Mobility Management for Cellular Telephony Networks

Yi-Bing Lin
National Chiao Tung University

Cellular telephony provides voice and data services to mobile users. This tutorial describes mobility management in cellular networks, including the details of different handoff schemes, location-tracking schemes, and call-delivery procedures.

In gaining more than 15 million subscribers over the past 10 years, cellular telephony services have become one of the telecommunications industry’s fastest-growing business sectors. These services facilitate the exchange of information (voice, data, video, image, and the like) for mobile users independent of time, location, and access arrangement. Their rapid growth demonstrates the great potential of cellular communication, which has made it the subject of widespread research interest.

In particular, cellular networks can serve as a general platform to build distributed computing applications (for example, mobile World Wide Web applications and database transaction systems) because they are typically connected to the Public Switched Telephone Network (PSTN) or Public Switched Data Network (PSDN), which makes the mobile services widely accessible to the public.

This tutorial discusses one of the most important issues in cellular communications: mobility management, or how to track the locations of the users and allow user movement during conversations.

Mobility management

A look at a cellular network’s architecture will help explain the issue. As Figure 1 shows, a set of base stations covers the cellular service area. The base stations serve the calls to or from the mobile phones (in the figure, the mobile phone is mounted on the vehicle) in their coverage areas (cells). Land links connect the base stations to mobile switching centers. An
tions, and initiates the handoff when some handoff criteria are met. In network-controlled handoff (for example, CT-2 Plus\(^6\) and AMPS, the Advanced Mobile Phone System\(^7\)), the surrounding base stations measure the signal from the mobile phone, and the network initiates the handoff when the criteria are met. In mobile phone-assisted handoff (for example, GSM), the network asks the mobile phone to measure the signal from the surrounding base stations. The network makes the handoff decision based on the report from the mobile phone.

The base stations involved in the handoff might be connected to the same MSC (inter-base station handoff) or two different MSCs (intersystem handoff or inter-MSC handoff).

**INTER-BASE STATION HANDOFF**
In the mobile phone-controlled strategy, inter-base station handoff occurs as follows (see Figure 2):

1. After the mobile phone detects the need for handoff, it temporarily suspends the conversation and initiates the handoff by signaling on an idle channel in the new base station. Then it resumes the conversation on the old base station.
2. On receipt of the signal, the switch transfers the encryption information to the selected idle chan-

---

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>Base station</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>IS-41</td>
<td>Interim Standard 41</td>
</tr>
<tr>
<td>GTT</td>
<td>Global title translation</td>
</tr>
<tr>
<td>HLR</td>
<td>Home location register</td>
</tr>
<tr>
<td>MIN</td>
<td>Mobile account number</td>
</tr>
<tr>
<td>MSC</td>
<td>Mobile switching center</td>
</tr>
<tr>
<td>PSTN</td>
<td>Public Switched Telephone Network</td>
</tr>
<tr>
<td>SCP</td>
<td>Service control point</td>
</tr>
<tr>
<td>SSP</td>
<td>Service switching point</td>
</tr>
<tr>
<td>SNC</td>
<td>Signaling System No. 7</td>
</tr>
<tr>
<td>STP</td>
<td>Signaling transfer point</td>
</tr>
<tr>
<td>VLR</td>
<td>Visitor location register</td>
</tr>
</tbody>
</table>

---

MSC is a telephone exchange specially assembled for mobile applications. It interfaces between the mobile phones (via base stations) and the PSTN.

Two aspects of mobility characterize a cellular telephony network:

- **Handoff** (small-scale mobility): When a mobile user is in a conversation, a radio link connects the mobile phone to a base station. If that user moves to another base station’s coverage area, the radio link to the old base station is disconnected, and a radio link to the new base station is required to continue the conversation. This process is called automatic link transfer or handoff.
- **Roaming** (large-scale mobility): When a mobile user moves from one location to another (for example, from New York City to Los Angeles), the system should be informed of the user’s current location. Otherwise, it is impossible to deliver the services to the mobile user.

Protocols such as the Electronic Industries Association/Telecommunications Industry Association (ELA/TLA) Interim Standard 41 (IS-41)\(^2\) and the Global System for Mobile Communications (GSM)\(^3\) support mobility management for cellular communications.

**Handoff**

Three strategies exist to detect the need for handoff. In mobile phone-controlled handoff (for example, the Digital European Cordless Telephone\(^4\) and Personal Access Communications System\(^5\)), the mobile phone continuously monitors the signal from the surrounding base sta-
nel of the new base station and sets up the new conversation path to the mobile phone through that channel. The switch bridges the new path with the old path and informs the mobile phone to transfer from the old channel to the new channel.

