Wireless Netw (2013) 19:1563–1575
DOI 10.1007/s11276-013-0553-5

Mitigating location management traffic via aggregation in multi-hop cellular networks

Rongsheng Huang · Hongxia Zhao · Yuguang Fang

Published online: 21 February 2013

Abstract In high mobility areas such as shopping malls and transportation stations, mobile users tend to move in and out in a bursty fashion, and hence location update (LU) and paging signaling cause substantial traffic burden to the cellular networks, leading to signaling congestion. This poses a great challenge to the system design for wireless cellular systems. Traditional cellular systems cannot cope with this situation very well. However, with the ad hoc operational mode recently introduced into the cellular systems, the added multi-hop relaying via mobile devices provides a new way to mitigate location management traffic. In this paper, based on this new architecture, we propose a novel scheme, called aggregative location management, which aggregates multiple location updates into group location updates. For the scenarios of high capacity transit (HCT) systems, we use this grouping scheme to alleviate the signaling traffic when mobile users in an HCT move into new location areas. For other scenarios with mass arrivals of LU requests, we develop a generic aggregative location management scheme in which LU requests can be first aggregated by designated mobile devices and then are periodically sent to the location registers. Performance evaluation is carried out and shows their significant effectiveness.

Keywords Multi-hop cellular networks · Ad hoc networks · Mobility management · Location management

Present Address:
R. Huang
Olympus Communication Technology of America, Inc.,
San Diego, CA, USA
e-mail: russell.huang@olympus.com

H. Zhao
Huawei Technologies Co., Ltd, Shenzhen, China

Y. Fang (✉)
Department of Electrical and Computer Engineering,
University of Florida, Gainesville, FL, USA
e-mail: fang@ece.ufl.edu

1 Introduction

In wireless cellular networks, timely and accurate location information of each mobile user is critical for normal system operations in terms of service delivery with desired quality of service (QoS). Location management (LM) is the function to manage the user location information. Here, each mobile user is required to update its location information to the system whenever it enters a new location area. Meanwhile, paging is the complementary process to track a mobile user. As it is observed, reducing the size of location area can indeed lower the paging cost, but at the price of increasing signaling burden for location updates (LUs). On the other hand, increasing larger location area does curtail the signaling traffic for LUs, but at higher paging signaling cost. Clearly, there is a tradeoff between LU traffic and paging traffic. Moreover, although location update and paging both consume scarce wireless resource, their costs may not be equivalent. Paging costs the paging channel resource and transmission power from base stations while LUs cost the access capability of the system and power consumption from mobile users. If LU requests overwhelm the RACH (random access channel), new calls and handoff calls may be blocked due to the blocking to the connection requests. Therefore, careful design is needed to balance the two kinds of signaling traffic while minimizing the overall cost. This problem has been extensively investigated for the traditional cellular networks in the past [1–4].

In many modern cities, population becomes denser and denser and so do the wireless devices and services. People tend to move in flocks due to their similar working and living patterns (mobility patterns). In addition, occasional big events or holiday shopping tend to attract large crowds to certain locations. To get to a location of attraction, people may take mass public transportation, especially in the metropolitan areas, because of various factors such as traffic concerns and parking. Thus, the moving mass can easily create large fluctuation on the population at such a location. Consequently, people living in a chain of crowded environments, such as office buildings, stores, subway stations, stadiums, theaters, etc., may experience frequent call connection blocking and call dropping due to mass location update traffic when switching locations areas, causing the control signaling congestion. For the cells along the location area boundaries, we observe that the moving mass of mobile terminals (MTs) launch signaling processes, typically the LU traffic, more or less at the same time and tends to congest the random access channels (RACHs), and thus new/handoff calls are constantly found to be rejected due to the RACH congestions although other channels, such as Dedicated control channels (DCCHs), has not been fully utilized. This problem becomes more serious for the emerging cellular systems because the location area sizes and cell sizes tend to be getting smaller. Researchers have already noticed this problem and proposed some solutions to address it. Wang et al. [5] propose a group LU scheme for cellular systems in which a group location update is performed by a leader mobile terminal instead of individual update by every normal mobile terminal. Zhang et al. [6] propose a buffering scheme to relieve the congestion caused by batch arrivals of LU requests. Our previous work in [7] also proposed a feasible solution to this problem for high speed moving group scenarios. However, although these schemes do curtail the location update traffic to some extent, they cannot solve this problem effectively. For the indispensable LU requests, the most effective way to mitigate the potential wireless congestion is to aggregate these messages and reduce the channel contentions over the RACH. Unfortunately, observing the architecture of the LM signaling in traditional cellular networks, we can see that the rigid one-hop wireless structure forbids the possible aggregation of wireless signaling traffic.

Fortunately, as cellular systems have evolved themselves into more integrated systems and have absorbed ad hoc mode into their normal operation, more effective solutions can be designed to solve this problem. The idea of integrating ad hoc mode into cellular systems comes from the motivation that cellular systems should become an all-IP platform so that all the service of emerging wireless networks (e.g., WiFi, WiMAX) can be incorporated together without much difficulty [8]. The integration can be either based on ad hoc systems, such as CAMA in [9], or based on cellular systems. Since our focus in this paper is on the LM issues, we consider only the cellular-based integration. The cellular-based integrated systems, usually called multi-hop cellular networks (MCNs) [10–13], are expected to bring the advantage of ad hoc networks into cellular systems, such as flexibility, robustness, self-organization, and low cost.

In MCNs, each MT can be others' relay node and can forward signaling and data packets for each other. Since the ad hoc links are allowed to use extra out-band spectrum different from the cellular spectrum, the signaling burden on cellular systems can be shifted to ad hoc spectrum and thus we can expect the relief on cellular signaling congestion. In this paper, we propose a grouping scheme to reduce the signaling traffic due to location management for the scenarios in high-capacity transit (HCT) systems. With our scheme, the LU requests from the moving-in-together MTs can be grouped together by special MTs or access points in the vehicle so that the congestion on the RACH can be greatly relieved due to the reduced RACH signaling. For other scenarios where MTs do not follow the same movement pattern but arrive at the same spot around the same time, control signaling congestion is also easily formed and the grouping approach cannot help much. We propose a generic aggregative scheme for these scenarios, which also takes advantage of the ad hoc links in MCNs.

Although aggregative approaches are not new in the literature, we are the first to propose to mitigate the LM traffic in MCNs by using aggregation in a way of incorporating the new feature, ad hoc mode, into the scheme design. In the proposals, we have given out performance analysis with respect to important parameters as well as the detailed protocol design. Finally, to evaluate our proposed scheme, we formulate the cost function in terms of LU and paging and determine the optimal location area size in MCNs. As a final remark, MCNs heavily rely on the cooperation among MTs in order to take advantage of ad hoc mode. This can be done by either installing operator-owned light-weight relaying stations or by designing operator-run incentive schemes to stimulate MTs to help the cellular systems relay packets. This issue is out of the scope of this paper and will be addressed elsewhere.

The rest of the paper is organized as follows. Section 2 surveys the related works. Section 3 describes the underlying system model. We present our grouping scheme and generic aggregation scheme in Sects. 4 and 5, respectively. Section 6 shows the LM optimization formulation and determine the optimal location area size and Sect. 7 provides the performance evaluation. Conclusions are drawn in the last section.

