Providing Multicast Short Message Services Over Self-Routing Mobile Cellular Backbone Network

Jen-Yi Pan, *Student Member, IEEE*, Wei-Tsong Lee, *Member, IEEE*, and Nen-Fu Huang, *Member, IEEE*

***Abstract*—The short message service (SMS), a bidirectional service for short alphanumeric (up to 160 bytes) messages, is a unique feature of GSM not found in older analog systems. The multireceivers short message traffic has increased amazingly over the years.**

In this paper, we propose a multicast SMS architecture over our backbone network. Then, we demonstrate some approaches to maintain the location information of SMS center consistent to the home location register and illustrate our scheme for mobile terminated short message transfer. Finally, we simulate and evaluate this architecture. Our simulation indicates that the proposed strong consistency approaches have optimal efficiency by adjusting parameters, and the proposed architecture efficiently provides self-routing capability and multicast functionality in our cellular backbone network. This study also provides a further insight on the issues of multicast wireless cellular backbone network and demonstrates a referable methodology to propose and analyze a multicast cellular backbone network, which can promote the technology of personal communication network.

***Index Terms*—Cellular core network, multidestination multicast, multiple-receivers short messages, shuffle-exchange network.**

I. Introduction

SINCE the first global system for mobile (GSM) communication network started operation in 1991, more than 100 countries have adopted the standard. Over 20 000 000 subscribers of GSM networks are now offered worldwide coverage, outstanding voice quality, and a variety of value-added services. These services include voice mail, call handling facilities, call line identification, and short message service (SMS).

The SMS, a bidirectional service for short alphanumeric (up to 160 bytes) messages, is a unique feature of GSM, not found in older analog systems [1]. With SMS, users are able to exchange alphanumeric messages with other users through digital cellular networks, almost anywhere in the world, within seconds of submission. These messages are transported in a store-and-forward fashion. There are two operation modes for SMS, point-to-point, and cell-broadcast, respectively. Point-to-point SMS can deliver a message from one subscriber to another and provide an acknowledgment of receipt to the sender. The other mode is cell-broadcast SMS, which disseminating messages, such as traffic conditions or news broadcasting, go to all the active handsets or mobile stations (MSs) in a cell.

Some industry observers presented several interesting phenomena in an SMS conference held in Cafe Royal, London, U.K., in September 2000 [2]. One of these phenomena is that the SMS traffic has increased amazingly these years. It is speculated that three reasons resulted in this phenomenon. The first one is SMS interoperability among different operators. The second one is prepaid mobile originated SMS on all networks. The final one is that several telecommunication companies proposed multireceivers short message value-added services, e.g., anonymous SMS chat, teletext chat, multiplayer games, and dating.

In addition to the value-added services mentioned above, let us imagine some interesting cases in end-user side. On Valentine's Day, a boy may send a short message to all his former girl friends to express his love. In another case, a team leader may inform all the members to attend an important meeting. Not only these cases are discussed in this paper, but also some innovative functions in the end-user side, e.g., Nokia FriendsTalk, are enthusiastically proposed by modern telecommunication industry. In all the above cases, someone requires to unicast the same message to each receiver.

Thus, some cellular handsets provide the function of multiple SMS sending (e.g., Nokia3310), which automatically sends the same short message to multiple receivers in one-to-one fashion. Nevertheless, the realistic cost depends on how many short messages are actually delivered, and the unicast SMS could not satisfy the requirement of multireceivers short message applications in efficient ways. This function is convenient for users, but takes more resource than that realized in multicast fashion, which is not yet implemented in today's GSM system.

Near future technology, general packet radio service (GPRS), has proposed a point-to-multipoint (PTM) services. The PTM service provides a transmission between a service requester and a receiver group. There are three different PTM services defined by the European Telecommunications Standards Institute (ETSI).

- PTM Multicast (PTM-M) [3].
- PTM Group Call (PTM-G) [4].
- IP Multicast (IP-M).

PTM-M provides unidirectional multicast transmission to a specific group within a geographical area specified by the service requester. The network has no knowledge of which subscribers are present within the geographical area, and the message reception is unreliable.

PTM-G provides a reliable (optional) transmission to a specific group within a geographical area specified by the service

Manuscript received February 27, 2002; revised May 29, 2002. This work was supported by the MOE Program for Promoting Academic Excellence of Universities under Grant 89-E-FA04-1-4.

J.-Y. Pan and N.-F. Huang are with the Department of Computer Science, National Tsing Hua University, Hsinchu, 300 Taiwan, R.O.C. (e-mail: nfhuang@cs.nthu.edu.tw).

W.-T. Lee is with the Department of Information Engineering, Feng-Chia University, Seatwen, Taichung, 407 Taiwan, R.O.C. (e-mail: wtlee@fcu.edu.tw).

Digital Object Identifier 10.1109/TVT.2002.807633

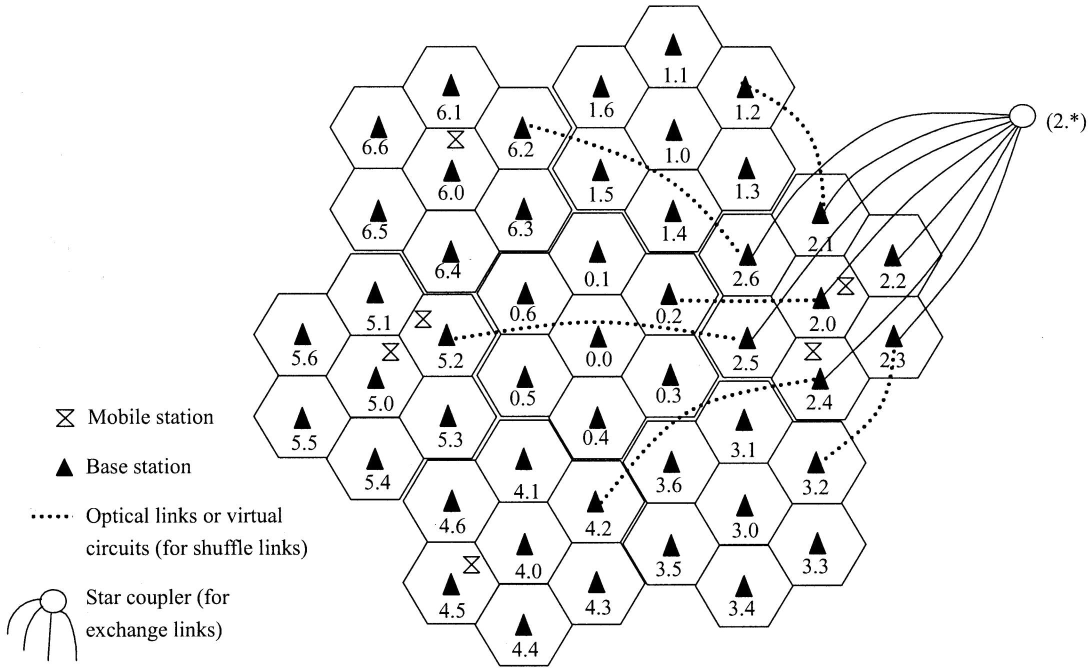

Fig. 1. Constructing the optical network from SBs.

requester, i.e., the one that requested the PTM-G service. The network has knowledge of which participants are present in the geographical area. PTM-G allows retransmission of lost messages and transfer can either be uni, bi, or multidirectional.

IP-M is a service defined as the part of the IP protocol suite and described in RFC 1920, RFC1458, RFC 1301, and RFC 1112.

Since GPRS provides such a powerful mechanism to realize multicast services, how to provide multicast ability over self-routing cellular backbone network is an important problem. Based on the above fact, we propose multicast SMS architecture over our backbone network and this architecture provides multicast functionality in an efficient way.

The rest of this paper is organized as follows. Section II highlights the architecture of our proposed backbone network. Section III presents proposed multicast SMS architecture over shuffle exchange backbone network. Section IV simulates and evaluates the performance of proposed SMS architecture and multidestination multicast backbone network. Concluding remarks are finally made in Section V.

II. Self-Routing Shuffle-Exchange Cellular Backbone Network Architecture

Our previous proposed backbone network is based on the shuffle-exchange network concept [5]. The network architecture is consisting of an optical backbone network connecting among base stations, and a wireless network connecting mobiles to base stations. We will briefly introduce our previous works in following subsections. The following Section II-A specifies the addressing method. Section II-B describes the basic unit of our backbone network. Section II-C discusses the construction and connections among basic units. Section II-D describes the routing scheme in our optical backbone network. Section II-E talks about the issue of psuedorouting nodes.

