Mobile Networks and Applications 4 (1999) 301–310

Handover management in Low Earth Orbit (LEO) satellite networks

Ian F. Akyildiz [a], Hüseyin Uzunalioğlu [b] and Michael D. Bender [c]

[a] *Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

[b] *Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733, USA*

[c] *Department of Defense, 9800 Savage Road, Fort Meade, MD 20755, USA*

Low Earth Orbit (LEO) satellite networks will play an important role in the evolving information infrastructure. Satellites in the low earth orbits provide communication with shorter end-to-end delays and efficient frequency usage. However, some problems need to be solved before LEO satellite systems can be successfully deployed. One of these problems is the handover management. The objective of this paper is to survey the basic concepts of LEO satellite networks and the handover research.

1. Introduction

Terrestrial wireless networks such as current cellular and Personal Communication Services (PCS) systems provide mobile communication services with limited geographic coverage. To provide global coverage to a diverse user population, a number of Low Earth Orbit (LEO) satellite networks have been proposed [14,28,35]. The LEO satellite networks can support both the areas with terrestrial wireline and wireless networks and the areas that lack any network infrastructure. In the former case, the satellite system could interact with the terrestrial wireless network to absorb the instantaneous traffic overload of these networks. In other words, mobile users could alternatively access either a terrestrial or a satellite network through dual-mode handheld terminals. In the latter application area, LEO satellites would cover regions where terrestrial wireline and wireless systems are economically infeasible because of rough terrain or insufficient user population.

First generation satellite networks utilized Geostationary Earth Orbit (GEO) satellites, which are located over the equator at an altitude of 35,786 km. At this altitude, a satellite circulates the earth in synchrony with the earth, i.e., a GEO satellite completes one turn around the earth approximately in 24 hours. As a result, the position of a GEO satellite is stationary with respect to a fixed observer on the earth surface. This is a good feature since the coverage area of a GEO satellite is also stationary. Moreover, a GEO satellite covers almost $1/3$ of the earth surface excluding the polar regions. Hence, three satellites are sufficient for global coverage. Large and stationary coverage area results from very high orbit altitude, which also results in certain disadvantages for mobile communications. First, the user terminals and the satellites have high power consumption for the communication. Second, the propagation delay between the mobile user and the satellite is too high for real-time multimedia communications. Third, high orbit altitude results in an inefficient use of the available frequency resources.

An alternative to GEO satellite systems is to utilize low earth orbit satellite systems. The major advantages of these new systems are low propagation delay, low power requirements in the user terminals and the satellites, simple user terminals, and efficient spectrum utilization using small coverage area for each satellite. Moreover, it is possible to route a connection using Intersatellite Links (ISL) without relying on terrestrial resources. However, in contrast to GEO systems, a number of mobility management problems occur in the LEO satellite systems. Mobility management in LEO satellite networks can be classified into:

- location management (registration and paging), and
- handover management.

Location management tracks and locates the user terminals for the incoming calls, while handover management allows a call in progress to continue without any distruption as the serving cell of the user is changing. Location management protocols deal with querying and storing information in location databases (registration) and sending paging signals to locate the user within the network (paging). As a result, many of the issues are not protocol dependent and can be applied to any of the mobile networks, i.e., similar algorithms can be used in terrestrial wireless networks and satellite networks. In contrast, handover algorithms in the satellite networks differ from those in the terrestrial wireless networks. This is because the handovers occur as a result of the satellite movement as explained in the following sections. In the terrestrial wireless networks, the handovers occur because of the user movement. Hence, there is a need for further research in the satellite handover management. We focus on the handover management issues in this paper.

The objective of this paper is to introduce the basics of the Low Earth Orbit satellite networks and to overview the handover problems and the suggested solutions in the literature. In section 2, we describe the basic concepts. In particular, we discuss the satellite system design criteria such as the selection of the orbit parameters, the coverage model, network connectivity, and connection routing model. In section 3, we cover the handover management. Since the handover management is a more mature research area for terrestrial mobile networks, we provide similarities and differences with these systems. Finally, in section 4, we give open research issues.

2. Basic concepts

When designing a satellite network, some decisions such as the selection of the orbit parameters, coverage model, the network connectivity, and routing model must be made. For example, the existing/proposed LEO satellite networks such as *Iridium* [14,16], *ICO* [22], *Teledesic* [12], *Globalstar* [12], and *Ellipso* [22] have basic architectural differences. Each design decision affects the capabilities of the system and the extent of the services that are provided to the users [8,13,17]. In this section, we want to emphasize on the basic building blocks and explain them. In section 2.1, we point out the trade-offs between different orbit types. We give examples from existing/proposed network architectures. In section 2.2, we introduce a model for coverage areas of the individual satellites. In section 2.3, we discuss the network connectivity models and a routing model.