2 Related works

There are several approaches aiming at reducing the LM signaling traffic. During network planning, the entire coverage is usually divided into location areas (LAs) consisting of multiple cells for location management. When an MT enters a new LA, it launches LU procedure. Paging is carried out in all cells within the current visited LA when an incoming call arrives and the terminated MT is in the IDLE state. For traditional cellular systems, Ma and Fang [14] provided a pointer forwarding scheme to reduce the location update traffic and alleviate the signaling burden on the HLR (home location register). Dynamic schemes can report the location updates more flexibly in various ways based on the information used for the design. Distance-based LM scheme requires the system to maintain the record on the location area an MT visited when the last LU was executed. When the moving distance exceeds a certain threshold, a new LU is triggered. This approach has been studied in [15]. Movement-based LM scheme is another dynamic LU approach [16], in which the LU is triggered when the movement of an MT reaches a threshold. Besides these, Misra et al. proposed an information theory-based LM scheme for vertical roaming users in [17]. For paging, Zang et al. [18] proposes to delay the paging process in a certain way to reduce the paging cost.

However, for the potential LU congestion problem, there are not much work done in cellular systems. Wang et al. propose a group location update scheme for cellular systems in [5]. By letting the group head report the location of its members, signaling traffic can be greatly reduced. However, even though we can utilize grouping method to curtail the location update traffic, we still consume wireless cellular resource to manage groups, which makes this method less attractive. In addition, the base station (BS) is not a good choice to take charge of the grouping maintenance because it is difficult and inefficient for BSs to group the mobile users with the same movement. Zhang et al. [6] propose a buffering scheme to relieve the congestion caused by mass arrivals of LU requests. By letting BSs buffer pending LU requests due to the exhaustion of DCCH (Dedicated Control Channel), LU retry attempts can be reduced and hence the congestion can be relieved to some extent. This method attempts to spread the LU requests along the time line by storing the arrived LU requests in BSs' buffers. However, since the BSs can only buffer the arrived LU requests, the congestion in the RACH cannot be effectively relieved with this buffering approach, which is the actual bottleneck in location management. Han et al. propose a similar group LM scheme for one-dimensional networks (transportation systems) in [19]. By specifying virtual visitor location register (VVLR) as the register of group information and an additional tier between VLR and HLR, the wireless access traffic and the signaling burden on HLR can be both mitigated. However, this scheme can only adapt to a dedicated system and cannot be applied to the general PLMN cellular systems. GSM-R approach is another work based on dedicated systems [20]. The focus is on accelerating handoff and cell re-selection procedure by predicting the locations of MTs moving along the railways.

Recently, we proposed a solution for this problem in [7]. As shown in Fig. 1, the massive arriving LU requests have been spread to a longer location update time period. In other words, the LU requests have been instructed to delay for random time lengths. When a large number of MTs is predicted to arrive at a certain location area, the corresponding BSs broadcast indications in the area so that the LU requests will be indicated to delay at least t_{min} which ensures the completion of handoff process. The maximum delay t_{max} is determined by the RACH bandwidth, the velocity and the size of moving MT groups. Although this solution requires BSs to collect and broadcast extra information corresponding to the moving MT groups, which necessitates further designing work, it can to some extent effectively mitigate the RACH congestion and ensure the completion of more time-stringent handoff processes.

Due to the single hop nature of traditional cellular systems, each MT's LU request inevitably contends for RACH, sooner or later. With the multi-hop wireless links in MCNs, signaling aggregation becomes a more attractive way to mitigate the RACH congestion.

3 System model

Before we present our proposed schemes, let us review some preliminaries about the LM. As shown in Fig. 2, in LTE (long term evolution) architecture, when each MT detects the change of location area, it reports the location information (the current visitor location register or VLR) to the HSS (home subscriber system, the HLR in LTE). It first contends for the access of the RACH to request for communication with the eNB (the enhanced node B, the BS in LTE) for channel grant. Beside LU traffic λ_{LU}, new calls λ_N and handoff calls λ_{HO} also contend for the RACH for such a transmission indication. The total traffic in the RACH λ_{RACH} can be expressed as Eq. (1). (Throughout this

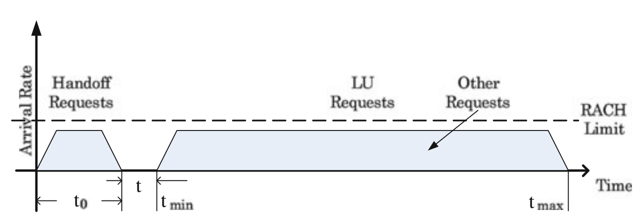


Fig. 1 RACH congestion solution in our previous work

paper, we assume different types of traffic all follow the Poisson process with different λs.)

$$\lambda_{RACH} = \lambda_{LU} + \lambda_{N} + \lambda_{HO} \quad (1)$$

If transmission over RACH is successful, the eNB assigns a DCCH for this request and the MT uses this assigned channel for further signaling. The signaling is then moving to wire-line until it reaches HSS or moving to the data traffic channels for further communications. For a LU signaling request, HSS responds to MME (mobility management entity, mobile switch center in LTE), eNB and then the MT. The DCCH is released when the process is over. Paging process starts when there is an incoming call or a downlink packet and the intended MT is in the IDLE status. Not knowing the exact cell location of the terminating mobile user, MME asks all eNBs within the location area where the mobile user lastly resides to page the MT on the PCCH (paging channel).

As mentioned before, the cost of LM is usually regarded as only the wireless cost (the consumption of wireless resource). Therefore, we ignore the cost of wireline signaling in our analysis. Among the wireless cost of LU and paging, since they consume the resource of different channels, they are not equivalent in the overall cost.

LU is not necessary in each cell because the location area consists of several cells and only when the location area changes is the location update carried out. However, paging process consumes the paging channel of all the cells within the location area. In our proposed schemes, the paging process in MCNs is not changed. Our effort here is to reduce the LUs over the RACH.

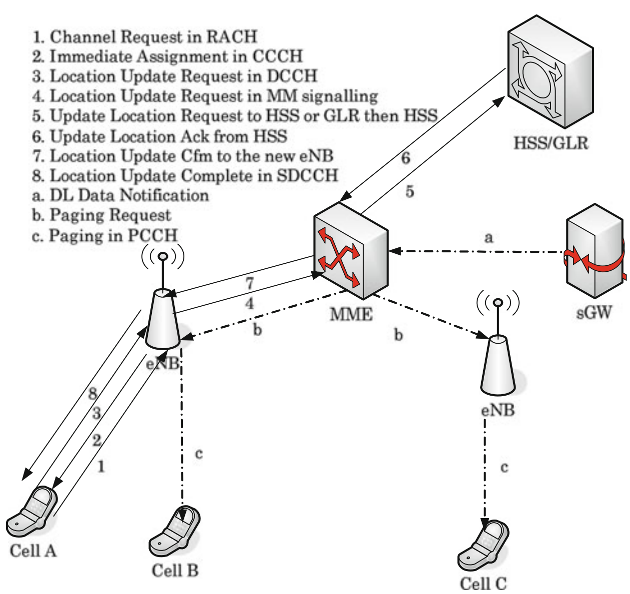


Fig. 2 Location management operational procedure

Although the proposed schemes take advantage of ad hoc links, which requires ad hoc network interface in addition to the cellular interface, they do not disable the terminals with only cellular interface. Old cellphones can normally operate in the proposed new system, only not able to take advantage of the benefits of grouping and aggregating. However, due to the license-free bands' properties, unpredicted interference and wireless failure might put the control signalling at risks. In any case, with careful design of the ad hoc signalling and with the cellular links as the last resort, using ad hoc links is no doubt valuable.

