Chapter 6

UMTS Mobility Management

This chapter describes the functions of the UMTS Network Architecture that manage communications to a mobile terminal. These functions are generic for all FPLMTS environments, including satellite networks, cordless pico-cells, macro and micro cells. Special attention is given to the interpretation of the UMTS Functional Models for satellite networks, including some hints and tips on optimum implementation. This information is not provided in the UMTS Functional Models because these implementations are examples of best practice at the current time used to demonstrate that the architecture to be standardized is valid. UMTS is being designed with the minimum standardized in functional models to future proof the standard as far as possible. What will be specified are the minimum functions and interfaces to allow networks to interact and co-operate to support terminal mobility. The UMTS standards will not be a "cook book" describing how to build a system in the way that the GSM standards tended to be.

This chapter introduces the UMTS concepts of terminal and user sessions and their registration before moving onto the mobility functions, location and domain updating and handover. Also described are call set-up and attach/detach functions. Some aspects of call control are briefly touched on but this subject is described in more detail in chapter 8.

6.1. Terminal and User Registration

UMTS is a single system to support all manner of communications on behalf of customers' applications. As mentioned earlier, the application is likely to shape the terminal rather than the specific communications networks supported. UMTS splits terminal and user registration to allow a user to simultaneously use multiple terminals (for example one for voice telephony and another for facsimile) and a terminal to simultaneously support a number of users (for example a video conferencing facility or an expensive specialist terminal shared by a number of employees).

When a mobile terminal powers up it first checks which users want to be registered on it. If there are no users requiring service then the terminal will do nothing but if there are users known to the terminal by their identification devices (called SIMs in GSM or ChargeCards in the PSTN) then the terminal will attempt to contact a UMTS network.

The terminal chooses the smallest cell network in the hierarchy of networks it can hear broadcast channels for, as described in section 2.4.4. The terminal attempts to register with whichever UMTS network it has chosen and after some minimal authentication the terminal is assigned a temporary mobile terminal identifier (TMTI) and encryption keys. The UMTS functional architecture only uses the TMTI to address the mobile terminal from this point on. Any permanent identification in the terminal would be checked only during terminal authentication, perhaps to blacklist stolen terminals, for example.

Once the TMTI is available the terminal's users can be registered so that they can receive incoming calls at the terminal. User authentication is much more rigorous than terminal authentication and requires that the user's identification device is in place in the terminal for the registration. Once registered, the user can remove their identification device, allowing other users to register and the identification device to be used for registering its user on other terminals for different services. The user's registration will remain valid on the terminal until de-registered by the user at the terminal or by the network[1].

User registration implies service provision and therefore the network will check whether the user has agreement to use the requested service on this network. If no agreements are in place then user registration will fail. The terminal may try a different network if one is available and it would probably have to re-register any other users with the new network as well unless it could maintain simultaneous communications with both networks.

At this stage the types of calls to be made are not yet known and so no preference for one network or another can be made on this basis. These choices are made at call set-up, although a terminal incapable of delivering any of the services offered by a network should not register with that network.

6.2. Location Update

Between user registration and call set-up a long time may pass and if the terminal is moving, the network needs to track that movement to ensure that it can page the terminal for an incoming call at any time. Location update's sole role is to provide enough information to the paging function to establish contact with a user's terminal on demand. The word "enough" is important here because location update and paging functions both centre around the CSS or FES and as they are designed to work together as a pair, the information that the location update function collects and stores can be uniquely tailored to the information the paging function needs, as no other FES or CSS will need to use it. It is important that these two functional entities always present a standard interface to other functions in UMTS (such as session set-up, call set-up, user registration, message services, etc...) but the interface between them is flexible.

Mobile terminals constantly receive and monitor the broadcast Paging Channel through a satellite spot beam or terrestrial cell of the network that they are registered in whilst they are inactive. From this, they can pick up *Page Messages* for incoming calls, timing information, their location area identity, the identity of the corresponding Access

---

[1] The security implications of this are still under study in ETSI and the ITU.

Channel and other system information. They can also measure the received power on this Paging Channel. The timing sequence of the Paging Channel periodically allows spare time for the mobile terminal to scan other Paging Channels to see if it can receive any other satellite beams' or terrestrial cells' Paging Channels, without missing any important information on the current Paging Channel, which just contains null data during this time. (This scanning of other channels is fairly demanding on frequency synthesizer technology because of the speed with which frequency lock needs to be acquired.) The mobile terminal measures received powers on any other Paging Channels that it could pick up and compares these with the received power of its current Paging Channel. The mobile terminal selects the highest powered Paging Channel from a terrestrial cell and if this is received with power greater than a threshold for acceptable reception ($power_{min}$) then this is chosen as the new channel. If the power is too low or if it fails to detect a terrestrial cells' Paging Channel then the best satellite Paging Channel is selected. Usually the new Paging Channel belongs to the same location area as the old one in which case the mobile terminal simply continues to monitor it. If the newly selected location area is different from the old one then the network needs to be informed so that it can forward incoming call pages through the correct Paging Channel. This is done by the terminal sending a *Location Update Message* through the new location area's Access Channel, informing the new CSS or FES of the terminal TMTI and the identity of the old location area so that the distributed databases' information can be updated. This process is known as location update.