A. Addressing Method

Every base station in the proposed network has a unique identifier. Suppose there is a base station, called X. The address of base station X can be presented as

$$X = (X_{D-1}X_{D-2}\cdots X_1X_0)$$
$$\forall i,\ X_i \in \{0,\ 1,\ \ldots,\ m-1\}. \tag{1}$$

The identifier is composed of D digits in base m [e.g., if $m = 7$, $D = 2$, possible identifiers are (0.0), (0.1), (1.2), (6.3)...]. Therefore, the total number of possible identifiers is

$$N = m^D. \tag{2}$$

B. Basic Unit

Star base (SB) is the basic unit of the proposed optical backbone network. An SB contains a fixed number of base stations and a star coupler, which connects those base stations. For example, an SB in Fig. 1 is composed of the star coupler (2,*) and the base stations (2,0), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6). Every base station belongs to one and only one SB in the proposed network. Thus, all base stations are geographically partitioned into several fixed-size groups, and each group serves a geographical area called location area (LA). The addresses of base stations in the same SB are only different at the least significant digit and the other digits are the same. Assume one SB contains m base stations in the following discussion. We define this concept formally as follows.

Assume the address of base station X is $(X_{D-1}$ $X_{D-2}\cdots X_1X_0)$ and the address of base station Y is

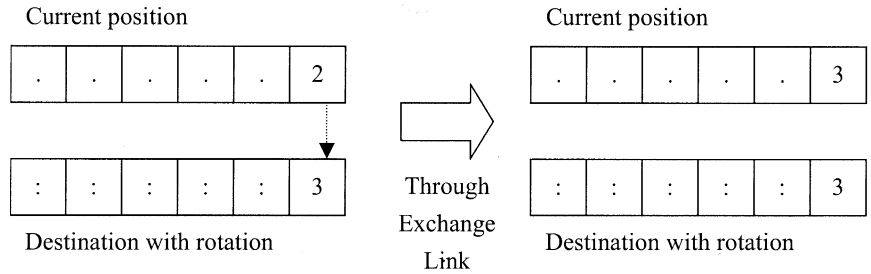

Fig. 2. Exchange action.

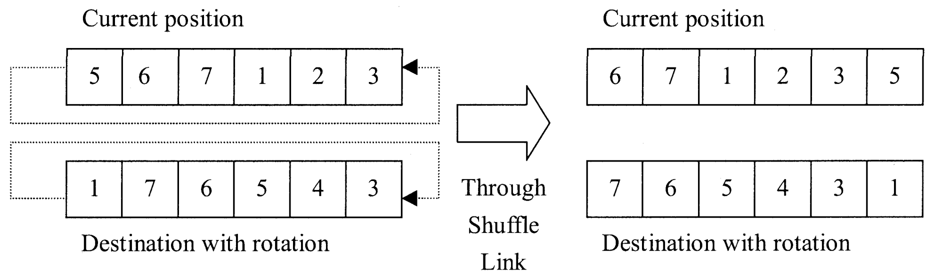

Fig. 3. Shuffle action.

$(Y_{D-1}Y_{D-2}\cdots Y_1Y_0)$. If $(X_{D-1}X_{D-2}\cdots X_1) = (Y_{D-1} Y_{D-2}\cdots Y_1)$, then X and Y are in the same SB.

C. Construction

Assume our proposed backbone network consists of N base stations and N/m SBs. Although base stations in an SB stands nearby, all the geographical locations of SBs are independent. Besides those base stations and star couplers, there are several shuffle links among base stations. The shuffle links may be physical optical links or virtual circuits. These links are unidirectional. Each shuffle link connects two base stations according the shuffle link rule as follows.

Assume the address of base station X is $(X_{D-1} X_{D-2}\cdots X_1X_0)$ and the address of base station Y is $(Y_{D-1}Y_{D-2}\cdots Y_1Y_0)$. If $(X_{D-1}X_{D-2}\cdots X_1X_0) = (Y_0Y_{D-1}Y_{D-2}\cdots Y_1)$, then there exists a shuffle link from X to Y.

Hence, these following components organize a shuffle-exchange network.

- Shuffle links $\leftarrow$ unidirectional optical links or virtual circuits.
- Exchange links $\leftarrow$ star couplers.
- Network nodes $\leftarrow$ base stations.

Fig. 1 exemplifies how to construct the optical backbone network from SBs and shows some of shuffle links when $D = 2$ and $m = 7$.

D. Routing

The routing actions consist of two types, i.e., exchange actions and shuffle actions.

Exchange actions make the least significant digits of packet current position equal to the corresponding digit of destination, i.e., the least significant digit of the rotation destination address. This action "conquers" one digit of the current packet address toward the destination. In our backbone network, one base station transmits the packet on its star coupler to the other base station according to the least significant digit of the rotation destination address. Consequently, that procedure forms the exchange action of routing, as shown in Fig. 2.

As shown in Fig. 3, shuffle actions rotate digits of current packet address to the left circularly, and form another address (that means, another place, another base station). Shuffle actions not only transmit packets, but also rotate the destination address. However, the rotation of destination address is not really to change the destination of a packet. This rotation just simplifies the operation of exchange action.

The unidirectional optical/virtual links between certain base stations realize the shuffle actions in our backbone network. These links are constructed in our network as mentioned above.

A 4×4 network is shown in Fig. 4. We will use this figure as an example in following discussion.

Table I demonstrates a routing example of proposed backbone network. The address in this example has two digits. The packet originates from the base station (0.2), and terminates at the base station (1.3). Initially, the least significant digit of the current packet address is 2 in step 0. Step 1 is an exchange action through star coupler. This step throws the packet from (0.2) to (0.3), and makes the least significant digit of the current packet address the same as the least significant digit of the destination. In step 2, the packet passes through the shuffle link, from base station (0.3) to (3.0). This shuffle link performs a shuffle action exactly. Step 3 exchanges the packet from (3.0) to (3.1) through the star base of (3.*). Finally, step 4 shuffles the packet from (3.1) to (1.3). Consequently, a routing scenario consists of several alternate exchange and shuffle actions. The packet approaches the destination step by step in each action. Generally, the routing process can be done in 2*D steps at most.

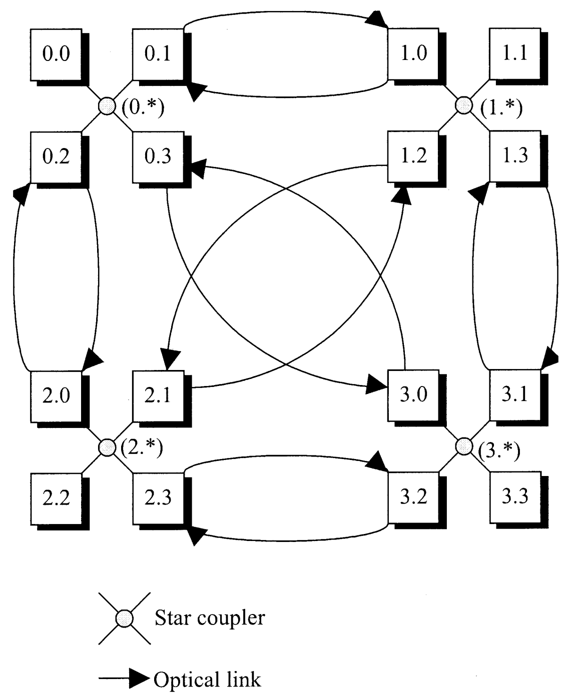

Fig. 4. A 4 × 4 proposed network.

TABLE I
A ROUTING EXAMPLE OF PROPOSED NETWORK

	Step0	Step1	Step2	Step3	Step4
Source	0.2	0.2	0.2	0.2	0.2
Current Pos	0.**2**	0.**3**(star)	3.**0**(link)	3.**1** (star)	1.**3**(link)
Dest Rotation	1.**3**	1.**3**	3.**1**	3.**1**	1.**3**
Destination	1.3	1.3	1.3	1.3	1.3

E. Pseudorouting Node

When the number of base stations is not sufficient to construct the complete shuffle-exchange networks, we can add some pseudorouting nodes to assist the routing operation of the whole network. Pseudorouting nodes have their own unique addresses, which are not source/destination addresses of any packet. These psuedonodes exist only for properly routing in such a shuffle-exchange network and they are not physical base stations extra to the system. These psuedorouting nodes attach to existing base stations; that is, some of base stations are responsible for routing of those psuedonodes, in addition to their original job. Hence, the completeness of shuffle-exchange networks can be achieved by insertion of these nodes.

For those psuedorouting nodes in an incomplete SB, they can attach to one of existing base station in the same SB, individually. If there is an SB that does not have any existing base station in it, called pseudo SB, there are two ways to overcome this situation. One is to attach the whole pseudo SB into one existing base station and the other is to map each psuedorouting nodes of the pseudo SB into an existing SB, one by one, no matter if the existing SB is complete. The former one might result in too much traffic going into one base station and the latter one might result in unnecessary traffic in the star coupler of the mapped SB.