3. After the mobile phone has transferred to the new base station, it signals the network and resumes conversation by using the new channel.

4. On receipt of the handoff-completion signal, the network removes the bridge from the path and releases resources associated with the old channel.

For the network-controlled handoff strategy, switching to the new path and rerouting the carried information occur simultaneously, and the whole process must be completed as quickly as possible to reduce the conversation-interrupt period.

If the new base station does not have an idle channel, the handoff call might be dropped (forced terminated). The forced termination probability is an important criterion in the performance evaluation of a cellular telephony network. Forced termination of an ongoing call is less desirable than the blocking of a new call attempt.

In current cellular telephony systems, the base station handles a handoff call exactly the same as it does a new call attempt (that is, it blocks the handoff call immediately if no channel is available). This is called the nonprioritization scheme. To reduce forced termination and to promote call completion, three schemes have been proposed.

- The **reserved channel scheme** is similar to the nonprioritized scheme, except that each base station reserves a number of channels for handoff calls.
- The **queueing priority scheme** exploits the characteristic overlap of adjacent coverage areas of base stations. This overlap forms a considerable area where either base station can handle a call. This is the **handoff area**. If no channel is available in the new base station during handoff, the new base station buffers the handoff request in a **waiting queue**. The mobile phone continues to use the channel with the old base station until either a channel in the new base station becomes available (and the handoff call is connected) or the mobile phone moves out of the handoff area (and the handoff call is forced terminated).
- The **subrating scheme** creates a new channel for a handoff call by subrating an existing call if no channel is available in the new base station. Subrating temporarily divides an occupied full-rate channel into two channels at half the original rate: one serves the existing call, and the other serves the handoff request. When the occupied channels are released, the subrate channels switch back to the full-rate channel immediately.

The last two schemes have not been implemented in existing cellular telephony networks. However, studies indicate that under certain conditions, these handoff schemes can significantly reduce the forced termination probability as well as the call incompletion (new-call blocking plus handoff-call forced termination) probability.

**INTERSYSTEM HANDOFF**

Our description of the intersystem handoff follows IS-41 (GSM follows similar procedures), and we assume network-controlled handoff. Figure 3 illustrates the trunk (voice or data circuit) connection before and after the handoff. A communicating mobile user moves out of the base station served by MSC1 and enters the area covered by MSC2. The handoff follows these steps:

1. MSC1 requests MSC2 to perform handoff measurement. MSC2 then selects a candidate base station, BS2, for handoff. That is, MSC2 finds a base station that covers the mobile phone and has a free radio channel to cover the call. MSC2 returns the signal-quality parameter values and other information to MSC1.

2. MSC1 checks if the mobile phone has made too many handoffs or if intersystem trunks are not available. If so, MSC1 exits the procedure. Otherwise, MSC1 asks MSC2 to set up a voice channel. Suppose that a voice channel is available in BS2, MSC2 asks MSC1 to start the radio link transfer.

3. MSC1 sends the mobile phone a handoff order. The mobile phone tries to synchronize to BS2. After the mobile phone connects to BS2, MSC2 informs...
MSC₁ that the handoff is successful. MSC₁ then connects the call path (trunk) to MSC₂ and completes the handoff.

In this process, MSC₁ is called the anchor MSC, and is always in the call path after the handoff (see Figure 4a). If the mobile phone moves back to MSC₁ again, the connection between MSC₁ and MSC₂ terminates (see Figure 4b). If the mobile phone moves to a third MSC (MSC₃), MSC₂ will be in the call path (see Figure 4c). When the mobile phone moves to the third MSC, the second MSC might be removed from the call path (that is, the link between MSC₂ and MSC₁ terminates, and MSC₁ connects to MSC₁ directly, as Figure 4d shows). This process is called path minimization.

**Roaming**

Two basic operations in roaming management are registration (a mobile phone informs the system of its current location) and location tracking (the system locates the mobile phone). Location tracking is required when the network attempts to deliver a call to the mobile user.

Our description of roaming management is based on IS-41. GSM shares the same roaming management concept, with more emphasis on security provisions. (The upcoming release of IS-41 has enhanced the security aspect.)

Both protocols propose roaming-management strategies that use a two-tier system of databases. When a user subscribes to the services of a cellular telephony system, it becomes the user’s home system. A subscription record
is created in the system's database called the home location register. Mobile phone identities are assigned to the HLR for record purposes such as mobile user information (for example, directory number, profile information, current location, and validation period). When the mobile user visits a different cellular service area (a visited system), a temporary record for the mobile user is created in that system's Visitor Location Register. The VLR is used to retrieve information for handling calls to or from a visiting mobile user.