2.1. Selection of the orbit parameters

Satellites are classified as Geosynchronous Earth Orbit, Medium Earth Orbit (MEO), and Low Earth Orbit (LEO) based on their respective altitudes from the earth surface. GEO satellites are located at an altitude of 35,786 km. At this altitude, the rotation period of a satellite is approximately equal to 24 hours. Thus, a satellite that is positioned over the equator is stationary with respect to a fixed point on the earth surface. This type of satellites are also referred to as *geostationary*. As mentioned before, GEO systems have large end-to-end delays and high power requirements in the user terminals and the satellites. As a result, it is very difficult to support mobile and interactive communications using GEO satellites. Furthermore, it is difficult to provide small spotbeams inside the satellite coverage area to achieve frequency reuse, since very large antennas in the satellite are needed to realize small spotbeams [12]. To limit the end-to-end delay and power requirements and to use the frequency resources more efficiently through frequency reuse, satellites that are located in lower altitudes can be used. The existence of two Van Allen radiation belts, which contain trapped electrons and protons above the earth atmosphere [4], leads to the classification of the satellite orbits as Medium Earth Orbit (MEO) and Low Earth Orbit (LEO). Van Allen radiation belts are located at altitudes ranging from 1,500 to 5,000 km and from 13,000 to 20,000 km. MEO satellites are located between two radiation belts while LEO satellites are located below the lower radiation belt between altitudes 500 km and 1500 km. As a result of the smaller altitude, MEO and LEO satellites circulate the earth in time periods shorter than 24 hours. Hence, these satellites are not stationary with respect a fixed observer on the earth surface. LEO and MEO satellites are classified also as *non-geostationary satellites*. The speed of the satellites increases with decreasing orbit altitude. Thus, LEO satellites move faster than MEOs. Low altitude provides small coverage areas for individual satellites. Hence, LEO satellite systems require larger number of satellites for global coverage compared to GEO satellites. Small coverage area can be utilized to increase the frequency reuse among the satellites to improve the bandwidth efficiency. Among the non-geostationary satellite systems proposed so far, as mentioned before, LEO satellites are more popular since they can provide lower end-to-end delay and require lower power consumption in the satellites and the user terminals compared to MEOs. However, mobility problems as described in section 3 are more challenging because of their higher speed. Among the existing/proposed nongeostationary systems, *Iridium* and *Teledesic* are LEO systems while *ICO* is a MEO system.

Satellites in the low and medium orbits can further be classified into two groups based on the shape of their orbits. An orbit can be either an elliptical or a circular one. In elliptical orbits, the earth is located in one of the two focal points of the elliptical orbit. The speed of the satellite is highest when it is located farthest from the earth. Similarly, satellite speed is smallest when it is closest to the earth. Elliptical orbits can be designed to increase the visibility period of the satellites over highly populated areas such as northern hemisphere. Among the proposed LEO networks *Ellipso* and *Molniya* utilize elliptical orbits. In *Ellipso*, satellites spend more time when they are serving northern hemisphere. When circular orbits are used, earth is located at the center of the orbit. The altitude of the satellite from the earth's center is constant during the satellite motion. The speed of the satellite is fixed during the rotation. *Iridium* [14,16], *ICO* [22], and *Teledesic* [12] utilize circular orbits. An orbit, either circular or elliptical, has an associated inclination angle that is the angle at which a satellite orbit is tilted relative to the earth's equatorial plane. If the inclination angle is equal to 90 degrees, the orbit is called a *polar orbit* while the others are referred to as *inclined orbits*. Polar orbits intersect over the poles. Hence, they provide maximal coverage over the poles. *Iridium* and *Teledesic* use circular polar orbits to achieve global coverage[1]. Figure 1 gives an example to a system with circular polar orbits. There are six orbits in the network shown in the figure, similar to the *Iridium* system. The satellites in orbits 1 and 6 are counter-rotating, while the satellites in the other adjacent orbit pairs are co-rotating. Since the speed of the satellites in a circular orbit is constant, if the satellites are located in equal distances over the orbit, the system configuration as shown in figure 1 repeats itself with

[1] Note that the orbits of these systems are slightly inclined. Still, these systems are classified as polar systems.

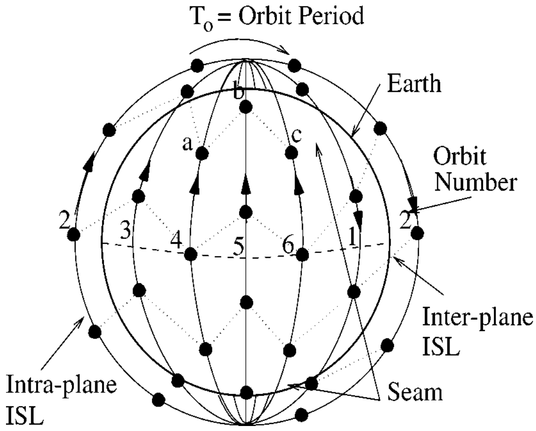

Figure 1. LEO satellite network.

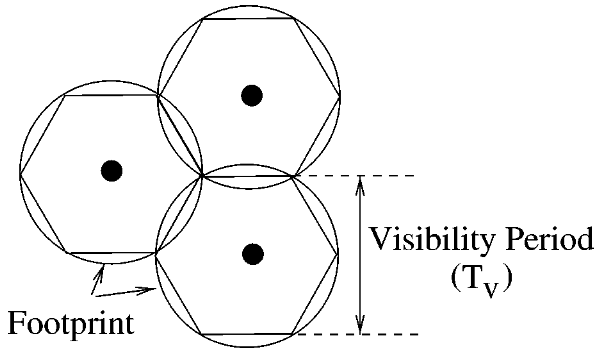

Figure 2. The footprints of the LEO satellites.

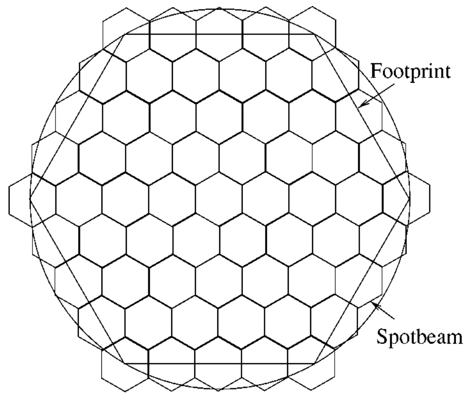

Figure 3. The spotbeams in a satellite footprint.

a period T_{O}, which is referred to as *orbit period*, i.e., if *Loc*(t) is a function that gives the location of the satellites at time t, then *Loc*(t) = *Loc*($t + T_{\mathrm{O}}$). With circular polar orbits, the network resources are inefficiently utilized due to the maximal coverage in the polar regions. To increase the network efficiency, inclined polar orbits can be utilized; however, systems with inclined orbits do not provide global coverage. In this paper, we mainly consider circular polar orbits as the reference model as shown in figure 1. The presented algorithms can be utilized for both orbits with no or little modifications to the algorithms.

