The Computer Journal Advance Access published December 16, 2009

© The Author 2009. Published by Oxford University Press on behalf of The British Computer Society. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

doi:10.1093/comjnl/bxp111

Enhanced IMS Handoff Mechanism for QoS Support over Heterogeneous Network

JIANXIN LIAO[1,2,*], QI QI[1,2], XIAOMIN ZHU[1,2], YUFEI CAO[2] AND TONGHONG LI[3]

[1] *State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China*
[2] *EBUPT Information Technology Co., Ltd., Beijing, China*
[3] *Technical University of Madrid, Madrid, Spain*
[*] *Corresponding author: liaojianxin@ebupt.com*

IP multimedia subsystem (IMS) is over IP network architecture, but mobile IP cannot directly support session mobility controlled by session initiation protocol-based signaling. The long signaling delay for session reestablishment in application layer always results in session interruptions during the handoff. Therefore, handoff poses a challenge for quality of service (QoS) maintenance in IMS that targets to offer real-time multimedia applications over wireless mobile networks. The existing approaches to solve this problem depend on the advance resource reservation and the optimization of handoff control. Unfortunately, big cost of the advance resource reservation in neighboring domains is a major problem that leads to a serious signaling load and a waste of wireless bandwidth. To solve this issue, we present an enhanced IMS handoff mechanism (EHM) based on user mobility prediction to save network resources by avoiding multiple useless advance reservations. In addition, to support the heterogeneous access networks in IMS domain, EHM evolves a network selective scheme to utilize the network resources more efficiently. The architecture of EHM and the advance QoS negotiation signaling are also presented. We model the cost, the handoff delay and the session blocking probability for EHM and the previous work. Analytical and simulation results show that EHM can enhance the handoff performance, such as reducing resource reservation cost greatly, decreasing session reestablishment delay and making good use of multiple access network resources.

Keywords: handoff; mobility prediction; network selection; QoS; IMS

Received 1 January 2009; revised 25 September 2009
Handling editor: Erol Gelenbe

1. INTRODUCTION

Since IP multimedia subsystem (IMS) combines high-speed mobile access with IP-based services and targets to offer real-time IP multimedia applications over wireless mobile networks [1–3], effective mobility management is studied to realize the seamless handoff. The handoff management in IMS assures the provision of sufficient network resources to user equipment (UE) for quality of service (QoS) guarantee; as well as, it enables a UE to keep attachment to the wireless network [4]. IMS offers a universal core for multiple access technologies, and hence seamless handoff in IMS implies that the customers moving among different areas cannot experience the existence of heterogeneous networks.

On the other hand, IMS employs session initiation protocol (SIP), a mobile management protocol in application layer, but handoff still poses a challenge for end-to-end QoS maintenance [5]. In the over IP network architecture of IMS, handoff control is triggered not only in the IP and transport layer, but also in the application layer, because IPv4 or IPv6 mobility management cannot directly support session mobility controlled by SIP-based signaling. When UE moves among IMS domains during session, it reissues the resource reservation in the new IMS access network. Thus, the large signaling delay in the application layer may result in not only poor service quality but also session interruptions [6]. The session reestablishment procedure includes the SIP signaling such as

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

re-INVITE, PRACK, 183 and common open policy service (COPS) messages that are used to negotiate QoS parameters and reserve resources in the access network [7]. Accordingly, the IMS handoff delay contains the signaling delay in application layer, and the processing delay for resource reservation. Therefore, it is certainly necessary to research the advance mechanism that makes QoS be negotiated and resource be reserved prior to the handoff.

IMS handoff research combines the optimization of handoff control signaling with the improvement of the resource reservation algorithm. Recently, some significant efforts, the improving of handoff signaling [8–11] and the proposals of cross-layer protocol [12–14] have been underway to decrease handoff signaling delay; also some advance resource reservation algorithms [4, 15, 16] are proposed to reduce the reservation delay. Nevertheless, the current architectures to optimize handoff management in IMS mainly involve the following three shortcomings. (i) The resources in the domains all around the current IMS area are reserved, which results in serious waste of signaling and bandwidth [4]. (ii) The assumed IMS core with only one access technology is impractical. The operators always divide the IMS core into several IMS domains according to location and administrative divisions, and each IMS domain contains multiple heterogeneous access networks. The current advance resource reservation algorithm can only be used in one access network [4, 9, 10, 12–14], whereas the choice of the access network before handoff according to operator's rule or network load balance policy is needed. (iii) The path of signaling and media transmitting contains the back-to-back user agent (B2BUA) entities with much address mapping for packet redirection, which become new bottlenecks in the network [13, 14].

This paper proposes a new approach, namely enhanced IMS handoff mechanism (EHM) based on mobility prediction and advance network selective resource reservation, to guarantee QoS when the UE is roaming among IMS domains. The proposed mechanism provides a number of advantages over the existing approaches. First, EHM employs a movement detection scheme to predict UE's next wireless attachment point. This saves network resources by avoiding multiple useless advance reservations. Second, before UE moves to a new IMS domain, the network can select a most appropriate access network and perform resource reservation. Accordingly, this mechanism not only focuses on when to trigger vertical handoff to improve QoS, but also considers all the available networks for the IMS handoff (either homogeneous or heterogeneous). Thus, it can choose the optimal network from all available candidates to utilize network resources more efficiently. Third, EHM guarantees that the UE's IP address is updated in the call session control functions (CSCFs) and the correspondent host (CH) through the session reestablishment after UE's domain handoff, without introducing the B2BUA entity. This assures that media data can be sent directly to the UE without IP address mapping. Finally, EHM only requires some enhanced functionalities in IMS and the UE, importantly, has no great changes on the existing IMS architecture.

The remainder of this paper is organized as follows. Section 2 surveys related work. Section 3 describes the current IMS handoff mechanism that supports end-to-end QoS. Section 4 proposes the EHM, with original contribution presented in detail. In Section 5, the cost of resource reservation, the handoff delay and the session blocking probability under load balance policy are analyzed and modeled by formulas. In Section 6, the numerical and simulation results of resource reservation cost, handoff latency and session blocking probability are presented to investigate the performance of the new mechanism. Finally, conclusion and future work are described in Section 7.

2. RELATED WORK

Improvement of handoff control to decrease latency and support end-to-end QoS, such as new signaling flows, cross-layer protocol, advance resource reservation and mobility prediction, is currently in progress. The work is not only within IMS but also in wireless network, mobile Internet and other non-IMS networks.

For the handoff control, there have been many researches on optimizing the handoff signaling flows in order to reduce the handoff delay [8, 9, 11, 15]. The work in [8] introduces the approach to share the registration information and call states for supporting IMS macro mobility (UE changes IP address). Nilanjan *et al*. [9] introduce a SIP-based architecture to support soft handoff for IP-centric wireless networks. Furthermore, a proactive signaling mechanism is proposed for minimizing handoff delay between access gateways for intra-administrative and inter-administrative mobility scenarios [11].

Huang *et al.* [10] utilize SIP mobility and propose an automatic IPv6 tunneling mechanism to support UE handoff between different networks. Bernaschi *et al.* [16] propose a cross-layer mechanism in which the prediction of bit rate change is used in the session layer. Moreover, Chen *et al.* [12] develop a cross-layer protocol, i.e. SIP mobile Stream Control Transmission Protocol (SmSCTP), and utilize the multi-homing mechanism to reduce handoff delay and to provide a better seamless handoff scheme. Similarly, Thanh *et al.* [13] propose a proxy-based mobile Stream Control Transmission Protocol (mSCTP) for establishing the signaling path before handoff, to realize fast handoff in an IMS heterogeneous environment. Also, Wang *et al.* [17] propose a novel transport-layer soft handoff mechanism based on a concurrent multi-path Stream Control Transmission Protocol (cmpSCTP) which is flow-oriented and switches the traffic to the new path progressively. In [14], Stefano *et al*. propose an application-layer solution for mobility management using SIP extensions and a mobility management server (MMS). The MMS duplicates the media flow and transmits them via the current and the new networks.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

As the IMS handoff triggers session reestablishment, the QoS renegotiation signaling delay is relatively long. The work in [8] only saves a part of the registration and session setup delay, and the soft handoff [9] and fast handoff [13, 14] all introduce B2BUA entities that may result in media transmitting congestion. Therefore, to reduce handoff delay, QoS negotiation along with resource reservation in the new access network during the session reestablishment should be performed in advance.

Advance resource reservation has also been investigated to accelerate the handoff procedure, but not limited in the IMS network. Yang *et al.* [4] propose a mobile QoS framework in IMS based on the concept of SIP multicast with the UE modeled as a transition in the multicast group membership. In addition, the proposed resource reservation algorithm allows the UE to reserve resources in the neighboring domains before handoff; and the reserved resources that can be temporarily exploited by other UEs are marked as inactive. The work reduces the handoff delay as well as obtains the more efficient use of the scarce wireless bandwidth.

On the other hand, Kyounghee *et al.* [15] make use of handoff prediction using layer 2 (L2) information to save network resources by avoiding multiple useless advance resource reservations. The introduced mechanism called selective advance reservations and resource-aware handoff (SARAH) direction establishes a pseudo-reservation between the old base station (BS) and the new one in advance and the pseudo-reservation is activated after handoff. Accordingly, the work in [18] enhances the next step in signaling protocol based on advance resource reservation for supporting host mobility in the mobile Internet. The approach makes use of handoff prediction to detect a crossover node and reserves network resources in advance along the new path that will be used after handoff. It significantly reduces session reestablishment delay caused by handoff.

Then to the point of handoff prediction, the algorithms are mainly proposed in L2, and mostly based on five methods [19]: received signal strength (RSS) with threshold [15, 20, 21], movement extrapolation [22], handover history data [23], mobility pattern, and distance from the access point (AP) or BS.

It is important to mention that, in the heterogeneous network, multi-access technology selection in the handoff decision has been considered. Guo *et al.* [24] present an adaptive multi-criteria vertical handoff decision algorithm. Furthermore, a policy based handoff is proposed, considering many factors such as monetary cost, offered services, network conditions and user preferences [25]. Liu *et al.* [26] propose a general handoff decision algorithm with consideration of both horizontal and vertical handoff in heterogeneous wireless networks.

Since there are so many network selection rules for the terminal, and the general algorithm cannot meet the operators' requirement, network-driven selective mechanisms have been considered for both the handoff sessions and the new arrival sessions. Asanga *et al.* [27] propose a policy engine that receives information from a multitude of sources, and makes handoff according to two kinds of information sources: policy information and environmental information. This policy processing and enforcing environment can be operated on the UE and IMS networks. For the policy-based network selection, the work in [28] proposes a mechanism in which new arrival sessions are assigned to the determined network in accordance with some rules. The network controlled QoS model in IMS is proposed in [29], in which the IMS network issues a request to the UE for making it connect to one of the access networks.

However, none of the previous proposals consider reserving resources in advance according to mobility prediction except for [15]. Nevertheless, the work in [15] is for L2 handoff which is not suitable for handoff in the application layer since the area of the IMS domain is much larger than the BS-serving area, and contains the heterogeneous networks. In the scenario of IMS inter-domain handoff, resource should be proactively reserved in the neighboring IMS domains during QoS negotiation through a mobility prediction algorithm, so as to reduce the handoff delay. Although the advance resource reservation algorithm [4] in IMS has comprehensively considered the heterogeneous network and can effectively use the bandwidth, the selection among multiple access networks in one IMS domain is not of concern.

