Location Management Optimization Schemes for Cellular System

Sudhir Agrawal
M.E. (CSE)
NITTTR
Sec-26, Chandigarh, India

Sarvpal H. Singh
Assistant Professor
MMM Engg. College
Gorakhpur, India

C. Ramakrishna
Associate Professor
NITTTR
Sec-26, Chandigarh, India

ABSTRACT
The ever-increasing demand of mobile system is posing a great challenge for efficient call delivery due to tremendous surge in numbers of mobile subscribers. In order to scale well with increased users, the efforts are underway to develop the new location management techniques as well as to optimize the existing ones so as to keep up-to-date information about current location of mobile phones without causing much traffic load in cellular network. Various location management optimization approaches are being explored and developed by research community in recent past. This papers sheds light on various such approaches in order to give a comprehensive overview of research registered in this field..

Keywords
Mobility Management, Location Management (LM), Location Area (LA)

1. INTRODUCTION
The cellular system has registered a rapid growth in world communication scenario now-a-days. The standard like GSM has captured around 75% of the world communication market through its widespread use in European countries, India etc. In GSM, a mobile unit (MU) is free to move around within the entire area of coverage. The movement is random and therefore the geographical location of mobile unit is unpredictable. In order to deliver a call to MU, it is necessary to have information about its current location. This requirement has been implemented in GSM standard through HLR/VLR architecture [1-3]. The entire process of the mobility management component of the cellular system is responsible for two tasks:(a) Location management- that is, identification of the current geographical location or current point of attachment of a mobile unit which is required by the MSC (Mobile Switching Center) to route the call- and (b) Handoff- that is, transferring (handing off) the current (active) communication session to the next base station, which seamlessly resumes the current session using its own set of channels. Thus, the entire process of location management is a kind of directory management problem where current locations of MU are maintained continuously by carrying out updations in HLR with the help VLR. The current point of attachment or location of a subscriber (mobile unit) is expressed in terms of the cell or the base station to which it is presently connected. The mobile units (called and calling subscribers) can continue to talk and move around in their respective cells; but as soon as both or any one of the units moves to a different cell, the location management procedure is invoked to identify the new location.

The provision of unrestricted mobility of mobile units creates a complex dynamic environment, and the location management component must be able to identify the correct location of a unit without any noticeable delay. The location management performs three fundamental tasks: (a) location update, (b) location lookup, and (c) paging. In location update, which is initiated by the mobile unit, the current location of the unit is recorded in HLR and VLR databases. Location lookup is basically a database search to obtain the current location of the mobile unit and through paging the system informs the caller the location of the called unit in terms of its current base station. Understanding the importance and role of location management, it is always desired to have efficient Location management so as to provide communication facilities to clients during mobility. Efforts are continued to explore new strategies to reduce the handshaking overheads, improve the bandwidth usage and facilitating the call to subscriber with better and better response time.

The continuous location trapping of MU has made it desirable to have the efficient location management schemes so that HLR updations' overhead can be minimized. The other related issue is the distribution of HLR to shorten the access path, which is similar to data distribution problem in distributed database systems. These issues have motivated research world to a great extent and recently a number of innovative location management schemes have appeared in the research world. This paper attempts to present a comprehensive overview of such approaches/techniques.

The location management approaches can broadly be classified as (1) Centralized approaches (2) Distributed approaches. Centralized approach keeps information only on one node in the mobile network. For example, existing location management standards, IS-41 and GSM are centralized approaches. Location lookup and update operations are simple in this case but they suffer from severe problems like congestion, central point failure etc. In Distributed approaches [9-14] user information is distributed among many nodes in the network. This has better stability in comparison to centralized approach but location lookup and update operations are somewhat complex in this case. However, many schemes have come forward to solve this problem. A variant of distributed approach is Hierarchical approach [9,11-13] which has been suggested to overcome the drawbacks of centralized approach. In this approach, location registers in the PCS network are arranged in hierarchical fashion. The topology resembles to a tree with a root level Location Register (LR) and its ascendant LRs while each leaf level LR performs location management operation for one zone of PCS network. Various approaches have been suggested regarding how to replicate user profile in the hierarchy. One key problem in hierarchical approach is what should be the topology of hierarchy.

