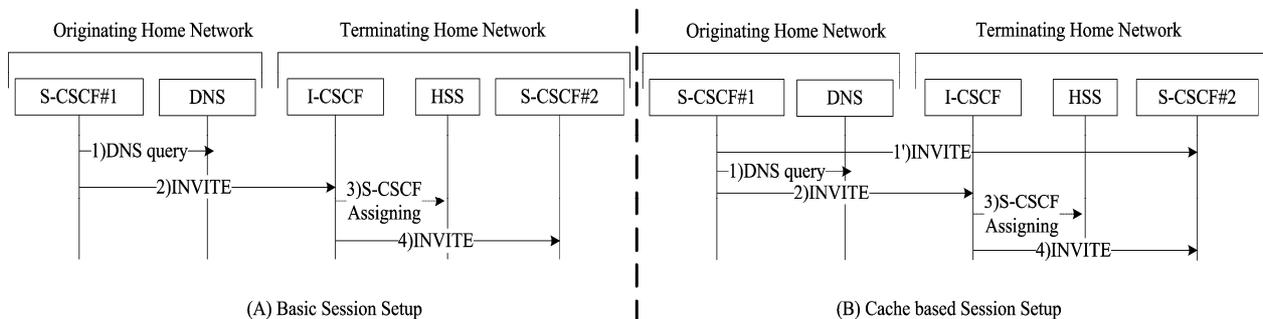


Figure 2. To find the terminating S-CSCF.

improved if we try to save a few messages for per session. Therefore, optimizing of IMS basic session setup mechanism is taken into account. Furthermore, by addressing the terminating S-CSCF from the originating S-CSCF fleetly we can decrease the session setup delay and save network resources.

#### 4. The Cache Based Session Setup Mechanism

This part describes the details of cache based session setup mechanism. The principles of this mechanism are: 1) If caller and callee are located in one S-CSCF serving area, their session will be established directly; 2) the originating S-CSCF forwards the 'INVITE' requests directly to the terminating S-CSCF, instead of the I-CSCF through DNS query; 3) the originating S-CSCF caches the terminating S-CSCF address information obtained from the previous session, and set a valid time for the cached information, which is an exponential distribution.

##### 4.1. Cache Based Local Routing Policy

According to the principles of the cache based session setup mechanism, the local routing policy is adopted during session setup. The originating S-CSCF first checks whether the callee UE has registered on the originating S-CSCF or not. If true, session is established directly. If originating and terminating users are not in one S-CSCF, the originating S-CSCF checks whether the current terminating S-CSCF addresses has been cached or not, then it can determine that the next-step is directly visiting the terminating S-CSCF or querying DNS for retrieving the I-CSCF addresses.

The details are as follows, see Figure 2 (B):

1) When the originating S-CSCF receives the session setup request from caller, it checks whether the callee is within the same S-CSCF currently (the UE learns which the S-CSCF will be serving it through the IMS registration). If originating and terminating users are in one S-CSCF, session is established directly; else goes to 2.

2) The originating S-CSCF queries local cache to look for whether there has been the address of the S-CSCF serving the callee in previous sessions. If not, the S-CSCF performs a DNS query to retrieve the I-CSCF address, and sends 'INVITE' request to the terminating S-CSCF via the I-CSCF for session establishment, otherwise goes to 3. The originating S-CSCF caches the terminating S-CSCF address from the first response message (e.g. '183' message). 1–4 steps in Figure 2 (B).

3) For the originating S-CSCF cached the terminating S-CSCF address information, the originating S-CSCF sends 'INVITE' request to the terminating S-CSCF to

establish session, 1' step in Figure 2(B)

4) When the terminating S-CSCF receives originating 'INVITE' request, re-registration process perhaps be initiated because of the terminating user's roaming. A new S-CSCF is selected again in terminating home network, and the S-CSCF (the current terminating S-CSCF) is no longer available with returning error message. The originating S-CSCF needs to re-initiate basic session setup process. 1–4 steps in Figure 2 (B).

