Module – 4 **10**

Mobile and Multimedia Networks: Cellular Internet Access: An Overview of Cellular Network Architecture, 3G Cellular Data Networks: Extending the Internet to Cellular subscribers, On to 4G:LTE,Mobility management: Principles, Addressing, Routing to a mobile node, Mobile IP, Managing mobility in cellular Networks, Routing calls to a Mobile user, Handoffs in GSM, Wireless and Mobility: Impact on Higher-layer protocols.

6.4.1 **An Overview of Cellular Network Architecture**

- In the description of cellular network architecture, the terminology of the *Global System for Mobile Communications (**GSM**)* standards is adopted.
- In the 1980s, Europeans recognized the need for a pan-European digital cellular telephony system that would replace the numerous incompatible analog cellular telephony systems, leading to the GSM standard .
- Europeans deployed GSM technology with great success in the early 1990s, and since then GSM has
- grown to be the 800-pound gorilla of the cellular telephone world, with more than 80% of all cellular subscribers worldwide using GSM.
- Cellular technology, is often classified the as belonging to one of several "generations."
- The earliest generations were designed primarily for voice traffic.
- First generation (1G) systems were analog FDMA systems designed exclusively for voice-only communication.
- These 1G systems are almost extinct now, having been replaced by digital 2G systems. The original 2G systems were also designed for voice, but later extended (2.5G) to support data (i.e., Internet) as well as voice service.
- The 3G systems that currently are being deployed also support voice and data, but with an ever increasing emphasis on data capabilities and higher-speed radio access links.

Cellular Network Architecture, 2G: Voice Connections to the Telephone Network

- The term *cellular* refers to the fact that the region covered by a cellular network is partitioned into a number of geographic coverage areas, known as **cells**, shown as hexagons on the left side of Figure 6.18.
- GSM has its own particular nomenclature.
- Each cell contains a **base transceiver station (BTS)** that transmits signals to and receives signals from the mobile stations in its cell.
- The coverage area of a cell depends on many factors,
  - including the transmitting power of the BTS,
  - the transmitting power of the user devices,
  - obstructing buildings in the cell,
  - and the height of base station antennas.

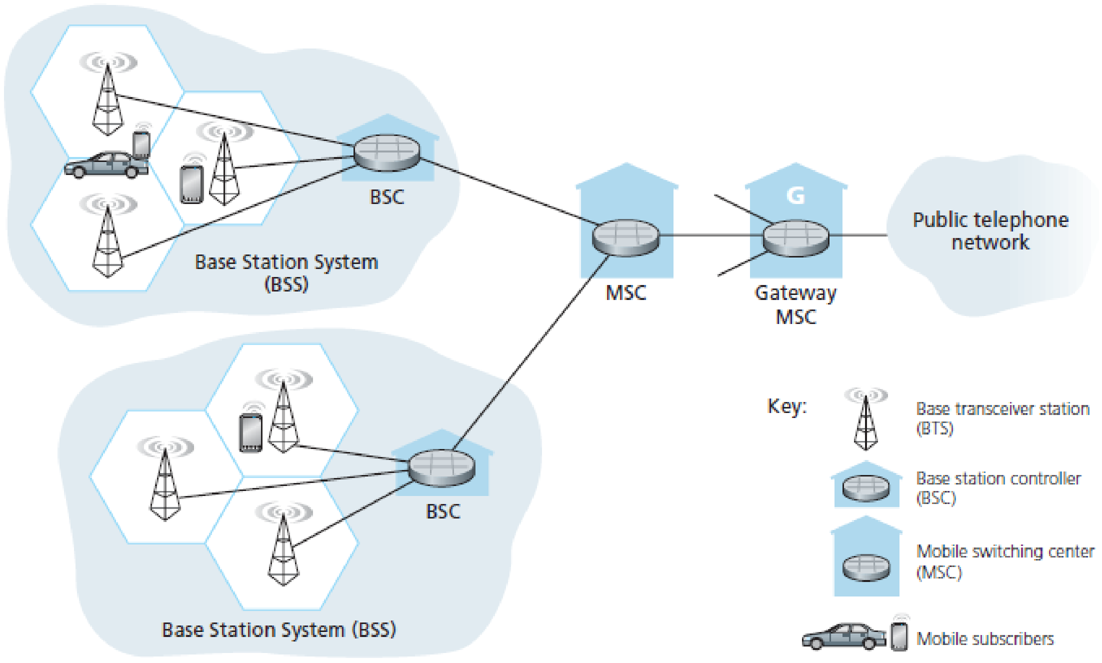


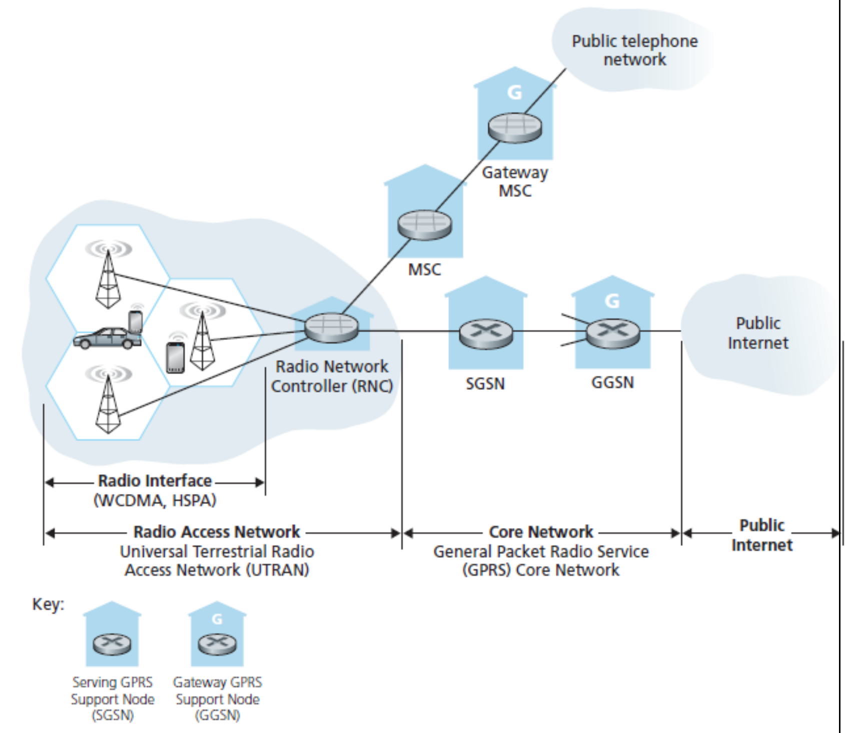
Figure 6.18 ♦ Components of the GSM 2G cellular network architecture

- Although Figure 6.18 shows each cell containing one base transceiver station residing in the middle of the cell, many systems today place the BTS at corners where three cells intersect, so that a single BTS with directional antennas can service three cells.
- The GSM standard for 2G cellular systems uses **combined FDM/TDM (radio)** for the **air interface**.
- With pure FDM, the channel is partitioned into a number of frequency bands with each band devoted to a call. Also with pure TDM, time is partitioned into frames with each frame further partitioned into slots and each call being assigned the use of a particular slot in the revolving frame.
- In combined FDM/TDM systems, the channel is partitioned into a number of frequency sub-bands; within each sub-band, time is partitioned into frames and slots.
- Thus, for a combined FDM/TDM system, if the channel is partitioned into *F* sub-bands and time is partitioned into *T* slots, then the channel will be able to support *F.T* simultaneous calls.
- The cable access networks also use a combined FDM/TDM approach.
- GSM systems consist of 200-kHz frequency bands with each band supporting eight TDM calls. GSM encodes speech at 13 kbps and 12.2 kbps.
- A GSM network’s **base station controller (BSC)** will typically service several tens of base transceiver stations.
- The role of the BSC is to allocate BTS radio channels to mobile subscribers, perform **paging** (finding the cell in which a mobile user is resident), and perform handoff of mobile users.
- The base station controller and its controlled base transceiver stations collectively constitute a GSM **base station system (BSS)**.
- The **mobile switching center (MSC)** plays the central role in user authorization and accounting (e.g., determining whether a mobile device is allowed to connect to the cellular network), call establishment and teardown, and handoff.
- A single MSC will typically contain up to five BSCs, resulting in approximately 200K subscribers per MSC. A cellular provider’s network will have a number of

MSCs, with special MSCs known as gateway MSCs connecting the provider's cellular network to the larger public telephone network.

6.4.2 **3G Cellular Data Networks: Extending the Internet to Cellular Subscribers**

The UMTS (Universal Mobile Telecommunications Service) 3G standards developed by the 3rd Generation Partnership project (3GPP) [3GPP 2012], a widely deployed 3G technology. Let's take a top-down look at 3G cellular data network architecture shown in Figure 6.19.