3. IMS HANDOFF MECHANISM

The overview of the IMS network and session handoff is shown in Fig. 1. The IMS core is on the top of the IP network, and bridges the divide of heterogeneous networks. With the advent of the B3G/4G network, the IMS will provide expanded services for large numbers of users. Thus, operators may implement the whole IMS network through several independent administrative domains with the proxy-call session control functions (P-CSCFs) as the entry points. Although in 3GPP specifications IMS is accessed by general packet radio service (GPRS), some researches extend the IMS access network to different technologies, in order to provide more colorful services to users through different access technologies, such as WLAN [22], WiMax [30, 31] and so on. Each access network contains an access gateway just as the gateway GPRS supporting node (GGSN) in the GPRS network with the capability of resource reservation and admission control. In Fig. 1, the access gateway of WLAN is called the WLAN access gateway (WAG), and the access gateway of WiMax is the access service network gateway (ASNG). At the same time, the UE is able to simultaneously access the various technologies.

3.1. IMS handoff procedure

First, the types of IMS handoff considered in this paper are concluded. On the one hand, according to Fig. 1, handoff in the IMS network can be classified as two types: intra-domain

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

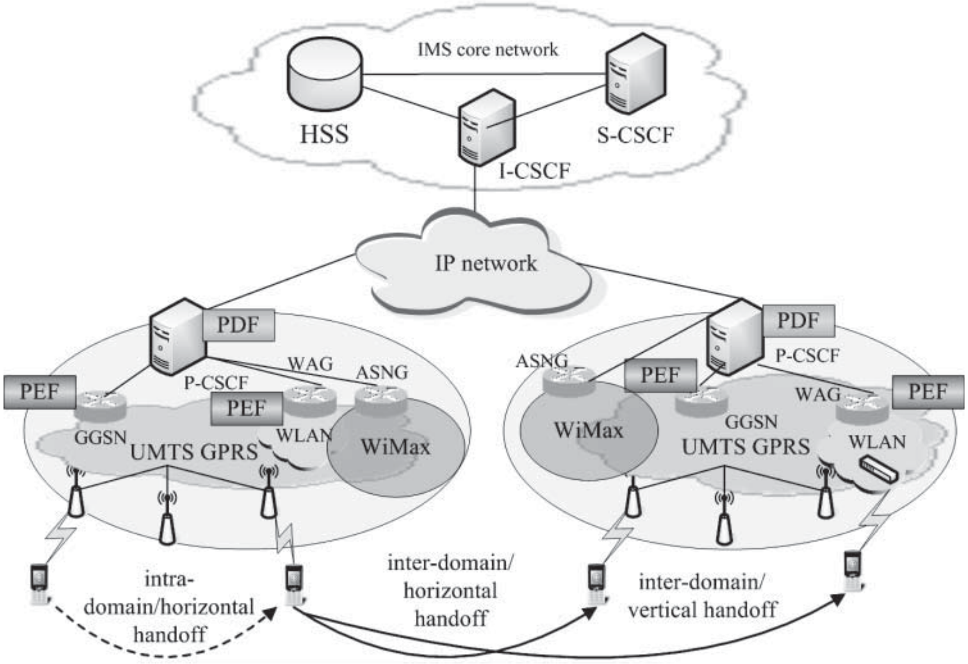


FIGURE 1. IMS network overview.

handoff and inter-domain handoff. If a user moves within one IMS domain (i.e. the area served by one P-CSCF), the handoff is only carried out in L2, which is transparent to IMS, i.e. intra-domain handoff. When a user roams among the domains, IMS inter-domain handoff is triggered. On the other hand, in heterogeneous access networks, handoff can be separated into two parts: horizontal handoff and vertical handoff [32]. A horizontal handoff is performed between different APs within the same link-layer technology such as when transferring an ongoing session from one service GPRS support node (SGSN) to another. A vertical handoff is a handoff between two access networks with different link-layer technologies, such as from GPRS to WiMax.

Second, the binding of the IP address during handoff is presented. In the IMS network, the UE's IP connectivity is established over the IP layer. Once the UE has an IP address, it informs the home serving-call session control function (S-CSCF) and the home subscriber server (HSS) about its new location. Then it can exchange SIP messages in the application layer, either directly or through a gateway, independent of the underlying network-access technology. The actual type of access network is not important for the IMS, but the IMS expects that mobility can be handled in the access network [8]. In other words, if the UE's IP address changes, the ongoing session has to be terminated and the long standard SIP-based session setup procedures have to be performed once more in the new IMS network [13].

Handoff is always caused by user's mobility. There are four cases as follows:

(1) Intra-domain horizontal handoff: a user crossing the network with the same access technology within one P-CSCF serving domain, and the binding IP address does not change. The intra-domain horizontal handoff is transparent to an IMS application layer, which means the actual type of access network and user's location are not important.
(2) Intra-domain vertical handoff: a user moving within one IMS domain but across networks with different access technologies, and the binding IP address changes. The intra-domain vertical handoff needs session reestablishment.
(3) Inter-domain horizontal handoff: a user enters a neighboring IMS domain, with the same access technology, but the P-CSCF is changed.
(4) Inter-domain vertical handoff: a user changes the P-CSCF as well as the access technology.

In cases 2–4, the binding IP address of the UE must change, which leads to a REGISTER and re-INVITE to all of the IMS core entities. After the UE gets a new IP address and completes handoff in L2 and layer 3 (L3) for setting up an IP media data path in the new network, SIP takes over to perform the application layer handoff to complete a new QoS negotiation procedure.

Third, the handoff signaling in IMS contains two procedures: registration and session re-establishment. Figure 2 shows the signaling flow that includes a series of phrases. A preliminary

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

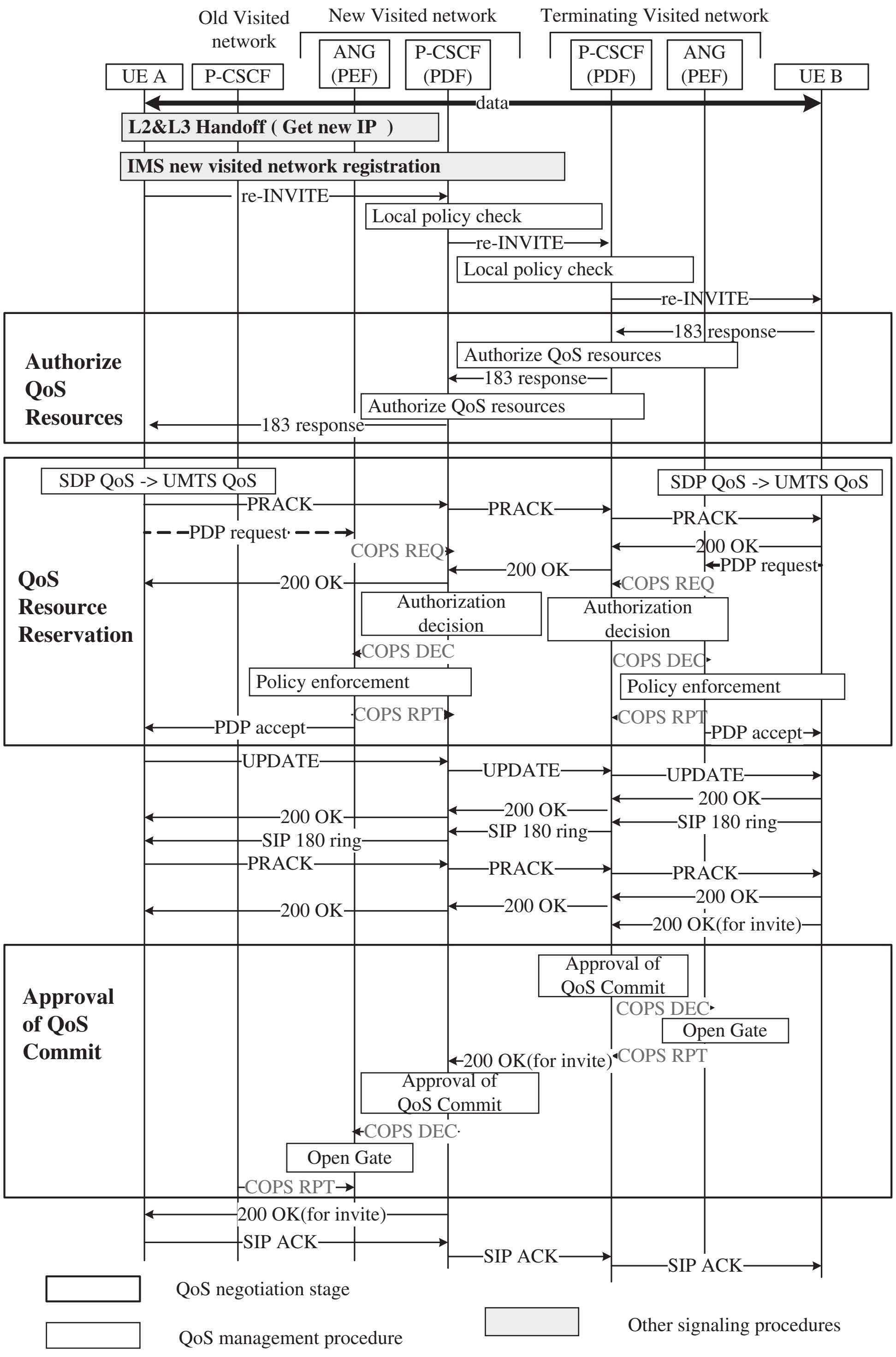


FIGURE 2. IMS inter-domain handoff procedure.

scenario is that UE A has established a session with UE B. Due to the mobility of UE A, it enters another IMS domain. Then the serving P-CSCF is changed, and the binding of a new IP address is triggered.

After handoff takes place, for location management in the IMS, it is essential that a binding is created by the S-CSCF between the public user identity and the IP address of the UE during the registration procedure. This makes it possible for the

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

requests from the other users to be routed from the S-CSCF to the new registered user [33]. Then the ongoing session of UE A is reestablished to inform UE B about UE A's new IP address and renegotiate session QoS parameters. During the session reestablishment, QoS negotiation is finished in the application level through SIP signaling, such as re-INIVTE, 183, PRACK, 180, 200 OK, along with COPS messages REQ, DEC and RPT. This procedure is the same as a new arrival session setup, and provides a logic path for transmitting the media data between UE A in the new visited network and UE B in the terminating visited network. The details are depicted in Fig. 2.

3.2. IMS QoS management

Fundamental QoS parameters for multimedia services are as follows: end-to-end delay, delay variation or jitter and packet loss rate. Since the handoff procedure always interrupts the ongoing session and the IMS requires to provide end-to-end QoS in a wireless access network, QoS management for handoff control contains two aspects.

One is QoS guarantee. A network handoff caused by user mobility during a session leads to end-to-end delay and a considerable packet loss rate which may affect the overall user perceived QoS. Thus, the implementation proposed should aim to minimize the application-level handoff delay, which contributes toward the end-to-end QoS [34].

The other important aspect is QoS negotiation during both IMS session initiation and session reestablishment. The session required QoS is depicted as the parameters in a session description protocol (SDP), including media codec, bit-rate and bandwidth. During the session establishment, these QoS parameters are exchanged by the two communication sides, and negotiated based on the capabilities of the terminal and access networks [5]. IMS QoS management is a policy-based model, including the policy decision function (PDF) and the policy execution function (PEF). The PDF is logically a centralized entity that makes the policy decision according to the rules and the dynamic or static information of the network. The PEF realizes the polices for the resources. The policy-based IMS QoS negotiation contains three stages: 'Authorize QoS Resource', 'QoS Resource Reservation' and 'Approval of QoS Commit' [2], shown in Fig. 2. The PDF carries out the mapping from SDP parameters to IP QoS parameters, authorizes the bearer media data and reserves QoS resources for IMS services in the access network. The access network gateway, e.g. GGSN, serves as the PEF and translates the policy to some measures such as scheduling, queuing, classifying, traffic policy and shaping in the access network to support end-to-end QoS.