##### 4.2. Adaptive Caching Algorithm

How to cache this information decides the accuracy of address query for the S-CSCF serving the terminating user. When the session setup request arrives, if cache information is effective, the originating S-CSCF hit the terminating S-CSCF with the most perfect performance. If the cached information in terminating S-CSCF is invalid or inexistence, basic session establishment procedure is needed, then at least the performance of cache based session setup is not worse than that of the basic one. But if cached information is outdated, after session arriving at a wrong location, it should re-establish in accordance with basic procedure, which is obviously the worst. Therefore, the effective of cached information has a great impact on probability of hitting terminating location in session establishment procedure.

Consequently, we design an adaptive caching algorithm in terms of the mobility patterns and the locality of call traffic rules, i.e. the probability that caller and callee locate in the same S-CSCF serving area is large. And we assume there is a data buffers in S-CSCF and the buffer size can satisfy system requirement.

Let  $G$  be the set of all out-area users.

$A = \{x|x : \text{the out-area users with new location information}\}.$

$B = G - A = \{y|y : \text{the out-area users with comparative mobile stabilization}\}.$

Define vector  $V$ . //  $V$  expresses the state of out-area user address information.

If  $V(x) = \text{TRUE}$

$x \in B;$

Else

$x \in A.$

End If

Define vector  $T(x)$  for the caching time of the information. //  $T(x)$  expresses the period from previous resetting to the present time.

Assume  $Z$  is the out-area user obtained in the session setup procedure.

If  $z \in G$

Set  $G = G \cup \{z\}, B = B \cup \{z\}, V(z) = \text{TRUE}, T(z) = 0;$

Else

Compare the address information of this procedure and the previous cached one.

If they are same &  $V(z)=FALSE$

Move the address information of user  $z$  from set A to set B;

Set  $V(z)=TRUE$ ,  $T(x)=0$ ;

Else If they are not same &  $V(z)=TRUE$

Move the address information of user  $z$  from set B to set A

Set  $V(z)=FALSE$ .

End If

End If

The judgment of cached information is:

Let the cache time threshold be  $D$ .

Let out-area callee as  $w$ .

If  $w \in G$

If  $V(w)=TRUE$

If  $T(w) < D$

Get the address information of user  $w$  from set B.

Else If  $T(w) > D$

Address as basic session setup procedure

End If

Else If  $V(w)=FALSE$

Session is established according to basic method;

End If

Else If  $w \in G$

Session is established according to basic method.

End If

## 5. Analytical Model

This section deduces the cache based session setup cost function and mean-delay function. For simplicity, we

have only analyzed the process shown in Figure 2, without considering the delay, the cost of UE, P-CSCF entities etc. Our definition of parameters is shown in Table 1.

According to [14],  $\alpha, \beta, \gamma$  are defined.  $\alpha$  means the probability of cache valid, i.e. the address of S-CSCF serving the terminating user has been cached in the originating S-CSCF and the called UE is in the terminating S-CSCF service area as it receives the session request.  $\beta$  means the probability of cache invalid: the originating S-CSCF hasn't cached the address information of S-CSCF serving the callee.  $\gamma$  means the probability of cache miss, i.e. the address of the terminating S-CSCF serving callee has been cached, but the callee no longer resides the terminating S-CSCF. According to the definition, we know  $\alpha + \beta + \gamma = 1$ . The cost function of basic session setup mechanism is given by

$$C_{basic} = C_{dns} + C_{s-i} + C_i + C_{i-s} + C_{s-cscf} \quad (1)$$

The mean delay function of basic session setup mechanism is:

$$D_{basic} = d_{dns} + d_{s-i} + d_i + d_{i-s} + d_{s-cscf} \quad (2)$$

The cost function of cache based session setup mechanism is:

$$C_{caching} = P_1 \times C_{s-cscf} + (1 - P_1) \times (\alpha \times (C_{s-s} + C_{s-cscf}) + \beta \times (C_{dns} + C_{s-i} + C_i + C_{i-s} + C_{s-cscf}) + \gamma \times (C_{s-s} + C_{dns} + C_{s-i} + C_i + C_{i-s} + C_{s-cscf})) \quad (3)$$

The mean delay function of cache based session setup mechanism is:

$$D_{caching} = P_1 \times d_{s-cscf} + (1 - P_1) \times (\alpha \times (d_{s-s} + d_{s-cscf}) + \beta \times (d_{dns} + d_{s-i} + d_i + d_{i-s} + d_{s-cscf}) + \gamma \times (d_{s-s} + d_{dns} + d_{s-i} + d_i + d_{i-s} + d_{s-cscf})) \quad (4)$$

Table 1. Parameters definition.

