A Multicast Mechanism for Mobile Multimedia Messaging Service

Ai-Chun Pang and Yuan-Kai Chen

Abstract—Based on the cell broadcast service architecture, this paper proposes an efficient multicast mechanism for the universal mobile telecommunications system to support multimedia messaging service (MMS). We define a new interface between the serving GPRS support node and the cell broadcast center to track the current locations of the multicast members. Then we describe the location tracking procedures (including attach, detach, and location update) of the multicast members and the multicast message delivery procedure. We use an analytic model to investigate the performance of our approach. This paper indicates that our MMS multicast mechanism outperforms the previous proposed approaches.

Index Terms—General packet radio service (GPRS), mobility management, multicast, multimedia messaging service (MMS), universal mobile telecommunications system (UMTS).

I. INTRODUCTION

Existing 2G systems support short message service (SMS) that allows mobile subscribers to send and receive simple text messages (e.g., up to 140 bytes in GSM). In the 2.5G systems (e.g., general packet radio service or GPRS) and the 3G systems (e.g., universal mobile telecommunications system or UMTS), multimedia messaging service (MMS) [12], [4] is introduced to deliver messages of sizes ranging from 30 to 100 K bytes. The content of an MMS can be text (just like SMS), graphics (e.g., graphs, tables, charts, diagrams, maps, sketches, plans, and layouts), audio samples (e.g., MP3 files), images (e.g., photos), video (e.g., 30-s video clips), and so on [12]. An example of the photo MMS is shown in Fig. 1. Fig. 2 illustrates the abstract view of an MMS architecture. In this architecture, the MMS user agent [Fig. 2(a)] resides in a mobile station (MS) or an external device connected to an MS, which has an application-layer function to receive the MMS. The MMS can be provided by the MMS value added service (VAS) applications [Fig. 2(b)] connected to the mobile networks or by the external servers [e.g., email server, fax server; see Fig. 2(d)] in the internet protocol (IP) network. The MMS server [Fig. 2(c)] stores and processes incoming and outgoing multimedia messages. The MMS relay [Fig. 2(e)] transfers messages between different messaging systems, and adapts messages to the capabilities of the receiving devices. It also generates charging data for the billing purpose. The MMS server and the relay can be separated or combined. The MMS user database [Fig. 2(f)] contains user subscription and configuration information. The mobile network can be a (wireless application protocol (WAP) [14]) based 2G, 2.5G, or 3G system [Fig. 2(g)]. Connectivity between different mobile networks is provided by the Internet protocol. For the illustration purpose, Fig. 3 considers UMTS (Release 99) [11] as the mobile network for multimedia messaging service. In this figure, the dashed lines represent signaling links, and the solid lines represent data and signaling links. The core network (CN) consists of two service domains, a circuit-switched (CS) service domain and a packet-switched (PS) service domain. In the CS service domain, UMTS connects to the public switched telephone network (PSTN; see Fig. 3(a)) through the mobile switching center (MSC; see Fig. 3(b)). In the PS service domain, UMTS connects to the packet data network (PDN; see Fig. 3(c)) through the serving GPRS support node (SGSN; see Fig. 3(d)) and the gateway GPRS support node (GGSN; see Fig. 3(e)). The SGSN in the PS domain plays a similar role as the MSC in the CS domain. The GGSN provides interworking with the external PDN, and is connected with SGSNs via an IP-based GPRS backbone network. In Fig. 3 the MMS server/relay connects to the GGSN through a WAP gateway. The UMTS terrestrial radio access network (UTRAN) consists of Node...
BS [the UMTS term for base stations; see Fig. 3(f)] and radio network controllers [RNCs; see Fig. 3(g)] connected by an ATM network. An MS or user equipment communicates with one or more Node Bs through the radio interface Uu based on the Wideband CDMA (WCDMA) radio technology [7].

In UMTS, short messages are delivered through the control plane of the CS domain. The short message is issued from a message sender (e.g., an MS or an input device) to a Short message service center [SM-SC; see Fig. 3(h)]. The SM-SC is connected to a specific MSC called the short message service gateway MSC (SMS GMSC). The SM-SC may connect to several mobile networks, and to several SMS GMSCs in a mobile network. Following the UMTS roaming protocol, the SMS GMSC locates the current MSC of the message receiver by querying the home location register [HLR; see Fig. 3(i)], and forward the message to that MSC. Then the MSC broadcasts the message to the UTRAN, and the corresponding Node Bs page the destination MS. Messages can be stored either in the subscriber identity module or in the memory of the mobile equipment for display on the standard screen of the MS.

On the other hand, multimedia messages can be delivered through either the user plane of the PS or the CS domain. Without loss of generality, we assume that multimedia messages are transmitted over the user plane of the PS domain. In the existing MMS architecture, the mechanisms for MMS unicast and broadcast are well defined. However, no efficient multicast mechanism has been proposed in the literature. Thus in this paper, we describe an efficient multicast mechanism for the PS-domain MMS. Then we use an analytic model to investigate the performance of our approach. Note that although the mechanisms for providing multimedia broadcast and multicast services (MBMS) in the UMTS PS domain were proposed in 3GPP TS 23.246 (this work is still in progress) [6], the 3GPP TS 23.246 approach cannot accommodate the standard UMTS network nodes as well as the standard UEs. In other words, the existing UMTS network nodes such as RNCs and GGSNs cannot operate in the 3GPP TS 23.246 architecture to support multimedia multicasting. A more detailed comparison between 3GPP TS 23.246 and our proposed approach will be presented in the next section.