In all mobile networks location update signalling is a trade-off against signalling required to page an imprecisely located mobile terminal for incoming calls. In satellite networks, where an antenna's radio coverage is much wider than in terrestrial networks, the optimum trade-off between precision location update and wide-area paging is different to that in terrestrial cellular networks. In all cases it must be remembered that this signalling is an expensive use of resources which traditionally is charged for as a network overhead (see section 3.5). Thus for any network to be economical the network must either be defined for the optimum trade-off at the outset or allow operators and equipment providers the flexibility to implement the optimum trade-off for their system design.

6.2.1. Location Areas

A location area is the smallest area unit used to locate a terminal when it is not engaged in a call. Within terrestrial cellular, a location area is defined by [Q.1001] as the area (cluster of adjacent cells) in which the mobile terminal can roam without having to perform a location update. Ultimately, the mobile terminal is known to be reachable through one of those terrestrial cells. The cell site antennas are static and location area re-selection can be performed as new cells are selected based on received signal strength of broadcast channels at the terminal.

Within the space segment of FPLMTS, a spot beam of a non-GEO satellite does not provide static coverage and there is relative motion between satellites and the Earth's surface. Therefore, it is not possible to permanently associate an area of the Earth's surface with the radio coverage of a specific spot beam's footprint. Because of the potential for LEO, MEO and HEO satellite motion, a different approach to location updates might be better.

6.2.2. Location Area Update in Satellite Networks

Fortunately it is not necessary to use the same location techniques. It is instructive to attempt an approach similar to the terrestrial cellular one, attempting to track the potential network-to-terminal link through satellite antennas as the satellites move. In this approach, which is called the "Direct Connection" approach in [SAINT15], satellite spot beams are considered equivalent to terrestrial cells. Each spot beam may be monitored by receiving its own unique broadcast signalling channel. In all non-GEO satellite systems those cells (and their corresponding location areas) are moving at high speed over the Earth's surface.

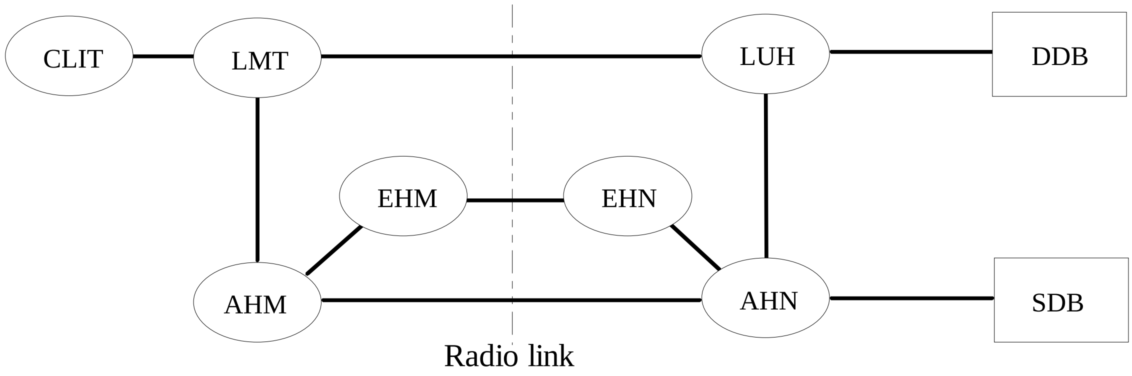
In idle mode, a mobile terminal monitors the broadcast channels from the different spot beams. The terminal maintains a list of the best broadcast channels received and initiates location update when the current spot beam moves away. The clear disadvantage of this approach is the tremendous quantity of location update signalling, which is the result of predictable motion of the satellites even when the mobile terminal is stationary. In a system such as the 769km altitude LEO, location update would occur about every 5 to 10 minutes for every mobile registered in the satellite network, generating unacceptable amounts of signalling.

A solution to the problem lies in recognizing that the location update system is only useful for the core UMTS network to identify a route by which an incoming call can be connected to its called mobile terminal. Therefore it is important for the core network to know which entry point to the satellite network it should route an incoming call through, specifically which FES is capable of contacting the mobile terminal. So long as this FES can contact a mobile terminal, there is no need for location update. This allows the satellite network operators to use whatever method they like to make contact with the mobile. A terminal constantly monitors the Paging Channel broadcast by the FES to determine if there are any incoming calls for it. The FES also includes its own unique location area identifier in the broadcast message. The mobile will lose the Paging Channel if it moves out of FES coverage and will then search for a new one and initiate location update. Note that it is always the terminal that initiates location update.

6.2.3. Location Update Functional Model

The LMT function continuously monitors the location area information transmitted by the base stations and FESs covering the terminal, and compares it with the location information already stored in the CLIT. If the LMT function decides that the terminal should access a new location area within the same domain, it will send a location update request to the network, indicating the new location area. If the new location area is in a different domain the terminal initiates domain update, which is similar to location update but additionally includes transfer of registration to the new domain, possibly requiring changes to users' registrations depending on access rights to the new domain. In the network, the terminal's data will be modified so the new location area is stored as the location of the terminal. Registration data of individual users is not updated because user's registration data refers only to the TMTI and not the user's location. Therefore update of just the terminal's data is sufficient.

In network terms, a location area is the address of a UMTS switch that knows sufficient about the mobile terminal to be able to page it. Exactly what the switch knows and how it found it out is unimportant, so long as the network is capable of routing signalling for incoming calls to that switch.