4 Grouping to save the signaling cost

In this section, we present our grouping scheme.

4.1 Overview

Among the wireless signaling congestion scenarios, many are caused by the arrivals of HCT vehicles. In HCT vehicles, many MTs move together along certain fixed routes. When an HCT vehicle arrives at a new location area, the MTs inside the HCT start to update their locations more or less at the same time. These LU requests can easily overwhelm the RACH and cause serious signaling congestions (wireless congestion for short). Apparently, the most direct way to solve this problem is to group these MTs and carry out LUs as a group to relieve the contentions over the RACH. Previous work, GLU [5], proposes to use grouping method in 3G and beyond systems to alleviate the LM traffic. Different from it, we propose a new group LM scheme for MCNs which utilizes the special MTs (selected cooperative personal MTs or operator-deployed lightweight relaying stations) for group maintenance. We assume these special devices, which we name it group head (GH), in each HCT vehicle, control the group maintenance and interact with each eNB they traverse. The incentive for MTs to interact with GHs (deposit LU requests to GHs) is that they can have higher probability to complete their LU process and will not miss prospective calls. Grouping can curtail the contending LU requests and increase the access probability. Moreover, higher access priority for the group LU can be designed to encourage MTs to use group LM. The automatic depositing of LUs can be done by software setting easily. By moving most of the group maintenance messages to the ad hoc links, which may use out-band spectrum, LM cost in MCNs can be greatly reduced and signaling traffic over the RACH can be significantly mitigated. However, to achieve this gain, the procedures need to be carefully designed.

In grouping approaches, there are two challenging problems the designers have to carefully investigate: how to form the groups and how to reduce the group maintenance signaling cost. Moreover, during the LU process, the cellular system may have to allocate new temporary identities to replace the old temporary identities and the grouping scheme should have the capability to fulfill this task.

4.2 Procedures of group location management

In MCNs, a GH, which has both the ad hoc interface and the cellular interface, does not need to be specially fabricated except that it needs to install extra software to take care of the grouping tasks and group message assembling. In addition, the core-networks need to support the group information storage and other corresponding operations. Users can configure their mobile devices to choose whether or not to associate with GHs and which GH to associate with. Automatic association can be configured via software and make users less concerned about how their cell phones are operated. We describe the group LM procedures based on automatic association. The accompanied problem with automatic association will be discussed. The group LM procedures include the following basic operations. An MT needs to execute the association if it is to join a group and associate with a GH, and dissociate with the GH whenever it leaves the group, which can be done by the GH unilaterally in our scheme. The GH shall update the group information with the cellular systems involved periodically. When a new LA is entered, the GH is required to launch the group LU to the system while MTs having association with the GH will not carry out any individual LUs. Before these basic operations are described, we first discuss how a group is formed.

4.2.1 Group formation

Due to the existence of GHs, the group formation seems to be an easy task. Each GH built in the vehicle can easily take over the MTs' LM task whenever they are close to or reside in the vehicle. However, if the group membership changes rapidly, the benefits resulting from the grouping will be diminished. In traditional cellular networks, it is difficult for any BS to form a relatively stable group because the user movement is unpredictable. In MCNs, with the GHs built in the vehicles, MTs with the same movement pattern can be easily grouped together based on signal strength. Unfortunately, GHs cannot distinguish the MTs inside the vehicle from the ones outside. Moreover, one MT may associate with a GH located in any other nearby vehicle. Thus, we need to find an effective way to form a more stable and reasonable group.

We propose to use a guarding period as the condition to start the association with any new group member. Denote the length of this period as T_{ag}. Each GH uses its ad hoc interface to communicate with its current members and at the meantime probes the existence of prospective members. A timer with the length of T_{ag} will be started when a new MT appears. If this MT is still in the range after T_{ag} expires, the GH will start the association process with the MT for group LM purpose and add the MT to the group if the MT agrees to join. Due to channel variation, the links between GHs and their group members can be interfered or even broken. Therefore, another guarding timer T_{dg} is used to prevent the group members from being dissociated with the GHs inadvertently.

These two guarding timers are used to make the group membership stable. The value of T_{dg} depends on the channel condition. Therefore, it is better to use a practical value for this timer or allow the software configuration. On the other hand, the value of T_{ag} determines the gain of using grouping approach. If T_{ag} is too small, many passing-by MTs can easily associate with the GHs, assuming that everyone configures the association as automatic mode. The unnecessary associations waste the channel resource and the gain of using groups decreases. If T_{ag} is too big, it is difficult for MTs to become group members, leading to a smaller group size. With a smaller group size, the gain of grouping is not fully exploited. We will leave the process of finding proper value of T_{ag} in the following section.

Besides the guarding timers, scenarios of co-existing signals from multiple GHs should be also considered. The strategy we propose is the simplest one. If an MT is previously associated with one GH, it will maintain the association even if some other GH has stronger signal strength. If the MT is not associated with any GH when multiple GHs are available, it chooses the one with the strongest signal strength. If the link outage duration reaches T_{dg}, both the GH and the MT deem themselves dissociated. Although in some scenarios this strategy will lead to an inefficient GH selection, such as when a passenger switches to a nearby HCT while maintaining the old association with the previous GH, the association will be corrected after two vehicles move apart with sufficient distance and the delayed association will not impact the system performance significantly.

4.2.2 GH LM association

When a passenger boards or moves close to an HCT vehicle, the MT he/she carries will receive indication messages from GHs residing in the vehicle for group association, instructing the MT to hand over the LM task. Following the indication, the MT with group association option shall start the association procedure as shown in Fig. 3. First, it sends a deposit request message to the GH using its ad hoc interface. After the guarding duration T_{ag}, the GH responds with its identity used by the MT

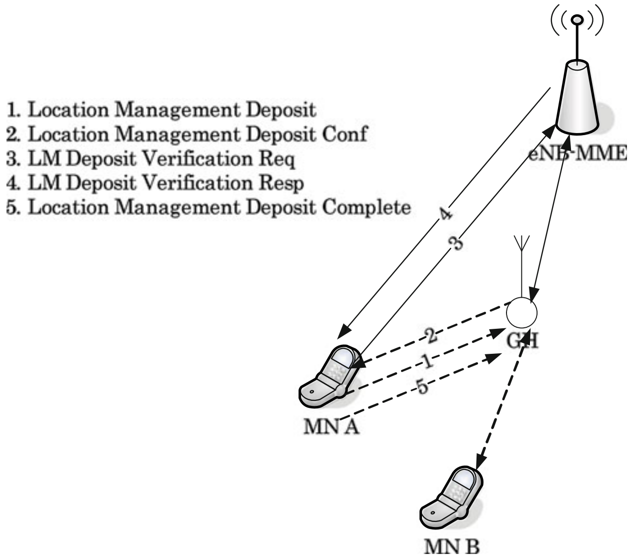


Fig. 3 Group association procedure

afterwards to verify the GH and report the association relationship to the eNB. With this process completed, the GH takes over the LM task of the corresponding MT, including the regular location update and periodic location update, and the system (eNB and MME) has the knowledge of the group and the associated MT. Notice that the communications between MTs and GHs are via ad hoc links and the communications between GHs/MTs and eNBs are via cellular links (Fig. 3).