II. EXISTING MULTICAST MECHANISMS FOR MOBILE NETWORKS

This section describes previously proposed approaches for mobile broadcasting and multicasting. We first introduce the

Similar mechanism can be used in the CS domain, and the details will not be presented in this paper.
concept of location area (LA) and routing area (RA). To track
the MSs, the cells (i.e., the coverage area of Node Bs) in the
UMTS service area are partitioned into several groups. To
deliver services to an MS, the cells in the group covering the MS
are paged to establish the radio link between the MS and the cor-
responding Node B. The location change of an MS is detected
as follows. The Node Bs periodically broadcast their cell identi-
cies. The MS listens to the broadcast cell identity, and compares
it with the cell identity stored in the MSs buffer. If the compar-
isation indicates that the location has been changed, then the MS
sends the location update message to the network. In the UMTS
CS and PS domains, the cells are partitioned into LAs and RAs,
respectively. An RA is typically a subset of an LA. Without loss
of generality, this paper assumes that an RA is equivalent to an
LA. The major task of mobility management [3] is to update
the location of an MS when it moves from one LA (RA) to an-
other. The location information is stored in the UMTS mobility
databases such as the HLR, the visitor location register (VLR)
and the SGSN. In the CS domain, the LA of an MS is tracked by
the VLR, and every VLR maintains the information of a group
of LAs. In the PS domain, the RA of an MS is tracked by the
SGSN and every SGSN maintains the information of a group of
RAs.

Several approaches have been proposed to provide
GSM/UMTS broadcast and multicast services. They are
described as follows.

**Approach I.** GSM voice group call service [2]: This ap-
proach can be used to support MMS when the voice calls
are replaced by multimedia messages. The GSM voice
group call service is provided through a broadcast mecha-
nism. Specifically, the call is delivered to all LAs when a
voice call is destined to the multicast members. Every LA
is paged even if no multicast member is in that area.

**Approach II.** iSMS [13]: In this approach, multicast is
achieved by sending a message to each individual member
in the multicast list. If \( n \) members are in an LA, then the
same message is sent \( n \) times to this LA.

**Approach III.** GSM/UMTS short message multicasting
based on multicast tables [9]: In this approach, the short
messages are only delivered to the LAs where the mul-
ticast members currently reside, and the LAs broadcast
the messages to these MSs. The LAs without multicast
members do not need to establish the communication link
for short message transmission.

Algorithm III utilizes the existing GSM/UMTS short message
architecture as shown in Fig. 4. For the demonstration purpose,
we assume that there are three VLRs in the GSM/UMTS
network: VLR1, VLR2, and VLR3. VLR1 covers location
areas LA1 and LA2. VLR2 covers location areas LA3 and
LA4. VLR3 covers location areas LA5 and LA6. To perform
multicast, the message sender first issues a short message to the
SM-SC, and the SM-SC sends the message to the SMS GMSC
associated with the multicast group [see in Fig. 4, (1)]. Then the
SMS GMSC queries the HLR to identify the MSCs where the
multicast members currently reside [see Fig. 4, (2)] and forward
the message to these MSCs [see Fig. 4, (3)]. Upon receipt of
the short message, the MSCs query the corresponding VLRs to
identify the LAs where the multicast members currently reside
[see Fig. 4, (4)] and page these LAs to establish the radio links
[see Fig. 4, (5)]. In Fig. 4 the message delivery path for SMS
multicast is \((1) \rightarrow (3) \rightarrow (5)\).

Two types of tables are utilized in this multicast mechanism.
A table \( MC_{V} \) is implemented in the HLR to maintain the ad-
resses of the VLRs and the numbers of multicast members re-
siding in the VLRs. A table \( MC_{V} \) is implemented in every VLR
to store the identities of the LAs and the numbers of multicast
members in these LAs. In Fig. 4, there are one multicast member
in LA2 and two multicast members in LA4. Thus, we have

\[
MC_{H}[VLR1] = 1
\]

\[
MC_{H}[VLR2] = 2
\]

and

\[
MC_{H}[VLR3] = 0.
\]

For VLR1, \( MC_{V}[LA1] = 0 \) and \( MC_{V}[LA2] = 1 \). For
VLR2, \( MC_{V}[LA3] = 0 \) and \( MC_{V}[LA4] = 2 \). For VLR3,
\( MC_{V}[LA5] = 0 \) and \( MC_{V}[LA6] = 0 \).
By using these tables, the locations of the multicast members are accurately recorded, and the short messages are only delivered to the LAs where the multicast members currently reside. Details of the location update and message delivery procedures can be found in [9]. Note that the PS-domain MMS multicasting cannot be realized by using this mechanism because the GSM/UMTS SMS architecture is only implemented in the CS domain. In the next section, we propose an efficient MMS multicast mechanism for the UMTS PS domain, which minimizes the number of multimedia messages sent to the RAs.

III. MMS MULTICAST APPROACH IV

Our multicast mechanism (Approach IV) is implemented in the UMTS PS domain, which utilizes the existing cell broadcast service (CBS) architecture [5]. Fig. 5 illustrates an example of the CBS architecture, which consists of two SGSNs: SGSN1 and SGSN2. SGSN1 covers routing areas RA1 and RA2. SGSN2 covers routing areas RA3, RA4, RA5, and RA6. We assume that RNC1 (covering RA1 and RA2) connects to SGSN1. Both RNC2 (covering RA3 and RA4) and RNC3 (covering RA5 and RA6) connect to SGSN2. The multimedia message is first delivered from the message sender to the cell broadcast entity [CBE; see Fig. 5, (1)]. The CBE sends the message to the cell broadcast center [CBC; see Fig. 5, (2)]. The CBC determines the RAs that should receive the multimedia message [see Fig. 5, (3)], and forward the message to the corresponding RNCs [see Fig. 5, (4)]. Then the RNCs multicast the multimedia message to the multicast members [see in Fig. 5, (5)]. In Fig. 5, the message delivery path for MMS multicast is (1) → (2) → (4) → (5). Like Approach III, a multicast table $MC_C$ is implemented in the CBC to maintain the identities of the RAs and the numbers of the multicast members in these RAs. In Fig. 5, there are one multicast member in RA2 and two multicast members in RA4. Thus, we have

\[
\begin{align*}
MC_C[RA1] &= 0, \\
MC_C[RA2] &= 1, \\
MC_C[RA3] &= 0, \\
MC_C[RA4] &= 2, \\
MC_C[RA5] &= 0, \\
MC_C[RA6] &= 0,
\end{align*}
\]