The remaining part of this paper has been organized as follows: Section 2 gives an overview about centralized approach used in IS-41 and GSM with optimizations/improvements suggested for it. Section 3 sheds light on distributed approaches followed by the final conclusion included in section 4.

2. CENTRALIZED APPROACH

This Section describes existing standard of location management followed by the approaches evolved to overcome drawbacks of existing standard.

The two popular standards currently used are GSM and IS-41. They make use of two types of Registers, home location register (HLR) and visitor location register (VLR), to store the location information of the mobile terminals. Figure 1 shows the basic architecture [15] under this two- level hierarchy.

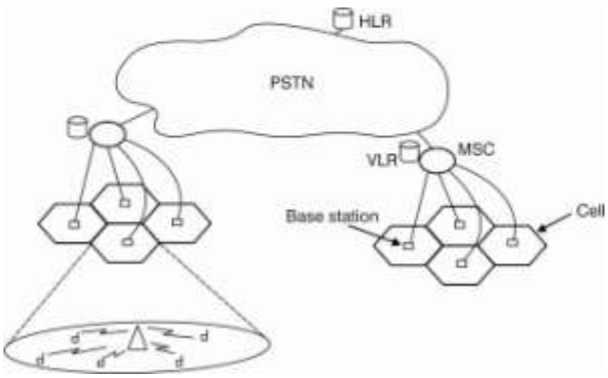

Figure 1: The Standard PCN architecture

The HLR stores the user profiles of its assigned subscribers. These user profiles contain information such as the type of services subscribed, the quality-of-service (QoS) requirements and the current location of the mobile terminals. Each VLR: stores replications of the user profiles of the subscribers currently residing in its associated RA. In order to effectively locate a mobile terminal when a call arrives, each mobile terminal is required to report its location whenever it enters a new RA. We call this reporting process location update. On receiving a location update message, the MSC updates its associated VLR and transmits the new location information to the HLR. We call this Register update process as location registration. The HLR will acknowledge the MSC for the successful registration and it will also deregister the mobile terminal at the VLR of old RA. In order to locate a mobile terminal for call delivery, the HLR is queried to determine the serving MSC of the target mobile terminal. The HLR then sends a message to this MSC which, in turn, will determine the serving base station of the mobile terminal by paging all cells within its associated RA.

This location tracking scheme requires the exchange of signaling messages between the HLR and the new and old MSC's whenever the mobile terminal crosses an RA boundary. This may result in significant traffic load to the network especially when the current location of the mobile terminal is far away from its HLR and the mobile terminal is making frequent movements among RA's. Besides, the HLR may experience excessively high database access traffic as it is involved in every location registration and call delivery. This may result in increased connection set up delay during periods of high network utilization.