4.2.3 Group information update

GHs are required to detect the existence of the MT members constantly. If an MT attempts to join the group, it follows the procedure of association. If a group member is lost, the GH is responsible for updating the group information periodically to the network.

Periodically, the updated group information in one MME will be broadcasted to its neighboring MMEs. Since HCTs always follow a certain path, the broadcast can be replaced with multi-cast to reduce the unnecessary communications. These information updates do not involve the HLR (or HSS). In this way, the group information can be synchronized to the MMEs along the traveling path. With this information ready beforehand, the wireless transmissions can be greatly reduced. This idea is similar to the shadow cluster in [21]. However, with the group under consideration, the signalling overhead in our scheme is far lower than the shadow cluster scheme.

4.2.4 Group location update

Upon the LM depositing, the security issues cannot be ignored. In this paper, we simply assume each MT sends its authentication keys along with its ID to the GH in a secure way. Security issues are out of scope of this paper and will be addressed elsewhere.

Upon arriving at a new location area, the GH launches the location update procedure with the group ID. If the constituent of the group is changed after the last report, a list of current members is sent out with the LU request. If there is no change in the group's constituent, a group ID is enough for the group location update. The new MME checks out the MTs requiring LU according to the previously received group information and then sends these LU requests to the corresponding HSS.

Since all group members stay in the same old LA before entering a new LA, their processes of temporary ID (TMSI in GSM) reallocation can be grouped as well. If some MTs are required to carry out some special processes, such as IMSI verification, the following procedures can be carried out individually. As of the system performance, the group LUs have mitigated the RACH congestion problem no matter whether the required procedures are carried out in a grouped fashion or not.

4.2.5 Dissociation

Upon leaving the coverage of a GH, an MT takes over the LM task just as normally done in cellular systems. This is triggered by the timeout of timer T_{dg}. When an HCT vehicle arrives at the destination, the GHs in the HCT vehicle carries out the dissociation process with all the group members clearing the information storage at MMEs.

4.3 Cost analysis

In this subsection, we conduct the performance evaluation for our proposed scheme and check how much gain we can have by using the proposed group LM scheme.

4.3.1 Signaling cost reduction

In MCNs, the cellular spectrum is more critical than ad hoc spectrum in that ad hoc mode is usually seen as an addition to cellular systems in MCNs while lack of available cellular bandwidth can directly cause call blocking or call dropping. As shown in Fig. 2, if we utilize the grouping method to report the location, massive location update requests can be combined as one. The utilization of grouping can first enhance the success rate of LU. As we know, LU requests contend for the RACH to setup new/handoff calls, mostly following the slotted-ALOHA protocol. Denote the LU traffic rate as R_{LU} and call connection traffic rate as R_c, respectively. The average group size is denoted by K. The gain of LU failure rate (we assume an

LU fails only because of the RACH congestion) can be expressed as follows.

$$\eta = \frac{1 - e^{-(R_c + R_{LU})}}{1 - e^{-(R_c + \frac{R_{LU}}{K})}} \quad (2)$$

We can see from Eq. (2) that, when the LU traffic overwhelms the RACH, the LU grouping can greatly improve the LU success rate as well as the success rate of new/handoff call setup.

In Eq. (2), the traffic of LUs is approximately reduced to K times when the grouping is executed. However, the exact wireless cost of LU should be analyzed more carefully in order to evaluate the performance quantitatively. We first investigate the wireless cost of regular LU without grouping in one location area with M MTs at unit time duration. The periodic LU cost is not included in this analysis for the simplicity of analysis. The average speed of the HCT vehicle is assumed to be $\bar{v}_{HCT}$ in the number of cells per unit time. We assume the location area consists of a cluster of cells, which forms a series of concentric hexagons with the radius of d cells. The expected regular LU signaling cost can be expressed as follows.

$$C_{rLU} = \frac{M \cdot \bar{v}_{HCT}}{2d - 1} \quad (3)$$

The LU cost means the average times of LUs during unit time among M MTs, assuming each MT updates its location after they move a $2d - 1$ distance.

With our grouping method, location update traffic is reduced K times along with the incurred MT-eNB registration traffic and the GH-eNB group information update traffic. Since ad hoc spectrum is not critical for mobile users to access the system due to possible harvesting of free out-band spectrum, we ignore the wireless cost in ad hoc spectrum. For analysis, we assume that all M MTs are moving in K groups together, creating the K times-reduction location update traffic, the group information is updated for each group once in every interval, and the association traffic for each MT once in every trip. $\bar{L}$ denotes the average trip length of passengers in the unit of number of cells associated with the GH. T_{gr} is the GHs' intervals for reporting group information update. ω_{gm} and ω_{ass} stand for the weights of group maintenance messages and individual association messages, respectively. Therefore, we have the wireless cost for group LM as follows.

$$C_{GLM} = \frac{M}{K} \cdot \frac{\bar{v}_{HCT}}{2d - 1} + \omega_{gm} \cdot \frac{M}{K} \cdot \frac{1}{T_{gr}} + \omega_{ass} \cdot \frac{M \cdot \bar{v}_{HCT}}{\bar{L}} \quad (4)$$

The first item in Eq. (4) stands for the group LU traffic cost, which is reduced K times compared to the traditional LM. Since different messages transmitted over wireless channels have different impact on system performance, we use different weights for them. This part has unit weight in order to be consistent with Eq. (3). The second item stands for the incurred traffic due to group information updating by GHs. The last item is the MTs' association cost, which needs only once during the whole trip. Obviously, with the aim of mitigating RACH congestion, ω_{gm} and ω_{ass} should be less than 1. It can be observed from Eq. (4) that when group size K and $\bar{L}$ are large, the total cost can be significantly reduced. We can easily conclude that when K is relatively larger, the wireless gain is mainly determined by the ratio of $\frac{\bar{L}}{\omega_{ass} \cdot (2d-1)}$. When $\bar{L}$ is much larger than $2d - 1$, the diameter of a location area, the wireless gain is determined by the value of K.

The wireless gain through grouping is hereby defined as follows.

$$\zeta_{GLM} = \frac{C_{rLU}}{C_{GLM}} \quad (5)$$

Besides this gain, grouping can greatly disperse the signaling traffic in time, leading to the mitigation of the RACH congestion, which is a more important benefit of grouping method.

4.3.2 Finding T_{ag}

As mentioned earlier, the wireless cost gain of the proposed grouping scheme depends on the value of T_{ag}. If T_{ag} is too small, the group size K can be large while the average trip length of the passengers $\bar{L}$ is smaller because it is easy to join the group and many non-passengers will be added in the group. On the other hand, when T_{ag} is too large, K will be reduced and $\bar{L}$ is larger. In this case, many short-term passengers might be excluded from the group. There is no available model which can define the functions of $\bar{L}$ and K in terms of T_{ag}. Obviously, they could be estimated from real data. Suppose we acquire the relationship, *i.e.* $\bar{L} = f(T_{ag})$ and $K = g(T_{ag})$, we can derive the optimum T_{ag} with Eqs. (3), (4) and (5).