2.2. Coverage model

The service area, i.e., the *footprint*[2], of a single satellite is a circular area on the earth's surface in which the satellite can be seen under an elevation angle equal to or greater than the minimum elevation angle determined by the link budget requirement of the system. To give an example, the diameter of the footprint of an *Iridium* satellite is around 4021 km. For complete coverage of the earth's surface, some overlapping between the footprints of the adjacent satellites is necessary. The largest possible *effective footprint* of a satellite is then equivalent to the largest hexagon inscribed into the footprint as shown in figure 2. The visibility period of a satellite, T_{V}, is defined as the maximum time duration that a ground terminal resides in the coverage region of a satellite and can directly communicate with that satellite. The visibility period of a typical LEO satellite is around 8–11 minutes [1]. If it is assumed that only one satellite is visible to a ground terminal (minimal coverage) at any time, it is trivial to show that the visibility period and the orbit period are identical, i.e., $T_{\mathrm{O}} = T_{\mathrm{V}}$. However, note that, because of the the circular coverage of the satellites, some overlapping between the footprints of different satellites is required to achieve global coverage. So, T_{V} can be slightly larger than T_{O}.

The footprints of the individual satellites are covered by smaller cells or *spotbeams* as shown in figure 3 to achieve frequency reuse inside the footprint. Identical frequencies can be "reused" in different spotbeams if the spotbeams are geographically separated to limit interference [19,27]. As an example, each footprint consists of 48 spotbeams with diameters 700 km in the Iridium system [16]. Spotbeams also have hexagonal shapes like footprints, however, they are much smaller in size. As a result, the maximum visibility time of a spotbeam is around 1–2 minutes [1].

LEO satellites move with respect to a fixed observer on the earth surface. Because of this non-stationary characteristic, the coverage area of a LEO satellite changes continuously. A single satellite in the ith polar orbit traces the coverage region of that orbit, R_i, as it circulates the earth. In other words, all satellites in the ith polar orbit have exactly the same coverage region, R_i. But, at a given time, each satellite in the ith orbit handles traffic from a portion of R_i. In general, the coverage regions of two adjacent orbits may overlap with each other as shown in figure 4. Note that the overlapping coverage regions of adjacent orbits are different than the overlapping coverage regions of adjacent satellites. The former results from the movement of the satellites along their orbits while the latter is due to the circular footprints of individual satellites.

2.3. Network connectivity and connection routing

Because of the use of a large number of satellites, it is very probable that the source and the destination terminals of a call have to be served via different satellites. Thus,

[2] Service area, coverage area, and footprint are used interchangeably in this paper.

the end-to-end communication route between two user terminals involves links between user terminals and their respective serving satellites, and a backbone network between the serving satellite pair. The link from a user terminal to a satellite is referred to as *uplink*, while the reverse link is referred to as *downlink*. The links between the user terminals and the satellites are also referred to as *user links*.

The backbone network can be *terrestrial* or *space-based*. In the terrestrial case as shown in figure 5, the satellite serving the source terminal forwards the data received from an uplink to a gateway that forwards the data using a terrestrial network to another gateway that is in communication with the satellite serving the destination terminal. Among the existing/proposed architectures, *Globalstar* and *Ellipso* route their calls through terrestrial networks.

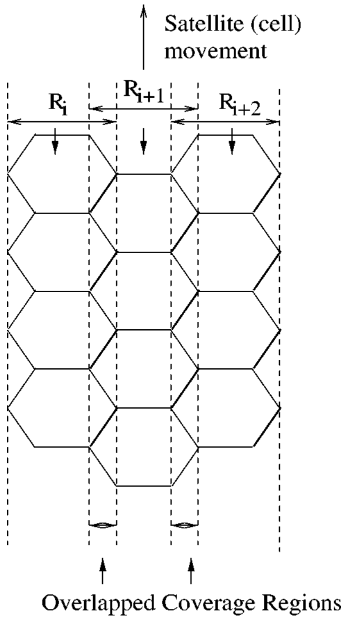

Figure 4. Overlapped coverage regions of adjacent orbits.

In space-based routing, calls are routed through links between satellites that are referred to as *intersatellite links* (ISL) as shown in figures 1 and 6. Space-based routing has advantages over terrestrial routing since terrestrial infrastructure can be destroyed because of wars or natural disasters. Moreover, space network and the terrestrial network can be operated independently by different service providers. Leading existing/proposed satellite architectures such as *Iridium* and *Teledesic* utilize space-based routing.

The network connectivity pattern is assumed to be space-based in this paper. There are two types of ISLs; *intra-plane* ISLs connecting satellites within the same orbit and *inter-plane* ISLs connecting satellites in adjacent orbits. Intra-plane ISLs can be maintained permanently. On the other hand, inter-plane ISLs would be temporarily switched off because of the change in distance and viewing angle between satellites in neighbor orbits. In the analysis reported in [34] for the Iridium system, it is concluded that only ISLs between latitudes of approximately 60° north or south would be maintained between counter-rotating orbits. The regions with latitudes higher than 60 degrees are labeled as *seams* in the example network model depicted in figure 1. The counter-rotating satellites going into the seam temporarily switch their ISLs to the satellites in the neighbor orbits off. Similarly, the satellites passing through polar regions switch their ISLs to the satellites in the neighbor orbits off [33]. Figure 7 depicts the satellites passing through a pole. The drawing reflects the top view, i.e., looking at the pole from a viewing position above the satellites. Satellites a, b, and c (also shown in figure 1) are moving toward the pole. Satellite b's left and right neighbors are satellites a and c, respectively. After passing the pole, the neighbors of satellite b swap their positions. The new satellite posi-