Public telephone network
G
Gateway MSC
MSC
Radio Network Controller (RNC)
SGSN
G
GGSN
Public Internet
Radio Interface (WCDMA, HSPA)
Radio Access Network
Universal Terrestrial Radio Access Network (UTRAN)
Core Network
General Packet Radio Service (GPRS) Core Network
Public Internet
Key:
G
Serving GPRS Support Node (SGSN)
Gateway GPRS Support Node (GGSN)

Figure 6.19 ♦ 3G system architecture

3G Core Network

The 3G core cellular data network connects radio access networks to the public Internet.

- The core network interoperates with components of the existing cellular voice network.
- Given the considerable amount of existing infrastructure in the existing cellular voice network, the approach taken by the designers of 3G data services is clear: leave the existing core GSM cellular voice network untouched, adding additional cellular data functionality in parallel to the existing cellular voice network.
- The alternative—integrating new data services directly into the core of the existing cellular voice network—would have raised the same challenges encountered earlier.
- There are two types of nodes in the 3G core network:
  - **Serving GPRS Support Nodes (SGSNs)**
  - **Gateway GPRS Support Nodes (GGSNs)**.
- (GPRS stands for Generalized Packet Radio Service, an early cellular data service in 2G networks; here we discuss the evolved version of GPRS in 3G networks).
- An SGSN is responsible for delivering datagrams to/from the mobile nodes in the radio access network to which the SGSN is attached.
- The SGSN interacts with the cellular voice network's MSC for that area, providing user authorization and handoff, maintaining location (cell) information about active mobile nodes, and performing datagram forwarding between mobile nodes in the radio access network and a GGSN.
- The GGSN acts as a gateway, connecting multiple SGSNs into the larger Internet.
- A GGSN is thus the last piece of 3G infrastructure that a datagram originating at a mobile node encounters before entering the larger Internet.
- To the outside world, the GGSN looks like any other gateway router; the mobility of the 3G nodes within the GGSN's network is hidden from the outside world behind the GGSN.

3G Radio Access Network: The Wireless Edge

- The 3G **radio access network** is the wireless first-hop network that we see as a 3G user.
- The **Radio Network Controller (RNC)** typically controls several cell base transceiver stations similar to the base stations that we encountered in 2G systems (but officially known in 3G UMTS parlance as a "Node Bs"—a rather non-descriptive name!).
- Each cell's wireless link operates between the mobile nodes and a base transceiver station, just as in 2G networks.
- The RNC connects to both the circuit-switched cellular voice network via an MSC, and to the packet-switched Internet via an SGSN. Thus, while 3G cellular voice and cellular data services use different core networks, they share a common first/last-hop radio access network.
- A significant change in 3G UMTS over 2G networks is that rather than using GSM's FDMA/TDMA scheme, UMTS uses a CDMA technique known as **Direct Sequence Wideband CDMA (DS-WCDMA)** within TDMA slots; TDMA slots, in turn, are available on multiple frequencies—an interesting use of all three dedicated channel-sharing approaches.
- This change requires a new 3G cellular wireless-access network operating in parallel with the 2G BSS radio network shown in Figure 6.19. The data service associated with the WCDMA specification is known as HSP (High Speed Packet Access) and promises downlink data rates of up to 14 Mbps.

6.4.3 **On to 4G: LTE**

With 3G systems now being deployed worldwide, can 4G systems be far behind?

The 4G Long-Term Evolution (LTE) standard put forward by the 3GPP has two important innovations over 3G systems:

• **Evolved Packet Core (EPC)**

- The EPC is a simplified all-IP core network that unifies the separate circuit-switched cellular voice network and the packet-switched cellular data network shown.
- It is an "all-IP" network in that both voice and data will be carried in IP datagrams.
- IP's "best effort" service model is not inherently well-suited to the stringent performance requirements of Voice-over-IP (VoIP) traffic unless network resources are carefully managed to avoid (rather than react to) congestion.
- Thus, a key task of the EPC is to manage network resources to provide this high quality of service. The EPC also makes a clear separation between the network control and user data planes, with many of the mobility support features that we will study in Section 6.7 being implemented in the control plane.
- The EPC allows multiple types of radio access networks, including legacy 2G and 3G radio access networks, to attach to the core network. Two very readable introductions to the EPC are [Motorola 2007; Alcatel-Lucent 2009].

• **LTE Radio Access Network.**

- LTE uses a combination of frequency division multiplexing and time division multiplexing on the downstream channel, known as orthogonal frequency division multiplexing (OFDM).

- (The term “orthogonal” comes from the fact the signals being sent on different frequency channels are created so that they interfere very little with each other, even when channel frequencies are tightly spaced).
- In LTE,each active mobile node is allocated one or more 0.5 ms time slots in one or more of the channel frequencies. Figure 6.20 shows an allocation of eight time slots over four frequencies. By being allocated increasingly more time slots (whether on the same frequency or on different frequencies), a mobile node is able to achieve increasingly higher transmission rates. Slot (re)allocation among mobile nodes can be performed as often as once every millisecond.
- Different modulation schemes can also be used to change the transmission rate; see the earlier discussion of Figure 6.3 and dynamic selection of modulation schemes in WiFi networks. Another innovation in the LTE radio network is the use of sophisticated multiple-input, multiple output (MIMO) antennas.
- The maximum data rate for an LTE user is 100 Mbps in the downstream direction and 50 Mbps in the upstream direction, when using 20 MHz worth of wireless spectrum.
- The particular allocation of time slots to mobile nodes is not mandated by the LTE standard. Instead, the decision of which mobile nodes will be allowed to transmit in a given time slot on a given frequency is determined by the scheduling algorithms provided by the LTE equipment vendor and/or the network operator.
- With opportunistic scheduling ,matching the physical-layer protocol to the channel conditions between the sender and receiver and choosing the receivers to which packets will be sent based on channel conditions allow the radio network controller to make best use of the wireless medium. In addition, user priorities and contracted levels of service (e.g., silver, gold, or platinum) can be used in scheduling downstream packet transmissions. In addition to the LTE capabilities described above, LTE-Advanced allows for downstream bandwidths of hundreds of Mbps by allocating aggregated channels to a mobile node.

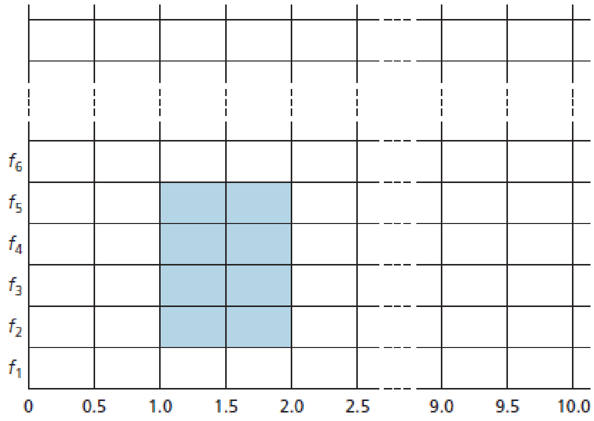


Figure 6.20 ♦ Twenty 0.5 ms slots organized into 10 ms frames at each frequency. An eight-slot allocation is shown shaded.

- An additional 4G wireless technology—WiMAX (World Interoperability for Microwave Access)—is a family of IEEE 802.16 standards that differ significantly from LTE. Whether LTE or WiMAX becomes the 4G technology of choice is still to be seen, but at the time of this writing (spring 2012), LTE appears to have significantly more momentum.

6.5 **Mobility Management: Principles**

In the broadest sense, a mobile node is one that changes its point of attachment into the network over time. Because the term *mobility* has taken on many meanings in both the computer and telephony worlds, it will serve us well first to consider several dimensions of mobility in some detail.

- *From the network layer’s standpoint, how mobile is a user?*
  - A physically mobile user will present a very different set of challenges to the network layer, depending on how he or she moves between points of attachment to the network.

  - At one end of the spectrum in Figure 6.21, a user may carry a laptop with a wireless network interface card around in a building. As we saw in Section 6.3.4, this user is *not* mobile from a network-layer perspective. Moreover, if the user associates with the same access point regardless of location, the user is not even mobile from the perspective of the link layer.
  - At the other end of the spectrum, consider the user zooming along the autobahn in a BMW at 150 kilometers per hour, passing through multiple wireless access networks and wanting to maintain an uninterrupted TCP connection to a remote application throughout the trip. This user is *definitely* mobile! In between these extremes is a user who takes a laptop from one location (e.g., office or dormitory) into another (e.g., coffeeshop, classroom) and wants to connect into the network in the new location. This user is also mobile (although less so than the BMW driver!) but does not need to maintain an ongoing connection while moving between points of attachment to the network. Figure 6.21 illustrates this spectrum of user mobility from the network layer's perspective.