The optimal placement and attachment of psuedorouting nodes, which assist to form a complete shuffle-exchange network, depends on the traffic pattern in the system and are beyond our discussion here.

III. PROPOSED MULTICAST SMS OVER SHUFFLE-EXCHANGE BACKBONE NETWORK

SMSs mean the services that enable the exchange of the short alphanumeric message between the short message entity (SME) and the cellular network. To provide the SMSs, the network should support the functions such as message collection, message management/administration, and message delivery. SMS center (SMSC) and the network would provide such functions.

The SMS can be generally divided into two stages—mobile originated SMS and mobile terminated SMS.

The mobile originated SMS is the service that the mobile user attempts to send a short message to some other entity or the SMSC. Since the short message cannot be transferred directly between SMEs, all the mobile originated short messages should go to the SMSC first. In this stage, we have no special implementation to reduce the corresponding cost and do not mention further more in this paper.

The second-stage mobile terminated SMS is to deliver the short message, which is saved in the SMSC to the appropriate mobile station (MS). For a mobile terminated SMS transfer, the control may be more complex than mobile originated SMS transfer, because of the change of the user location, and the SMSC requires communicating with home location register (HLR) to know the user location information and handset status. We focus on this stage and detail this issue over our backbone network as follows.

A. Traditional Unicast Solutions

A well-known traditional multireceivers solution for mobile terminated SMS is to repeat sending the same one short message as many times as the number of receivers. Fig. 5 roughly depicts this protocol flow in mobile terminated SMS over our backbone network. This is the simplest solution to complex multireceivers SMS, however, is not the most cost-effective solution because of this resources consumption as much as several unicast short messages consume.

B. Traditional Multicast Solutions

Most of existent multicast backbone technologies build a multicast route tree to keep the best packet multicast flows and the responsible network nodes for packet duplication, and assign a unique identifier to efficiently identify the tree, called a multicast address. These technologies are suitable for long duration communications, stationary communication entities, and limited communication groups, but encounter some difficulties in today's mobile network. There are several papers and specifications discussing the mobile issues in existent backbone networks [7], [8], like asynchronous transfer mode (ATM) tree migration, mobile IP tunneling, remote subscription, and so on.

However, these techniques needs to maintain states about multicast trees in network node and especially do a lot of efforts to adjust these states according to the change of mobile

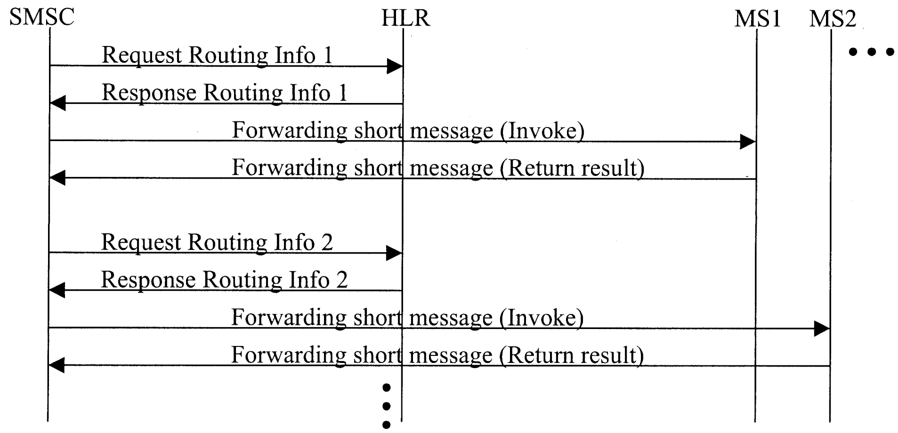

Fig. 5. Traditional unicast solution for multireceivers mobile terminated SMSs.

station status and locations. These states and corresponding maintenance cost complicate the design of protocol and network nodes, possibly result in difficulties of verification and debugging and, furthermore, may limit the scalability of the whole network system. In addition to the above facts, the next multicast tree state is generated according to not only the changes of mobile station location, but also the previous tree state. This results in difficulties in discussing the correlation between location management and tree migration. In the following subsection, we will propose an architecture, whose multicast routes are independent to the previous routes and this architecture eliminates not only the tree states in the backbone network, but also all the defects mentioned above.

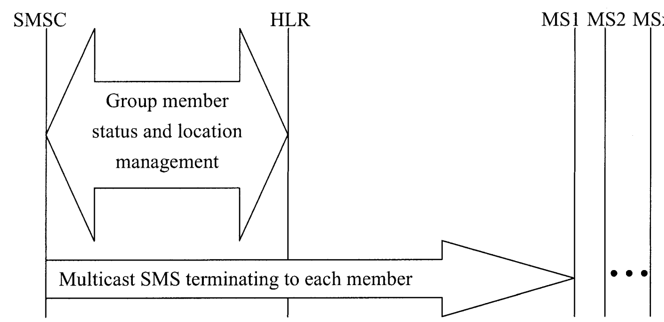

Fig. 6. Proposed multidestination multicast solution for multireceivers mobile terminated SMSs.

C. Proposed Multidestination Multicast Solutions

Based on the self-routing property of our backbone network, we proposed a multidestination multicast solution for multireceivers mobile terminated SMS. Instead of specifying a multicast address and maintaining a corresponding multicast tree in network nodes over backbone networks, we suggest filling multiple destinations in the packet destination field and, thus, eliminate the tree states in the backbone network. Because of the self-routing property in our backbone network, each destination address can be examined to determine the next hop in fixed time and this action could not be easily achieved in traditional packet backbone networks.

Moreover, the multicast process can be easily divided into two stages for modularity as Fig. 6. One is group member status and location management and the other is multicast mobile terminated short message transfer.

D. Group Member Status and Location Tracking

In the first stage, SMSC should collect all the locations of receivers of short messages and, thus, know where to send in the second stage. In fact, the problem of group member status and location tracking in SMSC is similar to the web cache consistency problem [6]. The home location register (HLR) always holds the latest status and locations of subscribers and SMSC roles a cache to hold near update information. For reducing unnecessary packet forwarding, the cache consistency of receivers' locations in SMSC must be maintained; that is, cached copies should be updated when the original information in HLR changes. We define weak consistency as the consistency model in which stale information of location might be used to multicast packets and strong consistency as the model in which, after a mobile station updates its location or status in HLR, no stale copy of the location information will be used to multicast packets.

1) Polling Every Time Approach: The polling every time (PET) approach simply sends a multiple-entity request for locations of receivers to HLR every time SMSC needs to multicast a short message to the receivers, as shown in Fig. 7. The approach has the advantage that it can be implemented easily in the existing signaling protocols and is strongly consistent for location information. The problem with this approach, however, is that the SMSC always needs to poll every time and results the considerable traffic in polling transaction, even though the location information has not changed for a long time.

2) Status Update Oriented Approach: The status update oriented (SUO) approach relies on the HLR to send out notifications when mobile terminals update location information, as shown in Fig. 8, and is a strong-consistency approach, too. It is similar to the invalidation approach of cache consistency mechanism used in the web cache. The HLR keeps track of all used location information of SMSC and, when the mobile terminals update location or status, the HLR sends out the location information notification to the SMSC. Upon receiving a notification, the SMSC updates the cached copy of location information. The

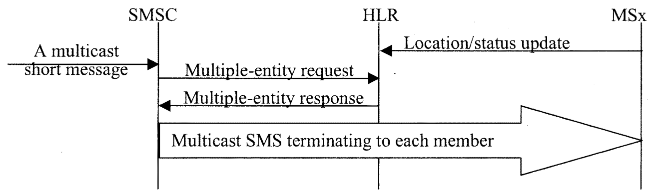

Fig. 7. Polling every time approach for group member status and location tracking.

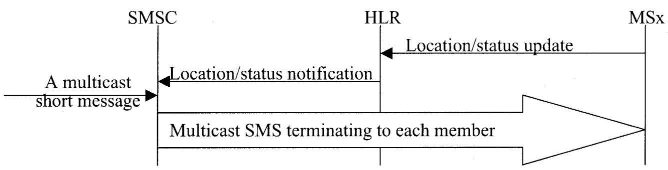

Fig. 8. Status update oriented approach for group member status and location tracking.

advantage of this approach is that it eliminates the stale copy problem with low cost when mobile terminals tend not to update information in the HLR. The disadvantage of this approach is that it may result heavy transaction traffic for cache consistency between the HLR and the SMSC when there are not many multicast short messages arrived at the SMSC.