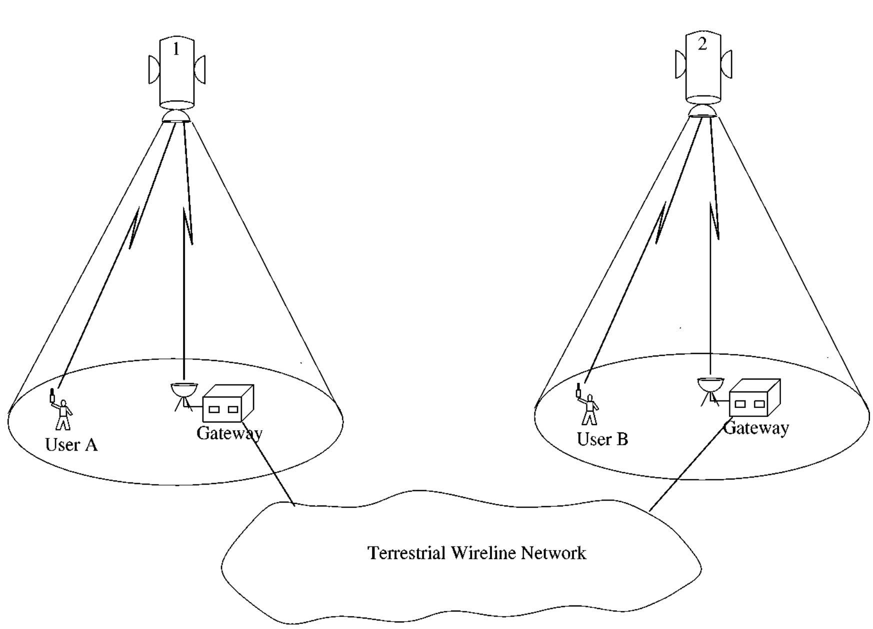

Figure 5. Low earth orbit satellite network with terrestrial routing.

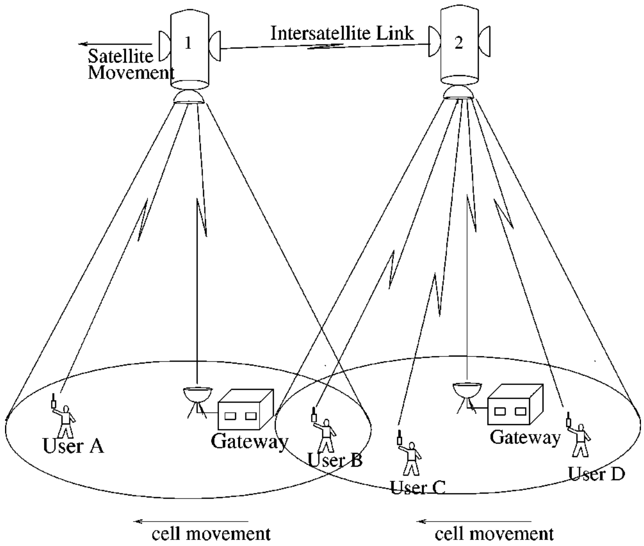

Figure 6. Wireless communication via Low Earth Orbit satellite network.

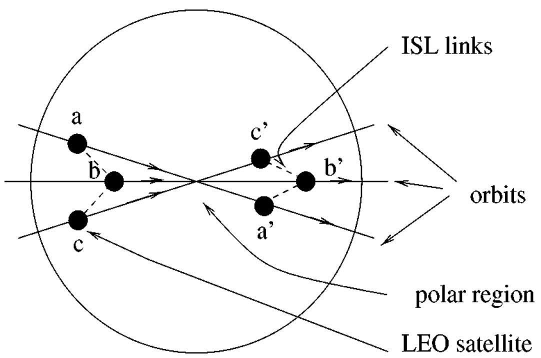

Figure 7. LEO satellites in polar region (top view).

tions are labeled as a', b', and c' in figure 7. During the transition, the ISL links a–b and b–c are turned off. Exact switching times of ISLs are system dependent and beyond the scope of our work. Without loss of generalization, we assume that a satellite passing just over the pole will switch its inter-plane ISL off until its neighbors swap their positions. In figure 7, satellite b turns off its ISLs to satellites a and c when it is just above the pole. The ISLs are restored when satellites a and c swap their positions, i.e., pass over the pole. The ISL connectivity change because of either the seam or the polar crossing results in a dynamic network topology. Thus, the routing algorithm used in the system should be able to cope with the network topology changes.

The route of a connection in a LEO satellite network is determined using a certain *optimality* criterion such as minimum hop or minimum cost [5]. The route established using such a criterion is referred to as an *optimal route*. Because of their high mobility, satellites are not used to determine the optimal route by themselves. Indeed, in the Iridium system [16], connection routes are determined by the gateways as shown in figure 6. As an example, assume that user A wants to communicate with user D with a handheld phone. User A sends a connection request to satellite 1. Upon receiving the connection request, satellite 1 forwards the request to a gateway. The connection route is computed by the gateway and is forwarded to satellite 1, which handles the signaling to establish the connection. The resulting path, in figure 6, is a two-hop route. However, note that the route could have more than two hops.