Associated problems: Since every location request as well as location registration are serviced through an HLR, it becomes overloaded. Due to this reason, traffic on the links leading to the HLR is heavy, which increases time required to establish connection to the MH. HLR is a single point of failure in the network as any HLR system failure causes all mobiles registered with HLR to be unreachable even though mobile host may be roaming and away from HLR region. The centralized approaches do suffer with an additional problem. Though the callee (mobile host to who is called) is located in the region near to the caller, its location information is needed to be obtained from HLR which can be placed over a long geographical distance. This is referred to as Tromboning problem. Various strategies have been suggested by researchers to optimize the centralized approach. The *Local Anchor (LA) Scheme*, proposed by Ho and Akyildiz [5], reduces the signaling traffic by using a local anchor (a VLR, a mobile user is currently visiting when he/she receives a call). In this scheme, a VLR close to the user is selected as the local anchor (LA) for the user. Whenever a user moves from one RA to another, it will perform location update to the LA. A LA for a mobile will change only when a call request to the mobile arrives; at the same time, the HLR is also updated via the registration process. When a call request terminating at this user is received by the HLR, the user can be traced to the LA. The LA scheme avoids update to HLR completely at the expense of the increase in local signaling traffic. The drawback of this scheme is that when the user keeps moving constantly without receiving any call, the updates to LA may become costly, a similar bottleneck as the HLR. For example, at the end of conferences or games, many people move away from one site without receiving calls, the LA for these people can become bottleneck. Another scheme, *Per-user Pointer Forwarding Scheme [6]* has been proposed by Jain and Lin. In this scheme, some updates to the HLR can be avoided by setting up a forwarding pointer from the previous VLR to the new VLR. When a call request to a mobile user arrives, the PCS network first queries the user's HLR to determine the VLR, which the user was visiting at the previous location update, then follows a chain of forwarding pointers to the user's current VLR to find the mobile user. The traffic to the HLR is decreased by using the pointer chain; however, the penalty is the time delay for tracking the user when a call to the user arrives. The longer the pointer chain, the less the signaling traffic, the longer the setup delay for finding the user. To avoid long setup delay, a threshold of the length of the pointer chain is used. The user needs to perform registration to the HLR after the chain threshold is reached. In order to overcome the drawbacks these two schemes, *two-level pointer forwarding strategy* was proposed by Ma & Fang [7]. Two kinds of pointers are used in this scheme. Some VLRs are selected as the Mobility Agents (MA), which will be responsible for location management in a larger area comparing to the RAs and can be geographically distributed.

Instead of always updating to the HLR, which would become the bottleneck otherwise, many location updates are carried out in the mobility agents. Thus, the two-level pointer forwarding scheme is designed to reduce the signaling traffic: pointers can be set up between VLRs as the traditional pointer forwarding scheme and can also be set up between MAs. The results has shown that this strategy can significantly reduce the network signaling traffic for users with low Call to Mobility Ratio (CMR)[6-7] without increasing much of the call setup delay. Ratnam, Matta and Rangrajan have suggested caching [10] for current standards. In existing HLR/VLR scheme, call is routed through MSC to LR in which callee is located. When a particular MSC receives a large number of calls to a particular mobile that belongs to a different home system, the signaling and database cost involved in setting up the call can be significantly reduced by caching the location information at the calling MSC. Each time when a call is attempted, the cached information is checked first. Since the access time in looking up an entry in the cached memory is very short (order of microseconds), checking the cached information for every call doesn't affect the performance of the MSC. In case of cache hit, call is directly routed to serving LR of callee. But in case of cache miss, HLR is needed to be contacted and call-establishing time will be longer than normal HLR/VLR scheme. Location information of MH is cached at an MSC if local call to mobility ratio (LCMR) maintained for the MH at the MSC is larger than a threshold derived from the link and database access cost of the network. LCMR is the ratio between the number of calls originating from an MSC to the number of times the MH changes its service area as seen by that MSC. A cached entry is invalidated after a threshold amount of time T since its last usage [10].

3. DISTRIBUTED APPROACHES

The approaches of this category can be further classified in the categories of fully and partially (hierarchically) distributed. A novel approach for efficient location management is proposed by Ratnam, Matta and Rangrajan [9] in which the idea is to fully distribute the location information across Location Registers (LRs). These LRs replace the centralized VLRs and HLRs which are found in current PCS networks.