According to Section 1, IMS handoff caused by the IP address changing or the P-CSCF varying is complicated. The procedure is carried out after an L2 and L3 handoff to deal with session reestablishment and QoS management. It includes many steps of both SIP and COPS signaling flows which may result in long handoff delay.

4. THE EHM FOR IMS

In this section, we refer to the features of the EHM different from those techniques [4, 15, 16] and present the key design elements of the EHM based on mobility prediction and advance selective resource reservation.

4.1. Challenges for EHM

To the four cases of handoff in Section 3.1, case 1 is not important to the IMS, and case 2 has already been researched deeply [4, 8, 9, 13, 14]. The EHM is proposed to optimize the inter-domain handoff, i.e. cases 3 and 4, by setting up the new media data path before handoff takes place in the application layer. Although the mechanism in [4] tries to finish the negotiation procedure in advance of the handoff for reducing IMS handoff delay in the application layer, the signaling and management cost of resource reservation in all neighboring areas is too big. To finish the QoS negotiation in advance and save the signaling traffic, we research on a mobility prediction algorithm. Also, in order to meet the multi-access requirement in the B3G/4G network [35], selective resource reservation method is proposed. Thus, before the IMS inter-domain handoff, the EHM should reserve resources in the appropriate access network of the domain that the UE most likely enters.

4.2. Architecture and functionalities in EHM

In the EHM architecture, we assume that the UE can communicate through multiple different kinds of network interfaces, and so it may take inter-domain handoff either between two networks with the same access technology or among heterogeneous networks. Furthermore, assume that the IMS domains located in various areas belong to the same operator, and so there is no service-level agreement during the handoff. The enhanced functionalities in the EHM are shown in Fig. 3. The framework of the EHM consists of a mobility management module including a client embedded in the UE and a server in the P-CSCF. Mobility management client is an important part in our handoff mechanism, and provides the information of handoff prediction; the MMS is responsible for getting predicted P-CSCF addresses through a dynamic host

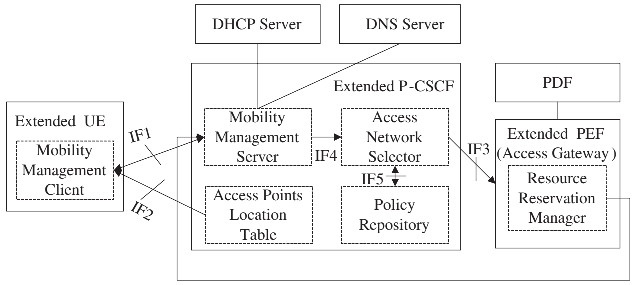

DHCP Server
DNS Server
Extended P-CSCF
PDF
Extended UE
Mobility Management Client
IF1
IF2
Mobility Management Server
IF4
Access Network Selector
IF3
Extended PEF (Access Gateway)
Access Points Location Table
IF5
Policy Repository
Resource Reservation Manager

FIGURE 3. Enhanced functionalities in EHM.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

configuration protocol (DHCP) discovery and a domain name system (DNS) query. The wireless access points location table stores the location topology of the wireless Aps within an IMS domain. An access network selector can choose an appropriate access network according to the network load level and the operator's policy stored in the policy repository if the active interfaces in the UE are more than one. A resource reservation manager in the PEF interacts with the PDF in its IMS domain, and executes the advance reservation algorithm. Also, IF1 to IF5 are new interfaces introduced in this mechanism. Their communication procedures are depicted in Section 4.3.

4.3. Advance QoS negotiation in EHM

In the EHM, handoff control signaling and advance QoS negotiation includes three phases: 'prepare handoff', 'advance resource reservation' and 'handoff procedure', shown as Fig. 4.

(1) When UE A arrives at a certain IMS network either mid-session or pre-session, it issues a SIP REGISTER message. In our EHM, the P-CSCF stores the topology of wireless APs within its IMS domain, so it can indicate to the UE which APs are located in the border area of the current domain. The 200 OK response of REGISTER is extended by adding a new header 'BS-TI' for carrying the topology information of the wireless APs, making reference to the definition method in RFC 3261 [36]. The BNF definition is as follows:

BS:="BS-TI" HCOLON
BSTI-parm*(SEMI BSTI-parm)
BSTI-parm := BS-Id COMMA Lev-Id
BS-Id := DIGIT
Lev-Id := DIGIT

Then the UE can get the list of BS IDs from the registration, such as 'BS-TI=001,0;002,0;0003,1'.

(2) When the mobility management client in UE A predicts that it may move to another IMS domain, advance QoS negotiation is triggered by a pre-REGISTER request. The procedure contains the new P-CSCF discovery and advance resource reservation.

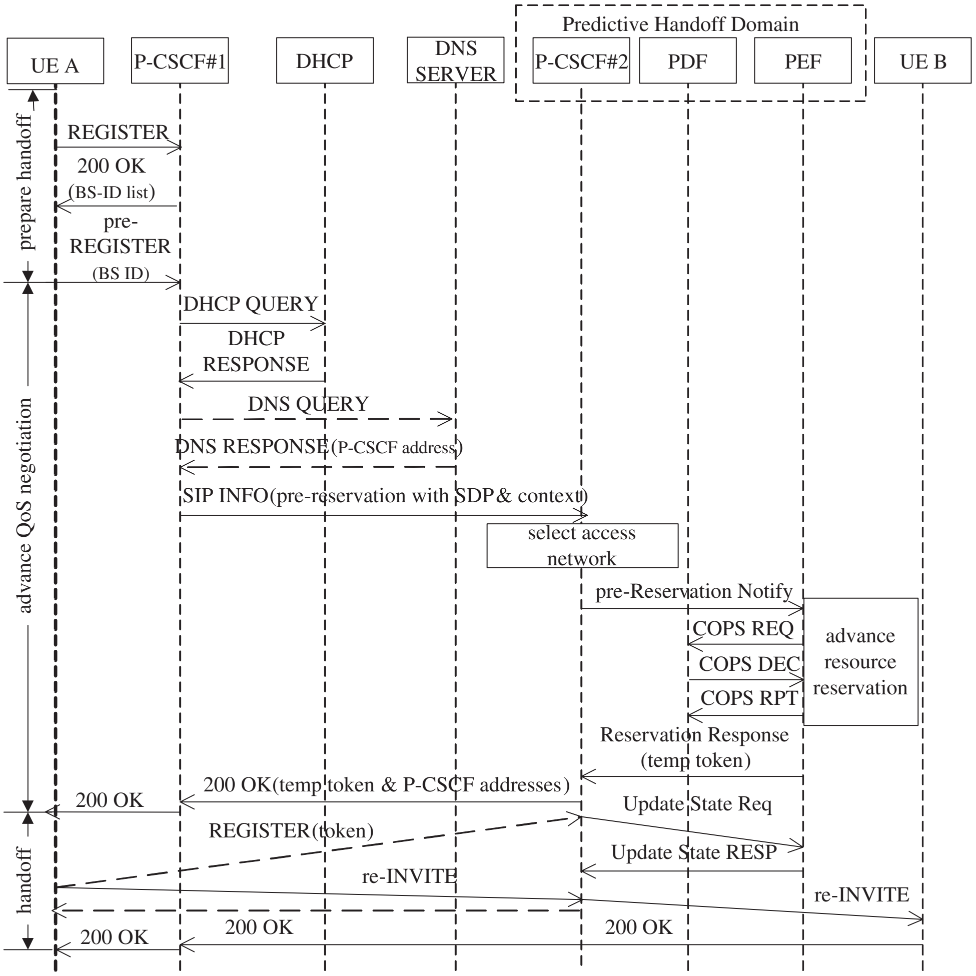


FIGURE 4. Predictive resource reservation signaling.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

First, pre-REGISTER is not a standard IMS SIP message but proposed in this EHM. Here, the pre-REGISTER message is extended by the 'BS-TI' header comparing with the standard REGISTER, in order to indicate the current P-CSCF (P-CSCF#1) to issue a request to obtain the predicted P-CSCF (P-CSCF#2).

Second, in the new P-CSCF discovery procedure, the MMS of the P-CSCF#1 gets the list of predicted APs' IDs by 'BS-TI' and sends a request to the DHCP server. According to the specification [7], it may request a list of fully qualified domain names of the new P-CSCF(s) and the IP addresses of the DNS servers. If the P-CSCF address(es) is not received in the DHCP Query/Response, a DNS query is performed to retrieve the new P-CSCF(s)' IP address(es) [7]. As the wireless APs in the list may be located in several IMS domains, more than one of the P-CSCF addresses will be obtained.

Third, after the P-CSCF#2 is discovered, the P-CSCF#1 notifies it to perform the advance resource reservation by the SIP INFO message. As the serving P-CSCF changes during handoff, a context should be reestablished between UE A and the P-CSCF#2, which contains session context, call states and parameters for the security association [8]. There is a proposed optimization that makes it possible to share the information between P-CSCF#1 and P-CSCF#2 [8]. According to this approach, the INFO message carrying the SDP and the context of the ongoing session is transmitted from P-CSCF#1 to P-CSCF#2.

Finally, P-CSCF#2 receives the SIP INFO message and communicates with the PEF of the selected access network by the pre-Reservation Notify message, to notify it to make advance resource reservation with the algorithm proposed in [4]. Furthermore, the access network allocates an IP address for the UE from the IP pool of the DHCP server. When P-CSCF#2 receives the success response, it adds the temp authorization token, its IP address along with the selected interface in the 200 OK message, and sends the message to UE A through P-CSCF#1. Thus, UE A gets the information of the new network it will connect to and can use this token to identify the media flow after it performs the handoff in the new IMS domain.

(3) When UE A enters the new IMS domain, it connects to the access network which has already been selected. At this moment, in the application layer, UE A sends the re-INVITE and REGISTER messages to UE B through P-CSCF#2 simultaneously. The REGISTER message contains the temp authorization token that is used to make the PEF update the state of the reserved resource. After the P-CSCF receives the success response of state update, it forwards the re-INVITE message to UE B.

In [4], a REGISTER_J message is used to update the UE's location in the SIP multicast group membership. Nevertheless, this approach is not consistent with the end-to-end connection of SIP, since the CH cannot get the new IP address of the UE and it requires an entity, i.e. root server (RS) to keep a large route table for the sake of forwarding the messages. If there is a great quantity of users in the network, the RS must be of heavy load. In the EHM, we follow the standard that a re-INVITE message is used along with the REGISTER message, to notify the CH the UE's new IP address. The REGISTER is the only message in the 'prepare handoff' phase.

In addition, please note that, for the resource reservation of UE B, pre-Reservation Notify is sent from P-CSCF#2 to UE B's visited network, and makes the P-CSCF of UE B reserve resources proactively. The resource is only marked as inactive, not occupied. When re-INVITE arrives at UE B, the gate in UE B's access network gateway is open, but the two procedures are omitted in Fig. 4. Thus, the EHM realizes the end-to-end QoS negotiation and resource reservation before handoff.

4.4. Mobility prediction algorithm

IMS mobility prediction for inter-domain handoff should be carried out before handoff takes place, probably as the UE enters the border area. When the UE moves toward the border of the current IMS domain, it can receive both signals from the access network of the current IMS domain and from the candidate access networks of the IMS domain likely to be entered after handoff. The signal coverage is shown in Fig. 5. When the UE moves across the border area, signal strength of the current access network becomes less than the signal strength of the new access network. Until the signal strength of the new access network exceeds the threshold, the handoff is performed.

Since IMS inter-domain mobility prediction is used to reserve resource proactively, the veracity of prediction is not very important and the prediction trigger point is earlier than the L2 and L3 prediction. Because the reserved bandwidth is only marked instead of occupied and the range of the IMS domain is larger than that of the BS or AP serving area, the mobility prediction algorithm in the EHM only considers the following factors: RSS, direction of UE movement and the coexist multiple access networks. Therefore, we should try to propose an effective mobility prediction algorithm for IMS advance resource reservation.