3. Handovers in LEO satellite networks

Satellite movement and dynamic connectivity pattern result in challenging mobility management problems in LEO satellite networks. Due to the movements of the satellites and accordingly the movements of their coverage area, footprints, a user terminal is served by a number of spotbeams and satellites during a connection. Transfer of a call to a new spotbeam or satellite is referred to as a *handover*. Handovers in satellite networks are classified as:

- *Intersatellite handover:* When the existing connection is transferred from a satellite to another satellite, an intersatellite handover occurs.
- *Link handover:* When the connectivity pattern of the network changes, ongoing calls passing through an ISL that is switched off need to be rerouted. This type of handover is referred to as a link handover.
- *Intrasatellite or spotbeam handover:* When the existing connection is transfered from a spotbeam to a neighbor spotbeam served by the same satellite, an intrasatellite or spotbeam handover occurs.

Handover problems have been investigated in the context of the terrestrial wireless and satellite networks.

3.1. Intersatellite handovers

Intersatellite handovers result in the change of the connection route since a new satellite is involved in the communication between two user terminals. As an example, user A is communicating with user B as shown in figure 8. Communication path consists of satellites 1 and 2 initially. Since the satellites are moving to the left, user B is going to be located in the footprint of satellite 3 shortly. Thus, satellite 3 should be included in the connection route to guarantee the continuation of the call. When a connection is admitted into the network, the system guarantees a certain *Quality of Service* (QoS) such as delay and delay variation (jitter). A connection route that provides required QoS is determined by the network, and the communication between user terminals is initiated. After an intersatellite handover, new connection route is required to provide original QoS to the user. Moreover, the connection rerouting should be performed very fast to avoid the interruption of the call. Thus, connection rerouting algorithm should be simple to implement, but it should also preserve the quality or optimality of the original connection route.

Intersatellite handovers are similar to *interswitch* handovers in terrestrial wireless networks. In a terrestrial wireless network, when a mobile user handovers to a

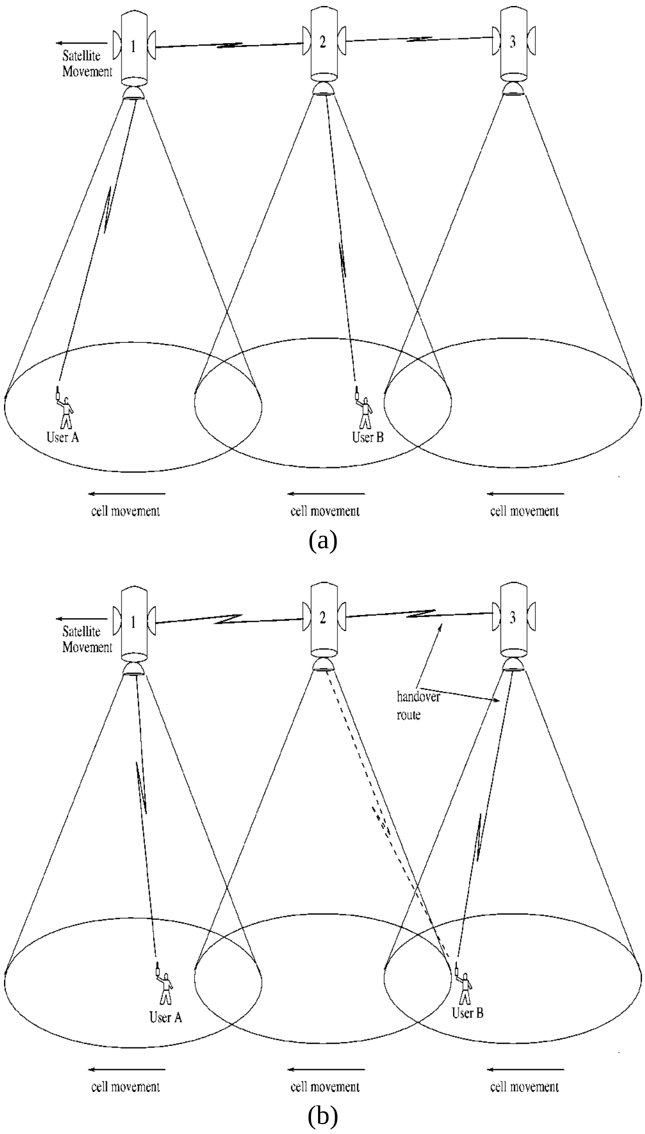

Figure 8. An intersatellite handover scenario: (a) The initial connection route between users A and B passes through satellite 1 and 2. (b) After user B handovers to satellite 3, the new connection route passes through satellites 1, 2, and 3.

cell connected to a Mobile Switching Center (MSC) different than its original MSC, an interswitch handover occurs. Interswitch handover algorithms in terrestrial wireless networks focus on the connection rerouting problem [2,3,20,29]. Basically, there are three connection rerouting approaches: full connection establishment [20], partial connection re-establishment [3,29], and multicast connection re-establishment [2]. Full connection establishment algorithms calculate a new optimum route for the call as if it is a new call request. The resulting route is always optimal; however, the call rerouting delay and the signaling overhead are high. To alleviate these problems, partial connection re-establishment algorithms re-establish certain parts of the connection route while preserving the remainder. This way route update process involves only local changes in the route and can be performed faster. However, the resulting route may not be optimal. In the multicast connection re-establishment algorithm, a virtual connection tree is created during the initial call admission process. The root of the tree is a fixed switching node, while the leaves are the MSCs which would serve the user terminal in the future. Using the multicast connection re-establishment method, when a call handovers to a cell with a new MSC, connection rerouting is immediate due to the already established routes. The disadvantage of this algorithm is that network resources can be underutilized as a result of resources allocated in the connection tree.