CLIT	Current location information in terminal	DDB	Distributed database
LMT	Location area monitor in terminal	LUH	Location update handler
EHM	Encryption handler in mobile	EHN	Encryption handler in network
AHM	Authentication handler in mobile	AHN	Authentication handler in network
		SDB	Security database

Figure 31   UMTS location update functional model

In a satellite system the location area is where the terminal thinks it is, because the terminal is responsible for initiating location update. The location area is therefore defined as formed by a FES's *instantaneous* coverage area so that a mobile terminal should update its location only if it loses the FES's Paging Channel and associated location area identity. Otherwise, from the network viewpoint, the location of the mobile is "somewhere within reach of the FES". The FES (either on its own or with the help of the mobile terminal) may provide a way of intelligently reducing the area over which it pages in the event of an incoming call. If this does happen, then it is an undefined network procedure, localized to between the FES and mobile terminal. The FES would be responsible for its own storage of data that will enable the paging function to page intelligently.

Implicit in this argument is the assumption that an FES is able to use satellite resources to maintain coverage of an area around it in which the mobile terminals that are registered with it are located, as was demonstrated in section 5.5. Note that for practical reasons, the location area is the instantaneous coverage of the FES rather than the GCA because this is the area over which the FES's broadcast channel can be received (refer to figure 22 in chapter 5 to see the difference). Because of the dynamics of a satellite network, the edges of an FES's coverage are only intermittently covered. The FES is programmed to always cover the GCA but minimize the transmission of its location area identity outside the GCA. In a satellite network designed to provide coverage throughout a region, FESs' GCAs will overlap in places and combine to cover the region with no gaps. If a mobile terminal is in overlapping FESs' coverages and location updates to an FES only intermittently covering its location (because the terminal is not in the FES's GCA), it will lose that FES's location area broadcast after a while and be forced to location update to another FES that covers its location properly. Hence a mobile terminal will always find an FES that guarantees its coverage after only a small number of location updates. By using large GCAs (section 5.6 shows how vast they can be), the need for location update would be very rare - perhaps only when the mobile terminal is first turned on and when the mobile terminal leaves or moves back into terrestrial coverage.

6.2.4. Satellite Location and Paging Strategies

Besides the "Direct Connection" approach discounted in section 6.2.2, Race Saint identified two more approaches [SAINT15] to location and paging. Using GCAs, [ARAKI] considers another two approaches for location update. These and some other approaches are presented here, demonstrating how the UMTS location update functional model accommodates all the approaches, their differences representing differing implementation choices. An objective has been to maintain compatibility with as broad a range of satellite systems as possible (see chapter 4). With suitable definitions of paging area and location area, all of these systems can work with the UMTS location update function, as can all terrestrial cellular networks. The use of identical functional models in the terrestrial and satellite segments will facilitate the integration of satellite components into UMTS.

As in the terrestrial segment, any incoming call will be routed to the FES that can guarantee that the mobile terminal, if working, is within the area the FES is covering with its broadcast channels. This section presents some alternatives for how the location area identity is broadcast by the FES and how paging is performed. Note that the area that will effectively be paged will always be bigger than the paging area because it is made up of spot beams including those which only part-cover the boundaries of the paging area.

6.2.4.1. Paging Area Covering Entire GCA of FES

With the FES at its simplest, the FES would transmit a paging message for the mobile through every spot beam that it is using to cover its GCA. If incoming calls occur infrequently, this may be acceptable but otherwise paging through these many spot beams is considered a waste of spectral and power resources.

Satellite network designers could use a number of techniques to reduce the number of spot beams paged, some examples of which are presented in the following alternatives.

6.2.4.2. Multiple Location Areas per FES

The GCA of an FES can be very big, as shown in section 5.6, so it may be convenient for the FES operator to split the area into two or more location areas, each with a distinct broadcast location area [D112]. As with the GCA, these location areas would be geographically fixed. For example, an FES covering southern Europe, the Mediterranean and the Middle East might split the location area along a border that is seldom crossed, say the middle of the Mediterranean sea. This then reduces the maximum area over which paging is necessary whilst increasing the location update signalling traffic only marginally. An additional overhead at the borders between location areas is that spot beams will need to broadcast the Location Area Identities (LAIs) of the location areas on both sides of the border. This is to guarantee that the LAI is broadcast throughout the location area all the time, which prevents mobiles from flip-flopping between location areas as spot beams switch from one location area to the next.

This approach could enable a satellite network to have location and domain areas of different sizes to map onto politically defined areas such as countries or regions. This may go some way to supporting the cross-border political aspects of Satellite FPLMTS - moving from one region to another could require location or domain update, allowing control of the services offered in each region.

Using multiple location areas does not preclude the use of other information about the mobile terminal's position to page only through spot beams where it is likely that the terminal is thereby avoiding paging throughout a location area. This is "intelligent paging". In each of the following alternatives the FES, possibly with help from the mobile, uses additional information to page the mobile in a paging area that is smaller than the location area. The network beyond the FES is not effected by this extra intelligence.

6.2.4.3. Paging Only Location Update Spot Beam's Position

A first approach to intelligent paging could be the FES identifying and recording the instantaneous size, shape and location (latitude, longitude) of the spot beam in which the mobile terminal last made contact (whether for location update, call set-up or any other reason), with a time-stamp. If there is an incoming call, the FES would page only those spot beams required to completely cover the recorded area in the first instance. In case the mobile does not respond the paging is repeated over a wider area, depending on the age of the time-stamp and on the mobility profile of the terminal [SAINT15].