4.3.3 Other prospective benefits

Besides group LU, grouping mechanism in MCNs can potentially improve the system performance in many ways. For example, GHs can help the adjacent cells to reserve resource for better QoS provisioning. GHs might provide an alternative way for a better handoff, and can provide in-vehicle data service. We can even use GHs in the HCT vehicles to provide a necessary hardware configuration for upcoming services by developing prediction algorithm.

5 Generic aggregative location management in MCNs

In the previous section, we focus on the scenario on HCT transporation systems. Our proposed group scheme can also be used to handle much more general scenarios, which is discussed in this section.

5.1 Overview

When big events are held, people tend to flock at the same location around the same time, which usually brings in wireless congestion on the RACH as well as the traffic jams. We cannot use our previous grouping scheme to solve the LU congestion because the MTs are not following the same route. However, we still can take advantage of the ad hoc feature of MCNs to aggregate the LU requests.

In this section, we propose to utilize the aggregation devices (ADs) along the location area boundaries to collect the location update requests via ad hoc links periodically. After each time interval T_{aggr}, these ADs combine the collected location update requests into one message and send to the eNBs so that the RACH congestion can be greatly mitigated. Each eNB is required to decide the T_{aggr} for the ADs connecting to itself according to the traffic load in the cell. To encourage mobile users to delegate their LUs to the ADs, T_{aggr} should be set smaller than the expected delay caused by possible transmission collisions and retransmissions. Shorter delay of LU means smaller probability of missing incoming calls. Therefore, ADs, which can provide smaller delay with a proper T_{aggr}, can attract MTs to delegate their location updates and thus mitigate the access traffic over the RACH. Since these ADs periodically access the system to carry out the LUs for multiple MTs, eNBs can further allocate fixed channel resource for them and then the repeated contention for RACH is mitigated. Next, we describe the generic scheme of aggregative LM in MCNs.

5.2 Aggregative LM scheme

An AD in this scheme can be any device with the aggregation software built in. All ADs and delegating MTs should have both ad hoc interface, usually a Wi-Fi card, and cellular interface. ADs should be authorized for the LM task by the system. Moreover, ADs cannot be the ones with high mobility in order to efficiently perform their aggregation tasks. When the ADs are in the positions where the LM aggregation is needed, they collect the delegation of location update requests through ad hoc links and launch LU procedure periodically, acting as the agents of corresponding MTs. To indicate the arriving MTs to delegate the location update task to ADs, each eNB sends out indications in the broadcast channel, as shown in Fig. 4. The indication messages should also include the list of the authorized ADs. Although security concerns should be included in this process, we do not cover them in our scheme due to the scope of this paper. Each MT searches

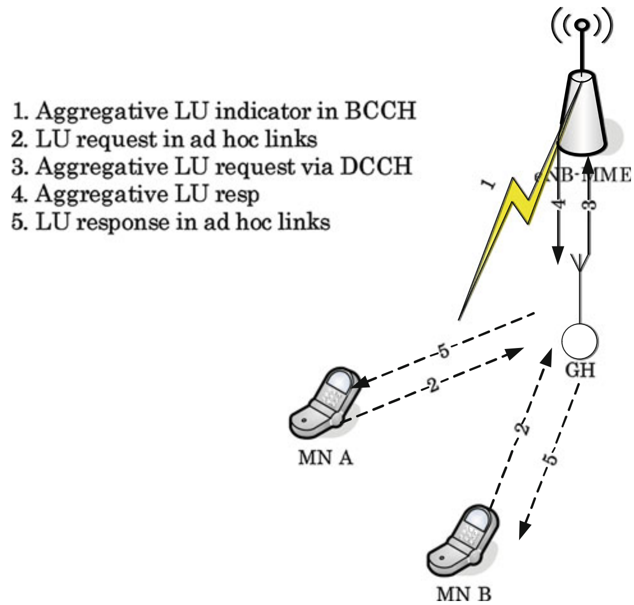


Fig. 4 Generic aggregative location management

for nearby ADs and decides whether or not to delegate LM according to the broadcasted information from both ADs and eNBs. How to indicate the LU congestion probability P_{col} in the RACH and how to decide the interval length of ADs' aggregation period T_{aggr} are two major design issues in this scheme.

ADs are selected by the eNBs in this scheme although more flexible and dynamic methods can be adopted in the future to manifest the benefits of ad hoc mode. T_{aggr} of each AD within the same cell is set to the same value by the eNB. Each cell adjusts the value of T_{aggr} according to the information of LU traffic and other traffic using the RACH during the last period. Each MT obtains the traffic information over the RACH from the BCCH. The collision probability P_{col} can be derived with the knowledge on the average packet duration in seconds, denoted as σ, and the arrival rate of the RACH traffic, denoted as λ_{RACH}, as show in Eq. (6).

$$P_{col} = 1 - e^{-\sigma \cdot \lambda_{RACH}} \qquad (6)$$

With the knowledge on the timeout for LUs t_{LU} and the transmission delay D_t, the expected delay for LUs over the RACH can be determined as well, as shown in Eq. (7). The second item in D_{RACH} is simply the product of the expected retransmission times and t_{LU}.

$$D_{RACH} = D_t + \frac{t_{LU} \cdot P_{col}}{1 - P_{col}} \qquad (7)$$

The total LU delay D_{LU} is the summation of D_{RACH} and wireline signaling delay D_{wire}, which can be expressed as

$D_{LU} = D_{RACH} + D_{wire}$. Furthermore, with the knowledge on the call-to-mobile ratio γ (the expected times of call to the mobile in unit time), we can derive the expected number of missed calls.

$$N_{missed} = D_{LU} \cdot \gamma \tag{8}$$

In order to determine the aggregation period T_{aggr}, we first need to find the constraints. D_{ALU} denotes the wireless delay of the proposed aggregative scheme, which is mainly due to the aggregation interval. D_{CSMA}, the CSMA access delay, also contributing to D_{ALU}, can be ignored when the traffic is light. D_{aggr} is the aggregation delay.

$$D_{ALU} = D_{CSMA} + D_{aggr} \tag{9}$$

D_{CSMA} can be expressed as Eq. (10), where P_f is the failure rate of one transmission in the CSMA scheme, W_i is the backoff window size of $(i+1)$th transmission, and *EIFS* is the timeout for one failed transmission. *Retrans* is the retransmission limit.

$$D_{CSMA} = D_t + \frac{W_0}{2} + \sum_{i=1}^{Retrans} (EIFS + \frac{W_i}{2}) \cdot P_f^i \tag{10}$$

Obviously, when D_{ALU} is smaller than D_{RACH}, *i.e.* $D_{ALU} < D_{RACH}$, MTs prefer delegating LUs to ADs.

When the access of ADs is not congested, the expected value of D_{aggr} is $T_{aggr}/2$. Therefore, we have the first constraint of T_{aggr} as Eq. (11).

$$T_{aggr} \leq 2 * \left(\frac{t_{LU} \cdot P_{col}}{1 - P_{col}} \right) \tag{11}$$

If an AD accepts too many LU delegation requests, the delay of accessing AD via ad hoc links, D_{CSMA}, can be so large that D_{ALU} is expected to be larger than D_{RACH} even if T_{aggr} is set to a small value. The MTs can choose other ADs or directly contend for the RACH instead.