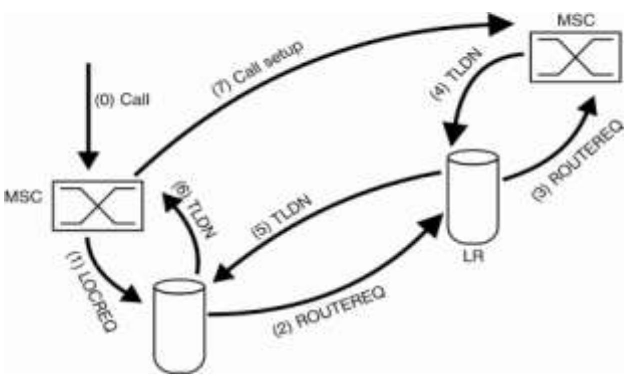

Figure 2: Call Delivery in Fully Distributed Location Management

Thus, this scheme proposes a unique feature of availability (fault-tolerance) by not having the concept of "home." Since there are no HLRs or VLRs in this system, each LR maintains the location information of not only the mobiles that are local to it, but also of other mobiles in the network, i.e., the location information of all mobile hosts are fully replicated in all the LRs. The LRs are distributed throughout the network. An LR may serve one or more MSCs just like the VLR in the PCS architecture. An LR/VLR could also co-exist with an MSC, and serves only that MSC. LRs function as both the location registry for the local mobile hosts as well as the lookup directory for the location of other mobile hosts. The type of location information maintained for a mobile host depends on whether the mobile is local to the LR or not. For local mobile hosts, LR maintains the id of the MSC that is currently serving the mobile. For mobile hosts that are not local, LR maintains the id of the LR where the mobile host currently resides. When a mobile registers with an LR, the new location information is disseminated to all other LRs in the network. This dissemination is carried out in parallel through the whole network so that the new location is very quickly updated at all LRs. When a call request arrives at the local LR, this LR can directly contact the serving LR (Figure 2), thus avoiding the tromboning problem present in the current IS-41 standard. The fully distributed location management scheme [9] requires that new location information about all mobiles be disseminated to all the LRs in the network. As the size of the network grows, location information dissemination not only consumes a significant portion of the network bandwidth but also consumes significant portion of LR resources to process large number of update messages. In addition, the gain of employing full dissemination diminishes with the size of the network. Full location information dissemination can be avoided by logically arranging LRs in a hierarchical fashion i.e. a tree structure. The idea is to divide the LRs into hierarchy of clusters, and confine location information dissemination to within the clusters as much as possible. In this scheme, the goal is not only to reduce the overhead of location management, but also to uniquely provide high availability through (selective) replication of location information.

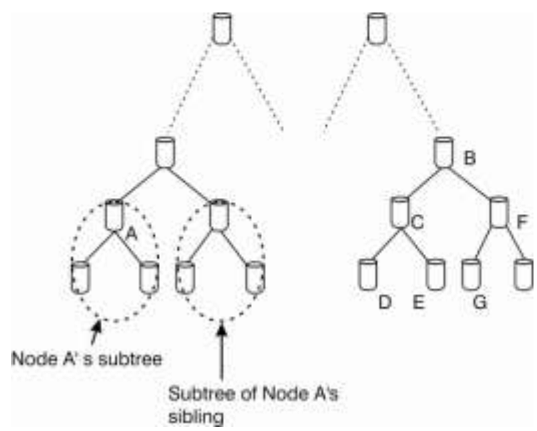

Figure 3: Conceptual diagram showing the hierarchical arrangement

Figure 3 shows the conceptual arrangement of the LRs in a hierarchical network. This approach uses a distributed location management. The mobile hosts are not associated with a home location register like in IS-41. Each LR maintains the location information of all the mobiles that are currently being served in the sub tree rooted from the LR. It also maintains the location of the mobiles that belong to the sub tree rooted from its sibling LRs. Note that the sub tree rooted from a leaf node contains only that leaf node. If a mobile host is being served by one of the descendants of an LR, then the LR maintains the ID of its immediate child LR, whose sub tree contains the mobile host, to track the mobile host.Tracking the LR serving a mobile host involves traversing the LR tree hop-by-hop until the serving LR is reached. If the location entry for a mobile host does not exist in an LR, then the tracking request is forwarded to the LR's parent LR. In this way the tracking request traverses the tree upwards until the LR which has the location information for the mobile host is reached. That LR forwards the tracking request to the LR pointed to by the location information. Here, location tracking traverses laterally. From there, it traverses downwards until the LR currently serving the mobile host is reached.