A handover rerouting algorithm, referred to as Footprint Handover Rerouting Protocol (FHRP) [31], has been proposed to handle the intersatellite handover problem. The protocol addresses the trade-off between the simplicity of the partial connection rerouting and the optimality of the complete rerouting. The FHRP is a hybrid algorithm that consists of the augmentation and the footprint rerouting phases. In the augmentation phase, a direct link from the new end satellite to the existing connection route is found. This way, the route can be updated with minimum signaling delay and at a low signaling cost. In case there is no such link with the required capacity, a new route is found using the optimum routing algorithm. In the Footprint Rerouting (FR) phase, connection route is migrated to a route that has the same optimality feature with the original route. The goal of the rerouting is to establish an optimum route without applying the optimum routing algorithm after a number of handovers. This property is significant because, in the ideal case, the routing algorithm computes a single route for each connection. The optimality of the original route is maintained after the FR phase. The FHRP requires the user terminals to store information about the connection route. The performance of the FHRP is compared with a static network and pure augmentation. In the former, the network nodes are fixed, hence there is no handover in the network. In the latter, the augmentation phase of the FHRP is applied during the intersatellite handovers; however, if a call is blocked during the path augmentation process, no rerouting attempt is made. The experimental results show that the FHRP performs very similar to the static network and substantially better than the pure augmentation algorithm in terms of call blocking probability. Moreover, handover calls have less blocking compared to the new calls. The FHRP algorithm is applicable to connection-oriented networks. The investigation of the intersatellite handovers in the connectionless LEO satellite networks is an open research area.

3.2. Link handovers

LEO satellite networks have a dynamic connectivity structure resulting from the satellite movement. Satellites near polar regions or counter-rotating satellites in seam turn off their links to other satellites in the neighbor orbits. Ongoing calls passing through these links need to be rerouted. This event is referred to as a *link handover*. Large number of rerouting attempts during the link handover would result in excessive signaling load in the network. Moreover, calls would be blocked because of the insufficient network resources in the newly established routes or large connection re-establishment delay.

The routing in the LEO satellite networks has been investigated in [33] with an emphasis on setting up routes between pairs of satellites to minimize the number of rerouting attempts during link handovers, i.e., optimization was performed for the routes between satellite pairs. Optimization process results in a unique route with minimum number of link handovers during a system period[3] for each satellite pair. All end user connections that are served by the same satellite pair use the same unique route. This algorithm reduces the link handover frequency; however, it can also congest some of the links, while it underutilizes some others. An optimal route between two satellite nodes is not necessarily optimum for a connection between two ground terminals since intersatellite handovers result in changing satellite end nodes for the connection. The optimization is needed for the route between two ground terminals. Moreover, the network connectivity pattern is assumed to be static in the reported simulation study. This assumption is not realistic in the LEO satellite environment.

Recently, the algorithm in [33] has been improved in [32] by introducing a *sliding window* mechanism. In the new algorithm, when optimization process is performed, the routes are determined such that minimum number of handovers occurs in a time window. By sliding the window, new routes are determined after each topology change. This algorithm uses a fixed window size and does not take the call statistics into account. As a result, its performance is sensitive to the relative magnitudes of the call duration and the window size. In [6,7], a LEO satellite network is modeled as a Finite State Automaton (FSA) by dividing the system period of the satellite network into equal-length intervals, where the system period is defined as the least common multiple of the orbit period and the earth period. In this approach, two satellites are defined to be visible from each other in a state if they are within line-of-sight throughout the state. The information about intersatellite visibility within a state is encoded into a visibility matrix. In this manner, the LEO satellite network in a state can be regarded as having a fixed topology. The purpose of the FSA algorithm is to determine an optimum link assignment (e.g., topological design) to make the best use of the limited number of ISL's in each satellite. The algorithm determines the optimum link assignments for each state using the visibility matrix. The optimal link assignment is defined as the one that yields the best performance when the optimal static routing is used. The FSA approach does not address the reduction of the number of rerouting attempts due to the link handovers. In contrast, more connections would need to be rerouted during the state changes of the FSA model since the link assignment is optimized only with respect to the traffic pattern.

In [30], a routing protocol, referred to as Probabilistic Routing Protocol (PRP), has been proposed to reduce the number of rerouting attempts during a link handover. The algorithm removes all the ISLs that are expected to experience a link handover during the lifetime of a connection from consideration for routing during the route establishment phase of a new call. Since the call holding time is a random variable, the connection lifetime can not be determined exactly. Instead the PRP finds the time duration in which the route will be used by the user terminals with a certain probability that is referred to as *target probability*. As a result, the route does not experience any link handover with the target probability. The performance evaluation results shows that a trade-off exists between the value of the target probability and the new call blocking rate.

3.3. Spotbeam handovers

A spotbeam handover, as shown in figure 9, involves the release of user links of the handover terminal in the current spotbeam and the allocation of new user links in the new spotbeam. Since both spotbeams are served by the same satellite, the spotbeam handover is performed by a single satellite, i.e., no other satellite is involved in the process. A handover call would be blocked if the necessary resources are not available in the new spotbeam. Spotbeam handovers are the most frequent type of handovers experienced in the LEO satellite networks because of the small spotbeams and high satellite speed. However, high satellite speed also ease the solution of the spotbeam handover problem, since the user mobility is negligible compared to the satellite speed, i.e., the mobility in the system can be approximated by the deterministic and constant movement of

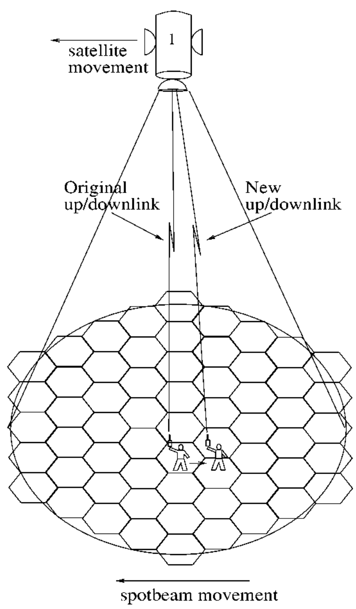

Figure 9. Spotbeam handover scenario.