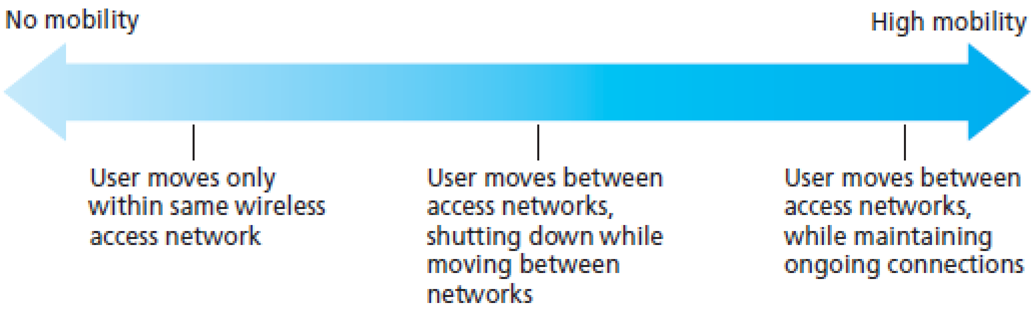

No mobility
High mobility
User moves only within same wireless access network
User moves between access networks, shutting down while moving between networks
User moves between access networks, while maintaining ongoing connections

Figure 6.21 ♦ Various degrees of mobility, from the network layer's point of view

- *How important is it for the mobile node's address to always remain the same?*
  - With mobile telephony, your phone number—essentially the network-layer address of your phone—remains the same as you travel from one provider's mobile phone network to another. Must a laptop similarly maintain the same IP address while moving between IP networks? The answer to this question will depend strongly on the applications being run. For the BMW driver who wants to maintain an uninterrupted TCP connection to a remote application while zipping along the autobahn, it would be convenient to maintain the same IP address. Recall from Chapter 3 that an Internet application needs to know the IP address and port number of the remote entity with which it is communicating.
  - If a mobile entity is able to maintain its IP address as it moves, mobility becomes invisible from the application standpoint. There is great value to this transparency—an application need not be concerned with a potentially changing IP address, and the same application code serves mobile and nonmobile connections alike. mobile IP provides this transparency, allowing a mobile node to maintain its permanent IP address while moving among networks.
  - On the other hand, a less glamorous mobile user might simply want to turn off an office laptop, bring that laptop home, power up, and work from home. If

the laptop functions primarily as a client in client-server applications (e.g., send/read e-mail, browse the Web, Telnet to a remote host) from home.

- *What supporting wired infrastructure is available?*
  - In all of our scenarios above, we've implicitly assumed that there is a fixed infrastructure to which the mobile user can connect—for example, the home's ISP network, the wireless access network in the office, or the wireless access networks lining the autobahn.
  - What if no such infrastructure exists? If two users are within communication proximity of each other, can they establish a network connection in the absence of any other network-layer infrastructure?
  - Ad hoc networking provides precisely these capabilities.This rapidly developing area is at the cutting edge of mobile networking research and is beyond the scope of this book. and the IETF Mobile Ad Hoc Network (manet) working group Web pages provide thorough treatments of the subject.

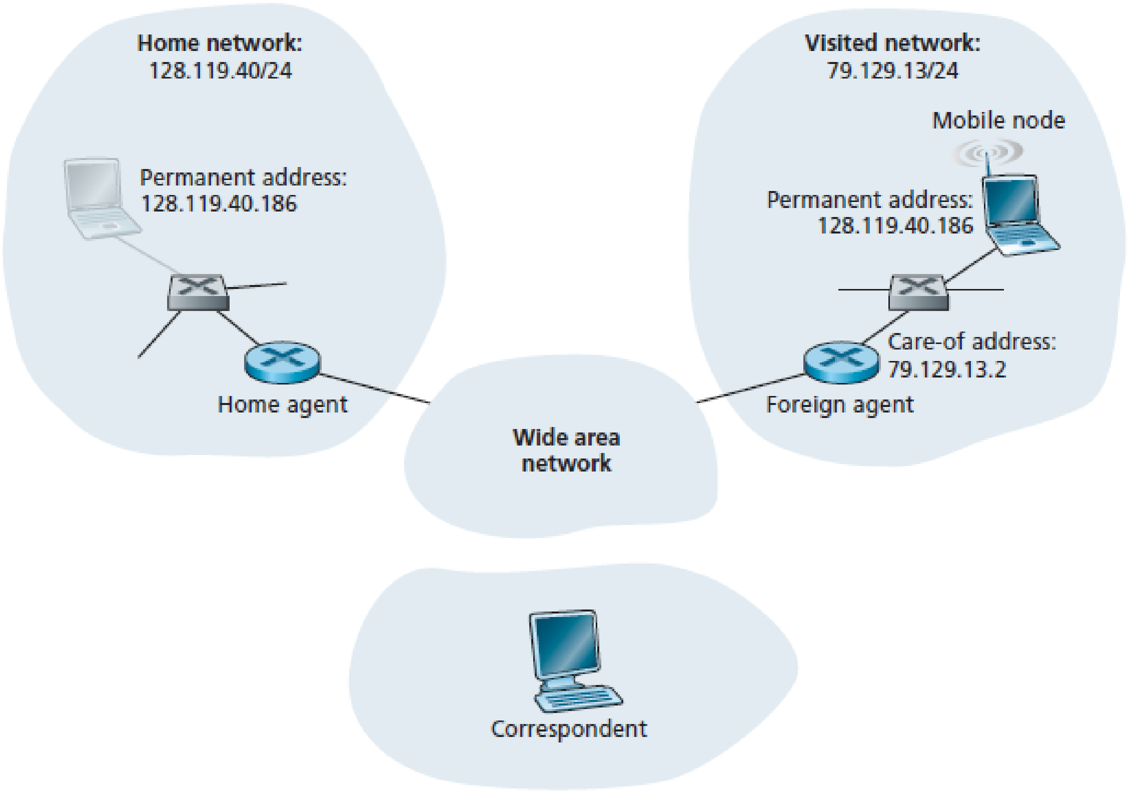


Figure 6.22 ♦ Initial elements of a mobile network architecture

- In order to illustrate the issues involved in allowing a mobile user to maintain ongoing connections while moving between networks, let's consider a human analogy. A twenty-something adult moving out of the family home becomes mobile, living in a series of dormitories and/or apartments, and often changing addresses. If an old friend wants to get in touch, how can that friend find the address of her mobile friend? One common way is to contact the family, since a mobile adult will often register his or her current address with the family .
- The family home, with its permanent address, becomes that one place that others can go as a first step in communicating with the mobile adult. Later communication from the friend may be either indirect (for example, with mail

being sent first to the parents' home and then forwarded to the mobile adult) or direct (for example, with the friend using the address obtained from the parents to send mail directly to her mobile friend).

- In a network setting, the permanent home of a mobile node (such as a laptop or smartphone) is known as the **home network**, and the entity within the home network that performs the mobility management functions discussed below on behalf of the mobile node is known as the **home agent**.
- The network in which the mobile node is currently residing is known as the **foreign** (or **visited**) **network**,and the entity within the foreign network that helps the mobile node with the mobility management functions discussed below is known as a **foreign agent**.
- For mobile professionals, their home network might likely be their company network,while the visited network might be the network of a colleague they are visiting.
- A **correspondent** is the entity wishing to communicate with the mobile node.Figure 6.22 illustrates these concepts, as well as addressing concepts consideredbelow. In Figure 6.22, note that agents are shown as being collocated with routers(e.g., as processes running on routers), but alternatively they could be executing on other hosts or servers in the network.

6.5.1 **Addressing**

- We noted above that in order for user mobility to be transparent to network applications, it is desirable for a mobile node to keep its address as it moves from one network to another.
- When a mobile node is resident in a foreign network, all traffic addressed to the node's permanent address now needs to be routed to the foreign network. How can this be done? One option is for the foreign network to advertise to all other networks that the mobile node is resident in its network.
- This could be via the usual exchange of intradomain and interdomain routing information and would require few changes to the existing routing infrastructure. The foreign network could simply advertise to its neighbors that it has a highly specific route to the mobile node's permanent address (that is, essentially inform other networks that it has the correct path for routing datagrams to the mobile node's permanent address; see Section 4.4).
- These neighbors would then propagate this routing information throughout the network as part of the normal procedure of updating routing information and forwarding tables. When the mobile node leaves one foreign network and joins another, the new foreign network would advertise a new, highly specific route to the mobile node, and the old foreign network would withdraw its routing information regarding the mobile node.