Symbol	Quantity	Value
$C_{dns} / d_{dns}$	The cost / mean time delay for performing a DNS querying	15u/15t
$C_{s-i} / d_{s-i}$	The cost / mean time delay for transmitting message from the originating S-CSCF to the terminating I-CSCF	10u/10t
$C_i / d_i$	The cost / mean time delay for one process of the S-CSCF assignment by the I-CSCF	20u/20t
$C_{i-s} / d_{i-s}$	The cost / mean time delay for transmitting message from the I-CSCF to the S-CSCF in the terminating network	5u/5t
$C_{s-s} / d_{s-s}$	The cost / mean time delay for transmitting message from the originating S-CSCF to the terminating S-CSCF	25u/25t
$C_{s-cscf} / d_{s-cscf}$	The cost / mean time delay of the S-CSCF	30u/30t
$P_1$	The probability of caller and callee in one S-CSCF	
$\alpha$	The probability of cache hit	
$\beta$	The probability of cache invalid	
$\gamma$	The probability of cache miss	

Then we try to derive probabilities of cache valid, miss, and invalid to (3), (4). Let the user residence time in S-CSCF as  $t_s$  which is assumed as an exponential distribution with parameter  $\lambda_s$ , and its probability density function is:

$$f_s(t) = \lambda_s e^{-\lambda_s t} \quad (5)$$

Denote by  $t_c$  the interval between two consecutive calls to the terminator, and  $t_m$  the interval between the arrival of previous call and the time as terminator move out the S-CSCF service area.  $f_c(t)$  and  $f_m(t)$  are density functions of  $t_c$  and  $t_m$ . We assume that the incoming call is a Poisson process, and then we have

$$f_c(t) = \lambda_c e^{-\lambda_c t} \quad (6)$$

According to the random observer property, we have

$$f_m(t) = \lambda_s \int_{r=t}^{\infty} f_s(r) dr = \lambda_s e^{-\lambda_s t} \quad (7)$$

We assume that the resident time of cached data is an exponential distribution with parameter  $\lambda_h$ , so

$$f_h(t) = \lambda_h e^{-\lambda_h t} \quad (8)$$

While caller initiates a session, if the address of S-CSCF serving the callee has been cached in the originating S-CSCF and the callee is still in cached S-CSCF, then the result is cache valid, and  $\alpha$  is

$$\begin{aligned} \alpha &= P[t_c < t_m \cap t_c < t_h] \\ &= \int_{t_m=0}^{\infty} \int_{t_c=0}^{\infty} \int_{t_h=0}^{\infty} f_m(t_m) f_c(t_c) f_h(t_h) dt_c dt_m = \frac{\lambda_c}{\lambda_s + \lambda_h + \lambda_c} \end{aligned} \quad (9)$$

If the originating S-CSCF has already removed the cached information of the address of S-CSCF serving the callee before an incoming call, the result is cache invalid  $\beta$ . The probability  $\beta$  is given by

$$\beta = 1 - P_r[t_c < t_h] = 1 - \int_{t_c=0}^{\infty} \int_{t_h=0}^{\infty} f_h(t_h) f_c(t_c) dt_c dt_h = \frac{\lambda_h}{\lambda_c + \lambda_h} \quad (10)$$

As  $\alpha + \beta + \gamma = 1$ , we can get

$$\gamma = 1 - \alpha - \beta = \frac{\lambda_h + \lambda_s}{\lambda_c + \lambda_h + \lambda_s} - \frac{\lambda_h}{\lambda_c + \lambda_h} \quad (11)$$

## 6. Performance Analysis

In this section, we firstly verify the validity of analytical model by using simulation experiments, and then we use numerical examples to investigate the performance of the proposed cache based session setup mechanism. For calculation convenience, let basic unit of cost be  $u$ , and basic unit of time delay be  $t$ . Parameter values are shown in Table 1. As the session setup cost of Equation (3) are the same as the session setup delay of Equation (4) in form, we only take the simulation experiments for the metric of session setup cost.