3) Polling With Piggybacked Adaptive Time-to-Live Approach: The polling with piggybacked adaptive time-to-live (PPATTL) approach handles the problem by adjusting a time-to-live (TTL) of the location information based on observations of the information update rate of mobile stations' in the HLR. If a mobile terminal has not update information for a long time, it tends to stay unchanged. Thus, in PPATTL, the SMSC keeps location information for a period of time according to its piggybacked TTL as shown in Fig. 9. However, this approach is a weak-consistency approach, and the simulation results in later section will show the stale percentage by adjusting TTL. The stale information will result in unbounded hops for forwarding packets to the current base station where the receiver mobile station stays. There is a tradeoff between maintaining cache consistency and unnecessary packets forwarding over the backbone network.

4) Polling With Rental Update Notification Approach: According to the previous discussion about polling and notification, we have two different approaches to combine polling and notification to achieve better performance than a single one. The first proposed approach, called the polling with rental update notification (PRN) approach, is modified from the PET and is also a strong-consistency approach. When the SMSC polls the HLR for multiple queries, it also estimates rental duration and tells the HLR at the same time. If there are any mobile terminals updating information in this duration, the HLR will notify the SMSC. Fig. 10 shows this protocol flow. For very short rental duration, this approach degrades to the PET and, for very long rental duration, this approach becomes the SUO.

5) Status Update Oriented With Suppressing Notification Duration Approach: The second approach, called status update oriented with suppressing notification duration (SUOSN)

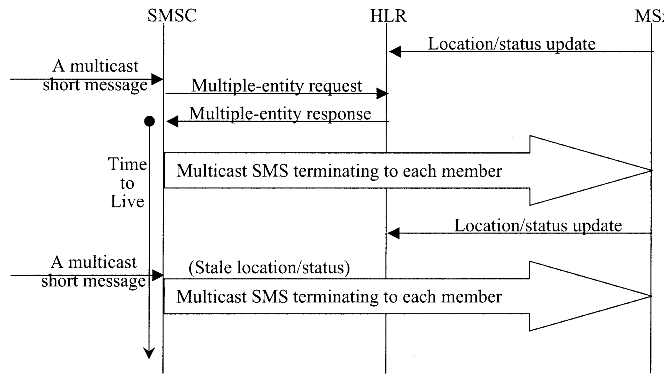

Fig. 9. Polling with piggybacked adaptive TTL approach for group member status and location tracking.

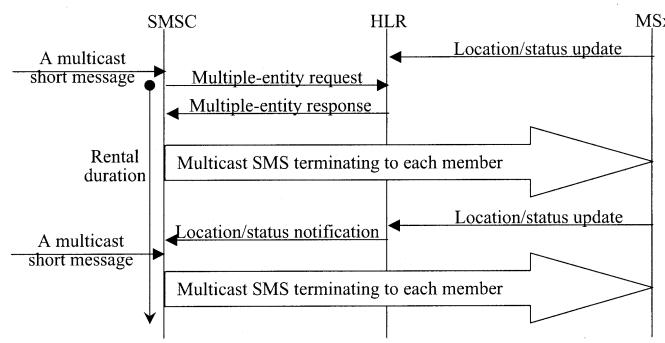

Fig. 10. Polling with rental update notification approach for group member status and location tracking.

approach, is based on SUO and is also a strong-consistency approach. Comparing to PRN, it has a different viewpoint to combine polling and notification. In addition to the operation of SUO, the SMSC conjectures a suppressing notification duration in which no short message might arrive and tells the HLR not to send any notifications in this duration for suppress the cache consistency traffic, as shown in Fig. 11. However, if any short message arrives at the SMSC and the SMSC requires to multicast this message, the SMSC needs to poll the HLR for the latest location information. For very short masking duration, this approach degrades to the SUO and, for very long rental duration, this approach becomes the PET.

E. Multidestination Multicast

The proposed multidestination multicast scheme routes multicast packets based on multiple destination fields rather than a single multicast address in the only one destination field. The multicast function of proposed backbone network exploits the broadcast characteristics of star couplers to duplicate the packets and then every base station checks the multiple destinations in a packet to determine whether to process the packet. We demonstrate a multicast example in Fig. 12.

Every rectangle means a packet in certain base station, whose address labeled on the top of the rectangle. Each arrow specifies the flow of packets through the shuffle (marked as "link")/exchange (marked as "star") links. The multiple destinations of one packet are noted within its rectangle.

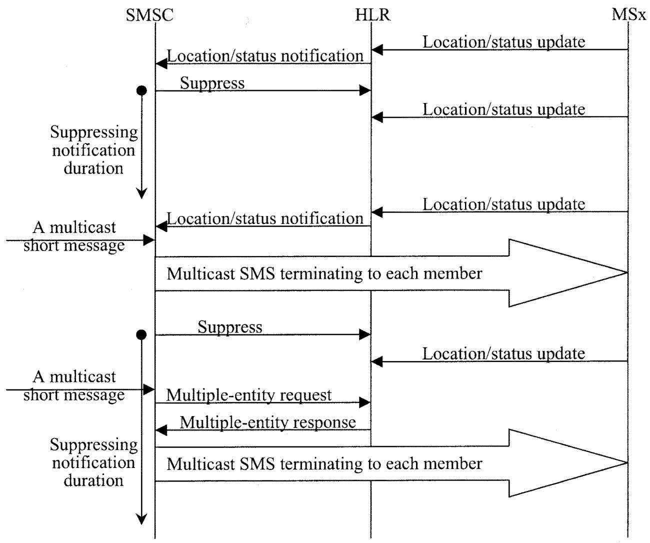

Fig. 11. Status update oriented with suppressing notification duration approach for group member status and location tracking.

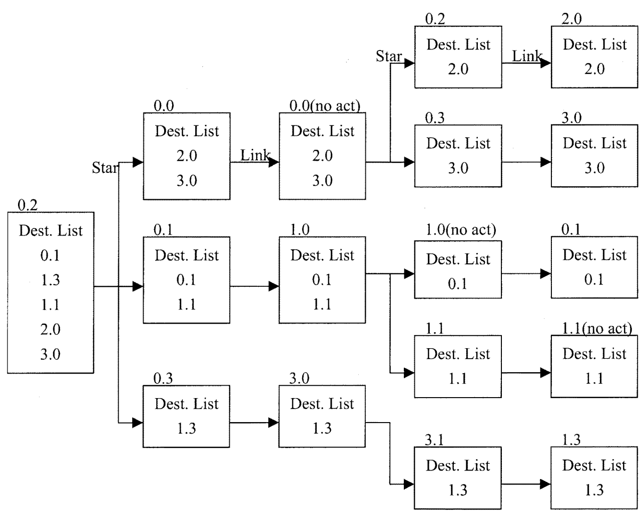

Fig. 12. A multidestination multicast routing example.

In the example, there is a multicast packet from (0.2), then going to (0.1), (1.3), (1.1), (2.0), and (3.0). Notably, the packet flow from (0.2) to (1.3), on the bottom side of Fig. 12, is the special case mentioned in Table I. In general, the multicast process can be done in 2*D steps (that is a limited hop count).

Because of this pre-planned routing scheme, it takes constant time to determine the next hop for a packet. Moreover, stateless multicast in this scheme does not need to pre-establish a multicast tree and maintain the tree in the core network. Hence, every multicast route for a packet is independent to previous packet routing and this architecture eliminates not only the tree states in the backbone network, but also simplifies the design of protocol and network nodes, which possibly results in difficulties of verification and debugging and, furthermore, may limit the scalability of the whole network system.

IV. Multicast Cost Metrics and Simulations

A. Evaluation Metrics

1) Group Member Status and Location Tracking: For evaluating group member status and location tracking, we define two metrics—the Effort ratio and the Efficient.

For the Effort ratio, we define

$$\mathrm{EffortRatio}_{\mathrm{Approach\,X}} = \frac{\text{The number of transactions to keep consistency in approach X}}{\text{The number of transactions to keep consistency in PET}} \quad (3)$$

where a transaction is defined as either a polling action (including request and response) or a notification (including informing and acknowledgment) between the HLR and the SMSC.

The Effort ratio is used to evaluate the effort of keeping consistency comparing to the basic approach, PET.

Before defining the Efficient, we define the waste, which used in the definition of the Efficient. A waste is defined as a redundant transaction, which can be reduced without affecting original operations.

Thus, for polling transactions, a waste is defined as a polling that does not change the cached location information in the SMSC. For example, a polling transaction happened at time $T1$ and then HLR update location information at time $T2$. If there are no location information changing between $T1$ and $T2$, then every polling transaction happened between $T1$ and $T2$ is a waste.