- This solves two problems at once, and it does so without making significant changes to the network layer infrastructure. Other networks know the location of the mobile node, and it is easy to route datagrams to the mobile node, since the forwarding tables will direct datagrams to the foreign network.
- A significant drawback, however, is that of scalability. If mobility management were to be the responsibility of network routers, the routers would have to maintain

forwarding table entries for potentially millions of mobile nodes, and update these entries as nodes move.

- An alternative approach (and one that has been adopted in practice) is to push mobility functionality from the network core to the network edge.A natural way to do this is via the mobile node's home network. In much the same way that parents of the mobile twenty somethingtrack their child's location, the home agent in the mobile node's home network can track the foreign network in which the mobile node resides.
- A protocol between the mobile node (or a foreign agent representing the mobile node) and the home agent will certainly be needed to update the mobile node's location.
- Let's now consider the foreign agent in more detail. The conceptually simplest approach, shown in Figure 6.22, is to locate foreign agents at the edge routers in the foreign network.
- One role of the foreign agent is to create a so-called **care-of address (COA)** for the mobile node, with the network portion of the COA matching that of the foreign network.
- There are thus two addresses associated with a mobile node, its **permanent address** (analogous to our mobile youth's family's home address) and its COA, sometimes known as a **foreign address** (analogous to the address of the house in which our mobile youth is currently residing).
- In the example in Figure 6.22, the permanent address of the mobile node is 128.119.40.186. When visiting network 79.129.13/24, the mobile node has a COA of 79.129.13.2. A second role of the foreign agent is to inform the home agent that the mobile node is resident in its (the foreign agent's) network and has the given COA.
- We'll see shortly that the COA will be used to "reroute" datagrams to the mobile node via its foreign agent.
- Although we have separated the functionality of the mobile node and the foreign agent, it is worth noting that the mobile node can also assume the responsibilities of the foreign agent. For example, the mobile node could obtain a COA in the foreign network (for example, using a protocol such as DHCP) and itself inform the home agent of its COA.

6.5.2 **Routing to a Mobile Node**

We have now seen how a mobile node obtains a COA and how the home agent can be informed of that address. But having the home agent know the COA solves only part of the problem. How should datagrams be addressed and forwarded to the mobile node? Since only the home agent (and not network-wide routers) knows the location of the mobile node, it will no longer suffice to simply address a datagram to the mobile node's permanent address and send it into the network-layer infrastructure.
Two approaches can be identified, which we will refer to as indirect and direct routing.

Indirect Routing to a Mobile Node

- Let's first consider a correspondent that wants to send a datagram to a mobile node.

- In the **indirect routing** approach, the correspondent simply addresses the datagram to the mobile node’s permanent address and sends the datagram into the network, unaware of whether the mobile node is resident in its home network or is visiting a foreign network; mobility is thus completely transparent to the correspondent.

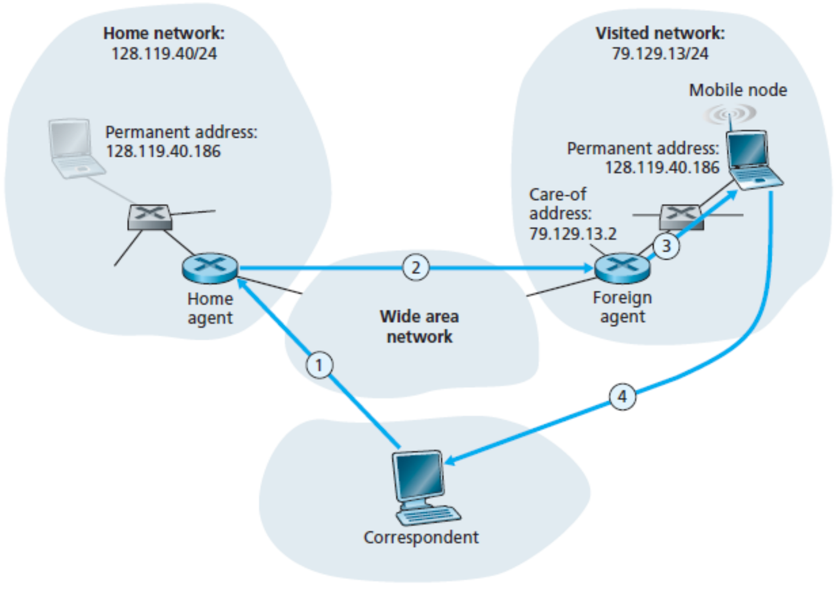


Figure 6.23 ♦ Indirect routing to a mobile node

- Such datagrams are first routed, as usual, to the mobile node’s home network.This is illustrated in step 1 in Figure 6.23.
- Let’s now turn our attention to the home agent. In addition to being responsible for interacting with a foreign agent to track the mobile node’s COA, the home agent has another very important function.
- Its second job is to be on the lookout for arriving datagrams addressed to nodes whose home network is that of the home agent but that are currently resident in a foreign network.
- The home agent intercepts these datagrams and then forwards them to a mobile node in a two-step process.
- The datagram is first forwarded to the foreign agent, using the mobile node’s COA (step 2 in Figure 6.23), and then forwarded from the foreign agent to the mobile node (step 3 in Figure 6.23).

- The home agent will need to address the datagram using the mobile node’s COA, so that the network layer will route the datagram to the foreign network.
- On the other hand, it is desirable to leave the correspondent’s datagram intact, since the application receiving the datagram should be unaware that the datagram was forwarded via the home agent.
- Both goals can be satisfied by having the home agent **encapsulate** the correspondent’s original complete datagram within a new (larger) datagram. This larger datagram is addressed and delivered to the mobile node’s COA.
- The foreign agent, who “owns” the COA, will receive and decapsulate the datagram—that is, remove the correspondent’s original datagram from within the larger encapsulating datagram and forward (step 3 in Figure 6.23) the original datagram to the mobile node.
- Figure 6.24 shows a correspondent’s original datagram being sent to the home network,an encapsulated datagram being sent to the foreign agent, and the original datagram being delivered to the mobile node.

- Note that the encapsulation/decapsulation described here is identical to the notion of tunneling, discussed in the context of IP multicast and IPv6.
- Let’s next consider how a mobile node sends datagrams to a correspondent.
- This is quite simple, as the mobile node can address its datagram *directly* to the correspondent (using its own permanent address as the source address, and the correspondent’s address as the destination address). Since the mobile node knows the correspondent’s address, there is no need to route the datagram back through the home agent. This is shown as step 4 in Figure 6.23.

Let’s summarize our discussion of indirect routing by listing the new network layer functionality required to support mobility.

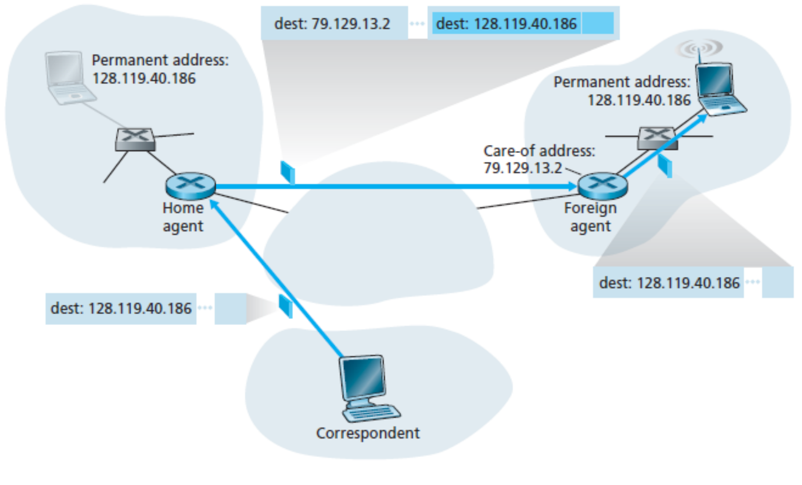


Figure 6.24 ♦ Encapsulation and decapsulation

- *A mobile-node–to–foreign-agent protocol*. The mobile node will register with the foreign agent when attaching to the foreign network. Similarly, a mobile node will deregister with the foreign agent when it leaves the foreign network.
- *A foreign-agent–to–home-agent registration protocol*. The foreign agent will register the mobile node’s COA with the home agent. A foreign agent need not explicitly deregister a COA when a mobile node leaves its network, because the subsequent registration of a new COA, when the mobile node moves to a new network, will take care of this.
- *A home-agent datagram encapsulation protocol*. Encapsulation and forwarding of the correspondent’s original datagram within a datagram addressed to the COA.
- *A foreign-agent decapsulation protocol*. Extraction of the correspondent’s original datagram from the encapsulating datagram, and the forwarding of the original datagram to the mobile node.