[3] System period is defined as the time interval where a satellite circulates the earth.

the satellites. Thus, the resource allocation algorithms can utilize this approximation to limit or reduce the handover call blocking probability.

Since blocking of a handover call is less desirable than blocking of a new call request, spotbeam handover algorithms give higher priority to handover calls. Handover prioritization techniques such as the use of the guard channels [15,26], handover queueing [11], handover queueing with dynamic channel allocation [10], and connection admission control algorithms [21,25] have been studied for non-geostationary satellite networks. Guard channels [15,26] are used to ensure that a number of channels is reserved for handover calls even when the new call arrival rate is high. In a system with guard channels, new call requests are rejected if the number of busy channels is larger than a certain threshold. The difference between the system capacity, in number of channels, and the threshold value is equal to the number of guard channels. The handover call blocking rate could be reduced by increasing the number of guard channels. However, the reservation of guard channels for handover calls increases the blocking rate for new arrivals. Hence, there is a trade-off between the handover call blocking and the new call blocking.

The handover queueing algorithm [11] rely on the overlapped coverage regions of the spotbeams involved in the handover process. When a user terminal is in the overlapped coverage region, handover process is initiated. If there is a channel available in the new spotbeam, this channel is allocated to the user terminal. Otherwise, the handover request is queued. When a channel becomes available, one of the calls in the handover queue is served. A handover call is blocked if no channel is allocated for the call in the new spotbeam when the power level received in the current spotbeam falls below the minimum power level for successful data transfer. Handover queueing reduces the handover call blocking ratio, however its performance depends on the new call arrival rate and the size of the overlapped coverage region. In the worst case, high call arrival rates or small overlapped coverage regions would result in a large handover call blocking rate.

A modification to [11] is to use dynamic channel allocation in addition to the queueing of the handover calls [10]. This algorithm performs well for low-to-moderate traffic levels. However, it requires channel reassignment after each call departure, which occurs very often because of the frequent handovers, resulting in extreme overheads for the LEO satellite networks.

In the connection admission control algorithm presented in [21], when a new call request arrives at a spotbeam, it is associated with a list of possible neighboring spotbeams that the user is going to visit with some probability in the future. A mobility reservation metric is updated for each spotbeam in the neighbor list to decrease the handover blocking probability during future handovers. Although the handover call blocking probability is decreased, the algorithm can not guarantee any upper bound. Moreover, because of the ad hoc nature of the reservation metric, the algorithm is conservative, i.e., it underutilizes the network. Finally, the algorithm is evaluated for a MEO satellite network, which has a very low handover probability. The applicability to a LEO network, which has a high handover probability, is questionable.

In the algorithm introduced in [25], fixed users, in addition to the mobile users, are considered. The algorithm eliminates handover call blocking for fixed users. The simulation results show that the presence of fixed user calls with the demanded quality of service leads to an increased satellite capacity requirement, growing linearly with respect to the proportion of fixed users. The algorithm requires that the network is scaled for peak traffic conditions. Moreover, there is no mechanisms to limit the handover call blocking probability for mobile users.

Spotbeam handovers are similar to *intraswitch handovers* in the terrestrial wireless networks. An intraswitch handover occurs if the handover is between cells connected to the same backbone switch, namely *Mobile Switching Center* (MSC). Once a mobile terminal is admitted to a terrestrial wireless network, the continuation of the call should be guaranteed by the network. The user terminal may handover to a neighbor cell. If resources are not available in the cell, the call is blocked. Hence, intraswitch handover algorithms try to reserve bandwidth in every cell for the future handover arrivals. In [24], when a call arrives at a cell, extra bandwidth is reserved in the neighbor cells before the call is admitted into the network. Although handover blocking probability is very low, the algorithm results in excessive new call blocking probability in the network, since most of the bandwidth reserved in the neighbor cells are not utilized. In [9,18], the direction of the user motion is predicted to identify the target handover cell. Bandwidth reservation is made in the target handover cell to avoid handover call blocking in the future. This algorithm results in a more efficient bandwidth usage. However, its performance depends on the success of the motion prediction algorithm.

In the distributed call admission algorithm introduced in [23], upon a call arrival at a spotbeam, the connection admission algorithm estimates the handover blocking rate at a time instant in the future using the information from neighbor cells. Based on the estimated handover blocking probability, the new call is admitted or rejected. The performance of this algorithm also depends on the success of the estimation process.

The performance of the intraswitch handover algorithms depend on the success of the motion estimation algorithms while, in LEO satellite network environment, mobility pattern is deterministic. Thus, handover algorithms that have been developed for terrestrial networks can easily be adapted to LEO satellite network environment.

4. Conclusions and suggestions for future research

In this paper, we presented the fundamental features of the low earth orbit satellite networks. Moreover, we surveyed the handover management algorithms and the proposed solutions in the literature. As we conclude, the majority of the related work covers circuit switched networks with voice traffic. Next generation satellite networks will serve multimedia traffic such as voice, video, and data. This type of traffic is more difficult to serve compared to the voice traffic in circuit-switched networks. First, each traffic type has different Quality of Service expectations. For example, video traffic reqiures bounded and small end-to-end delay and delay jitter; however, it can tolerate packet losses. In contrast, data traffic is very sensitive to the packet losses, but it is insensitive to delay and delay jitter. As a result, handover algorithms should guarantee individual QoS expectations of different traffic types during the handovers. Second, new routing algorithms are needed to serve traffic with QoS expectations. These algorithms should be developed based on the expected traffic characteristics, network architecture, connectivity structure, and handover characteristics. Third, multimedia traffic has longer call holding times compared to the traffic in the traditional circuit switched networks. In the existing work, user mobility and the earth's rotation are assumed to be negligible compared to the speed of the satellites. For multimedia connections, the earth's rotation may no longer be ignored. There is a need to investigate the effects of the user mobility and the earth's rotation. Besides the research problems due to the multimedia traffic, different system architectures such as orbit types and coverage models should be studied. In most of the existing work, it is assumed that the minimum number of satellites is used to achieve global coverage. Thus, the overlapped coverage areas of the neighbor satellites do not constitute a significant portion of the satellite coverage areas. To increase the resources available to the dense population areas, overlapped coverage areas can be utilized. This would also simplify the solution of the spotbeam handover management problem since the increased overlapping would enable better performance for handover queueing approach.