In effect the FES is keeping a record of the approximate location of the mobile terminal at every communication from the terminal to the FES. However, real location update invoking the functional model shown in figure 31 would only be necessary if the mobile terminal lost the broadcast Paging Channel of the FES and had to pick up the broadcast of another FES or another location area from the same FES.

6.2.4.4. Using a Terminal Position Fix to Reduce Paging Area

If a mobile terminal is capable of making the necessary measurements and calculations to fix its own position then the FES could record the measured position of the mobile, which would be more accurate than a spot beam's coverage area. The location update message from the mobile to the FES would be modified to include the latitude and longitude of the terminal and an uncertainty radius that determines the circular area where the terminal can be found at *any* time. The mobile then continuously monitors its own position and if it moves outside the uncertainty area it declared to the FES then it automatically performs a position update to declare its new uncertainty area [SAINT15]. From the FES's point of view it is a passive procedure with no signalling generated until the mobile has moved further than the uncertainty area radius. There is no signalling generated by the satellites' motions. If the mobile is still receiving the location area broadcasts of the FES then position update would remain local to the terminal and FES, since from the network point of view the terminal is still contactable through the same FES as before. Therefore the full location update model in figure 31 would probably not need to be used. However, if at any time the mobile lost the location area broadcast of the FES then it would search for and location update to a new FES, invoking the full location update procedure.

On an incoming call, the FES only needs to page the spot beams that cover the mobile's declared uncertainty area (or at least the part of it that is within the FES's location area). Depending on the mobility of the terminal and its user's incoming and outgoing call rates, the terminal might vary the uncertainty area radius to minimize either the paging area or the number of position updates in an attempt to minimize the total spectral resource consumed by this signalling. The quantity of paging signalling rises as the uncertainty radius increases whilst the quantity of position update signalling decreases, so there is clearly scope to optimize the uncertainty radius based on the mobility profile and call profile of the terminal's users.

Disadvantages of this approach are the extra complexity in the mobile terminals and the additional space resources needed to implement position fixing. A terminal capable of position fixing requires more functionality and processing power and is more expensive to produce than one that is not. However, position fixing might be a useful as a UMTS service for other purposes. Position fixing signals could either be supplied by a third party (like the US Navy's GPS system) or extra bandwidth on UMTS satellite broadcast channels could be reserved for the timing signal and satellite almanac details. Note that the accuracy required for position fixing as a UMTS service might be very different to that required for location management. The accuracy of a dedicated positioning system like GPS would be expensive to match.

6.2.4.5. Using Dual Satellite Coverage Position Fixing

If a mobile terminal is covered by two or more satellites (both in use by the FES) when it makes its location update then the FES could perform position fixing ranging measurements for itself. Using the relative measured delay and Doppler frequency shift through the different satellites, the terminal position could be calculated and stored along with an uncertainty area and time-stamp. As in section 6.2.4.3, this position would not be updated unless a call was set up or the mobile lost the FES location area broadcast. On an incoming call, the FES would page an area covering the uncertainty area or a wider area, depending on the age of the position time-stamp. Unlike terminal position fixing, this FES intelligence does not require any extra position fixing features in the mobile terminal. A periodical position update could be implemented to reduce uncertainty of old location updates by the FES signalling the mobile so that the mobile responds and position measurements can be made on the response.

6.3. Attach/Detach

The UMTS network architecture provides an attach/detach service that can be optionally used by network operators to prevent fruitless attempts to page a terminal that is temporarily inaccessible. Terminals can become inaccessible for long periods by being without power, inside metal briefcases or car boots and so on. In terrestrial cellular networks this attach/detach service is not likely to be used much but in satellite networks paging overhead could be large enough to make use of these functions attractive.

6.4. Call Set-up

Whilst inactive, mobile terminals constantly receive and monitor a broadcast Paging Channel for the location area in which they are registered. As well as paging information to announce incoming calls, the relevant Paging Channel contains information on which channel should be used as the Access Channel for both location updating and call set-up requests. This ensures that whenever the customer dials out or wants to answer an incoming call, the mobile terminal already knows which FES or base station to send the call request to, using a multiple random access protocol on the base station's or FES's common Access Channel. Once initial contact has been made by the mobile terminal the base station or FES assigns Uplink and Downlink Traffic Channels dedicated to the terminal for the call. In the call set-up procedure the terminal must tell the network what kind of services it requires to support the application the customer wants to use and the network needs to inform the application exactly what Traffic Channels and services it has been given so that the terminal can use them correctly. This signalling and negotiation aspect of call handling is known to be absent from the UMTS network architecture at the time of writing and so is studied in greater detail in chapter 8. Various strategies for assigning the Traffic Channels in satellite systems are described in chapter 7.

6.5. Satellite Handover

6.5.1. Handover Initiation

Once a call is in progress, handover should be initiated if the call's communications quality falls below a "try handover" threshold. Handover initiation is very simple - when either the C/I ratio or the received power of the Traffic Channel fall below their handover thresholds, $(C/I)_{\text{try handover}}$ and $\text{power}_{\text{try handover}}$ respectively, the mobile terminal sends a *Handover Request* to its currently serving FES using the Uplink Traffic Channel. Alternatively, if these uplink parameters fall below the same thresholds, the FES should initiate the handover procedure. Chapter 7 shows that assignment of the new channel may take several seconds after the process is started, so the thresholds should be set to initiate handover before the edge of beam conditions so that the call will not be dropped during this period. Having noted this, it is feasible and advantageous to use "fuzzy" thresholds that can be dynamically altered with traffic demand to effectively control the traffic that is accepted onto heavily loaded spot beams.