For notification transaction, a waste is defined as a notification whose location information is not used by a multicast short message delivery. For example, a notification $N1$ from the HLR to the SMSC happened at time $T1$ and then a notification $N2$ from the HLR to the SMSC happened at time $T2$. If there are no multireceivers, short messages required location information in $N1$ to send to multiple receivers between $T1$ and $T2$, then the notification $N1$ is a waste.

For the Efficient, we define

$$\mathrm{Efficient}_{\mathrm{Approach\,X}} = \frac{\text{The number of transactions} - \text{The number of waste}}{\text{The number of transactions}}. \quad (4)$$

In addition to the above two metrics, we require the Stale ratio, to evaluate a weak consistency approach. As Fig. 9 shows, if the cached location information in the SMSC is inconsistent to the information in the HLR, then this cached location information is stale and this information may cause unnecessary packet forwarding and unlimited hop counts in the backbone network.

For the Stale ratio, we define

$$\mathrm{StaleRatio} = \frac{\text{The number of multicast short messages terminating with stale information}}{\text{The number of multicast short messages}}. \quad (5)$$

2) Multidestination Multicasting: Before defining the metric to evaluate multidestination multicasting, we require some relative definitions, which used in the definition of the metric.

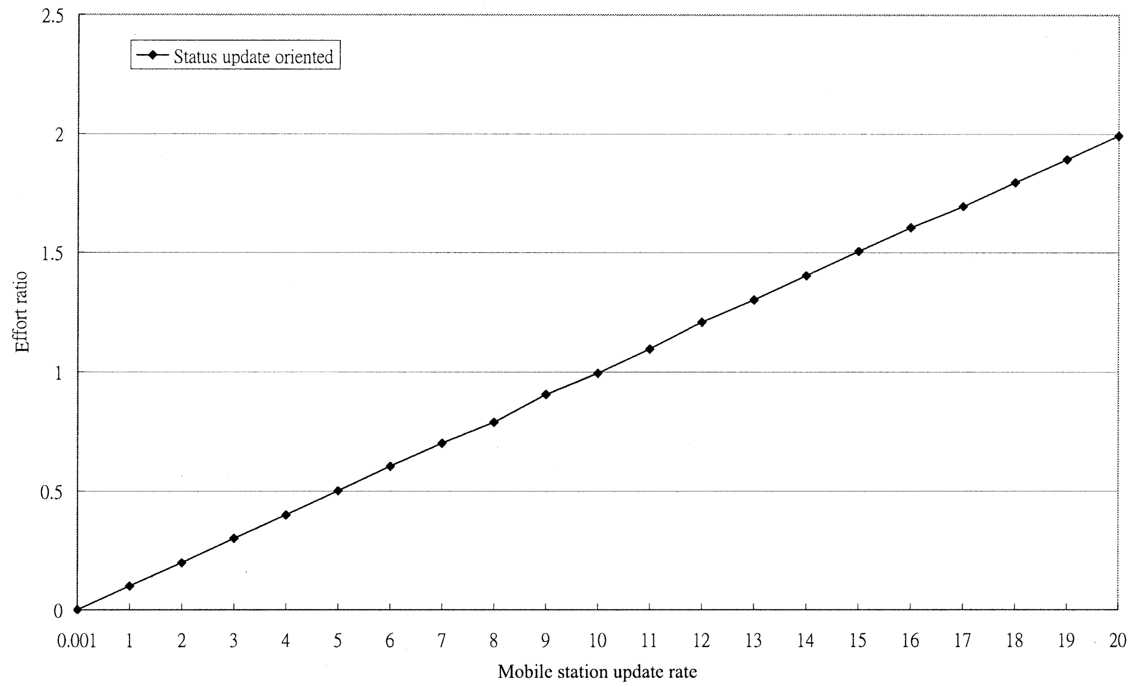

Fig. 13. Effort ratio of status update oriented approach.

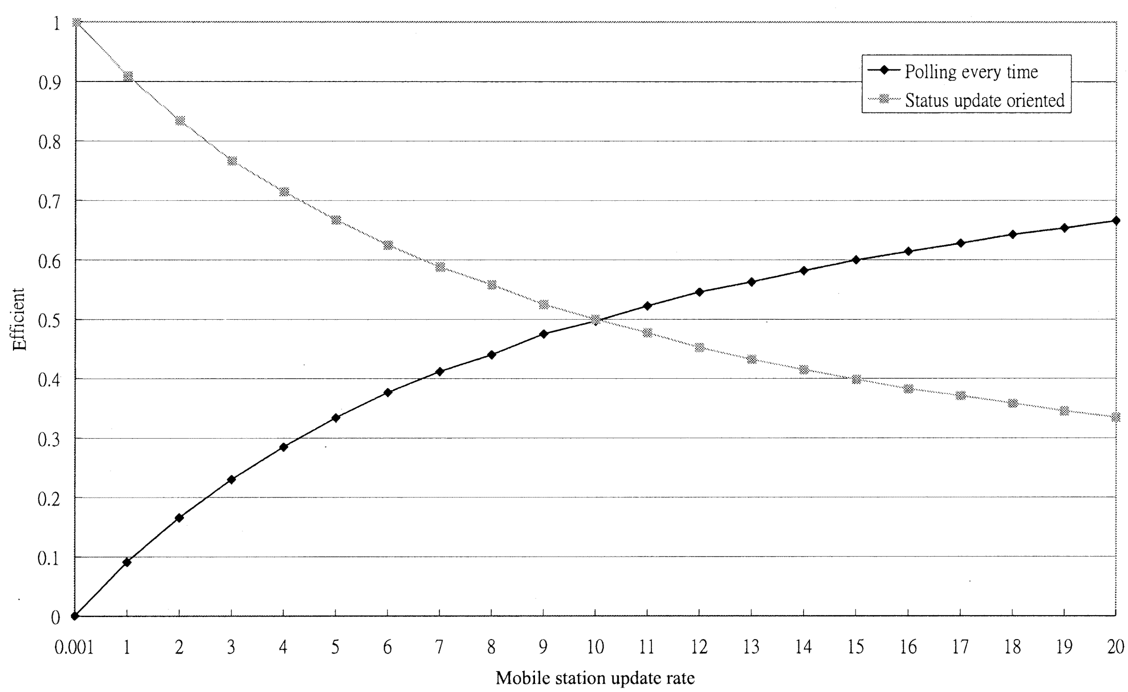

Fig. 14. Efficient of polling every time approach and status update oriented approach.

- Hop, noted by (X, Y), is defined as a packet forwarding action, e.g., shuffle action or exchange action, from the address X to the address Y.
- $Path(X, Y) \equiv$ {all the hops passed from X to Y }.
- $C_Q(X, Y, S, T)$ is the number of packets through the hop (X, Y) in approach Q, when sending a multidestination short message from the source address S to multiple destination addresses contained in the set T.

Furthermore

$$C_{\text{unicast}}(X, Y, S, T) = |\{t| \text{ for all destination } t \text{ in } T, (X, Y) \in \text{ Path } (S, t)\}| \quad (6)$$

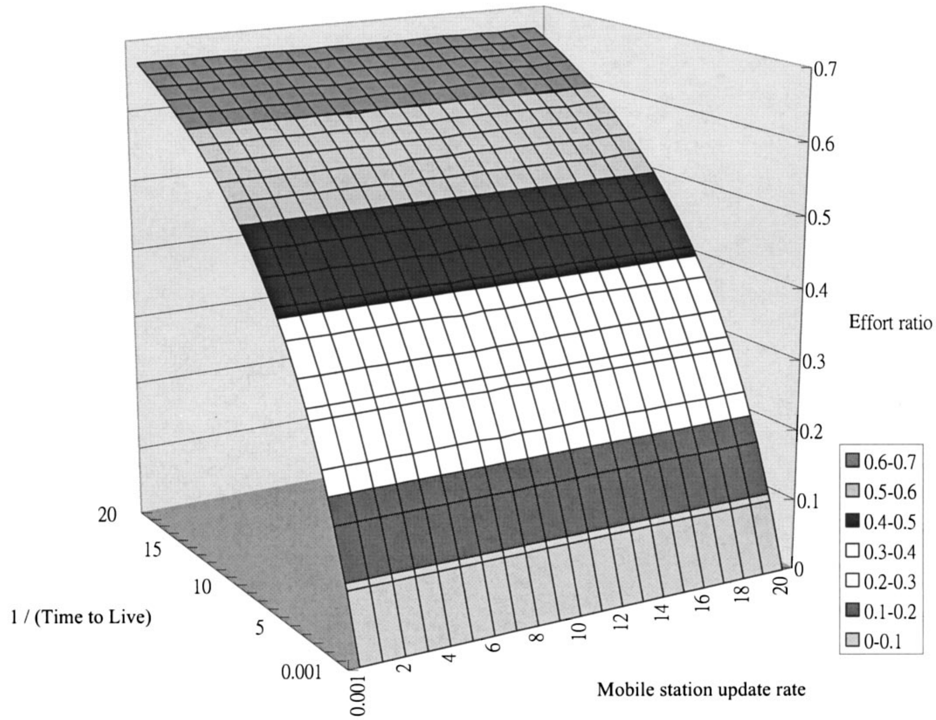

Fig. 15. Effort ratio of polling with piggybacked adaptive TTL approach.