Some of the current mobility prediction algorithms based on the RSS with threshold and movement extrapolation can be extended for IMS mobility prediction. In [15], the mobile host (MH) movement detection of the SARAH algorithm only considers the beeline moving; yet, in the IMS domain, users may move around or turn the corner or resort halfway. In these scenarios, the resources are reserved in more than one IMS neighboring domains, but will not be used in the future. The scenarios cannot happen in the L2 handoff control, but are possible in the IMS inter-domain handoff. Thus, a timer should be added to the mobile prediction algorithm to solve this problem. On the other hand, in the mobile network, the UE performs a handoff when the RSS of a neighboring cell exceeds the RSS of the current cell within a predefined threshold [21]. However, the RSS often changes gradually rather than breaks suddenly. Consulting the handoff prediction in [15]

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

and the handoff decision method in [21], we propose a novel algorithm based on the idea of detecting the changing of the RSS from neighboring cells, in order to perform advance IMS QoS negotiation.

For convenience, we assume the following:

(1) although in cellular networks, the wireless attachment points are referred to as base stations BSs, and in WLANs, they are called access points (APs), in our algorithm the wireless attachment points for different access networks are all called BS, only for convenience;
(2) UE can simultaneously detect L2 beacon frames from n wireless access networks of one IMS domain;
(3) The signal-to-noise ratio (SNR) of BS or AP is used to express the RSS value [15].

When the UE performs registration in an IMS domain, a mobility management client in the UE downloads the BS topology data from the attach points location table of the P-CSCF. The BS topology data within an IMS domain is divided into several levels. The outermost is 0-level, and from outside to inside, levels are added. When the UE is served by a 0-level BS, it is determined that the UE enters the border area, and then the mobility prediction algorithm is triggered to periodically gather the RSS of the UE. If the mobility management client finds the RSS value of a BS or several BSs from other domains increasing gradually and continuously, it considers that the UE is moving and there is a great probability of entering the neighboring domain (s). Last, the mobility management client sends the pre-REGISTER message to trigger the advance QoS negotiation procedure. The details of the algorithm are as follows:

```
1. Initialization:
  Create RSS record table SST: [BS id, RSS value];
  Create handoff perdition table PT: [BS id, increasing times,
    iFTrigger];
  Set all the observed RSS value and its increasing times
    as 0;
    iFTrigger = false; timer = 0;
2. Get all the observed RSS values of non-current IMS
    domains; //assumed as m RSS values
  For i=1 ··· m
    Calculate RSS variance;
    Set monitored BS_i and the RSS value to SST;
    If (RSS value of monitored BS_i $\neq$ 0 && variance of
     monitored BS_i > 0) // RSS is increasing gradually
      If monitored BS_i in PT
        PT[i].increasing times ++;
      Else
        PT[i].BS id=i;
        PT[i].increasing times=1;
        PT[i].iFTrigger=false;
      End If
    End If
  End For
3. For i=1 ··· size of PT
  If(PT[i].increasing times $\geq$ RSS increasing times threshold
   && PT[i].iFTrigger == false)
     Add BS_i to the prediction BS-ID List;
     PT[i].iFTrigger = true;
    End If
  End For
4. Send pre-REGISTER message with BS-ID list;
   If timer $\geq$ timer threshold
     Performs initialization;
   End If
```

The *RSS increasing times threshold* and *timer threshold* can be obtained from test. The time complexity of this algorithm is O $(2m)$, and the space complexity of this algorithm is O $(2m+3m+m)$. As the capability of the UE is enhanced gradually nowadays, this algorithm is practical. Through this mobility prediction algorithm, the EHM can forecast the time and the place of a roaming user's next move and reserve resource in the access network prior to the user's handoff from the current IMS domain to the new IMS domains.

4.5. Advance network selective resource reservation

An IMS network supports multiple access technologies including UMTS GPRS, WCDMA, WLAN, WiMAX etc. Figures 1 and 5 show the views of a typical heterogeneous IMS network where three access technologies coexist. There are three reasons for network selective resource reservation in IMS inter-domain handoff control. First, there are several differences of the bandwidths, load capabilities and bit rates of the access networks among IMS domains, even though they belong to the same operator. After an inter-domain handoff, the old access technology may not fit to the session. Furthermore, heterogeneous networks within an IMS domain provide a larger set of available resources than a single access network [4], and so a suitable choice of access technology can enhance the network usability and decrease the session blocking probability caused

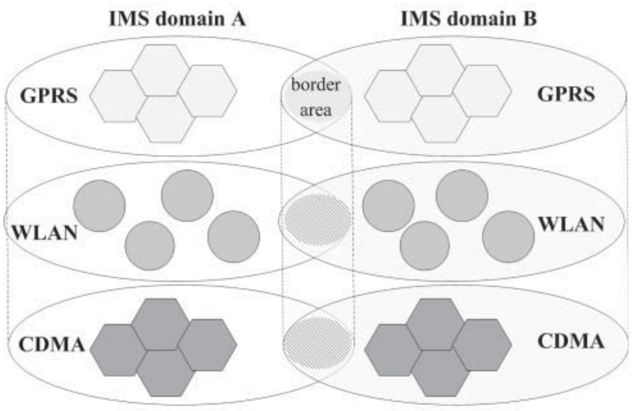


FIGURE 5. Multiple access networks coexist in IMS.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

by resource scarcity of one access network. In addition, future IMS networks require to be environment-aware. This means that lower-layer mechanisms and information should be available for IMS applications so that users and network operators can express their preferences [27].

Accordingly, the multimode UE faces the complicated problem to determine which network it should connect to before entering the new IMS domain. The appropriate selection of an access network can ensure the QoS required by both the sessions and the operators [28]. For this purpose, adequate information of each access network is needed before a selection is made, including precise understanding of the supported service types, system data rates, QoS requirements, communication costs and user preferences. Nevertheless, the UE cannot get enough information concerning its capability and the cost of several signaling interactions between the UE and network since the network availability changes from time to time. Fortunately, the network controlled QoS model in the IMS is proposed in [29], in which IMS network issues a request to the UE for making it access one of the access networks. In the EHM, resource is reserved in the access network before handoff takes place, and so the most suitable network should be selected in advance. Here, we also use the network controlling capability to select one of the access networks to reserve resource before the UE handoff to the new domain. As the resource reservation algorithm in [4] is smart and effective, we can make use of it and focus on how to realize the advance network selective scheme according to the operators' policy as well as different access networks' capabilities, bandwidths, supported service types and so on.

The advance network selective scheme is based on rules, i.e. network selection policy stored in the repository. Each rule in the repository has its corresponding selection policy logic (SPL) embedded in the access network selector. If there are more than one candidate access networks within the new IMS domain, the access network selector in the predicted P-CSCF triggers the SPL according to the rule and decides which access network is the best one. Also, during the execution of the SPL, the access network selector may communicate with other network entities to obtain the related information, for example, it gets the degree of occupation of each access network from the network monitor. The architecture of the network selective scheme is shown in Fig. 6.

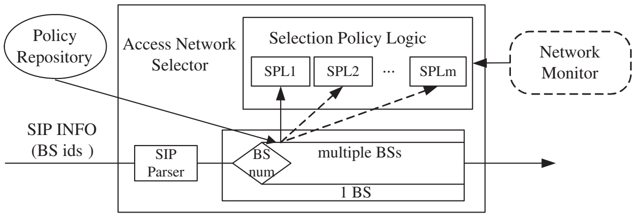


FIGURE 6. The architecture of network selective scheme.

Although there are lots of criteria that can be taken into account, we only list three of the rules to be considered during a UE handoff to a new domain as follows:

(1) User preferences: multimode UE has the capability of communicating through multiple interfaces simultaneously, but users may have their preference. If the IMS domain is configured with this rule, one more signaling between the UE and the predicted P-CSCF is transmitted. Then the access network selector informs the access network to reserve resource before the user enters the new IMS domain, in accordance with its preference.

(2) Network load balance: this policy intends to allocate handoff users to the access network that undergoes a lower load situation at a given time. This policy concentrates especially on the seamless handoff QoS along with the session continuity. Its final purpose is enhancing overall capacity as well as realizing load balance among different access networks.

(3) Operator policy: including service type, current time and so on. The policy can be composed of 1 to r sub-rules which are linked by means of logical expressions (AND, OR, NOT, etc.). Also, if a user has multiple communication sessions simultaneously, each session can perform handoff to a different network, respectively, according to the operator policy.

The detailed procedure of advance network selective resource reservation under the above policies is as follows.

```
1. Parse SIP INFO message;
2. Get BSs list and map each BS to the access networks;
3. If n==1 //only one candidate access network
     Inform the access network to reserve resources;
   Else
     Download the selection policy from repository;
4. Decision:
   If rule==1 //user preference
     Trigger the SPL1;
     Get the preference interface from UE;
     Inform the access network to reserve resources;
   End If
   If rule==2 // network load balance
     Trigger the SPL2;
     For i=1:n //obtain the occupation degree of all the access
     networks from Network Monitor
       Calculate the load level of monitored network i;
     End For
     Compare the load levels of networks 1 to n;
   Inform the lowest level access network to reserve resources;
   End If
   If rule==3 // operator policy
     Trigger the SPL3;
     Parse the SDP in SIP INFO message;
     Match the sub-rules until the selected network is
```

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

```
        determined;
      Inform the access network to reserve resources;
    End If
  ...
    End If
  5. Access network gateway reserves resources along with
      the algorithm in [4];
```

The advance network selective resource reservation guarantees that the user can directly get a most appropriate access method before handoff to another IMS domain. At the same time, the overall utilization of network resource will be optimized from the operator's perspective.

5. ANALYTICAL MODELING

In this section, we analyze the cost of advance resource reservation. It is very important to the whole network operation, and to the best of our knowledge, there is little research on this aspect until now. Moreover, the session incompletion probability under the load balance network selection policy is investigated, as it is one of the parameters of most concern to the operators for handoff control [4]. We define the parameters in Table 1.

For analytical convenience, we assume the following:

(1) The session arrival rate to an IMS domain λ_s, the new session arrival rate λ_0 and the handoff session arrival rate λ_h all follow the Poisson process [4]. Note that λ_s is not equal to the sum of λ_0 and λ_h, as some of the new arrival sessions and handoff sessions may not be served in the IMS domain due to the blocking queue.

(2) The session duration is exponentially distributed [11] with a mean value of μ^{-1}. Furthermore the mean residence time η^{-1} is the general continuous random variable with the probability density function of $f_m(x)$.

(3) There are n access networks that coexist in an IMS area and the max bandwidth of the access network N_i $(1 \leq i \leq n)$ is modeled as n_i users being served at most at the same time.

(4) The cost of resource reservation in each access network with different access technology is the same, and the mean cost value is c.

TABLE 1. Parameters for the analytical model.

Symbol	Definitions
λ_s	the session arrival rate to an IMS domain, except blocking or terminating sessions
μ^{-1}	the mean of session durations (holding time)
η^{-1}	the mean of residence time in an IMS domain
θ	the ratio of session holding time to domain residence time (called hold-to-residence)
λ_0	the new session arrival rate
λ_h	the handoff session arrival rate
n	the number of access networks coexist in an IMS domain
N_i	the access network in an IMS domain, $(1 \leq i \leq n)$
n_i	users being served at most at the same time in an access network $(1 \leq i \leq n)$
p_i	the terminal-driven probability of accessing to N_i for the new arrival session $(1 \leq i \leq n)$
c	the mean cost of resource reservation in each access network
d	the number of the neighboring areas around an IMS domain

5.1. Resource reservation cost

We provide an analysis of the resource reservation cost for our EHM, and compare it with the previous handoff mechanism (PHM) proposed in [4], which is one of the most optimizing advance resource reservation mechanisms in IMS. In this regard, we estimate the number of handoff times among IMS domains as N_h. Considering the timing diagram in Fig. 7, suppose that the UE resides in the R_0 domain at the beginning of session and t_s is the session holding time. During the session holding time, the UE visits other k domains and resides in the ith domain for a period t_{Mi} $(1 \leq i \leq k)$. Other variable definitions are depicted in Fig. 7.