6.5.2. Intra-Satellite Handover

In the UMTS Network Architecture, intra-satellite handover is regarded as an automatic feature of the FES to mobile terminal link rather than a UMTS network handover in which the core network becomes involved in re-routing and call control is changed. It is recommended that the FES be programmed to pre-empt these handovers, which can be planned for, and issue the *Handover Request* before C/I or received power become marginal. Because the positions of the mobiles on the ground are relatively stationary, the pattern of interference may be stable enough to hand a whole block of Traffic Channel Pairs from one spot beam to the next without disrupting mobile terminals in the block or other terminals nearby. Well-planned handovers such as this can be guaranteed interference-free and the mobile terminal can stay on its current Traffic Channel in the new spot beam without performing any checks on the channel itself. Such handover plans need co-ordinating between different FESs sharing the same satellite in neighbouring or even the same spot beams to ensure that changes made by one FES do not adversely affect another's traffic and to prevent large amounts of near-simultaneous signalling.

Intra-satellite handovers to the same Traffic Channel on a following spot beam are most transparent and rapid, requiring only that the Traffic Channel be routed through a different spot beam on the same satellite by re-configuring the beam forming network. Thus the channel frequency will remain the same, requiring no RF re-tuning, although there may be some phase discontinuity requiring carrier re-acquisition. The FES is in complete control of this operation so it is possible for it to predict handover to the beam following a mobile terminal's old spot beam and thus take advantage of the deterministic motion of the spot beams over the relatively stationary mobiles. In some cases the FES may need to request a change of Traffic Channel, even then there is guaranteed to be no change in path delay or Doppler effect. Even if RF re-tuning is required, the new Doppler offset is easily calculated.

In LEO constellations, these handovers between spot beams can be expected about one every minute.

6.5.3. Inter-Satellite Handover

If the satellite coverage footprint is moving away from the mobile terminal then an inter-satellite handover must be attempted. Initially it was assumed that the mobile terminal would find a new satellite in much the same way as it does in idle mode before call set-up by searching sequentially through the Paging Channels for one that it can receive clearly. A simulation was performed, described in chapter 7, to check this assumption. Acquiring a new satellite proved to be a very slow process, since the time taken to measure the received Paging Channel power is long because of the time taken to acquire the Doppler frequency shifted carrier. The real time typically taken for such a measurement is not clear so an estimated value of 500ms per power measurement was used for the simulation model in appendix A. Using this scheme, the simulations reported mean inter-satellite handover interruptions as high as 5 to 6s at operational traffic intensities. The maximum recorded interruption was 34.2s - it is unlikely that a customer would wait this long for his communications channel to return. This result led to consideration of the advantages of associating FESs with a number of satellites so that the geographical coverage of the FES is larger than the coverage footprint of just one satellite and inter-satellite handovers can be handled without changing FES, by the FES itself. This is the system of guaranteed coverage adopted for the UMTS Network Architecture, described in chapter 5.

To implement guaranteed coverage, a FES must predict the coverage areas and overlaps of different satellites' spot beams so it can select a new satellite for the mobile terminal and begin suggesting Traffic Channels and Doppler shift on the newly selected satellite, using the old satellite link. Any FES would be programmed to maintain complete coverage of a particular area, so if the mobile terminal is within that area then the FES will always use appropriate satellite beams to inter-satellite handover to. This provides an FES with the opportunity to semi-plan Traffic Channel reuse between satellites within its coverage area, although other FESs' coverage areas will overlap with this

FES's and they will also share some of the satellites. Some co-ordination between the FESs would be useful to ensure that they do not interfere with each other's mobile terminals. Inter-satellite handovers might be slightly slower than intra-satellite handovers if the Doppler frequency shift needs to be calculated for the new satellite before the mobile terminal can recommence transmission. This acquisition can be speeded up if the FES calculates the Doppler offset for the mobile terminal and suggests it with the channel assignment.

6.5.4. Rapid Inter-Satellite Handovers

In shadowed radio environments satellite diversity can be used to minimize link budget margin. Rapid inter-satellite handover is required when a mobile terminal first moves into a radio shadow from the old satellite. In these conditions the received power can drop by tens of dB in only milliseconds in which case handover must be either completed very rapidly or some kind of forward handover controlled through the new satellite must be used to recover the call after a short interruption. These drops in power are unpredictable, so the FES cannot plan a handover for them in advance. Shadowing will cause the fastest drops in channel quality. Therefore the values of $(C/I)_{\text{try handover}}$ and $\text{power}_{\text{try handover}}$ relative to $(C/I)_{\min}$ and $\text{power}_{\min}$ (the values at which the call must be dropped) need to be set to allow sufficient time to at least start the handover process before communications with the old satellite are lost. A practical solution might be for the FES to continuously provide an active mobile with an up-to-date list of alternative satellites ready for rapid handover if deep fades arise.

In summary, there will be two types of inter-satellite handover:

1. Those where the satellite coverage footprint is predictably moving away from the mobile. In these cases the FES will pre-empt the handover and plan handover to a following satellite. The signal will degrade slowly, allowing time for a simple handover controlled by the FES with very little effort by the mobile terminal. In a 769km altitude LEO constellation, such a handover would be expected every 5 or 6 minutes.