- As an example of how these pieces fit together, assume the mobile node is attached to foreign network A, has registered a COA in network A with its home agent, and is receiving datagrams that are being indirectly routed through its home agent.
- The mobile node now moves to foreign network B and registers with the foreign agent in network B,which informs the home agent of the mobile node’s new COA.
- From this point on, the home agent will reroute datagrams to foreign network B. As far as a correspondent is concerned, mobility is transparent—datagrams are routed via the same home agent both before and after the move.
- As far as the home agent is concerned, there is no disruption in the flow of datagrams—arriving datagrams are first forwarded to foreign network A; after the change in COA, datagrams are forwarded to foreign network B.
- But will the mobile node see an interrupted flow of datagrams as it moves between networks? As long as the time between the mobile node’s disconnection from network A (at which point it can no longer receive datagrams via A) and its attachment to

network B (at which point it will register a new COA with its home agent) is small, few datagrams will be lost.

- Recall that end-to-end connections can suffer datagram loss due to network congestion. Hence occasional datagram loss within a connection when a node moves between networks is by no means a catastrophic problem. If loss-free communication is required, upper-layer mechanisms will recover from datagram loss, whether such loss results from network congestion or from user mobility.
- An indirect routing approach is used in the mobile IP standard .

The indirect routing approach illustrated in Figure 6.23 suffers from an inefficiency known as the **triangle routing problem**—datagrams addressed to the mobile node must be routed first to the home agent and then to the foreign network, even when a much more efficient route exists between the correspondent and the mobile node. In the worst case, imagine a mobile user who is visiting the foreign network of a colleague. The two are sitting side by side and exchanging data over the network. Datagrams from the correspondent (in this case the colleague of the visitor) are routed to
the mobile user’s home agent and then back again to the foreign network!

Direct Routing to a Mobile Node

- **Direct routing** overcomes the inefficiency of triangle routing, but does so at the cost of additional complexity.
- In the direct routing approach, a **correspondent agent** in the correspondent’s network first learns the COA of the mobile node.
- This can be done by having the correspondent agent query the home agent, assuming that (as in the case of indirect routing) the mobile node has an up-to-date value for its COA registered with its home agent.
- It is also possible for the correspondent itself to perform the function of the correspondent agent, just as a mobile node could perform the function of the foreign agent. This is shown as steps 1 and 2 in Figure 6.25.
- The correspondent agent then tunnels datagram’s directly to the mobile node’s COA, in a manner analogous to the tunnelling performed by the home agent, steps 3 and 4 in Figure 6.25.
- While direct routing overcomes the triangle routing problem, it introduces two important additional challenges:
  - • A **mobile-user location protocol** is needed for the correspondent agent to query the home agent to obtain the mobile node’s COA (steps 1 and 2 in Figure 6.25).
  - • When the mobile node moves from one foreign network to another, how will data now be forwarded to the new

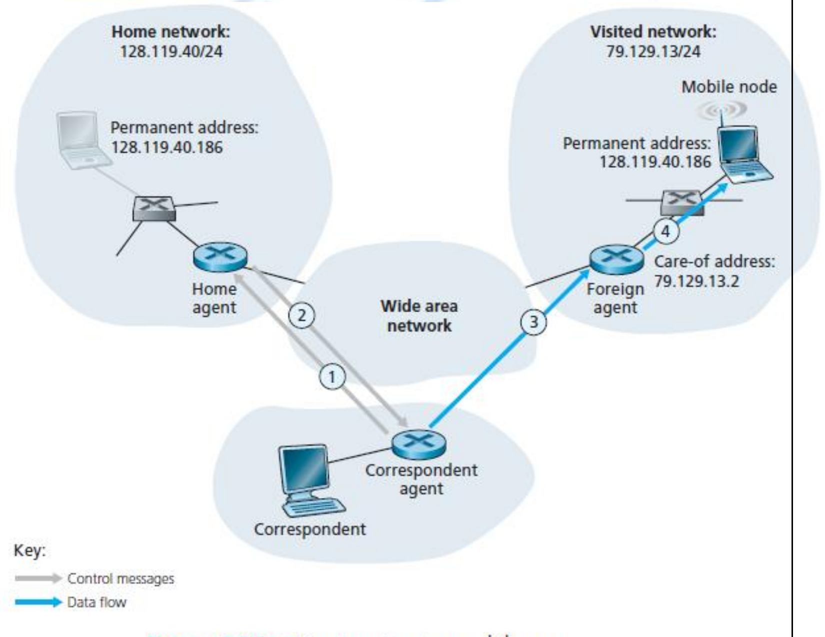


Figure 6.25 ♦ Direct routing to a mobile user

foreign network?

- In the case of indirect routing, this problem was easily solved by updating the COA maintained by the home agent. However, with direct routing, the home agent is queried for the COA by the correspondent agent only once, at the beginning of the session. Thus, updating the COA at the home agent, while necessary, will not be enough to solve the problem of routing data to the mobile node's new foreign network.
- One solution would be to create a new protocol to notify the correspondent of the changing COA.
- An alternate solution, and one that is adopted in practice in GSM networks, works as follows. Suppose data is currently being forwarded to the mobile node in the foreign network where the mobile node was located when the session first started (step 1 in Figure 6.26). The foreign agent in that foreign network where the mobile node was first found as the **anchor foreign agent**.
- When the mobile node moves to a new foreign network (step 2 in Figure 6.26), the mobile node registers with the new foreign agent (step 3), and the new foreign agent provides the anchor foreign agent with the mobile node's new COA (step 4).
- When the anchor foreign agent receives an encapsulated datagram for a departed mobile node, it can then re-encapsulate the datagram and forward it to the mobile node (step 5) using the new COA.
- If the mobile node later moves yet again to a new foreign network, the foreign agent in that new visited network would then contact the anchor foreign agent in order to set up forwarding to this new foreign network.

6.6 **Mobile IP**

The Internet architecture and protocols for supporting mobility, collectively known as mobile IP, are defined for IPv4.

Mobile IP is a flexible standard, supporting many different modes of operation (for example, operation with or without a foreign agent), multiple ways for agents and mobile nodes to discover each other, use of single or multiple COAs, and multiple forms of encapsulation.

The mobile IP architecture contains many elements including home agents, foreign agents, care-of addresses and encapsulation/decapsulation.

The mobile IP standard consists of three main pieces:

• ***Agent discovery***. Mobile IP defines the protocols used by a home or foreign agent to advertise its services to mobile nodes, and protocols for mobile nodes to solicit the services of a foreign or home agent.

•***Registration with the home agent***. Mobile IP defines the protocols used by the mobile node and/or foreign agent to register and deregister COAs with a mobile node's home agent.

• ***Indirect routing of datagrams***. The standard also defines the manner in which datagrams are forwarded to mobile nodes by a home agent, including rules for forwarding datagrams, rules for handling error conditions, and several forms of encapsulation.

Security considerations are prominent throughout the mobile IP standard. For example, authentication of a mobile node is clearly needed to ensure that a malicious user does not register a bogus care-of address with a home agent, which could cause all datagrams addressed to an IP address to be redirected to the malicious user.

Agent Discovery

A mobile IP node arriving to a new network, whether attaching to a foreign network or returning to its home network, must learn the identity of the corresponding foreign or home agent.

Indeed it is the discovery of a new foreign agent, with a new network address, that allows the network layer in a mobile node to learn that it has moved into a new foreign network.

This process is known as **agent discovery**.

Agent discovery can be accomplished in one of two ways:

- Agent advertisement
- Agent solicitation.

With **agent advertisement**, a foreign or home agent advertises its services using an extension to the existing router discovery protocol.

The agent periodically broadcasts an ICMP message with a type field of 9 (router discovery)

on all links to which it is connected.

The router discovery message contains the IP address of the router (that is, the agent), thus allowing a mobile node to learn the agent's IP address.

The router discovery message also contains a mobility agent advertisement extension that contains additional information needed by the mobile node.

Among the more important fields in the extension are the following:

• *Home agent bit (H).* Indicates that the agent is a home agent for the network in which it resides.

• *Foreign agent bit (F).* Indicates that the agent is a foreign agent for the network in which it resides.

• *Registration required bit (R).* Indicates that a mobile user in this network *must* register with a foreign agent. In particular, a mobile user cannot obtain a care of address in the foreign network (for example, using DHCP) and assume the functionality of the foreign agent for itself, without registering with the foreign agent.

• *M, G encapsulation bits.* Indicate whether a form of encapsulation other than IP in- IP encapsulation will be used.

• *Care-of address (COA) fields.* A list of one or more care-of addresses provided by the foreign agent.

In the example given below, the COA will be associated with the foreign agent, who will receive datagrams sent to the COA and then forward them to the appropriate mobile node.