### 6.1. Verify Analytical Results with Simulation Results

In our simulation, there are provided the IMS network topology consisting UEs, the P-CSCF, the I-CSCF, the S-CSCF and the HSS. The simulation signaling flows are the same as Figure 2. The situations of user roaming and session initiating are simulated by generating discrete-events, including three types: 1) call event; 2) cache update; and 3) UE roaming. To investigate the impact of various network parameters on the performance of the new mechanism, the probability of roaming and the probability of caller and callee in one S-CSCF service area are varied by using different simulation configurations.

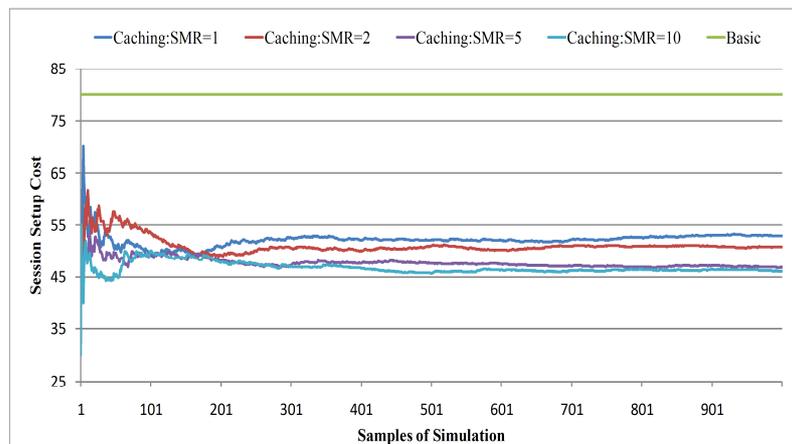


Figure 3. Simulation results of session setup cost.

**Table 2. Simulation and analytical results with difference of cost ( $P_l=0.5$ ).**

$\lambda_c$ (1/s)	SMR( $\lambda_c/\lambda_s$ )	$C_{caching}$		Error (%)
		Simulation	Analytical	
800	1	54.9999	55	0.00%
400	2	53.81361761	54.053	0.44%
266.66667	3	50.7577782	50.739	-0.04%
200	4	49.20909	48.862	-0.71%
160	5	47.78306	47.674	-0.23%
133.33333	6	47.1684	46.858	-0.66%
114.28571	7	46.3097	46.263	-0.10%
100	8	46.5301	45.811	-1.57%
88.88889	9	45.771	45.456	-0.69%
80	10	45.3963	45.169	-0.50%

Here, the  $\lambda_s$  is five time of  $\lambda_h$ , and the SMR (Session to Mobility Ratio), expressed as  $\lambda_c/\lambda_s$ , is varied from 1 to 10. Then the cost values of two session setup mechanisms are measured and the values of 1000 samples got from the experiment in different network configuration are shown in Figure 3. Moreover, Table 2 shows the session setup cost values of the cache based session setup mechanism in both simulation experiments and numerical results, respectively. The cost values between simulation and model have some discrepancy due to the number of random generated discrete-events. And the jitter of the simulation values is also depicted in the Figure 3. If several more session events generated during the simulation period, the cost value of simulation is bigger than that of analytical value, and vice versa. Although the values between simulation and model have some discrepancy, as the error rates are all under 3%, these experiments have verified that analytical model is consistent with the simulation results.

## 6.2. Session Setup Cost and Delay

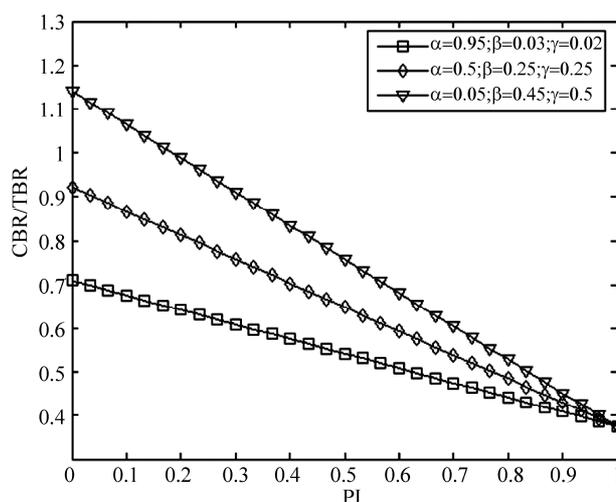
According to [14,15], we defined CBR (Cost Benefit Ratio) that is the ratio of  $C_{caching}$  to  $C_{basic}$  and TBR (Time Benefit Ratio) that is the ratio of  $D_{caching}$  to  $D_{basic}$ . We calculate session setup cost value and mean delay value in accordance with the formulas in Section 5. The conditions and results have been depicted in the figures.