Suppose that a mobile user's home system is at Morristown, New Jersey. When the mobile user moves from one visited system to another (for example, from New York City to Los Angeles), the user must register in the VLR of the new visited system (Step 1 in Figure 5). The new VLR then informs the mobile user's HLR of its current location (that is, the new VLR's address). The HLR sends an acknowledgment to the new VLR (Step 2 in Figure 5). The new VLR then informs the mobile phone of the successful registration (Step 3 in Figure 5). After Step 2, the HLR also sends a deregistration message to cancel the obsolete location record in the old VLR. The old VLR acknowledges the deregistration (Step 4 in Figure 5).

To originate a call, the mobile phone first contacts an MSC in the cellular telephony network. The call request is forwarded to the VLR, and is eventually connected to the called party through the PSTN.

If a wireline phone attempts to call a mobile subscriber, the call is forwarded to a switch (called the originating switch) in the PSTN. The originating switch queries the HLR to find the mobile phone's current VLR (Step 1 in Figure 6). The VLR returns a routable address to the originating switch through the HLR (Step 2 in Figure 6). Based on that address, a trunk is then set up from the originating switch to the mobile phone through the visited MSC (Step 3 in Figure 6).

This abstract view of roaming management is missing one part: the interactions between the mobile communications network and the PSTN. We'll now describe these interactions.

**The Interconnection Between the Cellular Telephony Network and the PSTN**

Typically, a Signaling System No. 7 network connects the cellular telephony network and the PSTN. Figure 7 shows the network elements that are involved in this interconnection. SS7 is an out-of-band signaling method, which separates the signaling (call setup, call release, and so on) network from the trunks. In Figure 7, the dashed lines represent the signaling links, and the solid line represents a trunk. The SS7 network has three distinct components:

- A service switching point is a telephone switch interconnected by SS7 links. The SSPs perform call
processing on calls that originate, tandem (pass through), or terminate at that SSP. The MSC is an SSP with specific functions for cellular communications.

- A service control point contains databases for providing enhanced services. An SCP accepts queries from an SSP and returns the requested information to the SSP. An example of an SCP is the 800-service database. In mobile applications, an SCP might contain an HLR or a VLR.

- A signal transfer point is a switch that relays SS7 messages between SSPs and SCPs. Based on the address fields of the SS7 messages, the STPs route the messages to the appropriate outgoing signaling links. To meet the stringent reliability requirements, the network uses pairs of STPs, as Figure 7 shows.

In this network, the trunks connect SSPs to carry data and voice information. The signaling links connect SCPs to STPs, and STPs to SSPs (SSPs and SCPs connect indirectly through STPs).

**REGISTRATION**

Figure 8 illustrates registration through the SS7 network. In this example, the mobile phone moves from VLR₁ to VLR₂.

1. MSC₂ launches a registration query to its VLR through STP₂ (assume that VLR₂ and MSC₂ are not collocated).
2. VLR₂ sends a registration message to HLR₄, the mobile phone’s HLR. VLR₂ might not know the actual address of HLR₄. Instead, VLR₂ sends the message with the mobile phone’s identity (the mobile identification number) to an STP (STP₃ in our example) that can translate the MIN into the HLR address.
3. STP₃ performs the MIN-to-HLR address translation by a table-lookup technique called global title translation (GTT). STP₃ then forwards the registration message to HLR₄.
4. After the registration, HLR₄ sends an acknowledgment back to VLR₂. Because VLR₂’s address is known, the acknowledgment can use a shortcut, bypassing STP₃.
5. After Step 3, HLR₄ sends a deregistration message to VLR₁ to cancel the obsolete record.

In Steps 2 through 5, the messages might visit several STPs before arriving at their destinations, and the registration process might generate heavy traffic in the SS7 network. So, reducing the registration traffic is desirable.

Two approaches have been proposed to reduce the cost of deregistration. Implicit deregistration does not delete obsolete VLR records until the database is full. If the database is full when a mobile phone arrives, the VLR deletes a record and reclaims the storage to accommodate the newly arrived phone. This approach requires a policy for selecting the record to replace. Otherwise, a current record might be replaced, and that information lost. This approach’s major advantage is that it sends no deregistration messages among the SS7 network elements.
In periodic re-registration, the mobile phone periodically re-registers to the VLR. If the VLR does not receive the re-registration message within a time-out period, it deletes the record. This approach creates only local message traffic between the MSC and the VLR. Furthermore, it generates no messages if the VLR and MSC are collocated.

To reduce registration traffic, a pointer-forwarding scheme has been proposed (see Figure 9). When a mobile phone moves from one VLR to another, this scheme creates a pointer from the old VLR to the new VLR (the move operation—see Figure 9a). No registration to the HLR is required. When the HLR attempts to locate the mobile phone for call delivery, the HLR traces the pointer chain (the find operation—see Figure 9b). After the find operation, the HLR points directly to the destination VLR.