The probability $P(k)$ that the UE moves across k domains during the period of a session is derived in two cases.

(1) For $k = 0$, i.e. the UE does not move out of the domain in which it begins the session, and so $P(k) = P(t_s \leq t_m) = \frac{1}{\theta}(1 - [1 - f_m^*(\mu)])$.

Here, $\theta = \frac{\mu}{\eta}$, $f_m^*(x) = \int_0^\infty f_m(x)e^{-\lambda x}\mathrm{d}x$ is the Laplace–Stieltjes transform for the $f_m^*(x)$ and $\frac{1}{\eta} = \int_0^\infty x f_m(x)\,\mathrm{d}x$.

(2) For $k > 0$, we get

$$P(k) = P(t_m + t_{M_1} + \cdots + t_{M_{k-1}} < t_s \leq t_m + \cdots + t_{M_k}) = \frac{1}{\theta}[1 - f_m^*(\mu)]^2[f_m^*(\mu)]^{k-1}.$$

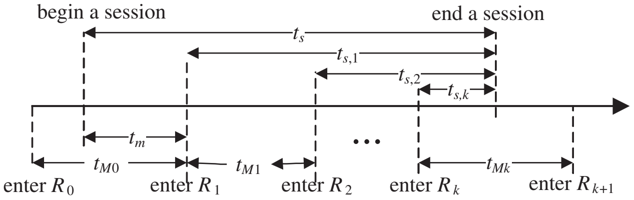


FIGURE 7. UE handoff timing diagram.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

The analysis method is based on the proposed mathematical model in [33, 37] although they analyze the UE moving times between two consecutive sessions. The interested reader is referred to [33] for details.

Then N_h, which is the mathematical expectation of handoff times during a session, can be obtained as follows:

$$N_h = \sum_{k=0}^{\infty} kP(k) = \sum_{k=1}^{\infty} \left(\frac{k}{\theta}\right) [1 - f_m^*(\mu)]^2 [f_m^*(\mu)]^{k-1} = \frac{1}{\theta}.$$

The number of arrival sessions during a period of time t is $N_s = \lambda_s t$.

Assume that there are d neighboring domains around an IMS domain, such as the hexagon network or the $n \times n$ mesh network. The mobility activities of the UE in the analysis model are described by the two-dimensional random walk model [4]. Thus, the UE directly moves to other neighboring IMS domains with the same probability $\frac{1}{d}$, such as situation (1), (5) and (6) in Fig. 8. If the UE turns the corner, such as situation (3), (4), (7) and (8) in Fig. 8, the probability is $(\frac{1}{d})^i$, and i is the number of the borders the UE has crossed continuously among the current IMS domain and the neighboring IMS domains. For example, when the UE moving in the non-shadowed area of situation (1), the RSS from domain A must become stronger and stronger, and so only domain A needs to reserve the resources with the probability of $\frac{1}{d}$. Furthermore, when the UE moves along the path like in situation (3), resource is reserved in the two non-shadowed domains with the probability of $(\frac{1}{d})^2$. The probability that the UE moves along the boundary of two neighbor domains, i.e. situation (2) in Fig. 8, is so small that it can be ignored. Then let t be the unit, and the total cost of the two mechanisms are given by

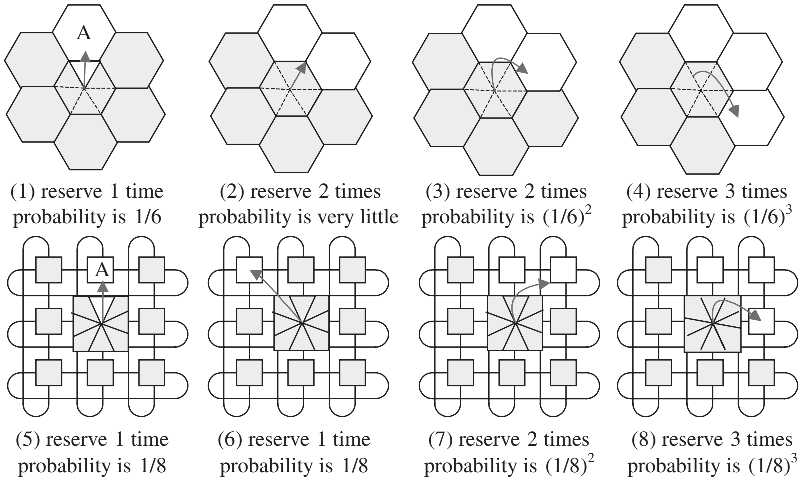
$$\mathrm{TCost}_{\mathrm{PHM}} = N_s \times N_h \times c \times d = \lambda_s \times (1/\theta) \times c \times d, \quad (1)$$

$$\begin{aligned} \mathrm{TCost}_{\mathrm{EHM}} &= N_s \times N_h \times c \times \left[\sum_{i=1}^{d} (1/d)^i \times i\right] \\ &= \lambda_s \times (1/\theta) \times c \times \left[\sum_{i=1}^{d} (1/d)^i \times i\right] \qquad (2) \end{aligned}$$

5.2. Handoff latency

In this section, we make an analytic comparison between the EHM, PHM proposed in [4] and mSCTP-based handoff mechanism (MHM) in [12, 13], in terms of handoff delay. Let $\mathrm{DH}_{\mathrm{PHM}}$, $\mathrm{DH}_{\mathrm{MHM}}$ and $\mathrm{DH}_{\mathrm{EHM}}$ be the handoff delays associated with the PHM, MHM and EHM, respectively. In the PHM, resources have been reserved in all the neighboring areas before handoff occurs, while in the MHM, QoS negotiation and resource reservation have finished after the backup path being established. When the UE performs the handoff, the three mechanisms only need to send messages to update the IP address on the path of SIP signaling. This indicates that when the UE moves into the coverage area of the new IMS domain, the GPRS attach and the packet data protocol context activation procedures are not required any more for the three mechanisms. Thereby, the IMS handoff delay only contains the SIP signaling delay.

$$\begin{aligned} \mathrm{DH}_{\mathrm{MHM}} &= D_{\mathrm{sip}} + D_{\mathrm{SGSN}} + D_{\mathrm{GGSN/PEP}} + D_{\mathrm{P\text{-}CSCF}} \\ &= D_{\mathrm{UE}} + 2D_{\mathrm{RLE}} + 6D_{\mathrm{I}} + D_{\mathrm{SGSN}} \\ &\quad + D_{\mathrm{GGSN/PEP}} + D_{\mathrm{P\text{-}CSCF}}, \qquad (3) \\ \mathrm{DH}_{\mathrm{PHM}} &= D'_{\mathrm{sip}} + D_{\mathrm{QoS}} \\ &= D_{\mathrm{UE}} + D_{\mathrm{RLE}} + D_{\mathrm{P\text{-}CSCF}} \\ &\quad + 8D_{\mathrm{I}} + D_{\mathrm{RS}} + D_{\mathrm{QoS}}, \qquad (4) \\ \mathrm{DH}_{\mathrm{EHM}} &= D''_{\mathrm{sip}} + D_{\mathrm{PEP}} \\ &= D_{\mathrm{UE}} + D_{\mathrm{RLE}} + D_{\mathrm{P\text{-}CSCF}} + 2D_{\mathrm{I}} + D_{\mathrm{PEP}}. \qquad (5) \end{aligned}$$



(1) reserve 1 time probability is 1/6

(2) reserve 2 times probability is very little

(3) reserve 2 times probability is $(1/6)^2$

(4) reserve 3 times probability is $(1/6)^3$

(5) reserve 1 time probability is 1/8

(6) reserve 1 time probability is 1/8

(7) reserve 2 times probability is $(1/8)^2$

(8) reserve 3 times probability is $(1/8)^3$

FIGURE 8. Mobility prediction-based resource reservation model.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

According to [4], an M/M/1 queuing model is assumed for the UE, P-CSCF servers and the other processing entities, e.g. RS and PEP. Their processing delays are assumed as exponential distribution with mean values of D_{UE}, $D_{\text{P-CSCF}}$, D_{PEP} and so on. We denote by D_{RLE} the transport delay of a signaling over an RLP link, while D_{I} is the Internet delay for transmitting SIP messages between two backbone points. In the MHM, the handoff delay is calculated as the round trip time of a SESSONSWITHCH message including transmitting delays from SGSN to GGSN, GGSN to P-CSCF, P-CSCF to CH and the processing delay in SGSN, GGSN and P-CSCF. In the PHM, the REGISTER_J message is transmitted from the current P-CSCF to the previous P-CSCF, forwarded by two RSs and the Merge Point with eight steps. However, in the EHM, when the REGISTER message arrives at the current P-CSCF, only a pair of messages is transmitted between the P-CSCF and the PEP, depicted in Fig. 4, and so the transmitting delay in the EHM is $2D_{\mathrm{I}}$.

5.3. Session incompletion probability

In this section, to present that the advance network selective scheme in the EHM can satisfy operators' economic requirements or performance requirements, we propose the analytical model based on the network-driven load balance selection policy and compare the session incompletion probability for the EHM with that for the current mechanism (CRM).

The new session blocking probability p_0, the forced termination probability p_f and the session incompletion probability p_{nc} are all deduced under the precondition of one access network [4], while there may be several heterogeneous networks coexisting in an IMS domain. We follow the deducing method in [4], and give the multi-network model of these measures. The blocking $M/G/n_i/n_i$ queuing model is used to analyze the blocking probability of each IMS access network. We denote by p_i the terminal-driven probability that the UE accesses to the network N_i for the new arrival session, decided by the terminal characteristics and user's preference.

Thus, for the CRM without a load balance policy in the heterogeneous IMS domain, the traffic intensity of N_i is given by

$$\rho_{\mathrm{i\_CRM}} = \frac{\lambda_{0i}+\lambda_{hi}}{\mu+\eta} = \frac{p_i(\lambda_0+\lambda_h)}{\mu+\eta}. \tag{6}$$

From Equation (6), the new arrival session blocking probability and the forced termination probability in N_i are given by

$$p_{0_i\_\mathrm{CRM}} = p_{fi\_\mathrm{CRM}} = \frac{({\rho_{i\_\mathrm{CRM}}}^{n_i}/n_i!)}{\sum_{t=0}^{n_i}({\rho_{i\_\mathrm{CRM}}}^{t}/t!)}. \tag{7}$$

Then in the whole IMS domain, there is

$$p_{0\_\mathrm{CRM}} = p_{f\_\mathrm{CRM}} = \sum_{i=1}^{n} p_i\, p_{0i\_\mathrm{CRM}} \tag{8}$$

According to deduction in [4], and taking use of Equation (8), the session incompletion probability is given by

$$p_{nc\_\mathrm{CRM}} = \frac{p_{0\_\mathrm{CRM}}}{1-(1-p_{0\_\mathrm{CRM}})[\eta/(\eta+\mu)]}. \tag{9}$$

In the EHM, the set of network selective rules is defined by

$$R = \{r_k \mid f_k(1)\to N_1,\ f_k(2)\to N_2,\\ \cdots,\ f_k(n)\to N_n, \forall k, k\in N\}. \tag{10}$$

Here, N is the set of positive integers. We denote by $r_k \in R$ a special rule for the handoff session to select the access network, with $f_k(i)$ as the feasibility conditions defined by network operators to select network N_i. Then $\alpha_i(r_k)$ with $r_k \in R$ is the probability that the arrival handoff user is assigned to the network N_i, which is determined by the selection rule r_k, and $\sum_{i=1}^{n}\alpha_i(r_k) = 1$. Moreover, different rules are corresponding to different α_i, and so $\alpha_i(r_k)$ is independent with $\alpha_i(r_l)$ when $r_k \in R, r_l \in R$ and $k \neq l$.