There is another constraint on the aggregation interval that the aggregated traffic should not go beyond the capacity. Although the periodic aggregated location update via ADs can bypass the contention over the RACH, there is still an upper-bound on acceptable LU traffic, denoted as $\hat{\lambda}_{LU}$, which is related to the traffic of new calls and handoff calls. N_{AD} denotes the number of ADs, λ_{LU} is the LU traffic, and δ is the aggregation rate of the LU traffic, then we obtain the second constraint expressed as follows.

$$\begin{aligned} &\frac{N_{AD}}{T_{aggr}} + (1-\delta) \cdot \lambda_{LU} \leq \hat{\lambda}_{LU} \\ &\Rightarrow T_{aggr} \geq \frac{N_{AD}}{\hat{\lambda}_{LU} - (1-\delta) \cdot \lambda_{LU}} \end{aligned} \tag{12}$$

Equation (12) indicates that, when all the LU traffic goes through ADs, *i.e.*, $\delta = 1$, larger aggregation interval can accept more LU traffic and reduce the wireless cost further.

With these two constraints, we can roughly determine the aggregation interval. It should be no greater than a value satisfying Eq. (11) in order to stimulate mobile users to use ADs. It also needs to be no smaller than another value satisfying Eq. (12) so that the aggregative traffic will not exceed the handling capability of the eNB. Although some ADs may not accept any LU delegation during some periods due to the skewed LU traffic distribution, we ignore in the analysis the possibility that these ADs would release the assigned channels and re-contend for the RACH when new LU traffic arrives. In the case that the value of T_{aggr} cannot satisfy both inequalities at the same time, we simply remove the first constraint [Eq. (11)], which means that the LU traffic has not congested the RACH and the delay via the RACH is acceptably small. Therefore, there is no need to stimulate the aggregation by blindly shortening the aggregation interval. It is also worth noting that only when there is RACH congestion is Eq. (11) required.

According to the value determined based on the above two inequalities, we obtain the wireless cost gain as follows.

$$\zeta_{ALU} = \frac{\lambda_{LU} \cdot T_{aggr}}{N_{AD} + (1-\delta) \cdot \lambda_{LU} \cdot T_{aggr}} \tag{13}$$

6 Location management cost analysis and application in MCNs

Although we only strive to reduce the wireless cost on location update, the proposed scheme can potentially change the overall design of LM. Before we can achieve this goal, we need to deepen our understanding of the LM. Previously, there are several papers formulating the cost function for LM, such as [2, 15, 16, 22]. However, observing that the RACH mitigation mostly deals with the linear movements of HCT vehicles and regular vehicles, we want a simpler cost function which simplifies the complicated mobility patterns and shows the relationship between LU cost (including LU and paging) and location area size straightforwardly.

To formulate the cost function for LM, we need to know two parameters. One is the call-to-mobile ratio, γ, which characterizes the probability that a mobile user is to be called. The other is the average moving speed, v, which captures how soon an MT is to change the location area on average. We assume that the location area consists of a cluster of cells which form a series of concentric hexagons as mentioned before. We denote the radius as d in the number of cells. For the convenience of understanding, we also define the average number of users in the location area, M, and the time interval for the cost evaluation, τ. The LM cost consists of the wireless cost for LU and the wireless cost for paging, as shown in Eq. (14).

$$C_{total} = \omega_{LU} \cdot C_{LU} + \omega_{paging} \cdot C_{paging} \quad (14)$$

To minimize the overall cost, we need to place different weights on these two parts. ω_{LU} and ω_{paging} stand for the weights for LU and paging, respectively. Apparently, since the RACH resource is more critical than PCCH as we mentioned before, we have the following relationship, $\omega_{LU} \gg \omega_{paging}$.

LUs happen when MTs detect the location area change. Assuming each MT moves straightly with the average speed v, we can expect an LU to occur at most after it moves across the diameter length of the location area, which is $2d - 1$. (Due to different mobility patterns, this value may not be the average value. However, we use this value since the scenarios under consideration in this paper are mostly about linear movement.) Therefore, for M MTs in τ duration, the wireless cost for LU can be expressed as

$$C_{LU} = \frac{v \cdot M \cdot \tau}{2d - 1} \quad (15)$$

Different from LU, paging will happen in each cell within the location area whenever there is an incoming call. Note that M and τ are in both equations and hence can be removed without impact on further analysis.

$$C_{paging} = \gamma \cdot M \cdot \tau \cdot (3d^2 - 3d + 1) \quad (16)$$

By finding the first and second derivative of C_{total}, we can obtain the minimum value of C_{total} and the corresponding d.

$$\frac{\partial C_{total}}{\partial d} = -\frac{2\omega_{LU} \cdot v}{(2d-1)^2} + \omega_{paging} \cdot \gamma \cdot (6d - 3) \quad (17)$$

$$\frac{\partial^2 C_{total}}{\partial d^2} = \frac{8\omega_{LU} \cdot v}{(2d-1)^3} + 6\omega_{paging} \cdot \gamma \geq 0 \quad (18)$$

Since the second derivative is positive, by letting the first derivative equal to 0 we can obtain d which minimizes C_{total}.

$$d_{opt} = \frac{\sqrt[3]{\frac{\omega_{paging} \cdot \gamma}{\omega_{LU} \cdot v}} + 1}{2} \quad (19)$$

7 Performance evaluation

In section, we carry out the comparison analysis and investigate the benefits of our proposed schemes in terms of wireless cost.

For the grouping scheme, we look into the grouping gains in terms of wireless signaling cost with different group sizes, K, and different average trip lengths, $\bar{L}$. The signaling cost is defined as in Eqs. (3) and (4). The signaling gain is defined as in Eq. (5). The basic system setting is given in Table 1.

By setting $\bar{v}_{HCT}$ to 1.5, 3.0 and 6.0, we calculate the gains of the signaling cost. Figure 5 shows the result. We can see that the gains increase with the group size K, however, not linearly. Larger $\bar{v}_{HCT}$ brings greater gain because the periodic group reporting occupies lower signaling cost compared with other $\bar{v}_{HCT}$ related signaling. Because cost for the individual association is not related to K, the gains do not have linear relationship to K.

Figure 6 shows the comparison of the gains with different average trip lengths $\bar{L}$. We set K to 10, 50 and 200 Apparently, longer $\bar{L}$ along with larger K brings greater signaling cost gains. When K is as large as 200, the gains almost increase linearly with $\bar{L}$. This means when the group size is large enough, the cost for LU and group reporting is relatively negligible, and thus longer average trip length directly increases the total gain.

Although big K and big $\bar{L}$ together bring great gains, they usually do not come together. We have mentioned earlier that in certain scenarios, when we set the association timer with a large value, the long average trip length can be acquired whereas with the group size sacrificed. Short association timer has opposite effect. Therefore, if we want to optimize the signaling cost gain, we need first gather real data and find the relationships among $K, \bar{L}$ and the timer.