2. Those where the mobile terminal is shadowed from the satellite by buildings, mountains or other clutter. In these cases other satellites may be visible and capable of maintaining communications and the FES is capable of providing mobile terminals with a list of frequencies through which it could try to re-establish contact with the FES. However, because of the rapid loss of signal and the short lengths of time in shadows, a facility to switch very rapidly between two or more satellites is useful. This is satellite diversity, a handover mechanism similar to CDMA's soft handover. It is supported in the UMTS network architecture by a macro diversity mechanism.

6.5.5. Satellite Diversity

UMTS macro diversity maintains two communications paths from the mobile terminal through the radio access system to the CSS or FES, where the signals are combined in some way. This combination could be time alignment (to compensate for delay) and addition of the signals or simply selection of the best signal, for example. The two Traffic Channel paths from FES to mobile are both fully assigned to that terminal for the periods using diversity, which in terrestrial networks has been shown to be up to 40% of the time. Despite apparently doubling the radio resource reserved for the mobile, the reduction in transmission power and hence in interference to other mobiles means that those resources can be reused closer to the terminal than the Traffic Channel for a single link could be.

To use macro diversity effectively a terminal is designed to be able to at least monitor channel quality on a second Traffic Channel besides the Traffic Channel it is using for communications. During communications it would be constantly hunting for a better channel to use as a target for macro diversity. The second, diverse Traffic Channel would be set up by handover request (perhaps prompted by the FES) through the current main communications satellite, if possible, to make the handover seamless.

So far in this section all handovers have been within the control of a single FES that hides the complexity of its communications from the core UMTS network completely. There is a possibility that the mobile terminal may move outside the FES's GCA in which case it will need to hand over to a new FES.

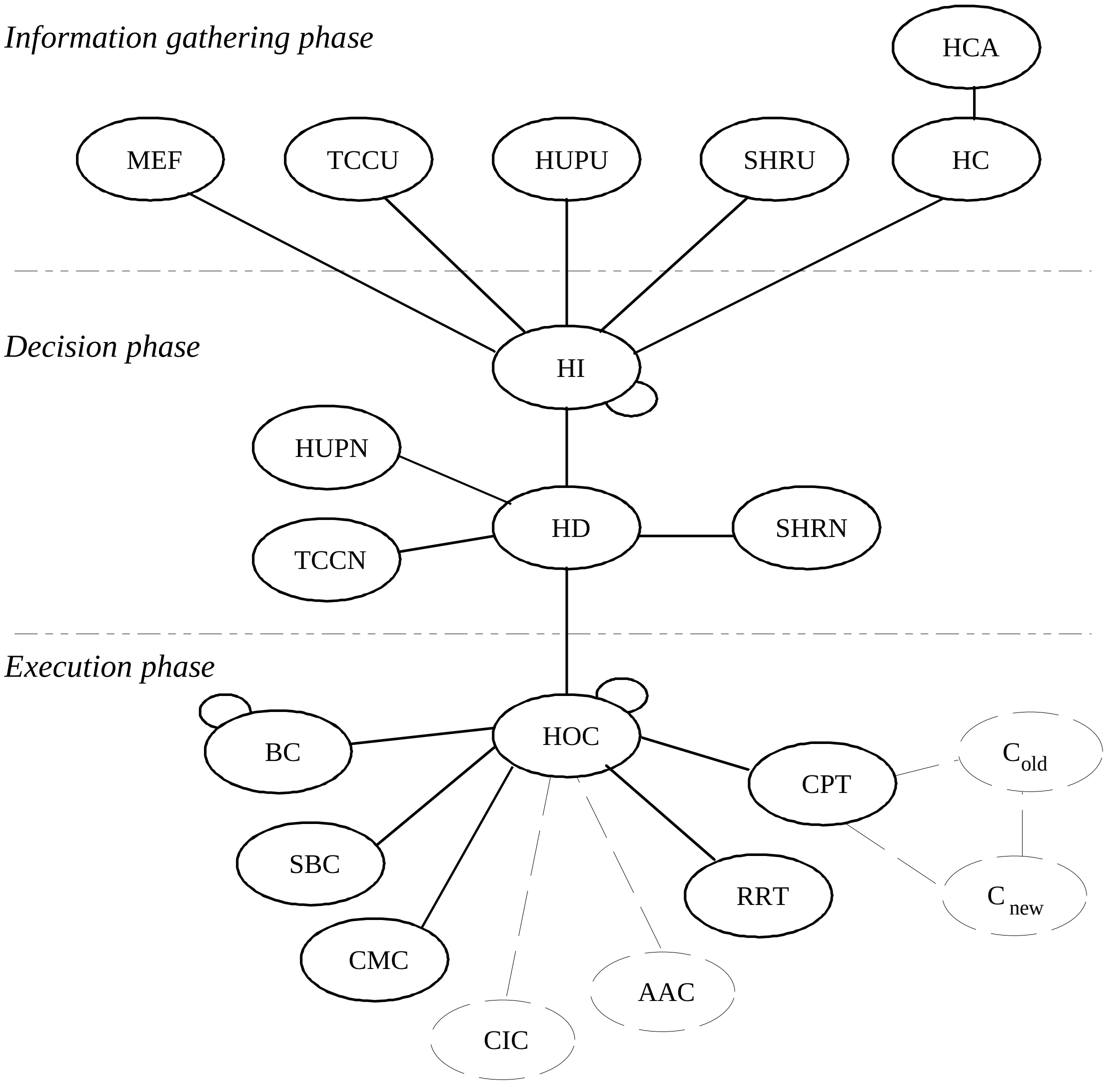
6.5.6. Inter-FES Handover

Handovers between FESs are not expected to occur very often because of the vast size of each FES's GCA. For most mobile terminals it would be difficult to travel far enough to move from inside to outside of a GCA during an active call. The two most likely reasons for inter-FES handovers are:

1. Mobile terminals switched on to immediately make a call. In this case a mobile near the edge of a GCA may mistakenly register to an FES that it can only temporarily communicate with. When this temporary communications channel begins to fade the FES will need to hand over to a FES guaranteeing coverage of the mobile terminal.