Assume the load balance selection rule is r_1. As the operators most care the service success probability, the goal of load balance selection policy is decreasing session incompletion probability p_{nc}. All the idle resource of access network from N_1 to N_n in an IMS domain can be modeled as an $M/G/n'/n'$ blocking queue with the handoff session arriving rate λ_h. We denote by n' the sum of the remaining bandwidth of all the access networks, which can be provided for handoff sessions.

$$n' = \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} p_i(1-p_{0_i\_\mathrm{EHM}})\lambda_0. \tag{11}$$

For r_1, α_i is deduced by taking Equation (11) into account:

$$\alpha_i(r_1) = \frac{n_i - p_i(1-p_{0i\_\mathrm{EHM}})\lambda_0}{n'}. \tag{12}$$

Thus, the traffic intensity of N_i is given by

$$\rho_{i\_\mathrm{EHM}} = \frac{\lambda_{0i}+\lambda_{hi}}{\mu+\eta} = \frac{p_i\lambda_0+\alpha_i(r_1)\lambda_h}{\mu+\eta}. \tag{13}$$

Furthermore, the new arrival session blocking probability in N_i is given by

$$p_{0_i\_\mathrm{EHM}} = \frac{({\rho_{i\_\mathrm{EHM}}}^{n_i}/n_i!)}{\sum_{t=0}^{n_i}({\rho_{i\_\mathrm{EHM}}}^{t}/t!)}. \tag{14}$$

The $p_{0_i\_\mathrm{EHM}}$ and $\alpha_i(r_1)$ can be calculated by the iterative algorithm in [4]. For the handoff sessions, the traffic intensity of the $M/G/n'/n'$ blocking queue is given by

$$\rho'_{\mathrm{EHM}} = \frac{\lambda_h}{\mu+\eta}. \tag{15}$$

Then we derive p_{f_EHM} from Equation (15) as follows:

$$p_{f\_\mathrm{EHM}} = \frac{({\rho'_{\mathrm{EHM}}}^{n'}/n'!)}{\sum_{t=0}^{n'}(\rho'_{\mathrm{EHM}}/t!)}. \tag{16}$$

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

Thus p_{nc_EHM} is calculated as follows:

$$p_{nc\_\mathrm{EHM}} = p_i p_{0i\_\mathrm{EHM}} + \left(\frac{\lambda_h}{\lambda_0}\right) p_f$$
$$= p_i \frac{(\rho_i{}^{n_i}/n_i!)}{\sum_{t=0}^{n_i} (\rho_i{}^{t}/t!)} + \left(\frac{\lambda_h}{\lambda_0}\right) \frac{(\rho'_{i\_EHM}{}^{n'}/n'!)}{\sum_{t=0}^{n'} (\rho'_{i\_EHM}/t!)}. \qquad (17)$$

6. PERFORMANCE SIMULATION AND EVALUATION

In this section, we first verify the validity of equations in Section 5 by using discrete-event simulation experiments, and then we use numerical examples to investigate the performance of the proposed EHM. In our simulation, the IMS network topology consists of several IMS domains, which is the same as in Fig. 3. Each IMS domain includes the P-CSCF, and three different access networks. The IMS core network includes the interrogating-call session control function (I-CSCF), the S-CSCF and the HSS. Then the session arrival, session department and session handoff events for simulating the session traffic and mobility behaviors of IMS users are defined. To investigate the impact of various network parameters on performance of the new mechanism, the session arrival rate λ_0 and λ_s, the mean of session holding time $1/\mu$ and the mean of UE residence time in domain $1/\eta$ are varied by using different simulation configurations.

6.1. Resource reservation cost

For simpleness and without loss of generality, we consider the IMS domains as a hexagon network as well as an 8 × 8 mesh network [4, 37]. A user may move to one of the neighboring IMS domains during the session with the mean residence time varying from 0 to 10 h. The mean session holding time $1/\mu$ is set to be 5 min [11], and the cost of each time of resource reservation is 10.

Table 2 shows the resource reservation cost for both the PHM and the EHM of simulation and analytic results, respectively. The values between simulation and analytic have some discrepancy due to the number of random generated discrete-events. For example, if several more handoff events are generated during the simulation period, the simulation result is bigger than the analytic result and vice versa. As the error rate is under 1%, these experiments have verified that the analytic model is consistent with the simulation results.

As shown in Fig. 9, the higher the ratio of hold-to-residence (θ), which means the weaker the mobility of the UE during session, the total cost values of the two mechanisms become lower. As θ decreases, extremely that UE has several handoff times during a session, the resource reservation cost for the PHM is very large, while the cost for the EHM is much less and tends to become smooth. Also, if there are more neighboring domains, such as modeling with an 8 × 8 mesh network, the advance resource reservation cost for the PHM becomes larger, but the cost for the EHM is similar to the cost for the EHM in the hexagon network model. Because in the PHM there is no prediction of UE movement, the access networks in all the neighboring domains perform resource reservation. Take the hexagon network model for example, when θ increases, UE handoff times grow, which results in much more resource reservation cost in six domains. However, in the EHM, the session at most reserves resources in six domains with very little probability $(\frac{1}{6})^6$, and usually only one domain performs reservation.

Moreover, the session arrival rate to the whole IMS network varies from 1 call per second (cps) to 100 cps, and the mean residence time in one IMS domain is set to be 10 h. Then total cost values of resource reservation for both the PHM and the EHM in hexagon networks and 8 × 8 mesh networks are shown

TABLE 2. Simulation and analytical results

$1/\eta$	Test	Model (PHM)	Error (%)	Test	Model (EHM)	Error (%)
60	50.36	50	−0.72	1.9925	1.9997	0.36
120	25.08	25	−0.32	0.9905	0.9999	0.94
180	16.51	16.667	0.94	0.6682	0.6666	−0.24
240	12.42	12.5	0.64	0.4986	0.4999	0.26
300	10.04	10	−0.40	0.3981	0.3999	0.45
360	8.31	8.3333	0.28	0.3316	0.3333	0.51
420	7.074	7.1429	0.96	0.2875	0.2857	−0.63
480	6.22	6.25	0.48	0.252	0.25	−0.80
540	5.55	5.5556	0.10	0.2215	0.2222	0.32
600	5.01	5	−0.20	0.2019	0.2	−0.95

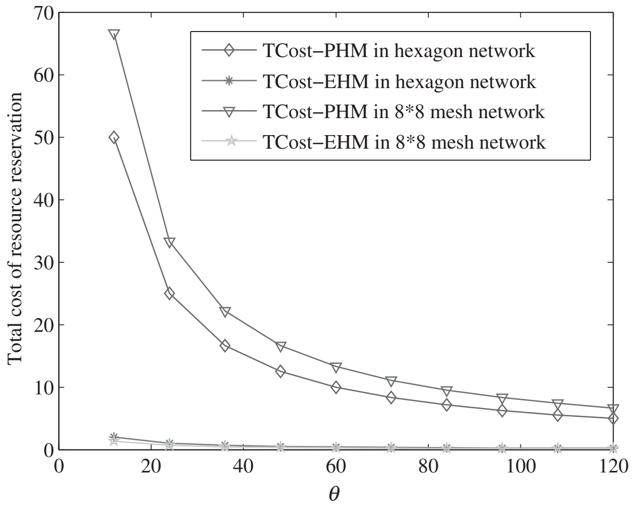


FIGURE 9. Total cost of resource reservation with θ.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

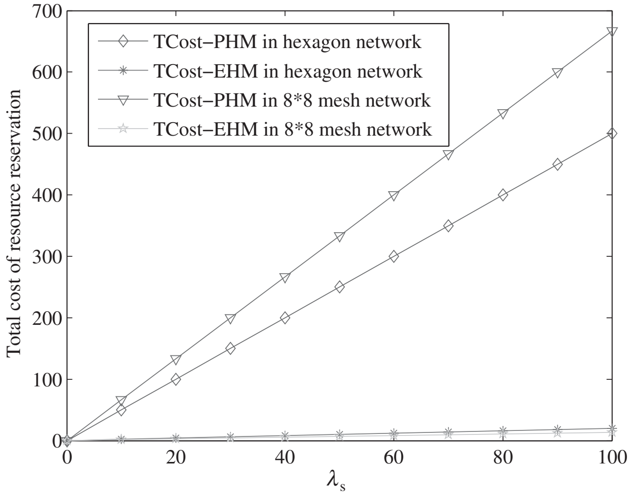


FIGURE 10. Total cost of resource reservation with λ_s.

in Fig. 10. Both the curves of the resource reservation cost become larger, because the increasing of carried sessions in the whole IMS network results in greater times of handoff. As the session arrival rate increases, the resource reservation cost for the PHM grows fast, while the cost for the EHM grows very slowly. Therefore, we can see that the EHM outperforms the PHM when the network operator wants to provide more services for large quantities of users.

6.2. Handoff delay

In [4], there are only some estimates instead of simulation results in terms of the handoff delay. Therefore, for the sake of performance comparison, we also simulate the handoff delay for the MHM, PHM and EHM in the same scenario. In our simulation, we create IMS domains as hexagon networks. According to [4, 38], we adopt the parameters in Table 3. The residence time for which the UE stays in an IMS domain is assumed to be an exponential distribution, and the handoff is generated as discrete-events at a rate of 10 per hour. Then the handoff delays for the three mechanisms are measured and the values of 300 samples, respectively, got from the experiment are depicted in Figs 11–13. The AS-IS values of the handoff delay vary just as shown in the figures and the mean values of them are also depicted.

TABLE 3. Values of delay.

Delay	Values (ms)	Delay	Values (ms)
D_{UE}	15	D_{RS}	30
$D_{\mathrm{P\text{-}CSCF}}$	30	D_{PEP}	30
D_{RLE}	85	D_{QoS}	6
D_{SGSN}	30	D_{I}	13

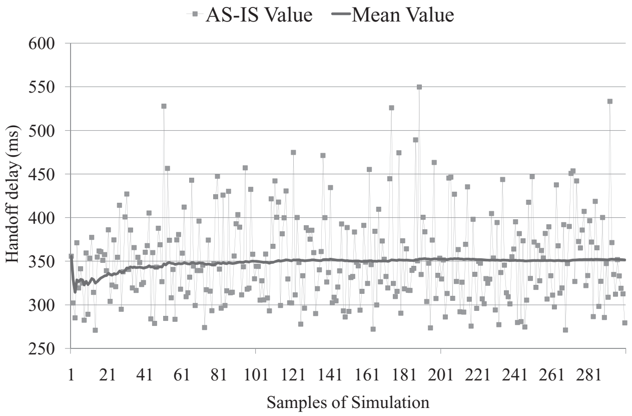


FIGURE 11. The handoff delay for MHM.

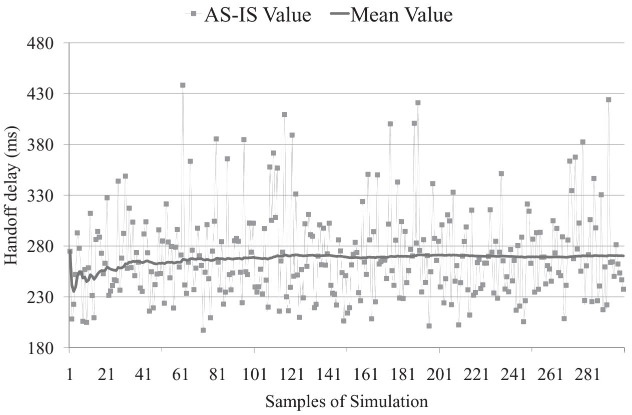


FIGURE 12. The handoff delay for PHM.