For the generic aggregation scheme, we need to first look into the interval T_{aggr}. Based on the longest allowable T_{aggr}, we then observe the gain of signaling cost. Before we evaluate the scheme, we list the parameters' setting as in Table 2.

The upper bound of T_{aggr} is defined in Eq. (11) and the lower bound is defined in Eq. (12). By changing the LU traffic load, we can observe the change of the range of the aggregation interval.

Figure 7 shows the results with different LU traffic load. We observe that the lower bound of the aggregation interval does not change much with the LU traffic. This is because only non-aggregated LU traffic, which is only a small proportion, affects the lower bound of the

Table 1 Parameters' setting for evaluating the grouping scheme

M	800
d	4 cell
ω_{gm}	0.2
ω_{ass}	0.02
T_{gr}	0.5 h
$\bar{v}_{HCT}$(if it is not the variable)	6.0 cell/h
$\bar{L}$ (if it is not the variable)	7.3 cell

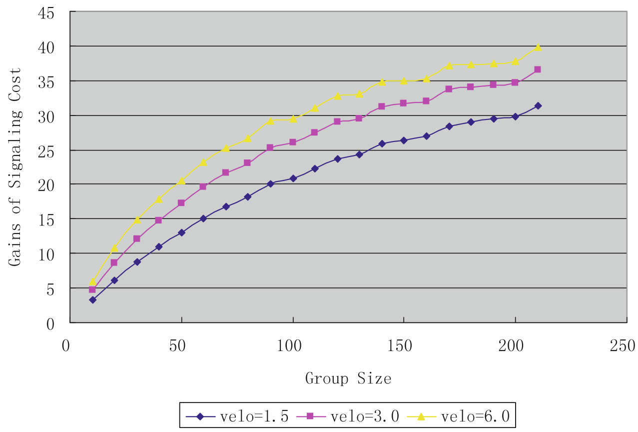


Fig. 5 Grouping gains of signaling cost with different group sizes (K)

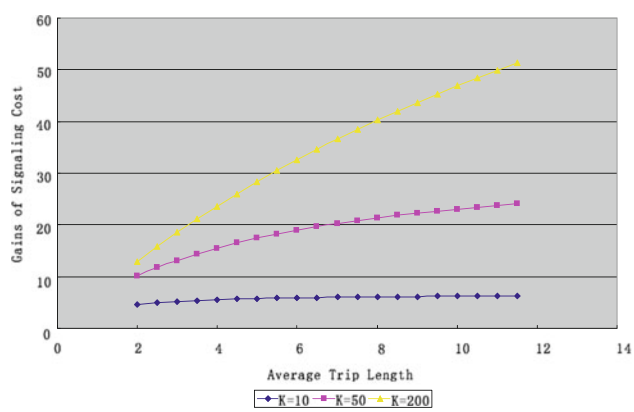


Fig. 6 Grouping gains of signaling cost with different trip lengths ($\bar{L}$)

Table 2 Parameters' setting for evaluating the generic aggregation scheme

λ_c	2.5 p/s
$\hat{\lambda}_{LU}$	6.0 p/s
σ	0.1 s/p
t_{LU}	5 s
δ	0.8
N_{AD} (if it is not the variable)	20

aggregation interval. The upper bound of the aggregation interval increases exponentially with the LU traffic in that it depends on the collision probability in the slotted-ALOHA.

Figure 8 shows the signaling cost gain for the generic aggregation scheme. This scheme does not bring other extra traffic and aggregates the LU traffic only. By setting the aggregation timer with the largest allowed value, we obtain the gains. It shows that fewer ADs can apparently bring significant gains. This is because fewer ADs can

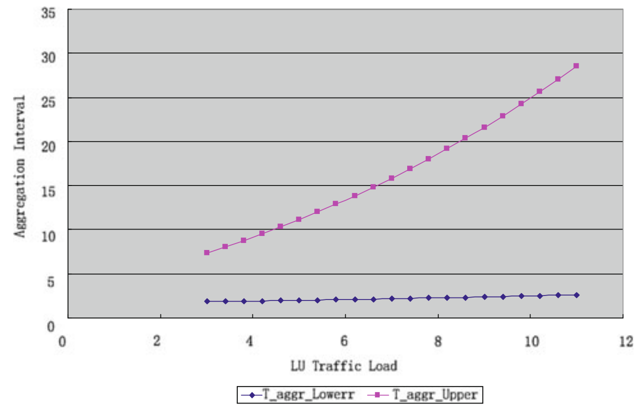


Fig. 7 The aggregation intervals with different LU traffic

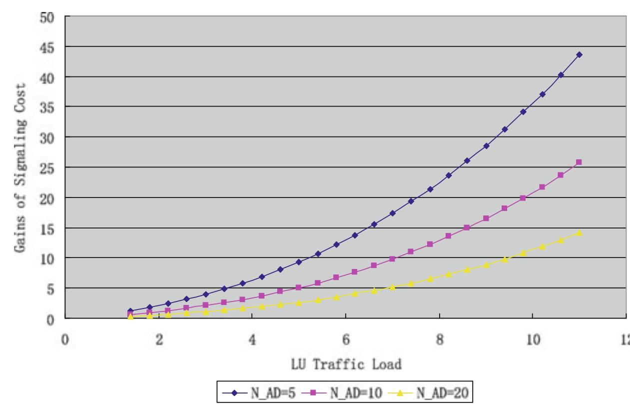


Fig. 8 The aggregation signaling gains with different LU traffic

aggregate more LU traffic. However, if there are not enough ADs, high aggregation rate δ cannot be guaranteed and more congestions at ADs will occur.

It is worth noting that this performance evaluation does not include periodic LU for the purpose of simplicity. However, periodic LU can be easily aggregated with our aggregative approach and thus higher signaling cost gain will be achieved.

8 Conclusion

Based on the multi-hop cellar network (MCN) architecture, in this paper, we have proposed two schemes to aggregate the LU traffic and thus mitigate the potential LU congestions over the random access channel. The grouping scheme is designed for the high capacity transportation systems where many MTs move together with HCT vehicles. By forming groups of moving-in-together MTs, this scheme can aggregate location updates into group LU updates, leading to less signaling traffic over the random access channel even with the incurred association messages and group maintenance messages. For the scenarios where MTs flock without regular paths, the proposed generic aggregation scheme provides a mechanism to aggregate the arriving location update requests and send them to the network in batches, which is apparently a more efficient way. The feasible value of the aggregation interval has been extensively discussed. Furthermore, for MCNs, due to the existence of ad hoc links, which can potentially use non-cellular spectrum, the location management can have lower wireless cost with the use of harvested out-band wireless spectrum. We expect that our proposed aggregative approach will provide significant performance gain for future wireless cellular networks.

Acknowledgments This work was partially supported by the National Science Foundation under Grants CNS-0916391 and CNS-1147813. The work of Y. Fang was also partially supported by the National Natural Science Foundation of China under grant No. 61003300.