To accurately record the current locations of multicast members, we define a new signaling interface between the CBC and the SGSN. This interface can be based on the Internet protocol or mobile application part (MAP) [3]. Table I shows the signaling message format in this interface. Six message types are defined: Attach Indication/Response, Detach Indication/Response, and RA Update Indication/Response. The Address Field 1 and Address Field 2 specify the addresses of the MSs current and previous RAs, respectively. In the attach, detach and location update procedures, the SGSN informs the CBC of the multicast member’s current location through these signaling messages. Detailed message flows are described in the following subsections.

A. Location Tracking of the Multicast Members

We provide a simplified description of the UMTS/GPRS attach, detach, and location update procedures, and show how the multicast table $MC_C$ is modified through these procedures. The complete description of the standard UMTS mobility management procedures can be found in [11] and [3].

1) Attach for a Multicast Member: When an MS powers on and attaches to the UMTS PS domain, the standard UMTS/GPRS attach procedure is performed to inform the network of the MSs presence. We use the multicast member MS1 in Fig. 5 as an example. In this example, MS1 attaches to SGSN1 with the message flow shown in Fig. 6, which consists of the following steps.

Step 1) MS1 initiates the attach procedure by sending the Attach Request message to SGSN1.

Step 2) The authentication function is performed between MS1, SGSN1, and the HLR.
Step 3) Through the standard UMTS RA update procedure, SGSN1 informs the HLR of the MS1’s current location and obtains the MS1’s subscription profile from the HLR.

Step 4) SGSN1 informs the CBC of the MS1’s RA identity (i.e., RA2 in Fig. 5) by sending the Attach Indication message. The CBC increments by 1. Then it acknowledges SGSN1 by sending the Attach Response message.

Step 5) SGSN1 and MS1 exchange the Attach Accept and the Attach Complete message pair to indicate that the attach procedure is complete.

In the above procedure, Steps 1)–3) and 5) are defined in the standard UMTS specifications [3], and Step 4) is executed if the mobile user is a multicast member. We note that the update of the multicast table is done in microseconds, which can be ignored as compared with the delays of message exchanges in the normal attach procedure.

2) Detach for a Multicast Member: After PS detach is executed, the MS will not receive the GPRS-based services. We use the multicast member MS1 in Fig. 5 as an example to illustrate the detach procedure. In this example, MS1 resides at RA2 that is covered by SGSN1. Simplified steps of the detach procedure are given as follows (see Fig. 7).

Step 1) MS1 detaches from the UMTS PS domain by sending the Detach Request message to SGSN1.

Step 2) Upon receipt of the MS1’s detach request, SGSN1 and the GGSN exchange the Delete PDP Context Request and Delete PDP Context Response message pair to deactivate the MS1’s packet data protocol (PDP) context. SGSN1 sends the Purge MS message to the HLR. This message indicates that SGSN1 has deleted the MS1’s mobility management (MM) and PDP contexts. The HLR acknowledges with the Purge MS Ack message. Then Steps 3) and 4) are executed in parallel.

Step 3) SGSN1 sends the Detach Indication message to the CBC to indicate that MS1 in RA2 has been detached from the network. The CBC decrements MC[i,RA2] by 1. The CBC replies the Detach Response message to SGSN1.

Step 4) If the MS1’s detach is not caused by power-off, SGSN1 sends the Detach Accept message to MS1. At the same time, SGSN1 initiates PS signaling connection release procedure to release the signaling connections between SGSN1 and the UTRAN, and between the UTRAN and MS1.

In the above procedure, Steps 1), 2), and 4) are defined in the standard UMTS specifications [3], and Step 3) is executed if the mobile user is a multicast member. Similar to what we mentioned in Section III-A.1, the cost of updating the multicast table can be ignored.

3) Location Update for a Multicast Member: When an MS moves into a new RA, the RA update procedure is performed. Two types of movements are defined in 3GPP 23.060 [3]: intra-SGSN movement and inter-SGSN movement. In inter-SGSN movement, the previous and current RAs are connected to different SGSNs. We use Fig. 5 as an example to illustrate the inter-SGSN movement, where the multicast member MS1 moves from RA2 to RA3. A simplified description of the location update procedure is given as follows (see Fig. 8).

Step 1) When detecting the RA location change, MS1 issues the Routing Area Update Request message to SGSN2.

Step 2) Through the standard UMTS SGSN context request procedure, SGSN2 obtains the MM and PDP contexts of MS1 from SGSN1.

2The PDP context contains the MS’s PDP address (e.g., the IP address), QoS parameters, and so on.

3The MM context contains the MS’s location information (e.g., the RA address), security parameters, and so on.
B. Mobile Multicast Message Delivery

We elaborate on how multimedia messages are multicasted in Approach IV by using the multicast table $M_C$. The procedure is described in the following steps (see Fig. 5).