2. Fast-moving terminals holding calls for long periods of time. Transatlantic flights and high speed trains could hold calls like this which are quite likely to cross GCAs of different FESs. Commercial flights may use FPLMTS pico or micro cells inside the cabin to communicate with the mobile terminals that are insulated from outside by the metal aircraft body. The aircraft still needs to use FPLMTS to connect back to the ground and may aggregate the calls in progress on the aircraft into one trunk channel back to the FES to simplify radio bearer control. This trunk radio bearer may need to be held by the aircraft for hours, in which case inter-FES handover is highly likely.

An inter-FES handover uses exactly the same functions as UMTS uses for handover between CSSs in terrestrial networks. The functional model is shown in figure 32 and its operation is described in the following sections.



MEF	Measurement Function (MT and FES)		
HC	Handover criteria (MT and FES)	HCA	Handover criteria adjustment (FES)
HI	Handover initiation (MT and FES)	HD	Handover decision (FES)
TCCU	Target cells and connections - user (MT)	TCCN	Target cells and connections - network (FES)
HUPU	Handover user profile - user (MT)	HUPN	Handover user profile - network (FES)
SHRU	Special handover request - user (MT)	SHRN	Special handover request - network (FES)
HOC	Handover control (network)	CPT	Control point transfer
BC	Bearer control	SBC	Switching and bridging control
CMC	Combining and multicasting control	RRT	Rerouting triggering
AAC	Authentication and access control	CIC	Confidentiality and integrity control

Figure 32   Handover functional model

6.5.6.1. Information Gathered

Five functions gather the information used by the HI (handover initiation) function to determine if handover is required or not.

1. *MEF (measurement function).* The MEF gathers measurements of the active communications link and translates these into a defined set of link quality parameters (for example C/I, received power, probability of shadowing, etc...).

These parameters are updated every second or so to remain representative of the rapidly changing radio conditions. There are, in fact, two instances of the MEF - one in the mobile terminal measuring downlink parameters and one in the FES or on board the satellite measuring uplink parameters. This is because radio interference is not always symmetrical on uplinks and downlinks, as will be seen in chapter 7.

2. *TCCU (target cells and connections - user).* The TCCU function continuously scans alternative radio links to compile a list of candidate cells (both terrestrial and satellite) to which this specific terminal could handover to. The list also contains the information (for example cell type, C/I, received power and traffic load on the cell) needed by the HI function to determine the preferred cell to handover to and order the alternatives. The TCCU function is performed by the mobile terminal and updates this list every second or so. This means that a multi-mode FPLMTS terminal that can access satellite and terrestrial networks must be able to scan both satellite and terrestrial links whilst maintaining the active call's communications to enable handover between satellite and terrestrial networks.

3. *SHRU (special handover request - user).* The SHRU function can force the HI to initiate a handover immediately when requested by a terminal's user.

4. *HUPU (handover user profile - user).* The HUPU function maintains a subset of the customer's profile, specifically all the information that is related to the handover process (quality of service preferences, profile of the terminal's mobility, access rights, service rights, priorities, operator preferences, etc...).

5. *HC (handover criteria).* The HC are the thresholds, such as $C/I_{\text{try handover}}$, C/I_{min}, $\text{power}_{\text{try handover}}$ and $\text{power}_{\text{min}}$, used in the HI algorithm. These thresholds will be different in each network, even in different parts of the same network. They are stored in the mobile terminal and in the radio access network and are modified by the HCA.

HCA (handover criteria adjustment). The HCA function sets and updates the handover criteria in individual terminals according to instructions it receives from the resources control and the network management functions. This allows the bearer control functions and the FES's network management functions to alter the thresholds for handover, making handover initiation a "fuzzy" control algorithm. By controlling handover thresholds, the FES or terrestrial base station can effectively alter the sizes of cells and so control the amount of traffic demand in a particular cell. The HCA function would be located in the satellite access network outside the mobile, most likely in the FES but maybe on board an intelligent switching satellite.

6.5.6.2. Handover Decision

HI (handover initiation). The HI function in the mobile terminal compares the active link measurements (supplied by the terminal's MEF) and the candidate cells' information (supplied by the TCCU) with the handover initiation criteria (supplied by the HC), taking into account preferences in the HUPU information and SHRU requests, to identify the need for a handover. When handover is needed, it issues a *Handover Request* to the HD, passing on all the available information needed for the HD to identify the actions required. This would include identifying the handover control point, the type of handover required (hard handover or to add or drop a connection for macro diversity) and the handover completion time constraints.

There is also a HI function in the network (at the FES or on the satellite) monitoring the uplink quality parameters from the network's MEF and comparing them with the network's HC thresholds. The two HIs effectively act as one function distributed across the radio link to enable it to monitor both ends of the link. Thus the HI function in the network is used to pass on any *Handover Requests* from the mobile HI function to the HD function which is always located in the network.