The mobile user will select one of these addresses as its COA when registering with its home agent.

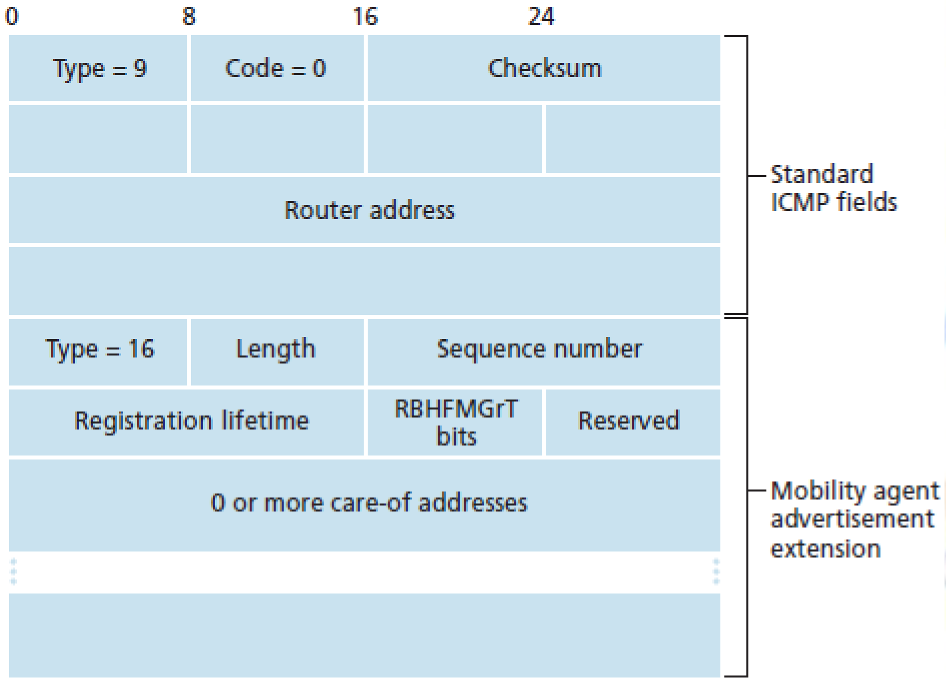
Figure below illustrates some of the key fields in the agent advertisement message.

Agent solicitation :

A mobile node wanting to learn about agents without waiting to receive an agent advertisement can broadcast an agent solicitation message.

Agent solicitation message is an ICMP message with type value 10.

An agent receiving the solicitation will unicast an agent advertisement directly to the mobile node, which can then proceed as if it had received an unsolicited advertisement.



Registration with the Home Agent

Once a mobile IP node has received a COA, that address must be registered with the home agent.

This can be done either via the foreign agent (who then registers the COA with the home agent) or directly by the mobile IP node itself.

Consider the former case below. Four steps are involved.

1.Following the receipt of a foreign agent advertisement:

A mobile node sends a mobile IP registration message to the foreign agent.

The registration message is carried within a UDP datagram and sent to port 434.

The registration message carries a COA advertised by the foreign agent, the address of the home agent (HA), the permanent address of the mobile node (MA), the requested lifetime of the registration, and a 64-bit registration identification.

The requested registration lifetime is the number of seconds that the registration is to be valid.

If the registration is not renewed at the home agent within the specified lifetime, the registration will become invalid.

The registration identifier acts like a sequence number and serves to match a received registration reply with a registration request.

2. The foreign agent receives the registration message & records the mobile node's permanent IP address.

The foreign agent now knows that it should be looking for datagrams containing an encapsulated datagram whose destination address matches the permanent address of the mobile node.

The foreign agent then sends a mobile IP registration message (again, within a UDP datagram) to port 434 of the home agent.

The message contains the COA, HA, MA, encapsulation format requested, requested registration lifetime, and registration identification.

3. The home agent receives the registration request and checks for authenticity and correctness.

The home agent binds the mobile node's permanent IP address with the COA;

In the future, datagrams arriving at the home agent and addressed to the mobile node will now be encapsulated and tunneled to the COA.

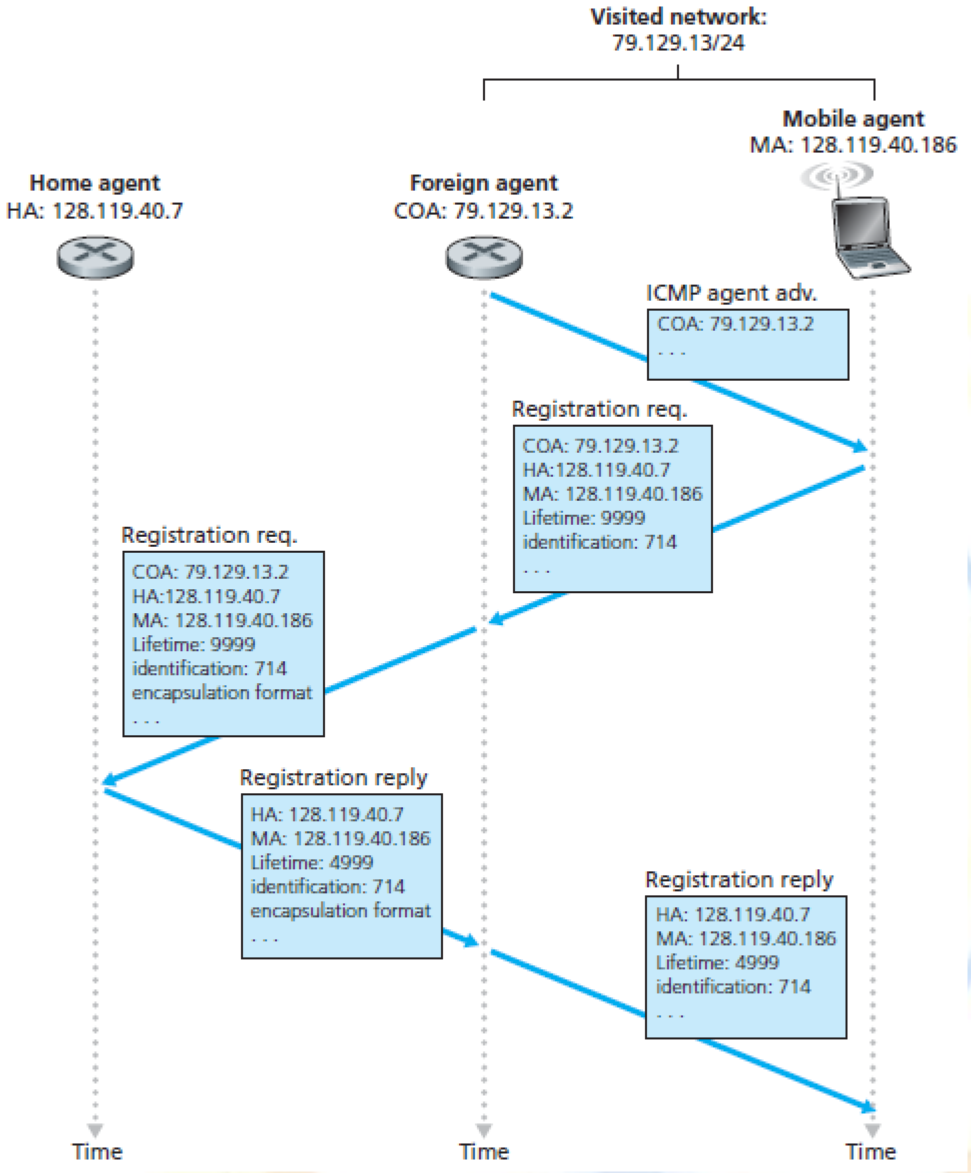
The home agent sends a mobile IP registration reply containing the HA,MA, actual registration lifetime and the registration identification of the request that is being satisfied with this reply.

4. The foreign agent receives the registration reply and then forwards it to the mobile node.

At this point, registration is complete and the mobile node can receive datagrams sent to its permanent address.

Figure below illustrates these steps.

Note that the home agent specifies a lifetime(4999) that is smaller than the lifetime requested by the mobile node(9999).



A foreign agent need not explicitly deregister a COA when a mobile node leaves its network.

This will occur automatically, when the mobile node moves to a new network (whether another foreign network or its home network) and registers a new COA.

6.7 **Managing Mobility in Cellular Networks**

GSM adopts an indirect routing approach.

First routing the correspondent’s call to the mobile user’s home network .

From home network to the visited network.

In GSM terminology, the mobile users's home network is referred to as the mobile user's **home public land mobile network (home PLMN)**.

The home network is the cellular provider with which the mobile user has a subscription (i.e., the provider that bills the user for monthly cellular service).