Compared result curves are given in Figure 4 upon three different probabilities of  $\alpha, \beta, \gamma$ , with ratio results at y-axis and x-axis expressing the probability changing of the caller and callee in one S-CSCF. If  $P_l$  increases, the ratio of CBR and TBR becomes smaller. According to y-axis, along with the increasing of cache valid probability, the advantages of cache based session setup mechanism become more obvious. When the probability

of cache miss is bigger as  $\gamma=0.5$ , the performance of cache based session setup mechanism is worse than the performance of basic session setup mechanism ( $CBR/TBR>1$ ), because of session re-establishment after requests arriving the wrong location. However, we note that  $CBR/TBR$  has more than 1 just upon  $P_l < 0.18$ .

Considering the locality of call traffic (the number of local users calling local users is a large part total call number), the probability of callers and callees not in the same S-CSCF is smaller, so this situation will occur less. Therefore, generally speaking, cached based session setup mechanism outperforms the basic one.

According to the definition of SMR, the small value of SMR means the high mobility that UE has, vice versa. The curves varying along with the variable value of SMR and different  $P_l$  have been given in Figure 5. We can see that CBR and TBR decrease as SMR increases, which means caching contributes to decrease the session

**Figure 4. Comparison of CBR/TBR with caller and callee in one S-CSCF.**

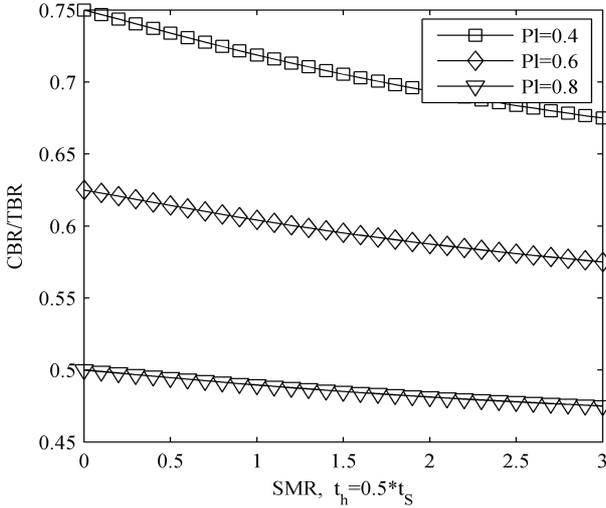


Figure 5. Comparison of CBR/TBR with  $t_h = 0.5t_s$ .

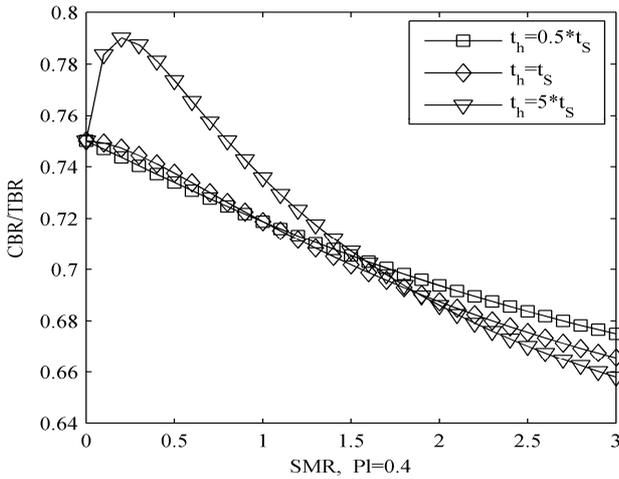


Figure 6. Comparison of CBR/TBR with  $P_l = 0.4$ .

delivery cost and mean delay if UEs have low mobility. Figure 6 presents the curves of CBR and TBR with three different multiple values of residence time and the cached data  $t_h$ . Study from y-axis, when the value of SMR is small (high mobility), the longer the caching time is, the greater the values of CBR and TBR are, and as the value of SMR is higher (low mobility) and the caching time is shorter, the values of CBR and TBR become the greater.