HD (handover decision). The HD function receives its *Handover Request* from the HI or SHRN functions and determines the location of the handover control point. The control point is a node in the network that has access to both old and new bearers to the mobile terminal, where the eventual switch-over between them will happen. It then determines the service available from the candidate cell by negotiation with radio resource management in the candidate cells, using information from the TCCN and HUPN. The HD function makes the final decision about whether to proceed with the handover request using this information.

✓ If the HD decides to proceed, it instructs the HOC function to execute the handover. It provides information about the position of the new connection point, which may be different to the one suggested by the HI function following the HD function's negotiations. It also provides handover completion time constraints to the HOC.

✗ If the HD decides not to proceed, it returns to the HI an indication that handover is not advisable at this time and that the poor quality service that the mobile terminal is receiving will have to be tolerated temporarily.

TCCN (target cells and connections - network). The TCCN performs a similar function to the TCCU but it works in the FES or on the satellite. The TCCN function's location in the network gives it the advantage over the TCCU of having access to the future pattern of satellite orbits which enables it to predict which connections are improving and which are degrading at the time. A TCCN in a terrestrial network might also know the identities of its neighbouring cells and so ease handover to these and may also know the identity of an FES that includes the terrestrial cell in its GCA. This would always ensure clean handover even if the terminal's TCCU function is not good enough to have identified these options. The TCCN also knows about the availability of resources in the target cells and provides this information to the HD function.

SHRN (special handover request - network). The SHRN forces a handover whenever the network side of the radio link requires one, even if the mobile terminal is content with its link measurements. In satellite networks this function could be used by the FES to force mobile terminals from spot-beam to spot-beam and from satellite to satellite to compensate for the predicted motion of the satellites. These intra-satellite and inter-satellite handovers will be extremely common and do not involve a change of FES. Therefore the HD would nominate the FES itself as the handover control point (or perhaps the satellite if the intelligence for this kind of function is on board the satellites). The TCCN would be primed with the target spot beams and satellites to ensure that the handover goes to the correct connection with no initiative required from the mobile terminal.

HUPN (handover user profile - network). The HUPN maintains a subset of the customer's service profile that relates to decisions made by the HD function. As this information is mainly used to negotiate acceptable degradation of service, it will mostly consist of quality of service requirements and how these can be varied.

6.5.6.3. Handover Execution

HOC (handover control). The HOC function co-ordinates the interactions with the BC, SBC, CMC, CPT, RRT, CIC and AAC functions and returns the result (handover completed or not) to the HD function. It also provides this information to the TMN functions in the radio access system for the FES operator's network management purposes.

If the handover is between two different networks (for example between satellite and terrestrial or two different operators' networks) then there will be HOC functions in each network which interact with each other and with the BC, SBC, CMC, CPT, RRT, CIC and AAC functions in their own networks. It is this interaction between HOC functions that is crucial to seamless handover between different FPLMTS networks, including from terrestrial to satellite networks and vice versa. So long as this interface from HOC to HOC can be standardized and the TCCU and TCCN functions can identify the possibility of handover to satellite and terrestrial networks then customers can be assured that their calls will roam through all of FPLMTS' different networks faultlessly, subject to the terminal having the right air interface emulations and the customer having the right service agreements.

BC (bearer control). The BC function receives instructions to establish or release bearers from the HOC. It establishes bearers with the defined bandwidth or releases them and returns the result (success or failure) to the HOC.

SBC (switching and bridging control). The SBC function controls bridging and switching, that is establishment and release of a bridge between a number of bearers and the switching from one bearer to another. Again, it returns the success or otherwise of the bridge or switch to the HOC function.

CMC (combining and multicasting control). The CMC function controls the combining and multicasting functions in a macro diversity group. It controls the set-up, release and change of these nodes. As usual, it returns the result of the operation to the HOC function.

RRT (rerouting triggering). The RRT function is in the core UMTS network and triggers the core network to re-route bearers from the old to the new local exchange.

CPT (control point transfer). The CPT function transfers the call control point from one node (C_{old}) in the network to another (C_{new}). Call control is usually at the originating local exchange and this would change with inter-FES handover, for example.

6.5.7. FES to CSS and CSS to FES Handover

As mentioned above, the use of a standard UMTS handover functional model by the FES ensures that the FES can interface with CSSs in exactly the same way as it does with other FESs for inter-FES handover. This makes handover from FES to CSS and from CSS to FES possible.

To recap, it is the standardized interaction between HOC functions that allows handovers between different networks to be controlled by two UMTS compliant HOC functions, one in each network. As well as having this interface standardized, the TCCU function in a customer's mobile terminal needs to be able to identify the possibility of handover to both satellite and terrestrial networks. This is crucial for FES to CSS handovers because the FES will not be able to know the layout of terrestrial coverage throughout its GCA and so the TCCN function cannot supply a list of candidate cells for the mobile to handover to. Handovers from CSS to FES may rely less on the TCCU function because a terrestrial network's TCCN function can suggest a target FES that the network planners know guarantees coverage of the cell's region.

Handovers between satellite and terrestrial networks will not be seamless because of the large changes in link delay and the physical separation of the elements in the networks makes control of the handover slow. However, even in the worst cases the call will resume within a few seconds, less time than it would take for the customer to redial.