Step 1) Multimedia message sender issues the message to the CBE.

Step 2) CBE forwards the message to the CBC.

Step 3) CBC searches the multicast table $M_C$ to identify the routing areas $RA_i$ where the multicast members currently reside (i.e., $M_C[RA_i] > 0$ in the CBC). In Fig. 5, $i = 2$ and 4.

Step 4) CBC sends the multicast message to the destination RNCs (i.e., RNC1 and RNC2 in Fig. 5) through the Write Replace message defined in 3GPP 23.041 [5].

Step 5) RNCs deliver the multimedia messages to the multicast members in the RAs following the standard UMTS call broadcast procedure.

In the above procedure, it is clear that the multimedia messages are only delivered to the RAs that contain multicast members.

C. Comparing Approach IV and 3GPP TS 23.246

This subsection compares Approach IV with the 3GPP TS 23.246 approach [6]. The following issues are discussed.

Backward Compatibility. In the 3GPP TS 23.246 approach, the existing network nodes such as RNCs, SGSNs, and GGSNs are modified to exercise multimedia multicasting. Furthermore, the standard 3GPP TS 23.060 procedures (e.g., attach and location update) performed by the existing UEs should be accordingly revised. However, minor modifications (i.e., a new signaling interface between SGSNs and CBC) to the standard UMTS architecture and procedures are required in Approach IV.

Core-network Resource Consumption. The message delivery path for Approach IV in the core network is $\langle CBC \leftrightarrow RNC \rangle$ while 3GPP TS 23.246 utilizes the path $\langle GGSN \leftrightarrow SGSN \leftrightarrow RNC \rangle$ to transmit multimedia messages. The different message delivery paths for these two approaches result in different bandwidth consumptions. Usually, a longer path implies more bandwidth consumption for message delivery. Thus the 3GPP TS 23.246 approach consumes more resources than Approach IV.

Roaming Support. In Approach IV, 3G users cannot enjoy the multimedia multicast service when roaming from home networks to visited networks. Nevertheless, this problem can be solved in the 3GPP TS 23.246 approach.

Location Accuracy. In 3GPP TS 23.246, MBMS contexts for multicast users are recorded at RNCs, SGSNs, and GGSNs. By using the MBMS contexts at RNCs, the number of multicast users in each cell can be known, and thus multimedia messages are only sent to the cells that contain multicast members. However in Approach IV, scattering of multicast members is only tracked at the RA level.

IV. ANALYTIC MODELING

This section describes an analytic model to investigate the four multicast approaches in the previous sections. In terms of multicast message delivery, Approaches I-III are defined in the CS domain, where the SMS GMSC and the MSC are involved. Approach IV is defined in the PS domain, which involves the SGSN and the CBC. Without loss of generality, we assume that an LA in the CS domain is exactly the same as an RA in the PS domain. That is, the number of LAs in an UMTS system is the same as that of RAs. In the remainder of this paper, we use the term “LA” to represent “RA” in the UMTS PS domain for the description purpose.

We first model how the multicast members are distributed in the LAs of an UMTS system. Assume that the LAs are classified into $I$ categories. For $1 \leq i \leq I$, there are $N_i$ LAs of class $i$. Let $k_i$ be the expected number of multicast members in a class $i$ LA.
Fig. 9. Comparison of analytic and simulation results ($T_i$ has a Gamma distribution with variance $\nu$). (a) $k_i = 0.5$. (b) $k_i = 5$. (c) $k_i = 50$.

and $\pi_i(j)$ be the probability that there are $j$ multicast members in a class $i$ LA. Probabilities $\pi_i(j)$ are derived as follows. We assume that the multicast members enter a class $i$ LA with rate $\lambda_i$, and the multicast members currently reside at the LA for a period $T_i$ that has a general distribution with mean $1/\mu_i$. The aggregate multicast member arrivals can be approximated by a Poisson stream where $\lambda_i = k_i \mu_i$, and $\pi_i(j)$ can be derived from the $M/G/\infty$ model. That is, for $j \geq 0$, we have

$$
\pi_i(j) = \left( \frac{\lambda_i}{j!} \right) e^{-\lambda_i} = \frac{(k_i)^j e^{-k_i}}{j!} \quad \text{where} \quad \rho_i = \frac{\lambda_i}{\mu_i}.
$$

Fig. 9 plots the $\pi_i(j)$ curves obtained from both (2) and the simulation experiments where $k_i = 0.5, 5$, and $50$. The simulation model is similar to the one we developed in [10], and the details are not presented. In the simulation model, we assume that $T_i$ has a Gamma distribution with different variances $\nu = 0.01/\mu_i^2$, $1/\mu_i^2$, and $100/\mu_i^2$. The Gamma distribution is selected because it can approximate many other distributions as well as experimental data [8]. Fig. 9 indicates that the analytic and simulation results are consistent, and $\pi_i(j)$ is not affected by the residence time distribution (specifically, the variance $\nu$ of the Gamma distribution).

The multicast message delivery costs of Approaches I-III are measured by the number of short messages delivered from the SMS GMSC to the MSCs and the number of short messages transmitted from the MSCs to the LAs. For Approach IV, the cost is measured by the number of messages transmitted from the CBC to the LAs. Note that the cost definition here is different from that in [9]. In [9], the costs of the three approaches for GSM/UMTS short message multicasting are measured by
the number of paging messages sent from MSCs to the LAs. On the other hand, the performance for the multicasting approaches in this paper is evaluated based on the number of multicast messages delivered from SMS GMSC/CBC (via MSCs) to the LAs/RAs.