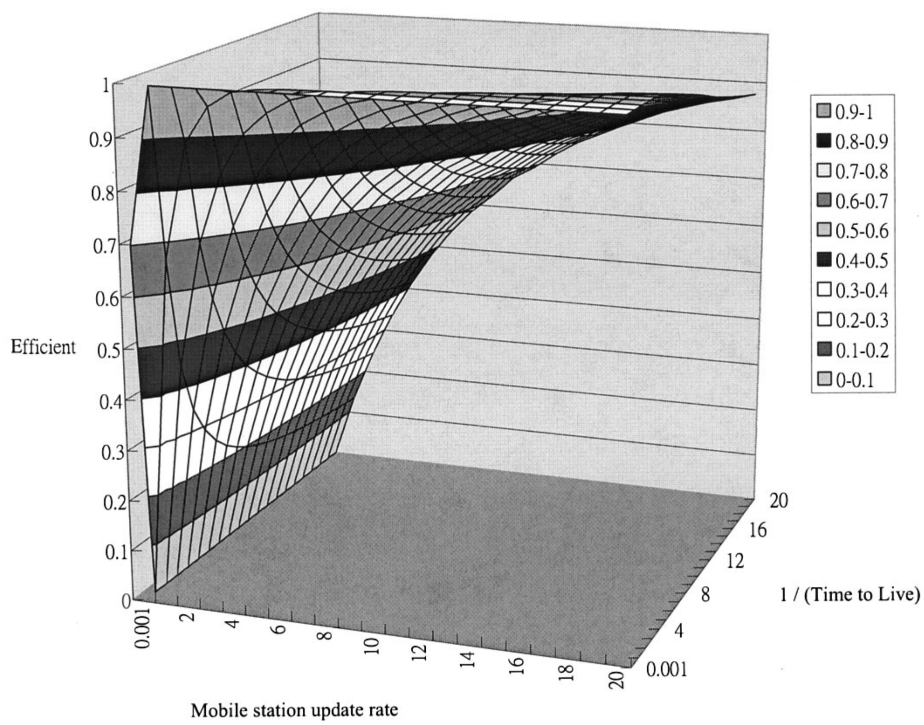

Fig. 16. Efficient of polling with piggybacked adaptive TTL approach.

and

$$C_{\text{multicast}}(X, Y, S, T) = \begin{cases} 1, & \text{if } (X, Y) \in \bigcup_{t \in T} Path(S, t) \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Finally, we define the Traffic cost to evaluate unicasting for each destinations and multicasting approaches

$$\text{TrafficCost}_{\text{Approach Q}}(S, T) = \sum_{\text{for all possible hop } (X, Y)} C_{\text{Approach Q}}(X, Y, S, T). \quad (8)$$

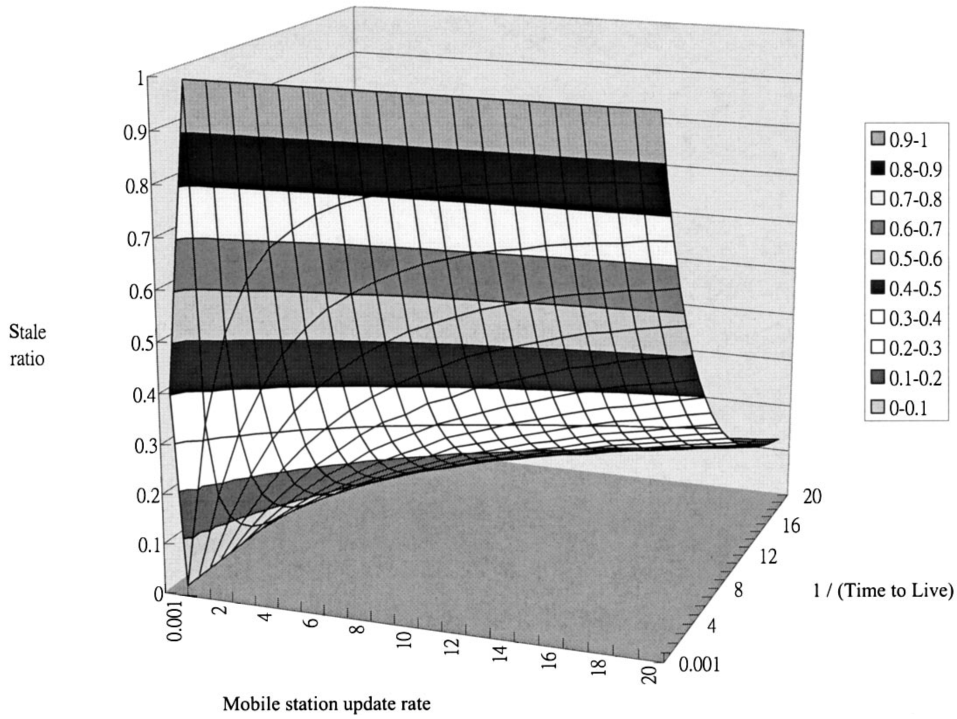

Fig. 17. Stale ratio of polling with piggybacked adaptive TTL approach.

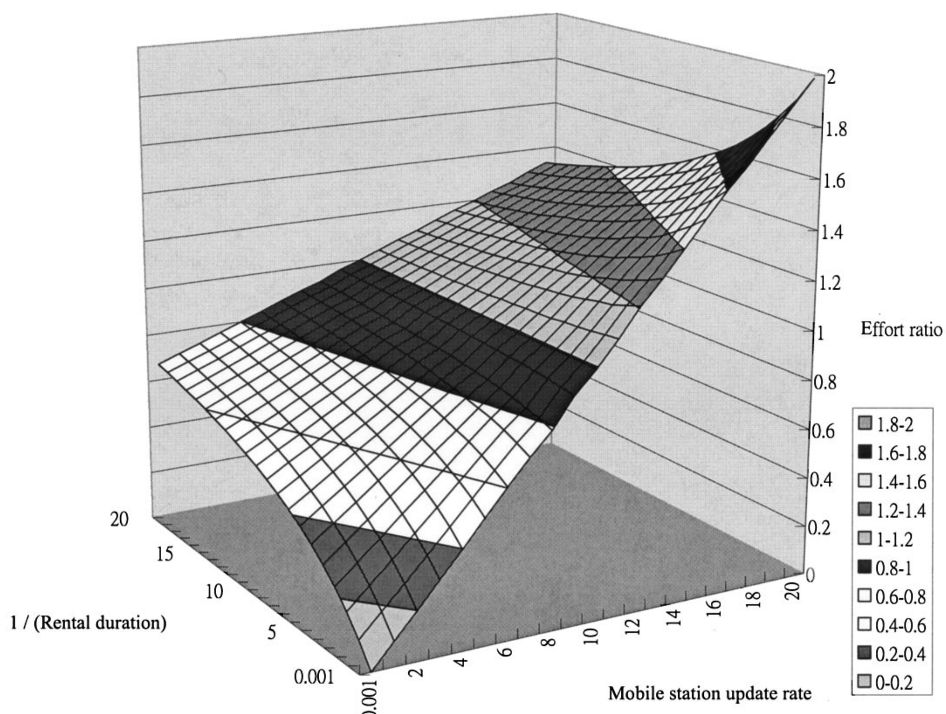

Fig. 18. Effort ratio of polling with rental update notification approach.

B. *Simulation Results for Several Group Member Status and Location Tracking Approaches*

The models of our simulation are listed as follows.

- Each simulation proceeds for 10 000 time units to reach stable states.
- The arrival process of multireceivers short messages is a Poisson arrival process and the average arrival rate is λ_{sm}, supposed to be ten per time units in the following simulations.
- The arrival process of mobile station update information event is a Poisson arrival process, and the average arrival rate is λ_{ms}.
- The time-to-live is a deterministic time called T_{tl}.
- The rental duration is a deterministic time called T_r.