References

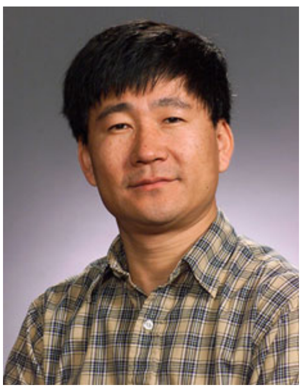
1. Ma, W., &, Fang, Y. (2004). Dynamic hierarchical mobility management strategy for mobile IP networks. *IEEE Journal on Selected Areas in Communications (Special Issue on All-IP Wireless Networks,) 22*(4), 664–676.
2. Fang, Y. (2003). Movement-based location management and tradeoff analysis for wireless mobile networks. *IEEE Transactions on Computers (Special Issue on Wireless Internet), 52*(6), 791–803.
3. Fang, Y. (2002). General modeling and performance analysis for location management in wireless mobile networks. *IEEE Transactions on Computers , 51*(10), 1169–1181.
4. Ma, W., &, Fang, Y. (2002). Two-level pointer forwarding strategy for loc ation management in PCS networks. *IEEE Transactions on Mobile Computing, 1*(1), 32–45.
5. Wang, F., Tu, L., Zhang, F., &, Huang, Z. (2005). Group location update scheme and performance analysis for location management in mobile network. In IEEE 61st semiannual vehicular technology conference, Stockholm, Sweden.
6. Zhang, Y., &, Fujise, M. (2007). Location management congestion problem in wireless networks. *IEEE Transactions on Vehicular Technology, 56*(2), 942–954.
7. Zhao, H., Huang, R., &, Fang, Y. (2011). Resolving RACH congestion for high speed moving group in wireless networks. In ICC 2011 Kyoto, Japan.
8. 3GPP Group (2005). All-IP network (AIPN) feasibility study. 3GPP TS 22.978 V7.1.0.
9. Bhargava B., Wu X., Lu Y., &, Wang W. (2004) Integrating heterogeneous wireless technologies: A cellular aided mobile ad hoc network (CAMA). *ACM Special Issue of the Journal on Special Topics in Mobile Networking and Applications (MONET), 9*(4), 393–408.
10. Lin, Y., &, Hsu, Y. (2000). Multi-hop cellular: A new architecture for wireless communications. In Infocom'00, Tel-Aviv, Israel.
11. Luo, H., Ramjee, R., Sinha, P., Li, L., &, Lu, S. (2003). UCAN: A unified cellular and ad-hoc network architecture. In Mobicom'03, San Diego, CA.
12. Cavalcanti D., Agrawal D., Cordeiro C., Xie B., &, Kumar A. (2005) Issues in integrating cellular networks WLANs, and MANETs: A futuristic heterogeneous wireless network. *Wireless Communications, IEEE, 16*(3), 30–41.
13. Law, L., Krishnamurthy, S., &, Faloutsos, M. (2008). Capacity of hybrid cellular-ad hoc data networks. In Infocom'08, Phoenix, AZ.
14. Ma, W., &, Fang, Y. (2005). A pointer forwarding based local anchoring (POFLA) scheme for wireless networks. *IEEE Transactions on Vehicular Technology, 54*(3), 1135–1146.
15. Wong, V., &, Leung, V. (2001). An adaptive distance-based location update algorithm for pcs networks. *IEEE Journal of Selected Areas in Communications, 19*(10), 1942–1952.
16. Akyildiz, I., Ho, J., &, Lin, Y. (1996). Movement-based location update and selective paging for pcs networks. *IEEE/ACM Transactions on Networking (TON), 4*(4), 629–638.
17. Misra, A., Roy, A., &, Das, S. (2008). Information-theory based optimal location management schemes for integrated multi-system wireless networks. *IEEE/ACM Transactions on Networking (TON) 16*(3), 525–538.
18. Zang, H., &, Bolot, J. (2007). Mining call and mobility data to improve paging efficiency in cellular networks. In MOBICOM'07, Montreal, Canada.
19. Han, I., &, Cho, D. (2004). Group location management for mobile subscribers on transportation systems in mobile communication networks. *IEEE Transactions on Vehicular Technology (TVT), 53*(1), 181–191.
20. Kastell, K., Bug, S., Nazarov, A., &, Jakoby, R. (2006). Improvements in railway communication via GSM-R. In vehicular technology conference IEEE 63rd, Melbourne, Vic.
21. Levine, D., Akyildiz, I., &, Naghshineh, M.A (1997). Resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept. *IEEE/ACM Transactions on Networking (TON), 5*(1), 1–12.
22. Li, J., Kameda, H., &, Li, K. (2000). Optimal dynamic mobility management for pcs networks. *IEEE/ACM Transactions on Networking (TON), 8*(3), 25–33.

Author Biographies

Rongsheng Huang received his B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1996 and 1999, respectively. He received his Ph.D. degree in the Electrical and Computer Engineering from University of Florida in 2011. From 1999 to 2001, he was working with Huawei Technologies Co. Ltd. as an R&D engineer on GPRS and 3G projects. From 2002 to 2005, he was working with UTStarcom Reseach Center, Shenzhen, China, as a senior engineer and team leader on 3G project. He is now working with Olympus

Communication Technology America, Inc. since 2011. His research interests are in the area of media access control, protocol and architecture for wireless networks.

Hongxia Zhao received a B.S. in Electrical Engineering and M.S. in Electrical Engineering from the University of Electronic Science and Technology of China (UESTC), in 1994 and 1997 respectively. She is an engineer at the Huawei Technologies Co., Ltd. Her main area of research interest is wireless networks.



Yuguang Fang received a Ph.D. degree in Systems Engineering from Case Western Reserve University in January 1994 and a Ph.D. degree in Electrical Engineering from Boston University in May 1997. He was an assistant professor in the Department of Electrical and Computer Engineering at New Jersey Institute of Technology from July 1998 to May 2000. He then joined the Department of Electrical and Computer Engineering at University of Florida in May 2000 as an assistant professor, got an early promotion to an associate professor with tenure in August 2003 and to a full professor in August 2005. He holds a University of Florida Research Foundation (UFRF) Professorship from 2006 to 2009, a Changjiang Scholar Chair Professorship with Xidian University, Xi'an, China, from 2008 to 2011, and a Guest Chair Professorship with Tsinghua University, China, from 2009 to 2012. He has published over 300 papers in refereed professional journals and conferences. Dr. Fang received the National Science Foundation Faculty Early Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002, and is the recipient of the Best Paper Award in IEEE International Conference on Network Protocols (ICNP) in 2006 and the recipient of the IEEE TCGN Best Paper Award in the IEEE High-Speed Networks Symposium, IEEE Globecom in 2002. Dr. Fang is also active in professional activities. He is a Fellow of IEEE and a member of ACM. He served as the Editor-in-Chief for IEEE Wireless Communications and serves/served on several editorial boards of technical journals including IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Wireless Communications Magazine and ACM Wireless Networks. He was an editor for IEEE Transactions on Mobile Computing and currently serves on its Steering Committee. He has been actively participating in professional conference organizations such as serving as the Steering Committee Co-Chair for QShine from 2004 to 2008, the Technical Program Co-Chair for IEEE INFOCOM'2014, the Technical Program Vice-Chair for IEEE INFOCOM'2005, Area Technical Program Committee Chair for IEEE INFOCOM (2008–2013), Technical Program Symposium Co-Chair for IEEE Globecom'2004, and a member of Technical Program Committee for IEEE INFOCOM (1998, 2000, 2003–2008), ACM Mobihoc (2008–2009), ACM Mobicom (2001), IEEE ICDCS (2004).