Let $M_1$ be the number of MSCs in the UMTS system and $M_2$ be the number of LAs covered by each MSC (in Fig. 4, $M_1 = 3$ and $M_2 = 2$). That is

\begin{equation}
M_1M_2 = \sum_{i=1}^{l} N_i. \tag{3}
\end{equation}

Let $S'$ be the number of multicast members in the UMTS system. In Fig. 4, $S' = 3$. Let $N$ be the number of LAs where the multicast members currently reside (it is clear that $N \leq M_1M_2$), and $P$ be the number of MSCs where the multicast members currently reside. In Fig. 4, $N = 2$ and $P = 2$. We first derive the expected multicast cost for Approach IV, and then derive the costs for Approaches I-III. We assume that the delivery cost from the SM-SC to the SMS GMSC (in Approaches I-III) and the delivery cost from the CBE to the CBC (in Approach IV) are the same. This cost will not be considered in our analysis.

- In Approach IV, the multimedia messages are only delivered from the CBC to the LAs where the multicast members currently reside. Thus the expected multicast cost is the expected number $E[N]$ of LAs where the multicast members reside.

For $1 \leq i \leq l$, let $n_i$ be the number of class $i$ LAs where the multicast members currently reside (it is clear that $0 \leq n_i \leq N_i$). From (2) the expected multicast cost $C_{IV}$ is expressed as

\begin{equation}
C_{IV} = E[N] = \sum_{n_1=0}^{N_1} \cdots \sum_{n_l=0}^{N_l} (n_1 + \cdots + n_l) \binom{N_1}{n_1} \binom{N_2}{n_2} \cdots \binom{N_l}{n_l}
\end{equation}

\begin{equation}
\times \left[ 1 - \pi_1(0) \right]^{n_1} \left[ \pi_1(0) \right]^{N_1-n_1} \times \left[ 1 - \pi_2(0) \right]^{n_2} \left[ \pi_2(0) \right]^{N_2-n_2} \times \cdots \times \left[ 1 - \pi_l(0) \right]^{n_l} \left[ \pi_l(0) \right]^{N_l-n_l}
\end{equation}

\begin{equation}
= A + B \tag{5}
\end{equation}

where

\begin{equation}
A = \sum_{n_1=0}^{N_1} \cdots \sum_{n_l=0}^{N_l} (n_2 + \cdots + n_l)
\end{equation}

\begin{equation}
\times \prod_{i=2}^{l} \left\{ \binom{N_i}{n_i} \left[ 1 - \pi_i(0) \right]^{n_i} \left[ \pi_i(0) \right]^{N_i-n_i} \right\}
\end{equation}

\begin{equation}
\times \sum_{n_1=0}^{N_1} \binom{N_1}{n_1} \left[ 1 - \pi_1(0) \right]^{n_1} \left[ \pi_1(0) \right]^{N_1-n_1}
\end{equation}

\begin{equation}
= \sum_{n_1=0}^{N_1} \cdots \sum_{n_l=0}^{N_l} (n_2 + \cdots + n_l)
\end{equation}

\begin{equation}
\times \prod_{i=2}^{l} \left\{ \binom{N_i}{n_i} \left[ 1 - \pi_i(0) \right]^{n_i} \left[ \pi_i(0) \right]^{N_i-n_i} \right\}
\end{equation}

\begin{equation}
B = \left\{ \prod_{i=2}^{l} \left( \sum_{n_i=0}^{N_i} \binom{N_i}{n_i} \left[ 1 - \pi_i(0) \right]^{n_i} \left[ \pi_i(0) \right]^{N_i-n_i} \right) \right\}
\end{equation}

\begin{equation}
\times \left\{ \sum_{n_1=0}^{N_1} \binom{N_1}{n_1} \left[ 1 - \pi_1(0) \right]^{n_1} \left[ \pi_1(0) \right]^{N_1-n_1} \right\}
\end{equation}

\begin{equation}
= N_1 [1 - \pi_1(0)] \left\{ \sum_{n_1=0}^{N_1} \binom{N_1}{n_1} [1 - \pi_1(0)]^{n_1-1} \left[ \pi_1(0) \right]^{N_1-n_1} \right\}
\end{equation}

\begin{equation}
\times \left[ \pi_1(0) \right]^{N_1-n_1-(n_1-1)} \tag{7}
\end{equation}

Let $n_i^* = n_i - 1$. From (2), for any $i$, $\pi_i(0) = e^{-k_i}$.

Therefore, (7) is re-written as

\begin{equation}
B = N_1 (1 - e^{-k_1}) \left\{ \sum_{n_1=0}^{N_1-1} \binom{N_1-1}{n_1-1} \left[ 1 - \pi_1(0) \right]^{n_1-1} \left[ \pi_1(0) \right]^{N_1-n_1} \right\}
\end{equation}

\begin{equation}
\times (1 - e^{-k_1})^{n_i^*} (e^{-k_1})^{(N_1-1)-n_i^*} = N_1 (1 - e^{-k_1}) \tag{8}
\end{equation}

From (6) and (8), (5) is rewritten as

\begin{equation}
C_{IV} = \sum_{i=1}^{l} N_i (1 - e^{-k_i}) \tag{9}
\end{equation}

- In Approach I, the messages are delivered from the SMS GMSC to every LA even if no multicast member is in that LA. It is clear that the expected multicast cost $C_I$ is expressed as

\begin{equation}
C_I = M_1 + M_1M_2 \tag{10}
\end{equation}

where $M_1$ represents the number of messages sent from the SMS GMSC to the MSCs (i.e., the total number of MSCs in the UMTS system) and $M_1M_2$ represents the number of messages sent from the MSCs to the LAs (i.e., the total number of LAs in the UMTS system).