The visited PLMN referred as the **visited network**, is the network in which the mobile user is currently residing.

Following are the responsibilities of the home and visited networks:

• The home network maintains a database known as the **home location register** (**HLR**), which contains the permanent cell phone number and subscriber profile information for each of its subscribers.

Importantly, the HLR also contains information about the current locations of these subscribers.

If a mobile user is currently roaming in another provider's cellular network, the HLR contains enough information to obtain an address in the visited network to which a call to the mobile user should be routed.

A special switch in the home network, known as the **Gateway Mobile services Switching Center (GMSC)** is contacted by a correspondent when a call is placed to a mobile user.

GMSC here is referred as **home MSC**.

• The visited network maintains a database known as the **visitor location register (VLR)**.

The VLR contains an entry for each mobile user that is *currently* in the portion of the network served by the VLR.

VLR entries thus come and go as mobile users enter and leave the network.

A VLR is usually co-located with the mobile switching center (MSC) that coordinates the setup of a call to and from the visited network.

In practice, a provider's cellular network will serve as a home network for its subscribers and as a visited network for mobile users whose subscription is with a different cellular provider.

6.7.1 **Routing Calls to a Mobile User**

The following steps, as illustrated in Figure above, describes routing calls to a mobile user :

1. The correspondent dials the mobile user's phone number.

   This number itself does not refer to a particular telephone line or location .

   The leading digits in the number are sufficient to globally identify the mobile's home network.

The call is routed from the correspondent through the PSTN to the home MSC in the mobile's home network. This is the first leg of the call.

2. The home MSC receives the call and interrogates the HLR to determine the location of the mobile user.

   In the simplest case, the HLR returns the **mobile station roaming number (MSRN),** referred as the roaming number.

   Note that this number is different from the mobile's permanent phone number, which is associated with the mobile's home network.

   The roaming number is temporarily assigned to a mobile when it enters a visited network.

   The roaming number serves a role similar to that of the care-of address in mobile IP and, like the COA, is invisible to the correspondent and the mobile.

   If HLR does not have the roaming number, it returns the address of the VLR in the visited network.

   In this case, the home MSC will need to query the VLR to obtain the roaming number of the mobile node.

3. Given the roaming number, the home MSC sets up the second leg of the call through the network to the MSC in the visited network.

   The call is completed, being routed from the correspondent to the home MSC, and from there to the visited MSC, and from there to the base station serving the mobile user.

HLR obtains information about the location of the mobile user in the following way :

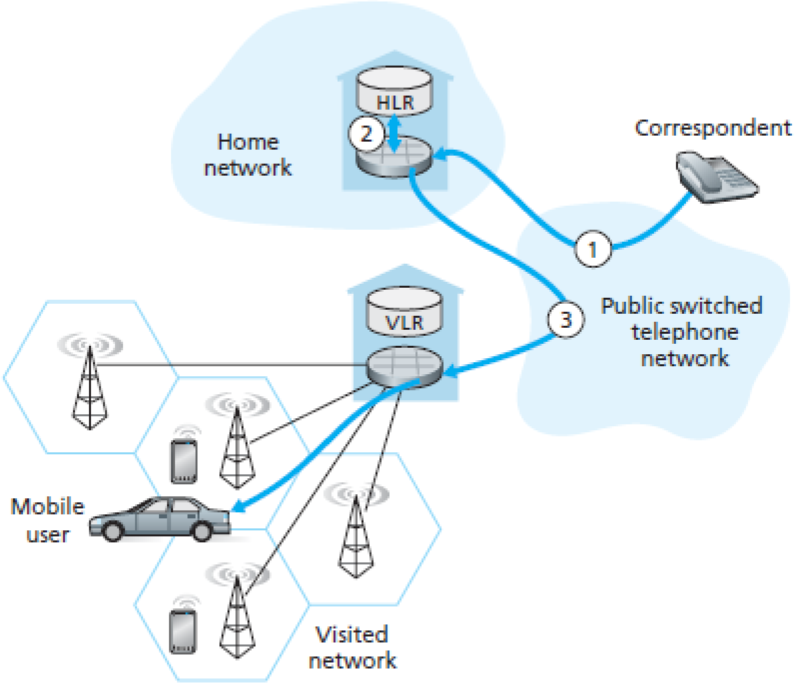
When a mobile telephone is switched on or enters a part of a visited network that is covered by a new VLR, the mobile must register with the visited network.

It is done through the exchange of signaling messages between the mobile and the VLR.

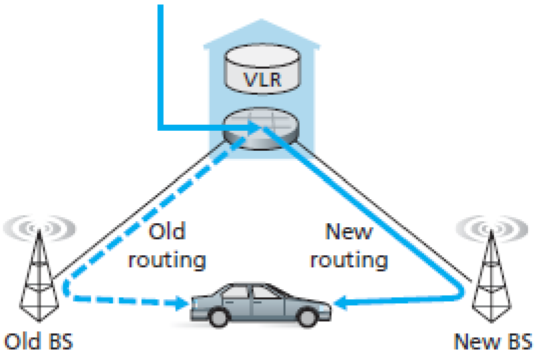
The visited VLR, in turn, sends a location update request message to the mobile's HLR.

The message informs the HLR of either the roaming number at which the mobile can be contacted, or the address of the VLR

As part of this exchange, the VLR also obtains subscriber information from the HLR about the mobile and determines what services (if any) should be accorded the mobile user by the visited network.



6.7.2 Handoffs in GSM



A **handoff** occurs when a mobile station changes its association from one base station to another during a call.

As shown in Figure above, a mobile's call is initially (before handoff) routed to the mobile through one base station (referred as the old base station) and after handoff is routed to the mobile through another base station (referred as the new base station).

A handoff between base stations results not only in the mobile transmitting/receiving to/from a new base station, but also in the rerouting of the ongoing call from a switching point within the network to the new base station.

Initially assume that the old and new base stations share the same MSC, and that the rerouting occurs at this MSC.

There may be several reasons for handoff to occur, including :

(1) the signal between the current base station and the mobile may have deteriorated to such an extent that the call is in danger of being dropped.

(2) a cell may have become overloaded, handling a large number of calls.

This congestion may be alleviated by handing off mobiles to less congested nearby cells.

While it is associated with a base station, a mobile periodically measures the strength of a beacon signal from its current base station as well as beacon signals from nearby base stations that it can "hear."

These measurements are reported once or twice a second to the mobile's current base station.

Handoff in GSM is initiated by the old base station based on these measurements, the current loads of mobiles in nearby cells.

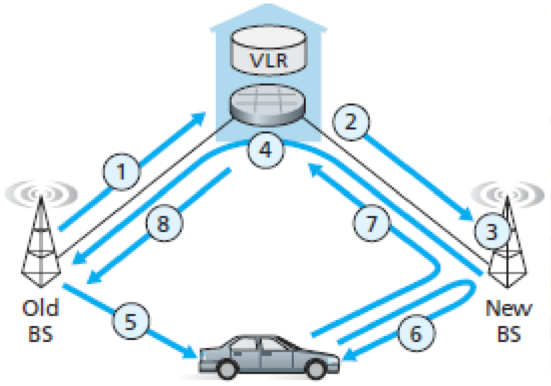


Figure illustrates the following steps involved when a base station does decide to hand off a mobile user:

1. The old base station (BS) informs the visited MSC that a handoff is to be performed and the BS (or possible set of BSs) to which the mobile is to be handed off.

2. The visited MSC initiates path setup to the new BS, allocating the resources needed to carry the rerouted call and signaling the new BS that a handoff is about to occur.

3. The new BS allocates and activates a radio channel for use by the mobile.

4. The new BS signals back to the visited MSC and the old BS that the visited- MSC-to-new-BS path has been established and that the mobile should be informed of the impending handoff. The new BS provides all of the information that the mobile will need to associate with the new BS.

5. The mobile is informed that it should perform a handoff. Until this point, the mobile has been unaware that the network has been laying the groundwork (e.g., allocating a channel in the new BS and allocating a path from the visited MSC to the new BS) for a handoff.

6. The mobile and the new BS exchange one or more messages to fully activate the new channel in the new BS.

7. The mobile sends a handoff complete message to the new BS, which is forwarded up to the visited MSC. The visited MSC then reroutes the ongoing call to the mobile via the new BS.

8. The resources allocated along the path to the old BS are then released.

Mobile moves to a BS that is associated with a *different* MSC than the old BS as shown in figure :

GSM defines the notion of an **anchor MSC**.

The anchor MSC is the MSC visited by the mobile when a call first begins; the anchor MSC thus remains unchanged during the call.