Figure 4 shows hierarchical location management structure discussed by Suh, Choi & Kim [10] having R layers. The hierarchically distributed servers have database for keeping information of mobile's location. The mobile enrolls in one of the layer1 servers called mobile's home server. To the ancestor servers of the mobile's home server, the mobile is regarded as home mobile. A server regards a mobile as foreign mobile which is located under server's coverage but whose home server is outside this server. The mobile identification (ID) is composed of hierarchical area address and terminal number. Server on layer 1 has location information about address of the parent server on layer 2 and address of mobile's residing LA and server on layer i (i=2,3….R) has location information about address of parent server on layer i+1, address of the server on layer i-1 under which the mobile is and address of the server on layer i-1 under which the mobile is. Consequently, information stored at the servers represents 'tracking path' for each mobile from its home server to its currently residing LA. The terminal movement from the old LA to the new LA triggers the location updates at the servers on hierarchical structure.

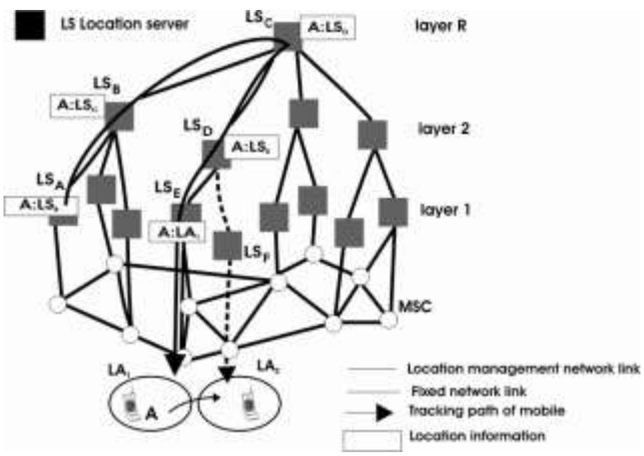

Figure 4 : Hierarchical location management system structure

For call delivery, there are different strategies depending on how to forward the call to the callee on hierarchical nodes. In first scheme referred to as *Home server first scheme* (HSF), the call is passed directly to callee's home server. This scheme is advantageous when the mobiles have a tendency of staying around their home server areas. Since mobile ID consists of hierarchical area address, the address of callee's home server can be easily found at caller's switch. In second scheme called *Least common ancestor* (LCA) server first scheme when a mobile-terminated call occurs, caller's switch does not figure out the switch of callee's current location. Thus caller's switch forwards the call to the appropriate ancestor server depending on how far it is away from callee's home. According to the LSF scheme, the call is passed directly to the LCA server of the caller's switch and callee's home server. The address of LCA server can be easily found by comparing the area addresses of caller's layer 1 location server and callee's home server. The third scheme called *Distributed LSF* scheme is based on 2 layered location management structure at which location servers are located. Layer 2 servers are logically connected together with each other as shown in figure 5. Mobile layer 2 ancestor server is called layer 2 home server and contain information like address of layer 2 server under which mobile is located, address of layer server under which the mobile is located at this moment and address of layer 1 server under which the mobile is located at this moment. The layer 1 server has same information as that of previous schemes. Thus when a mobile moves between LAs of the different layer 1 location servers, the location update occurs at the new layer 1 location server and the layer 2 location server(s).

According to distributed LSF strategy, when a call occurs, the call is forwarded to the closest home server of callee. In other words, when a call is initiated under callee's layer 1 home server, it is forwarded to callee's layer 1 home server. When a call is initiated under the area but callee's layer 1 home server's coverage, it is forwarded to layer 2 home server of the callee. Then, the call is forwarded to callee's current location server through the tracking path.