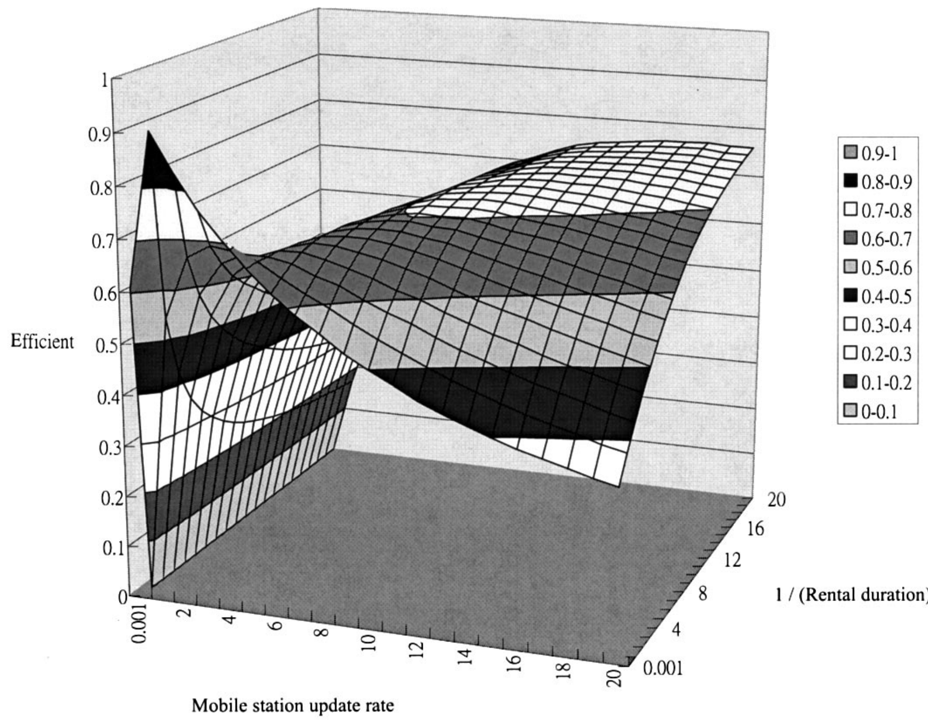

Fig. 19. Efficient of polling with rental update notification approach.

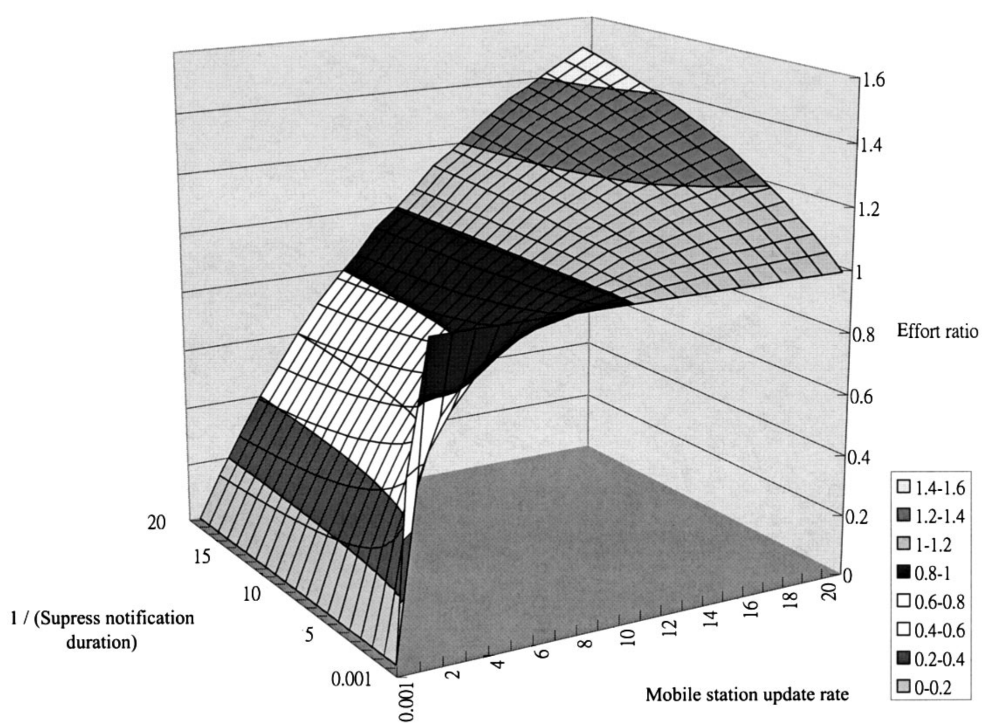

Fig. 20. Effort ratio of status update oriented with suppressing notification duration approach.

- The suppressing notification duration is a deterministic time called T_s.

Fig. 13 shows the Effort ratio of the SUO and the Effort ratio is linearly proportional to the rate of mobile terminal updating information. When the update rate λ_{ms} is greater than λ_{sm}, the Effort ratio is greater than one and this fact represents the SUO takes more efforts to maintain cache consistency than the PET does.

Fig. 14 shows when λ_{ms} is low, the SUO almost has no waste in maintenance on cache consistency and the PET almost has

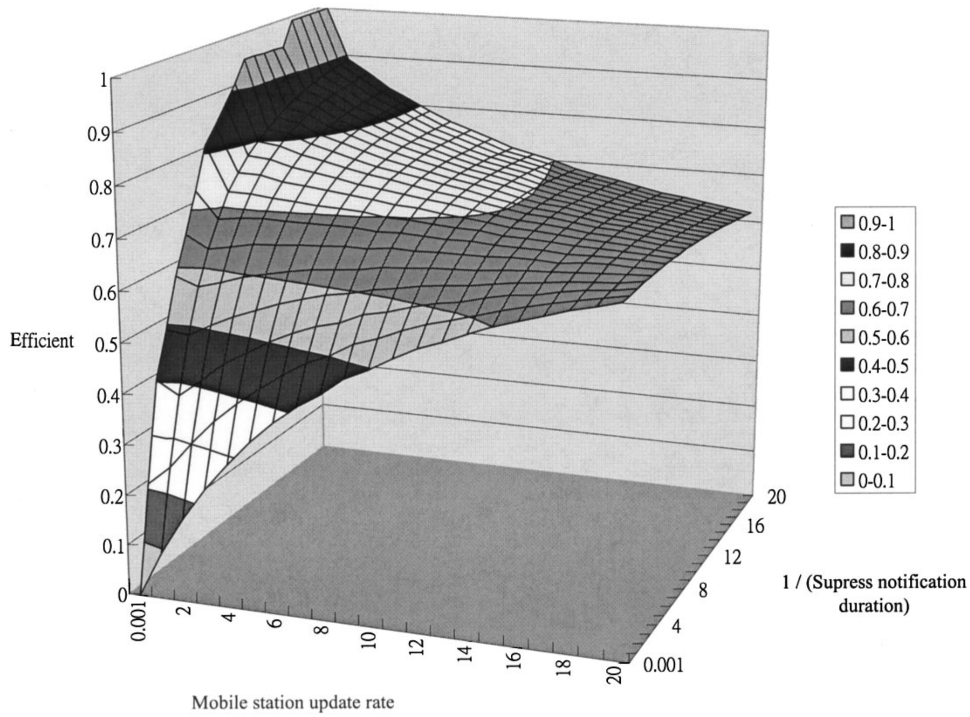

Fig. 21. Efficient of status update oriented with suppressing notification duration approach.

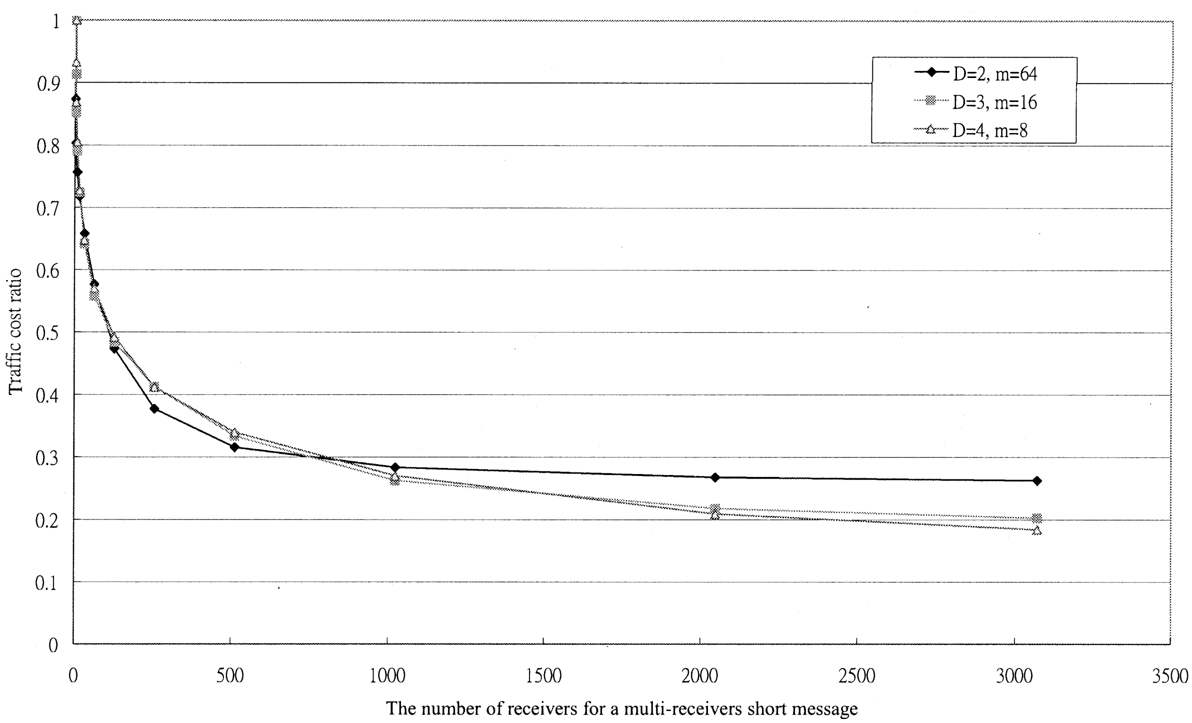

Fig. 22. The average multicast-to-unicast traffic cost ratio versus the number of receivers for a multireceivers short message for a 4096-node network for different values of SB size—m—and number of digits—D.