Let $\theta_I$ be the expected cost for Approach I normalized by the cost of Approach IV. From (3), (9), and (10), we have

\begin{equation}
\theta_I = \frac{C_I}{C_{IV}} = \frac{M_1 + M_1M_2}{E[N]}
\end{equation}

\begin{equation}
= \left( \sum_{i=1}^{l} N_i + M_1 \right) \left[ \sum_{i=1}^{l} N_i (1 - e^{-k_i}) \right]^{-1} \tag{11}
\end{equation}

- In Approach II, the multicast messages are sent from the SMS GMSC to every individual member in the multicast list. If there are $S$ multicast members in the UMTS system, the expected multicast message delivery cost $C_{II}$ for this approach is expressed as

\begin{equation}
C_{II} = 2E[S] \tag{12}
\end{equation}
where the expected number of messages delivered from the SMS GMSC to the MSCs is $E[S]$ and the expected number of messages transmitted from the MSCs to the LAs also is $E[S]$. Thus (12) is expressed as

$$C_\Pi = 2 \left( \sum_{i=1}^{l} N_i k_i \right).$$  \hfill (13)

Let $\theta_\Pi$ be the expected cost for Approach II normalized by the cost of Approach IV. From (9) and (13), we have

$$\theta_\Pi = \frac{C_\Pi}{C_\text{IV}} = \frac{2E[S]}{E[N]} = 2 \left( \sum_{i=1}^{l} N_i k_i \right) \left( \sum_{i=1}^{l} N_i (1 - e^{-k_i}) \right)^{-1}. \hfill (14)$$

- In Approach III, the short messages are only sent from the SMS GMSC to the LAs where the multicast members currently reside. Thus the expected multicast message delivery cost $C_\text{III}$ is expressed as

$$C_\text{III} = E[P] + E[N]$$  \hfill (15)

where $E[P]$ represents the expected number of messages delivered from the SMS GMSC to the MSCs and $E[N]$ represents the expected number of messages delivered from the MSCs to the LAs.

Let $\theta_\text{III}$ be the expected cost for Approach III normalized by the cost of Approach IV, then from (9) and (15) we have

$$\theta_\text{III} = \frac{C_\text{III}}{C_\text{IV}} = \frac{E[N] + E[P]}{E[N]}. \hfill (16)$$

To derive $E[P]$, we consider two cases for the distribution of class $i$ LAs in an UMTS system.

**Distribution A:** The LAs of each class are uniformly distributed among the $M_1$ MSCs. That is, for $1 \leq i \leq l$, there are $N_i/M_1$ class $i$ LAs in each MSC. Without loss of generality, we assume that $N_i$ is a multiple of $M_1$. From (2), the expected number $E[P]$ is derived as follows:

$$E[P] = \sum_{m=0}^{M_1} m \Pr[P = m]$$

$$= \sum_{m=0}^{M_1} m \left( \frac{M_1}{M_1} \right)^m \left[ 1 - \prod_{i=1}^{l} \left( \frac{N_i}{M_1} \right) \right]^m$$

$$\times \left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_1} \right]^{M_1 - m}$$

$$= M_1 \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_1} \right]$$

$$\times \left\{ \sum_{m=1}^{M_1} \left( \frac{M_1 - 1}{M_1} \right)^m \left( \frac{M_1 - m}{m - 1} \right) \right\}$$

$$\times \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_1} \right]^{m-1}$$

From (9), (16) and (17), we have

$$\theta_\text{III} = \left\{ \sum_{i=1}^{l} \left[ N_i (1 - e^{-k_i}) \right] \right\} + M_1 \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_1} \right]$$

$$\times \left\{ \sum_{m=1}^{M_1} \left( \frac{M_1 - 1}{M_1} \right)^m \left( \frac{M_1 - m}{m - 1} \right) \right\}$$

$$\times \left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_1} \right]^{M_1 - 1}.$$  \hfill (18)

**Distribution B:** All LAs covered by one MSC are of the same class. That is, for $1 \leq i \leq l$, there are $N_i/M_2$ MSCs covering class $i$ LAs, which is called the class $i$ MSC. Without loss of generality, we assume that $N_i$ is a multiple of $M_2$. Let $m_i$ be the number of class $i$ MSCs where the multicast members currently reside (it is clear that $0 \leq m_i \leq (N_i)/(M_2)$). Thus from (2), we have

$$E[P] = \sum_{m_1=0}^{M_1} \cdots \sum_{m_l=0}^{M_1} (m_1 + \cdots + m_l) \left( \frac{N_i}{M_2} \right)^{m_1} \left( \frac{N_i}{M_2} \right)^{m_2} \cdots \left( \frac{N_i}{M_2} \right)^{m_l}$$

$$\times \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_2} \cdots$$

$$\left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{M_1 - m_l} \left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_l}$$

$$= \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_1} \cdots \sum_{m_l=0}^{M_1} (m_1 + \cdots + m_l) \left( \frac{N_i}{M_2} \right)^{m_1} \left( \frac{N_i}{M_2} \right)^{m_2} \cdots \left( \frac{N_i}{M_2} \right)^{m_l}$$

$$\times \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_2} \cdots$$

$$\left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{M_1 - m_l} \left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_l}$$

From (9), (16) and (17), we have

$$\theta_\text{III} = \left\{ \sum_{i=1}^{l} \left[ N_i (1 - e^{-k_i}) \right] \right\} + M_1 \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]$$

$$\times \left\{ \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_1} \cdots \sum_{m_l=1}^{M_1} (m_1 + \cdots + m_l) \left( \frac{N_i}{M_2} \right)^{m_1} \left( \frac{N_i}{M_2} \right)^{m_2} \cdots \left( \frac{N_i}{M_2} \right)^{m_l}$$

$$\times \left[ 1 - \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_1} \left[ \pi_1(0)^{M_2} \right]^{N_i/M_2 - m_2} \cdots$$

$$\left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{M_1 - m_l} \left[ \prod_{i=1}^{l} \pi_i(0)^{N_i/M_2} \right]^{m_l}$$

We note that the normalized costs $\theta_1, \theta_\Pi$ and $\theta_\text{III}$ are not affected by the RA residence time distribution of the multicast members.
and decreases, the expected value typically ranges from 1000 to 5000. Thus, the costs of Approaches I-IV are dominated by the multicast message delivery.