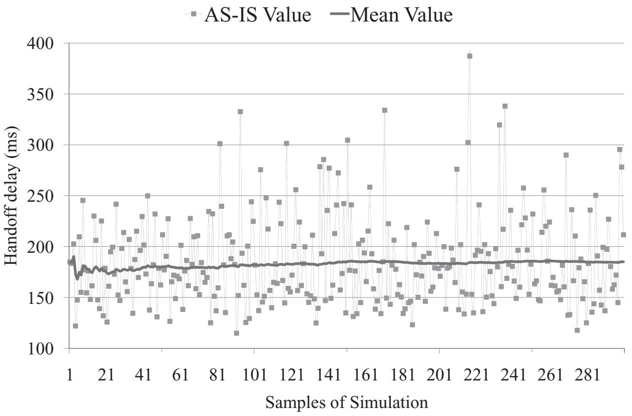


FIGURE 13. The handoff delay for EHM.

According to Equation (3–5), the mean delays of the three mechanisms with the parameters above is calculated as $\mathrm{DH}_{\mathrm{MHM}} = 353$, $\mathrm{DH}_{\mathrm{PHM}} = 270$ and $\mathrm{DH}_{\mathrm{EHM}} = 186$. Thus, Figs 11–13 show that the analytic model and simulation experiments are consistent for the mean handoff delay.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

From Figs 11–13, the proposed EHM shows a smaller delay compared with the previous ones, due to a series of efficient improvements, including less transmission of SIP messages, especially for the wireless interface, and the omitting of the QoS status update.

6.3. Session incompletion probability

As we have deduced the session incompletion probability under the load balance network selection policy for the IMS domain connected with multiple access networks, the discrete-event simulation experiments are presented to verify the Equations (9) and (17). The parameters are adopted as in [4]: $\lambda_s = 30$, $\mu = 1$, $\eta = 2$, $n_1 = 10$, $n_2 = 20$, $n_3 = 20$, and $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{2}$, $p_3 = \frac{1}{3}$. Then the p_{nc} for the CRM and the EHM are measured and 10% of 1600 samples obtained from the experiment are depicted in Fig. 14. At the beginning of the experiment, the channels are not all occupied, and so p_{nc} is very small. Nevertheless, it rises along with the simulation time until it becomes smooth. According to Equations (9) and (17), the session incompletion probabilities for two mechanisms with the parameters above are calculated as $p_{nc\_\text{CRM}} = 0.56$ and $p_{nc\_\text{EHM}} = 0.48$. Figure 14 shows that simulation experiment values are a little bigger than the calculated p_{nc} values.

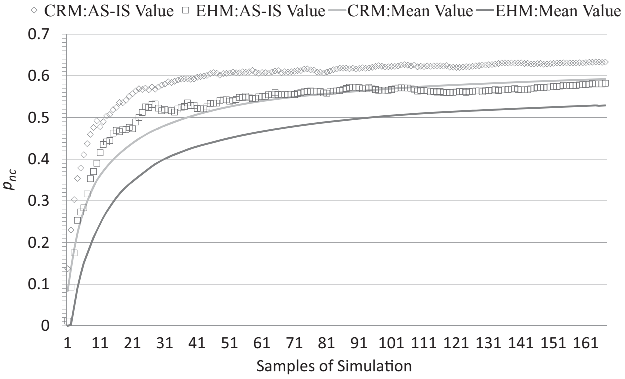
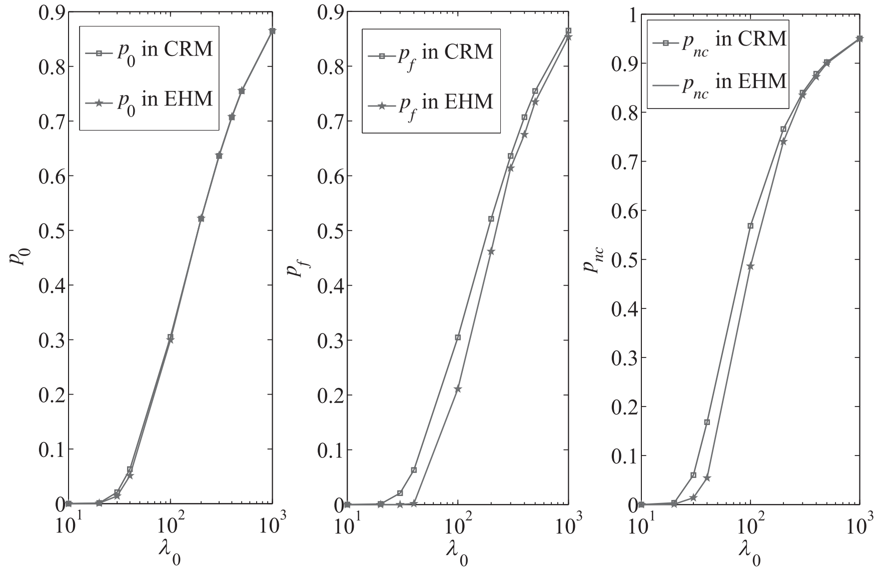


FIGURE 14. Simulation values: p_{nc} for CRM and EHM.

It is important to mention that the load balance policy is performed in advance of the handoff, and so the numbers of users provided by the network monitor may be different from the numbers at the moment the handoff takes place. Thus, the advance network selection is unable to achieve the most ideal situation, i.e. the numerical results. However, as the difference is under 10%, these experiments can verify that analytical results match simulation results. Thus, the equations of p_0, p_f and p_{nc} in Section 5.3 can be used to analyze the performance measures.

Figure 15a–c shows the effects of λ_0 on the output measures p_0, p_f and p_{nc} for both the CRM and the EHM. When λ_0 is small, all of the IMS access networks are not overloaded, and both a new session and a handoff session have a high probability of finding an idle channel, and hence p_0, p_f and p_{nc} are nearly 0. From Fig. 15a, for the new session, there is no impact on the blocking probability if the load balance network selection policy is applied to the handoff session. From Fig. 15b and c, p_f and p_{nc} for the EHM are lower than those for the CRM, especially when λ_0 is less than 100. When the whole IMS domain is in the relatively normal load level, the advantage of selective network



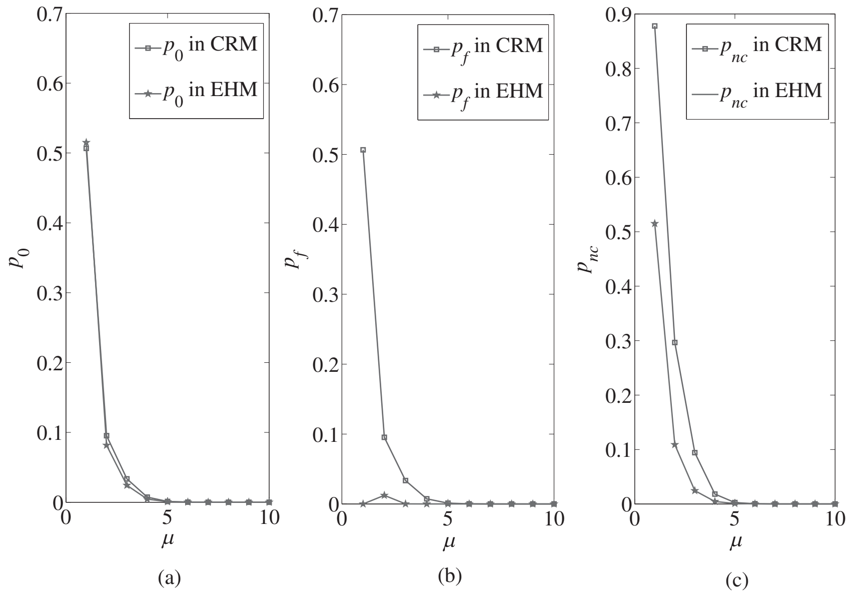
(a) (b) (c)

FIGURE 15. p_0, p_f and p_{nc} with λ_0 ($\mu = 1$, $\eta = 2$).

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

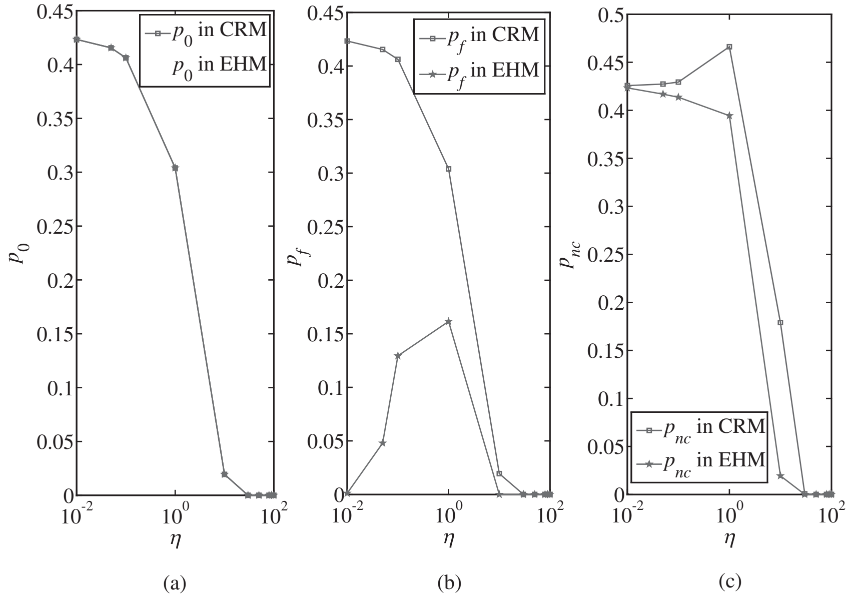
resource reservation is obvious. The reason is that the load balance network selection policy enables handoff sessions to be assigned to the access network that is under relatively low load level, which leads to the more effective use of resources in multiple access networks. However, when all of the access networks have nearly no idle channels, the p_{nc} for the two mechanisms are large, more than 90%.

Let μ vary from 1 to 10, i.e. $1/\mu$ from 1 to 0.1; we can get the curves of p_0, p_f and p_{nc} for both the CRM and the EHM to the changing of session holding time, shown in Fig. 16. Also in Fig. 16a, the load balance policy has no effect on the new session blocking probability. When the session holding time $1/\mu$ increases, it results in longer time of channel occupation and greater handoff times during the sessions. In this case, i.e. Fig. 16b, the p_f for the EHM almost does not grow, while the p_f for the CRM increases fast, for the reason that the load balance policy has more candidate networks to choose in the EHM. Thus, from Fig. 16b and c, we can see that the p_f and p_{nc} for the EHM outperform those for the CRM.



(a) (b) (c)

FIGURE 16. p_0, p_f and p_{nc} with μ ($\lambda_0 = 100$, $\eta = 6$).



(a) (b) (c)

FIGURE 17. p_0, p_f and p_{nc} with η ($\mu = 1$, $\lambda_0 = 80$).

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

Figure 17 depicts the curves of p_0, p_f and p_{nc} for the two comparing mechanisms along with the changing of η. The curves of p_0 for the EHM and the CRM are about the same along with the changing η in Fig. 17a. From Fig. 17b, we see that $\eta < \mu$ implies that the user stays in one IMS domain for a relatively long time. Thus, in the case of $\eta < \mu$, i.e. $\eta < 1$, as η increases, more handoff sessions occur and p_f for the EHM grows. In contrast, $\eta > \mu$, i.e. $\eta > 1$, users always moving to the neighboring IMS domains during session holding time, and so the released channel can be used to serve other coming sessions. Therefore, p_f decreases as η increases. In the two cases, Fig. 17b and c shows that the p_f and p_{nc} for the EHM are smaller than those for the CRM due to the load balance policy.