Depending on the memory capacities of the VLRs, the pointers in the obsolete chain might or might not be deleted. To limit the pointer traversal time in the find operation, this scheme can perform the registration procedure in Figure 8 for every \( k \) move operations. In other words, the number of pointers visited in the find operation will be limited by \( k \). The pointer-forwarding scheme is inappropriate when the net cost of pointer creation and pointer traversal is greater than the cost of accessing the HLR. This scheme has not been implemented in an existing mobile communication network, but performance studies indicate that it can significantly reduce network traffic in many cases.

**Call Delivery**

Similar to the registration process in Figure 8, the call-delivery process might require several STPs and a GTT to access the HLR (Step 1 in Figure 10). The process might visit several STPs to obtain the routable address from the VLR (Steps 2 and 3 in Figure 10).

To reduce the call-delivery traffic, one proposed scheme maintains a cache in the SSPs (see Figure 11). Another possibility is to maintain the cache in the STP that performs GTTs (STP, in our example). A cache entry consists of two fields: the mobile phone's MIN and the address of the phone's current VLR. The cache contains entries for mobile phones recently accessed from the SSP. When the calling party originates a call to a mobile phone, the SSP first checks if the cache entry for the mobile phone exists. There are three possibilities:

- **Case 1.** The cache entry does not exist. The call-delivery procedure in Figure 10 is performed.
- **Case 2.** The cache entry exists and is current. The SSP directly accesses the VLR, as shown in Figure 11.
- **Case 3.** The cache entry exists but is obsolete. The procedure detects that the cache entry is obsolete.
if the queried VLR's response is negative. The call-delivery procedure in Figure 10 is performed. The cache scheme can use implicit deregistration and periodic re-registration, but might not detect the obsolete cache information until it pages the mobile phone.

Because the cache information might be obsolete, this scheme requires heuristics to determine whether it will use the information to locate the mobile phone. In one technique, the SSP estimates the probability that Case 2 is true (the cache-hit ratio). If the probability is high, the SSP considers the entry current and enables it. Otherwise, the SSP disables the entry.

Another heuristic determines an entry's obsoleteness based on the period that a mobile phone resides in a VLR (as indicated in the cache entry). If the cache entry indicates that the mobile phone has stayed in a VLR for a period longer than a threshold, the SSP assumes that the entry is obsolete. The SSP can adjust the threshold in real time based on the cache-hit statistics. If Case 3 is more likely to occur than Case 2, the cache scheme is inappropriate. Although this scheme has not been implemented in an existing network, performance studies indicate that it can significantly reduce the call delivery cost in many examples.

Figure 12 provides a crossword puzzle to refresh the reader's memory of the terms described in this article. Figure 13 gives the solution.

Many distributed applications (for example, distributed file systems and distributed location-tracking systems) can be built on top of cellular networks. Cellular telephony provides user mobility for wide-area-network distributed applications. However, cellular telephony is circuit-switching-oriented. In some cases,
packet-switching environments might fit distributed systems better. So, an important research direction for distributed mobile systems is to integrate the infrastructure of packet-switching data networks such as the Internet (that is, mobile Internet Protocol[3]) and cellular telephony. With this combination, a distributed application can use cellular telephony’s highly available wireless resources and mobility databases, while transporting information through an efficient packet-switching data network.

Figure 13. The crossword puzzle solution.

ACKNOWLEDGMENTS
I thank Ravi Jain for his comments on the early draft of this article. This work was supported in part by National Science Council, R.O.C., under Contract NSC 86-2213-E-009-074.

REFERENCES

Yi-Bing Lin is the deputy director of the Microelectronics and Information Systems Research Center at National Chiao Tung University. His research interests include the design and analysis of personal communications services networks, mobile computing, distributed simulation, and performance modeling. He is on the editorial board of Int'l J. Communications and Computer Simulation Modeling and Analysis. He is an associate editor of ACM Trans. Modeling and Computer Simulation, the International J. Computer Simulation, IEEE Networks, and Simulation. He is also a subject-area editor for J. Parallel and Distributed Computing and ACM Mobile Computing and Communication Rev., and a columnist for ACM Simulation Digest. He is the guest editor for the ACM/Baltzer Winet (wireless network) special issue on personal communications, and the IEEE Trans. Computers special issue on mobile computing. Lin is a senior member of the IEEE. He received his BSEE from National Cheng Kung University in 1983, and his PhD in computer science from the University of Washington in 1990. Contact him at the Dept. of Computer Science & Information Engineering, National Chiao Tung Univ., Hsinchu, Taiwan, R.O.C., liny@csie.nctu.edu.tw.