no necessary polling transactions. However, while λ_{ms} growing up, these two lines meet each other and represent the PET is better than the SUO when λ_{ms} is higher than λ_{sm}.

Fig. 15 depicts the Effort ratio of the PPATTL, a weak consistency approach. In this approach, the Effort ratio is independent to λ_{ms}, and only related to $1/T_{tl}$. The fact means the consistency effort of this approach not directly depends on λ_{ms}, unless adjusting $1/T_{tl}$.

Fig. 16 depicts the Efficient of the PPATTL, which relates to λ_{ms} and $1/T_{tl}$ at the same time. While λ_{ms} rising and $1/T_{tl}$ decreasing, this approach tends to have fewer waste transactions.

In contrast to the efficient, Fig. 17 illustrates the Stale ratio also increasing when λ_{ms} rising and $1/T_{tl}$ decreasing, which means the more stale information may be used to deliver short messages and results in unnecessary packet forwarding and unlimited hop counts.

For the PRN, Fig. 18 reveals the Effort ratio comparing to the PET. It demonstrates this approach degrades to the PET when $1/T_r$ is very large and becomes the SUO when $1/T_r$ is very small.

Comparing to the PPATTL, the PRN takes variable efforts depending on not only $1/T_r$, but also λ_{ms}.

Fig. 19 displays the efficient of PRN and shows that the optimal Efficient can be achieved by adjusting $1/T_r$ according to λ_{ms}.

For the SUOSN, Fig. 20 depicts the relation of the effort ratio λ_{ms} and $1/T_s$. Similar to the PRN, this approach also degrades to the SUO when $1/T_s$ is very large and becomes the PET when $1/T_s$ is very small.

Fig. 21 shows the Efficient of the SUOSN, and also reveals that the optimal Efficient can be achieved by adjusting $1/T_s$ according to λ_{ms}.

C. Simulation Result for Multicast Terminating

The models of our simulation are listed as follows:

- Each simulation proceeds for 1000 times to generate enough random samples.
- The total number of base stations in the backbone network is 4096.
- The number of digits in an address is D.
- The number of base stations in an SB is m.
- The number of receivers for a multireceivers short message is a deterministic number in each simulation called Num_R.
- The location of SMSC is uniformly distributed over all the address space.
- The destination of receivers is uniformly distributed over all the address space.

Fig. 22 shows the average multicast-to-unicast ratio of Traffic cost for the same multireceivers short message input. It illustrates the more the number of receivers of a short message, the cheaper the Traffic cost than unicast case hyperbolically. Another interesting fact is, in the case of small number of short message receivers, the shorter digits of address take smaller Traffic ratio than the longer digits do, but take larger Traffic ratio in the case of large number of receivers.

V. Conclusion

Based on the above presentation, the proposed multicast architecture provides a multicast SMS over self-routing mobile cellular backbone network.

This architecture does not need to maintain the states of multicast trees in network nodes, and simplifies the design of protocol and network nodes, possibly also simplifies the verification and debugging of the backbone network, and furthermore, may be scaled up easily when the whole network system spreads out and the number of subscribers explodes.

This architecture divides this service into two stages and conquers them separately. Moreover, the simulation result shows the efficiency of different approaches in the stage of group member status, location tracking, and the traffic cost in the stage of multidestination multicast terminating.

After the study, we proposed a multicast architecture, which provides a multicast SMS over self-routing mobile cellular backbone networks. We also proposed several approaches for group member status and location tracking and these approaches adapt to the arrivals of mobile terminal update information by adjusting parameters.

Furthermore, we demonstrate self-routing backbone network philosophy, which can prompt the deployment of future generation of wireless personal communication networks.

Acknowledgment

The authors would like to thank the anonymous reviewers of the paper for their very helpful and constructive criticism. Their suggestions have enabled them to improve the quality of this work.

References

[1] C. Peersman, S. Cvetkovic, P. Griffiths, and H. Spear, "The global system for mobile communications short message service," *IEEE Trans. Personal Commun.*, vol. 7, pp. 15–23, June 2000.

[2] (2000, Sept.) SMS Presentations. GSM World, London, U.K.. [Online]. Available: http://www.gsmworld.com/technology/sms/presentations.shtml

[3] General Packet Radio Service (GPRS). Point to multipoint multicast service description; Stage 2. GSM 03.61. [Online]. Available: http://www.etsi.org

[4] General Packet Radio Service (GPRS). Point-to-multipoint group call; Stage 2. GSM 03.62. [Online]. Available: http://www.etsi.org

[5] J.-Y. Pan, W.-T. Lee, and N.-F. Huang, "A novel routing algorithm for WDM-based micro-cellular wireless system," *Comput. Commun., Elsevier Sci.—Special Issue: Current Issues in High Speed Networks*, vol. 24, no. 17, pp. 1714–1725, Nov. 2001.

[6] P. Cao and C. J. Liu, "Maintaining strong cache consistency in the World Wide Web," *IEEE Trans. Comput.*, vol. 47, pp. 445–457, Apr. 1998.

[7] N.-F. Huang, K.-S. Chen, H.-C. Chao, and J.-Y. Pan, "A distributed multicast tree migration scheme for ATM based personal communication networks," in *Proc. IEEE GLOBECOM'96 Conf.*, London, U.K., Nov. 1996, pp. 1054–1058.

[8] C. R. Lin and K. M. Wang, "Mobile multicast support in IP networks," in *Proc. IEEE INFOCOM'00 Conf.*, Tel-Aviv, Israel, Mar. 2000, pp. 1664–1672.

Jen-Yi Pan (S'96) received the B.S. degree in computer science from the National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., in 1995. He is currently working toward the Ph.D. degree in the same department.

His research interests include the wavelength division multiplexing (WDM) networks, backbone networks of wireless communication, and wireless Internet.

Wei-Tsong Lee (S'90–M'95) received the B.E., M.E., and Ph.D. degrees in electrical engineering from National Cheng-Kung University, Taiwan, R.O.C., in 1984, 1986, and 1995, respectively.

He is currently an Associate Professor with the Department of Computer Science and Information Engineering, Feng-Chia University, Taichung, Taiwan. His research interests include performance and architecture of high-speed and broadband network, multimedia applications on Internet, microprocessor systems, and interface design.

Nen-Fu Huang (S'85–M'86) received the B.S.E.E. degree from National Cheng Kung University, Taiwan, R.O.C., in 1981, and the M.S. and Ph.D. degrees in computer science from National Tsing Hua University, Taiwan, R.O.C., in 1983 and 1986, respectively.

From 1986 to 1994, he was an Associate Professor with the Department of Computer Science, National Tsing Hua University, Taiwan, R.O.C. From 1997 to 2000, he was the Chairman of Department of Computer Science, National Tsing Hua University, Taiwan, R.O.C. Since 1994, he has been a Professor of Department of Computer Science, National Tsing Hua University, Taiwan, R.O.C. He is the Guest Editor of the Special Issue on Bandwidth Management on High-Speed Networks for Computer Communications. Since 1997, he has been the Editor of the *Journal of Information Science and Engineering*. His current research interests include wavelength division multiplexing (WDM) networks, mobile networks, next generation Internet protocols, network management, and network security.

Dr. Huang received the Outstanding Teaching Award from the National Tsing Hua University in 1993 and 1998 and the Outstanding University/Industrial Cooperating Award from the Ministry of Education, R.O.C., in 1998. He also served as a Program Chair for the 14th International Conference on Information Networks (ICOIN-14) on 2000.