Throughout the call's duration and regardless of the number of inter-MSC transfers performed by the mobile, the call is routed from the home MSC to the anchor MSC, and then from the anchor MSC to the visited MSC where the mobile is currently located.

When a mobile moves from the coverage area of one MSC to another, the ongoing call is rerouted from the anchor MSC to the new visited MSC containing the new base station.

Thus, at all times there are at most three MSCs (the home MSC, the anchor MSC, and the visited MSC) between the correspondent and the mobile.

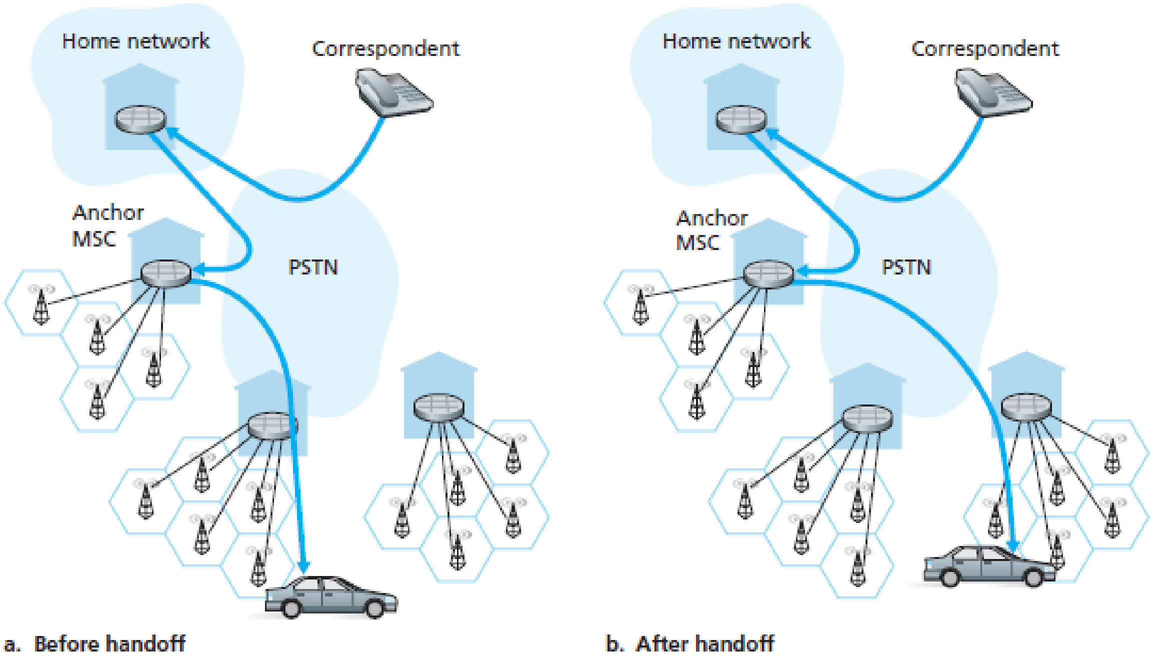
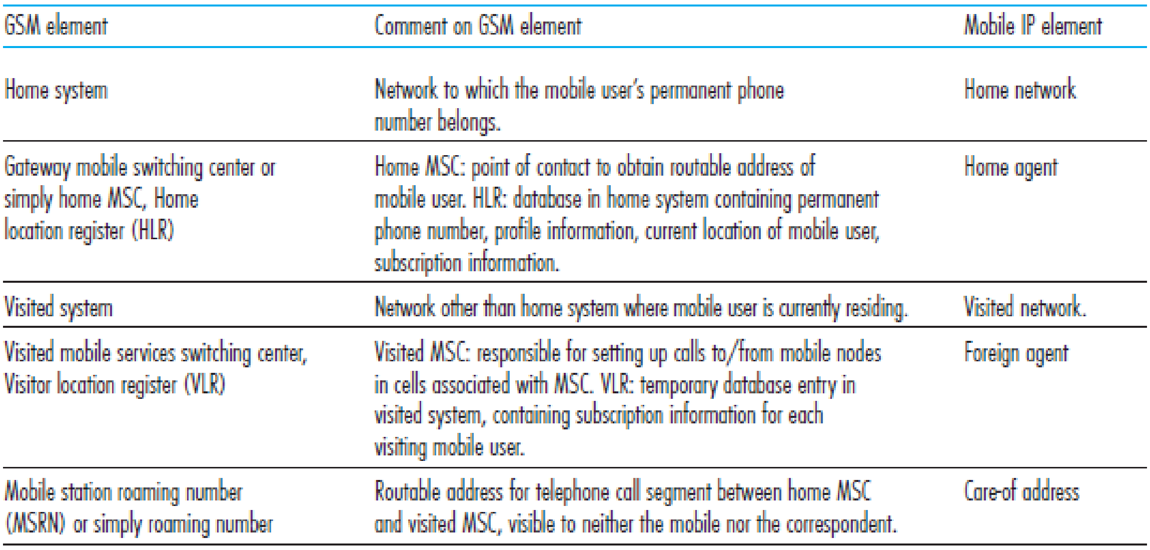


Figure above illustrates the routing of a call among the MSCs visited by a mobile user.

GSM mobility management is compared with mobility management in Mobile IP.

The comparison in Table below indicates that although IP and cellular networks are fundamentally different in many ways, they share common functional elements and overall approaches in handling mobility.



GSM element	Comment on GSM element	Mobile IP element
Home system	Network to which the mobile user's permanent phone number belongs.	Home network
Gateway mobile switching center or simply home MSC, Home location register (HLR)	Home MSC: point of contact to obtain routable address of mobile user. HLR: database in home system containing permanent phone number, profile information, current location of mobile user, subscription information.	Home agent
Visited system	Network other than home system where mobile user is currently residing.	Visited network.
Visited mobile services switching center, Visitor location register (VLR)	Visited MSC: responsible for setting up calls to/from mobile nodes in cells associated with MSC. VLR: temporary database entry in visited system, containing subscription information for each visiting mobile user.	Foreign agent
Mobile station roaming number (MSRN) or simply roaming number	Routable address for telephone call segment between home MSC and visited MSC, visible to neither the mobile nor the correspondent.	Care-of address

6.8 **Wireless and Mobility: Impact on Higher- Layer Protocols**

Wireless networks differ significantly from wired network.

The network layer provides the same best-effort delivery service model to upper layers in both wired and wireless networks.

Similarly, if protocols such as TCP or UDP are used to provide transport-layer services to applications in both wired and wireless networks, then the application layer should remain unchanged.

TCP retransmits a segment that is either lost or corrupted on the path between sender and receiver.

In the case of mobile users, loss can result from either network congestion (router buffer overflow) or from handoff (e.g., from delays in rerouting segments to a mobile's new point of attachment to the network).

In all cases, TCP's receiver-to-sender ACK indicates only that a segment was not received

intact;

The sender is unaware of whether the segment was lost due to congestion, during handoff, or due to detected bit errors.

In all cases, the sender's response is the same—to retransmit the segment.

TCP's congestion-control response is *also* the same in all cases—TCP decreases its congestion window.

By unconditionally decreasing its congestion window, TCP implicitly assumes that segment loss results from congestion rather than corruption or handoff.

Bit errors are common in wireless networks than in wired networks.

When such bit errors occur or when handoff loss occurs, there’s really no reason for the TCP sender to decrease its congestion window (and thus decrease its sending rate).

Three broad classes of approaches are possible for dealing with this problem:

• ***Local recovery***. Local recovery protocols recover from bit errors when and where (e.g., at the wireless link) they occur.

• ***TCP sender awareness of wireless links***. In the local recovery approaches, the TCP sender is unaware that its segments are traversing a wireless link.

An alternative approach is for the TCP sender and receiver to be aware of the existence of a wireless link, to distinguish between congestive losses occurring in the wired network and corruption/loss occurring at the wireless link, and to invoke congestion control only in response to congestive wired-network losses.

• ***Split-connection approaches***. In a split-connection approach the end-to-end connection between the mobile user and the other end point is broken into two transport-layer connections:

1. From the mobile host to the wireless access point.
2. From the wireless access point to the other communication end point (assumed as a wired host).

The end-to-end connection is thus formed by the concatenation of a wireless part and a wired part.

Split TCP connections are thus widely used in cellular data networks.

Consider the effect of wireless and mobility on application-layer protocols.

Here, an important consideration is that wireless links often have relatively low bandwidth.

As a result, applications that operate over wireless links, particularly over cellular wireless links, must treat bandwidth as a scarce commodity.

For example, a Web server serving content to a Web browser executing on a 3G phone will likely not be able to provide the same image-rich content that it gives to a browser operating over a wired connection.

Although wireless links do provide challenges at the application layer, the mobility they enable also makes possible a rich set of location-aware and context-aware applications .