In the remainder of this section, we focus on the analysis of the multimedia message delivery cost. We consider two classes of LAs. A class 1 LA has the multicast user traffic $\rho_1 = \lambda_1/(\mu_1) = \delta$, and a class 2 LA has the user traffic $\rho_2 = \lambda_2/(\mu_2) = 1/\delta$. A larger $\delta$ value implies that it is more likely to find at least one multicast members in a class 1 LA, and to find no multicast member in a class 2 LA. Let $\alpha$ be the portion of class 1 LAs. That is, $N_1 = \alpha M_1 M_2$, and $N_2 = (1 - \alpha) M_1 M_2$. In this paper, we consider $M_1 = 10$ and $M_2 = 10$. For other $M_1$ and $M_2$ values, similar results are observed and will not be presented in this paper.

**Delivery Cost for Multicast Message.** According to the discussions in the previous section, the multicast message delivery costs for Approaches I-IV are $\beta C_1, \beta C_{II}, \beta C_{III}$ and $\beta C_{IV}$, respectively, where $\beta$ is the ratio of the multicast message size to the signaling message size.

Based on 3GPP 24.008 [1], the $\beta$ value typically ranges from 1000 to 5000. Thus, the costs of Approaches I-IV are dominated by the multicast message delivery.

In this paper, we consider two classes of LAs. A class 1 LA has the multicast user traffic $\rho_1 = \lambda_1/(\mu_1) = \delta$, and a class 2 LA has the user traffic $\rho_2 = \lambda_2/(\mu_2) = 1/\delta$. A larger $\delta$ value implies that it is more likely to find at least one multicast members in a class 1 LA, and to find no multicast member in a class 2 LA. Let $\alpha$ be the portion of class 1 LAs. That is, $N_1 = \alpha M_1 M_2$, and $N_2 = (1 - \alpha) M_1 M_2$. In this paper, we consider $M_1 = 10$ and $M_2 = 10$. For other $M_1$ and $M_2$ values, similar results are observed and will not be presented in this paper.

**Comparison of Approaches I, II, and IV.** By using (11) and (14), Fig. 10 plots $\theta_I$ and $\theta_{II}$ as functions of $\alpha$, where $\delta = 10, 100$ and 1000. To explain the phenomena shown in Fig. 10, we first observe the following facts.

**Fact 1**  From (10), it is clear that the expected multicast cost $C_I$ for Approach I is fixed.

**Fact 2**  As $\alpha$ or $\delta$ increases, the expected number $E[S]$ of multicast members increases, and the multicast cost $C_{II}$ for Approach II increases [see (12)].

**Fact 3**  As $\alpha$ increases or $\delta$ decreases, the expected number $E[N]$ of LAs where the multicast members

![Fig. 10. $\theta_I$ and $\theta_{II}$ performance.](image-url)
reside increases, and the multicast cost $C_{IV}$ for Approach IV increases [see (4)].

From Facts 1) and 3), $\theta_I$ increases as $\alpha$ decreases or $\delta$ increases (see the solid curves in Fig. 10). Based on Facts 2) and 3), $\theta_{II}$ increases as $\delta$ increases (see the dashed curves in Fig. 10). The dashes curves in Fig. 10 also indicates that $\theta_{II}$ increases as $\alpha$ increases. The phenomenon is explained as follows. As $\alpha$ increases, both $E[N]$ and $E[S]$ increase [see Facts 2) and 3)]. However, $E[N]$ does not increase as fast as $E[S]$ does. Therefore from (14), $\theta_{II}$ increases as $\alpha$ increases. We note that Approach IV significantly outperforms Approach I when $\alpha < 0.2$ (that is, when there are less than 20% of class 1 LAs in the system). Approach IV significantly outperforms Approach II when $\alpha > 0.1$.

Comparison of Approaches III and IV. Based on (18) and (20), Fig. 11 plots $\theta_{III,A}$ and $\theta_{III,B}$ against $\alpha$ and $\delta$. This figure indicates that both $\theta_{III,A}$ and $\theta_{III,B}$ increase as $\alpha$ decreases. Specifically, Approach IV significantly outperforms Approach III when $\alpha < 0.2$ (i.e., the number of class 2 LAs is large). The reason is given below. From (16), $\theta_{III}$ can be rewritten as

$$\theta_{III} = 1 + \frac{E[P]}{E[N]}$$  \hfill (21)

where $E[P]$ is the expected number of messages delivered from the SMS GMSC to the destination MSCs and $E[N]$ is the expected number of messages sent from the destination MSCs to the LAs. From the previous discussions, we observe the following fact.

Fact 4: As $\alpha$ decreases or $\delta$ increases, the number of LAs without any multicast member increases and $E[N]$ decreases [see Fact 3]). However, it is not very likely that all LAs in an MSC do not have any multicast member. Therefore, $E[P]$ does not decrease as fast as $E[N]$ does, and the $E[P]$-to-$E[N]$ ratio increases.