The above analysis shows that the EHM with the advance network selection load balance policy can effectively decrease the session blocking probability. Furthermore, the EHM can help operators make better use of network resources, especially for the IMS environment with lots of domains and multiple access networks.

7. CONCLUSION

Wireless networks are evolving to provide users with a wider set of access technologies with different capabilities and properties to choose. Nowadays, the IMS which provides an enabling, standardized multimedia architecture is widely applied through independent domains along with multi-access networks. As mobile computing is more and more widespread, mobility support for this IMS architecture becomes very important. Then there is the demand for effective handoff management in the IMS application layer to provide seamless mobile communications, less handoff delay, smaller cost of control and better resource utilization.

In this paper, we discuss the issue of inter-domain handoff in heterogeneous IMS networks. In particular, an EHM is introduced to solve the problem of the big cost of advance resource reservation in all the neighboring areas and make good use of multi-access network resources. The EHM is based on our proposed enhanced IMS functionalities, advance QoS negotiation flows, as well as two important parts: the mobility prediction algorithm coupled with the advance network selective scheme. The mobility prediction algorithm can detect the UE's movement through BS signal strength values in L2. The network selective resource reservation scheme chooses the suitable access network for users who will enter the new IMS domain. We present the details of the proposed approach, and evaluate its performance through simulation experiments. Our simulation results demonstrate that the EHM can effectively reduce the advance resource reservation cost comparing with the previous work; the optimized signaling is able to accelerate the handoff procedure; and the network selective scheme is able to utilize the available multi-access network resource efficiently. In conclusion, the proposed EHM is extremely beneficial for the heterogeneous IMS environment with frequency inter-domain handoff and limited access bandwidth.

The packet loss rate during the handoff is not analyzed in this paper, as we focus on signaling control. Nevertheless, the parameters related to media packet loss are indeed an important QoS factor that greatly concerns the user. The analysis and evaluation of these issues are our future work.

FUNDING

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); (4) Nature Science Fund of China (60902051).

REFERENCES

[1] 3GPP TS 23.002 V.8.2.0 (2007) *Network architecture*. 3rd Generation Partnership Project (3GPP), Valbonne, France.

[2] 3GPP TS 23.228 V.8.3.0 (2007) *IP multimedia system (IMS)*. 3rd Generation Partnership Project (3GPP), Valbonne, France.

[3] 3GPP TS 23.218 V8.0.0 (2007) *IP multimedia (IM) session handling; IM call model*. 3rd Generation Partnership Project (3GPP), Valbonne, France.

[4] Yang, S.R. and Chen, W.T. (2008) SIP multicast-based mobile quality-of-service support over heterogeneous IP multimedia subsystems. *IEEE Trans. Mob. Comput.*, **7**, 1297–1310.

[5] 3GPP TS 23.207 V7.0.0 (2007) *End-to-end Quality of Service (QoS) concept and architecture*. 3rd Generation Partnership Project (3GPP), Valbonne, France.

[6] Munasinghe, K. and Jamalipour, A. (2008) Interworking of WLAN-UMTS networks: an IMS-based platform for session mobility. *IEEE Commun. Mag.*, **46**, 184–191.

[7] 3GPP TS 24.228 V5.15.0 (2006) *Signaling flows for the IP multimedia call control based on SIP and SDP*. 3rd Generation Partnership Project (3GPP), Valbonne, France.

[8] Renier, T., Kim, L.L., Castro, G. and Schwefel, H.P. (2007) Mid-session macro-mobility in IMS-based Network. *IEEE Veh. Technol. Mag.*, **2**, 20–27.

[9] Nilanjan, B., Arup, A. and Sajal, K.D. (2006) Seamless SIP-based mobility for multimedia applications. *IEEE Netw.*, **20**, 6–13.

[10] Huang, S.M., Wu, Q., Lin, Y.B. and Yeh, C.H. (2006) SIP mobility and IPv4/IPv6 dual-stack supports in 3G IP multimedia subsystem. *Wirel. Commun. Mob. Comput.*, **6**, 585–599.

[11] Said, Z. and Admela, J. (2008) A Simple Signaling Mechanism for Seamless Inter-operator Mobility in All-IP Networks. *Proc. Int. Conf. 5th IEEE Consumer Communications and Networking Conf. (CCNC) 2008*, Las Vegas, NV, USA, January 10–12, pp. 381–385. IEEE Press.

[12] Chen, Y.H., Chiu, K.L. and Hwang, R.H. (2007) SmSCTP: SIP-Based MSCTP Scheme for Session Mobility over WLAN/3G Heterogeneous Networks. *Proc. Int. Conf. IEEE Wireless*

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010

Communications and Networking Conference (*WCNC*) *2007*, Hong Kong, March 11–15, pp. 3307–3312. IEEE Press.

[13] Thanh, N.H., Hung N.T., Lan, T.N. and Thomas, M. (2008) Msctp-Based Proxy in Support of Multimedia Session Continuity Support and QoS for IMS-Based Networks. *Proc. Int. Conf. 2nd International Conference on Communications and Electronics* (*ICCE*) *2008*, HoiAn, Vietnampp, June 4–6, pp. 162–168. IEEE Press.

[14] Stefano, S., Andrea, P. and Chiara, M. (2008) SIP-based moblity management in next generation networks. *IEEE Wirel. Commun.*, **15**, 92–99.

[15] Kyounghee, L., Myungchul, K., Chansu, Y., Ben, L. and Hong, S. (2006) Selective advance reservations based on host movement detection and resource-aware handoff. *Int. J. Commun. Syst.*, **9**, 163–184.

[16] Bernaschi, M., Cacace, F., Iannello, G. and Vellucci, M. (2008) Mobility management for VoIP on heterogeneous networks: evaluation of adaptive schemes. *IEEE Trans. Mob. Comput.*, **6**, 1035–1047.

[17] Wang, J.Y., Liao, J.X. and Zhu, X.M. (2008) Latent handover: a flow-oriented progressive handover mechanism. *Comput. Commun.*, **31**, 2319–2340.

[18] Lee, S., Kim, M., Lee, K., Seol, S. and Lee, G. (2008) Seamless QoS Guarantees in Mobile Internet using NSIS with Advance Resource Reservation. *Proc. 22nd Int. Conf. Advanced Information Networking and Applications* (*AINA*) *2008*, GinoWan, Okinawa, Japan, March 25–28, pp. 464–471. IEEE Computer Society Press.

[19] Kwon, H., Yang, M.J. and Park, A.S. (2008) Handover Prediction Strategy for 3G-WLAN Overlay Networks. *Proc. Int. Conf. IEEE/IFIP Network Operations and Management Symposium* (*NOMS*) *2008*, Salvador, Bahia, Brazil, April 7–11, pp. 819–822. IEEE Press.

[20] Kousalya, G., Narayanasamy, P., Park, J.H. and Kim, T.H. (2008) Predictive handoff mechanism with real-time mobility tracking in a campus wide wireless network considering ITS. *Comput. Commun.*, **31**, 2781–2789.

[21] Christopher, C. and Andreas P. (2006) MBMS Handover Control for Efficient Multicasting in IP-Based 3G Mobile Networks. *Proc. IEEE Int. Conf. Communications* (*ICC*) *2006*, Istanbul, Turkey, June 11–15, pp. 2112–2117. IEEE Press.

[22] Soh, W.S. and Kim, H.S. (2003) QoS provisioning in cellular networks based on mobility prediction techniques. *IEEE Commun. Mag.*, **41**, 86–92.

[23] Yu, F., Wong, V.W.S. and Leung, V.C.M. (2005) Performance enhancement of combining QoS provisioning and location management in wireless cellular networks. *IEEE Trans. Wirel. Commun.*, **4**, 943–953.

[24] Guo, Q., Zhu, J. and Xu, X.H. (2005) An Adaptive Multi-criteria Vertical Handoff Decision Algorithm for Radio Heterogeneous Network. *Proc. IEEE Int. Conf. Communications* (*ICC*) *2005*, Seoul, Korea, May 16–20, pp. 2769–2773. IEEE Press.

[25] Zhu, F. and Janise, M. (2006) Multiservice vertical handoff decision algorithms. *EURASIP J. Wirel. Commun. Netw.*, **2**, 2006, 52–52.

[26] Liu, M., Li, Z.C., Guo, X.B. and Dutkiewicz, E. (2008) Performance analysis and optimization of handoff algorithms in heterogeneous wireless networks. *IEEE Trans. Mob. Comput.*, **7**, 846–857.

[27] Asanga, U., Koojana, K., Carmelita, G., Frank, P. and Laurensius, T. (2007) NetCAPE: enabling seamless IMS service delivery across heterogeneous mobile networks. *IEEE Commun. Mag.*, **45**, 84–91.

[28] Xavier, G., Jordi, P., Oriol, S. and Ramon A. (2008) A Markovian approach to radio access technology selection in heterogeneous multiaccess/multiservice wireless networks. *IEEE Trans. Mob. Comput.*, **7**, 1257–1270.

[29] Corici, M.I., de Gouveia, F.C. and Magedanz, T. (2007) A Network Controlled QoS Model over the 3GPP System Architecture Evolution. *Proc. The 2nd Int. Conf. Wireless Broadband and Ultra Wideband Communications 2007*, Sydney, Australia, August 27–30, pp. 39–39. IEEE Press.

[30] Jiao, W.H., Chen, J.F. and Liu, F. (2007) Provisioning end-to-end QoS Under IMS over a WiMAX architecture. *Bell Labs Tech. J.*, **12**, 115–121.

[31] Kumudu, S.M. and Abbas, J. (2008) An Architecture for Mobility Management in Interworked 3G Cellular and WiMAX Networks. *Proc. Symp. Wireless Telecommunications* (*WTS*), Pomona, California, USA, April 24–26, pp. 291–297. IEEE Press.

[32] Nidal, N., Ahmed, H. and Hossam, H. (2006) Handoffs in fourth generation heterogeneous networks. *IEEE Commun. Mag.*, **44**, 96–103.

[33] Cao, Y.F., Liao, J.X., Qi, Q. and Zhu, X.M. (2009) A two-tier location management mechanism for IMS. *High Technol. Lett.*, **15**, 155–161.

[34] Chitra, B. and Khalid A.B. (2007) Towards a User-Centric and Quality-Aware Multimedia Service Delivery. *Proc Int. Conf. Next Generation Mobile Applications Services and Technologies* (*NGMAST*) *2007*, Cardiff, Wales, UK, September 12–14, pp. 36–42. IEEE Press.

[35] Hui, S.Y. and Yeung, K.H. (2003) Challenges in the migration to 4G mobile systems, *IEEE Commun. Mag.*, **41**, 54–59.

[36] Rosenberg, J. *et al.* (2002) *IETF RFC 3261 SIP: Session Initiation Protocol*. The Internet Engineering Task Force (IETF).

[37] Lin, Y.B. (1997) Reducing location update cost in a PCS network. *IEEE/ACM Trans. Netw.*, **5**, 25–33.

[38] Melnyk, M. and Jukan, A. (2006) On Signaling Efficiency for Call Setup in All-IP Wireless Network. *Proc. IEEE Int. Conf. Communications* (*ICC*) *2006*, Istanbul, Turkey, June 11–15, pp. 1939–1945. IEEE Press.

Downloaded from http://comjnl.oxfordjournals.org at Macau University of Science and Technology on July 21, 2010