### 6.3. Query Accuracy Analysis

Let  $\alpha = 0, \beta = 0, \gamma = 1$ ,  $CBR/TBR = 1$ , then  $P_l = 0.333$ . In the worst situation  $n$ , i.e. cached information is outdated. After session arrives at wrong location, session

should be re-established in accordance with basic procedure. When  $P_l \leq 0.333$ , the ratio of CBR and TBR is equal to 1 or less than 1. This shows that enlargement of S-CSCF serving area conduces to improve the performance of session setup mechanism based on caching. We define the query accuracy as  $\eta = \alpha + \beta$ , then we get  $\eta + \gamma = \alpha + \beta + \gamma = 1$ . From the above analysis, the performance of session setup mechanism based on caching can be improved by increasing the query accuracy  $\eta$ . When caller and callee are in one S-CSCF serving area, no information required to cache and the performance of session setup procedure based on caching must be better than the basic. Then we discuss the required query accuracy when caller makes a call to the user in other S-CSCF serving area. For description convenience, we name the callee in this situation as out-area user.

From the definition of  $\eta$  and the Equations (9) and (10), we can derive the expression as follow:

$$\eta = \alpha + \beta = 1 - \gamma = \frac{\lambda_c}{\lambda_c + \lambda_h} \quad (12)$$

As shown in Figure 7, as SMR increasing,  $\eta$  increases. This is the query accuracy is added along with the value of CBR/TBR decreasing, i.e. the performance of session setup procedure based on caching is improved. Upon  $t_h = 0.5t_s$ , when SMR equals 0.333, the value of  $\eta$  can be got as 0.957, i.e. the probability of cache hit is large. And  $\eta$  decreases a little along with the SMR increasing. Upon  $t_h = 5t_s$ , when SMR equals to 0.333, the value of  $\eta$  is 0.592.  $\eta$  increases fast along with the SMR rising, and the query accuracy  $\eta$  is as large as 0.90 when SMR is 8.33. Thus, comparing of the two situations, to the out-area users with less mobility, the

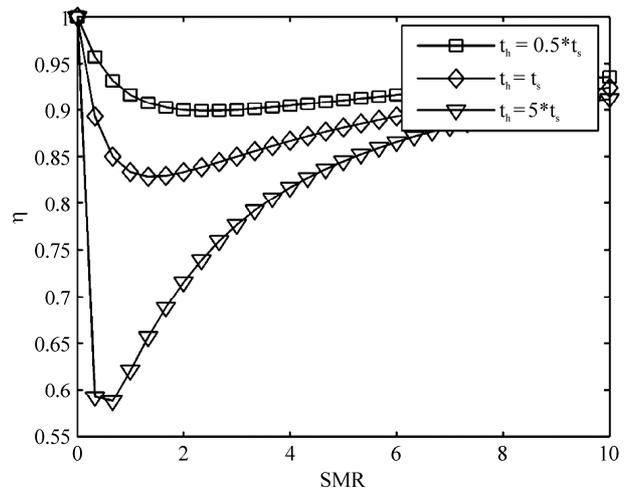


Figure 7. Query accuracy.

session setup mechanism based on cache has advantages of performance. The query accuracy will increase and the performance of cache based session establishment procedure will be improved while the effective time of cached data is prolonged.

## 7. Conclusions

How to decrease the network load and session delay is an important issue for designing and deploying IMS network. This article proposes: a cache based session setup mechanism, mainly through improving signaling flows in session establishment and introducing the caches to reduce network load and session delay. As it mostly perfects signaling flows, with no change of IMS core network architecture, no change or adding to terminal signaling, the cost of this improvement is comparatively smaller. Therefore, this innovation is practical for building of IMS network.

## 8. Acknowledgements

This work was jointly supported by: 1) National Science Fund for Distinguished Young Scholars (No. 60525110); 2) National 973 Program (No. 2007CB307100, 2007CB307103); 3) Development Fund Project for Electronic and Information Industry (Mobile Service and Application System Based on 3G).