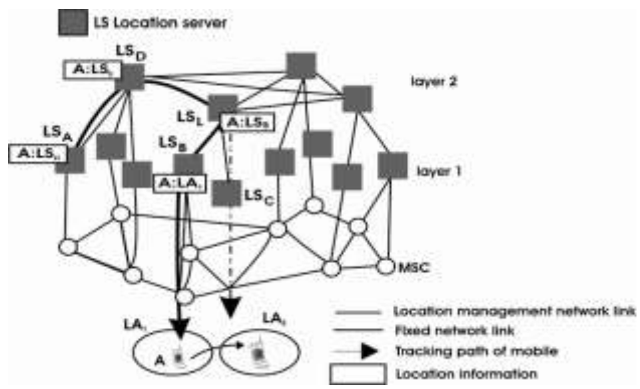

Figure 5: Call Delivery procedure of LSF scheme

4. CONCLUSION

As an outcome of above discussion, the need of binding a user to a HLR is not needed when hierarchical location databases are used. The user can be located by querying the databases in the hierarchy. In the worst case, an entry for the user will be found in the database at the root. The hierarchical scheme leads to reductions in communication cost when most calls and moves are geographically localized. In such cases, instead of contacting the HLR of the user that may be located away from the user's current location, a small number of location databases in the user's neighborhood are accessed. However, the number of location databases that are updated and queried increases relative to HLR/VLR scheme. Thus hierarchical scheme has advantage of no pre assigned HLR for user and support for locality. Instead, the hierarchical scheme has disadvantages of increased number of database operations and communication messages and increased load and storage requirements at higher levels. The schemes overviewed in above sections differ in many aspects such as their nature, analysis, location search strategy and update policies etc. It is clear that the centralized and fully distributed approaches have the poorer scalability than hierarchical approach. The centralized approach does suffer from the problem of fault tolerance as compare to distributed approach. Consideration of resource consumption as an performance metric concludes that centralized approach proves to be better if the movement of MU is local while in fully distributed approach irrespective to MU's movement, resource consumption is more. The resource consumption in hierarchical approach is proportional to the number of levels used in hierarchy. Limit in number of levels in hierarchy may lead to improvement in bandwidth utilization as well as reduction in unnecessary updates and deregistration messages.

5. REFERENCES

[1] EINTIA, "Cellular radio-telecommunications intersystem operations," Technical Report IS-41 Revision B, EIA/TIA, 1991.

[2] S. Mohan and R. Jain, "Two user location strategies for personal communications services," IEEE Personal Commun., pp. 42-50, First Quarter 1994.

[3] M. Mouly and M. B. Pautet, "The GSM system for mobile communications," Palaiseau, France, 1992.

[4] R. Jain, Y. B. Lin, and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, Oct. 1994.

[5] J. Ho and F. Akiyildiz, "Local Anchor Scheme for reducing Signaling Cost in Personal Communication Networks," ," IEEE*/ACM Transaction on Networking.*, vol. 4, no. 5, pp. 709-725 Oct. 1996.

[6] R. Jain and Y. B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impact of PCS," *ACM-Baker J. Wireless Network*, vol. 1, no. 2, pp. 197-210, July 1995.

[7] W. Ma, Y. Fang, "Two Level Pointer Forwarding Strategy for Location Management in PCS Networks," *IEEE/ACM Trans. On Mobile Computing*, vol. 1, no. 1, January - March 2002.

[8] S. Rangarajan, K. Ratnam, and A. T. Dahbura, "A fault-tolerant protocol for location directory maintenance in mobile networks," in Proc. of the 25th International Symposium on Fault-Tolerant Computing, June 1995.

[9] K. Ratnam, I. Matta, and S. Rangarajan, "A fully distributed location management scheme for large PCS," *Tech. Rep. BU-CS-1999-010*, Computer Science Department, Boston University, Boston, MA 02215, August 1999.

[10] K. Ratnam, I. Matta, and S. Rangarajan, "Analysis of caching-based location management in personal communication networks," in IEEE ICNP '99, 1999.

[11] Bong-sue Suh, Jin-seek Choi, Jae-kyoon Kim, "Design and performance of hierarchical location management strategies for wireless mobile communication system" Computer Communication, 23 (2000) 550-560.

[12] Anantharam, V., M.L. Honig, U. Madhow and V.K. Wei, 1994. Optimization of a database hierarchy for mobility tracking in a personal communications network. Performance Evaluation, 20: 287-300.

[13] Mao, Z. and C. Douligeris, 2004. A distributed database architecture for global roaming in next- generation mobile networks. *IEEE/ACM Transactions* on Networking, 12: 146-160.

[14] Wang, J.Z., 1993. A fully distributed location registration strategy for universal personal communication systems. IEEE J. Selected Areas Commun., 11: 850-860.

[15] Vijay Kumar, Mobile Databases, JOHN WILEY & SONS, INC., PUBLICATION.