From Fact 4, both $\theta_{III,A}$ and $\theta_{III,B}$ increase as $\alpha$ decreases. When $\alpha = 0$, we have $\theta_{III,A} = \theta_{III,B} \approx 2$. On the other hand, when $\alpha = 1$, we have $\theta_{III,A} = \theta_{III,B} = 1.1$.

Fig. 11(a) shows that as $\delta$ increases, $\theta_{III,A}$ increases. This phenomenon is explained in the following two cases.

Case A.1 ($\alpha = 0$). In this case, the system only has class 2 LAs. As $\delta$ increases, the $E[P]$-to-$E[N]$ ratio in (21) increases and therefore $\theta_{III,A}$ increases (see Fact 4).

Case A.2 ($\alpha > 0$). In Distribution A of Approach III, the class 1 LAs are uniformly distributed among the $M_1$ MSCs and it is likely that the expected number $E[P]$ is equal to $M_1$ when $\delta$ is large. Therefore as $\delta$ increases, the expected number $E[N]$ decreases [see Fact 3]) and $\theta_{III,A}$ increases (from (21)).

Fig. 11(b) shows the effect of $\delta$ on Distribution B of Approach III. This effect is explained in the following two cases.

Case B.1 ($\alpha = 0$). $\theta_{III,B}$ increases as $\delta$ increases. This phenomenon is similar to that in Case A.1.
TABLE III
\( \theta_{\text{III},B} \) PERFORMANCE (\( \alpha = 0.1 \))

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( E[P_1] )</th>
<th>( E[N_{m1}] )</th>
<th>( E[P_2] )</th>
<th>( E[N_{m2}] )</th>
<th>( \theta_{\text{III},B} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>5.69</td>
<td>8.56</td>
<td>1 + ( \frac{5.69}{8.56} ) = 1.36</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.86</td>
<td>0.89</td>
<td>1 + ( \frac{0.86}{0.89} ) = 1.17</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>0.09</td>
<td>0.09</td>
<td>1 + ( \frac{0.09}{0.09} ) = 1.11</td>
<td></td>
</tr>
</tbody>
</table>

Case B.2 (\( \alpha > 0 \)). \( \theta_{\text{III},B} \) increases as \( \delta \) decreases. We rewrite (21) as

\[
\theta_{\text{III},B} = 1 + \frac{E[P_1] + E[P_2]}{E[N_{m1}] + E[N_{m2}]}.
\]  

In (22), \( E[P_1] \) and \( E[P_2] \) are the expected numbers of class 1 and class 2 MSCs where the multicast members reside. \( E[N_{m1}] \) and \( E[N_{m2}] \) are the expected numbers of class 1 and class 2 LAs where the multicast members reside. Table III lists the values for \( E[P_1], E[P_2], E[N_{m1}], E[N_{m2}] \) and \( \theta_{\text{III},B} \) for \( \alpha = 0.1 \), and \( \delta = 10, 100, \) and 1000. The table indicates that \( E[P_1] \) and \( E[N_{m1}] \) are not affected by \( \delta \) for \( \delta \geq 10 \). On the other hand, both \( E[P_2] \) and \( E[N_{m2}] \) decrease as \( \delta \) increases. The net effect is that \( \theta_{\text{III},B} \) decreases as \( \delta \) increases. We note that this result is opposite to Fact 4.

VI. CONCLUSION

This paper proposed an efficient multicast mechanism for the PS-domain MMS. Our multicast approach is implemented based on the CBS architecture. The CBC is a standard UMTS network node defined in 3G 23.041 [5]. We proposed a new interface between the CBC and the SGSN to track the current locations of the multicast members. Then we described the location tracking procedures (including attach, detach and location update) of the multicast members and the multicast message delivery procedure. The implementation and execution of the multicast table are very efficient. Therefore the cost for updating this table can be ignored compared with the standard mobility management procedures. We proposed an analytic model to investigate the performance of our approach. This study indicated that in terms of MMS multicast message delivery cost, our approach outperforms the previously proposed approaches. As a final remark, the multicast table mechanism is an US and an ROC pending patents.

ACKNOWLEDGMENT

The authors would like to express our sincere thanks to Prof. Y.-B. Lin for his valuable advice on this paper. They also would like to thank the two anonymous reviewers. Their comments have improved the quality of this paper.

REFERENCES


Ai-Chun Pang was born in Hsinchu, Taiwan, R.O.C., in 1973. She received the B.S., M.S., and Ph.D. degrees in computer science and information engineering from National Chiao Tung University (NCTU), Hsinchu, in 1996, 1998, and 2002, respectively. She joined the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, as an Assistant Professor in 2002. Since August 2004, she has been an Assistant Professor with the Graduate Institute of Information Networking and Multimedia, National Taiwan University, Taipei. Her research interests include design and analysis of personal communications services network, mobile computing, voice-over-IP, and performance modeling.

Yuan-Kai Chen received the B.S.C.S.I.E. and M.S.C.S.I.E. degrees from National Chiao Tung University (NCTU), Hsinchu, in 1989 and 1991, respectively, and the Ph.D. degree in computer science and information engineering from NCTU in 2003. In 1991, he joined the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd. and was involved in the implementation of SONET/SDH multiplexer and the development of ADSL transceiver. In 1998, he was with the 3G trial team. Since then, he has been involved in the design of the radio access network, mobile packet switched data and multimedia services, and the study of mobile network evolution. His research interests include design and analysis of personal communications services (PCS) network, 3G networks, wireless Internet, mobile computing, and performance modeling.