## 9. References

- [1] H. Y. Dong, S. Kim, C. Nam, *et al.*, "Fixed and mobile service convergence and reconfiguration of telecommunications value chains," *Wireless Communications*, IEEE, Vol. 11, No. 5, pp. 42–47, October 2004.
- [2] 3GPP TS 23.228, V.8.2.0, IP Multimedia System (IMS), September 2007.
- [3] K. Shuang and F. C. Yang, "Study on enhanced 3G core network architecture," *ACTA ELECTRONICA SINICA*, Vol. 34, No. 7, pp. 1189–1193, July 2006.
- [4] 3GPP TS 22.228, V8.2.0. Service requirements for IP multimedia core network subsystem (IMS), October 2007.
- [5] 3GPP TS 23.218, V7.7.1. IP Multimedia(IM) session handling; IM call model, September 2007.
- [6] 3GPP TS 24.228, V5.15.0, Signaling flows for the IP multimedia call control based on SIP and SDP, October 2006.
- [7] 3GPP TS 24.229, V8.1.0. IP multimedia call control protocol based on SIP and SDP, September 2007.
- [8] K. Umschaden, I. Miladinovic, S. Bessler, and I. Gojmerac, "Performance optimizations in UMTS switched call control," *Fifth IEE International Conference on 3G mobile communication technologies*, pp. 73–177, 2004.
- [9] G. Foster, M. I. Pous, D. Pesch, A. Sesium, and V. Kenneally, "Performance estimation of efficient UMTS packet voice call control," *Proceedings of the IEE Fall Vehicular Technology Conference*, pp. 1447–1451, 2002.
- [10] H. Jiang, B. Lu, and L. M. Li, "A novel call setup mechanism in a GSM network," *Journal of China Institute of Communications*, Vol. 23 No. 8, pp. 52–58, October 2002.
- [11] 3GPP TS 23.119, V7.0.0, Gateway Location Register (GLR), July 2007.
- [12] X. Jiang and I. F. Akyildiz, "A novel distributed dynamic location management scheme for minimizing signaling costs in mobile IP," *IEEE Transactions on Mobile Computing*, Vol. 1, No. 3, pp. 163–175, July–September 2002.
- [13] W. Ma and Y. Fang, "Improved distributed regional location management scheme for mobile IP," *IEEE Proceedings for Personal, Indoor and Mobile Radio Communications, (PIMRC'03)*, Beijing, China, Vol. 3, pp. 2505–2509, September 7–10, 2003.
- [14] C. W. Pyo, J. Li, and H. Kameda, "A caching scheme for dynamic location management in PCS network," *IEEE 58th Vehicular Technology Conference*, Lbaraki, Japan, Vol. 2, pp. 761–765, 2003.
- [15] H. Zhang, J. X. Liao, and X. M. Zhu, "A novel call setup mechanism based on three-tier databases in 3G," *Journal of Electronics & Information Technology*, Vol. 29, No. 6, pp. 1290–1294, June 2006.
- [16] H. Fathi, S. S. Chakraborty, and R. Prasad, "Optimization of SIP session setup delay for VoIP in 3G wireless networks," *IEEE Transactions on Mobile Computing*, Vol. 5, No. 9, pp. 1121–1132, September 2006.
- [17] M. Melnyk and A. Jukan, "On signaling efficiency for call setup in all-IP wireless networks," *IEEE International Conference on Communications (ICC)*, Vol 5, pp. 1939–1945, June 2006.
- [18] A. Munir, "Analysis of SIP-based IMS session establishment signaling for WiMax-3G networks," *Fourth International Conference on Networking and Services (icns)*, pp. 282–287, 2008.
- [19] Y. B. LIN and M. H. TSAI, "Caching in I-CSCF of UMTS IP multimedia subsystem," *IEEE Transactions on Wireless Communications*, Vol. 5, No.1, pp. 186–192, January 2006.
- [20] M. T. Alam and Z. D. Wu, "Comparison of session establishment schemes over IMS in mobile environment," *2005 Fifth International Conference on Information, Communications and Signal Processing*, pp. 638–642, December 2005.
- [21] J. Rosenberg and H. Schulzrinne, *Session Initiation Protocol (SIP): Locating SIP Servers*, RFC 3263, IETF, June 2002.