References

[1] F. Abrishamkar and Z. Siveski, PCS global mobile satellites, IEEE Communications Magazine 34(9) (September 1996) 132–136.

[2] A.S. Acampora and M. Naghshineh, An architecture and methodology for mobile-executed handoff in cellular ATM networks, IEEE J. on Selected Areas in Communications 12(8) (October 1994) 1365–1374.

[3] B.A. Akyol and D.C. Cox, Rerouting for handoff in a wireless ATM networks, IEEE Personal Communications 3(5) (October 1996) 26–33.

[4] J.M. Benedetto, Economy-class ion-defying ICs in orbit, IEEE Spectrum 35(3) (March 1998) 36–41.

[5] D. Bertsekas and R. Gallager, *Data Networks* (Prentice Hall, 1992).

[6] H.S. Chang, B.W. Kim, C.G. Lee, Y.H. Choi, S.L. Min, H.S. Yang and C.S. Kim, Topological design and routing for low-earth orbit satellite networks, in: *Proc. of IEEE GLOBECOM* (1995) pp. 529–535.

[7] H.S. Chang, B.W. Kim, C.G. Lee, S.L. Min, Y.H. Choi, H.S. Yang and C.S. Kim, Performance comparison of static routing and dynamic routing in low earth orbit satellite networks, in: *IEEE VTC '96* (1996) pp. 1240–1243.

[8] G. Comparetto and R. Ramirez, Trends in mobile satellite technology, Computer 30(2) (February 1997) 44–52.

[9] S.K. Das, R. Jayaram and S.K. Sen, An optimistic quality of service provisioning scheme for cellular networks, in: *Proc. of the 17th Int. Conf. on Distributed Computing Systems* (1997) pp. 536–542.

[10] E. Del Re, R. Fantacci and G. Giambene, Call blocking performance for dynamic channel allocation technique in future mobile satellite systems, in: *IEE Proceedings* (1996) pp. 289–296.

[11] E. Del Re, R. Fantacci and G. Giambene, Handover requests queueing in low earth orbit mobile satellite systems, in: *Proc. of the 2nd European Workshop on Mobile/Personal Satcoms* (1996) pp. 213–232.

[12] P.R. Giusto and G. Quaglione, Technical alternatives for satellite mobile networks, in: *Proc. of the 1st European Workshop on Mobile/Personal Satcoms* (1994) pp. 15–27.

[13] L.S. Golding and L.C. Palmer, Personal communications by satellite, Int. J. of Satellite Communications 10(5) (September–October 1992) 283–291.

[14] J.L. Grubb, IRIDIUM overview, IEEE Communications Magazine 29(11) (1991).

[15] D. Hong and S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures, IEEE Trans. on Vehicular Technology 35(3) (August 1986) 77–92.

[16] Y.C. Hubbel, A comparison of the IRIDIUM and AMPS systems, IEEE Network Magazine 11(2) (March/April 1997) 52–59.

[17] A. Jamalipour, *Low Earth Orbital Satellites for Personal Communication Networks* (Artech House, 1998).

[18] D.A. Levine, I.F. Akyildiz and M. Naghshineh, A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept, IEEE/ACM Trans. on Networking 5(1) (February 1997) 1–12.

[19] V.H. MacDonald, The cellular concept, The Bell Systems Technical J. 58(1) (January 1979) 15–43.

[20] M. Marsan, C.-F. Chiasserini, R. Lo Cigno, M. Munafo and A. Fugamalli, Local and global handovers for mobility management in wireless networks, IEEE Personal Communications 4(5) (October 1997) 16–24.

[21] I. Mertzanis, R. Tafazolli and B.G. Evans, Connection admission control strategy and routing considerations in multimedia (non-geo) satellite networks, in: *Proc. IEEE VTC '97* (1997) pp. 431–435.

[22] B. Miller, Satellite free mobile phone, IEEE Spectrum 35(3) (March 1998) 26–35.

[23] M. Naghshineh and M. Schwartz, Distributed call admission control in mobile/wireless networks, IEEE J. on Selected Areas in Communications 14(4) (May 1996) 711–717.

[24] C. Oliviera, J.B. Kim and T. Suda, Quality of service guarantee in high speed multimedia wireless networks, in: *Proc. IEEE ICC '97* (1997) pp. 728–734.

[25] J. Restrepo and G. Maral, Providing appropriate service quality to fixed and mobile users in a non-geo satellite-fixed cell system, in: *Proc. of the Second European Workshop on Mobile/Personal Satcoms* (1996) pp. 79–96.

[26] G. Ruiz, T.L. Doumi and J.G. Gardiner, Teletraffic analysis and simulation of mobile satellite systems, in: *IEEE Vehicular Technology Conference* (1996) pp. 252–256.

[27] G.L. Stüber, *Principles of Mobile Communication* (Kluwer Academic, 1996).

[28] M.A. Sturza, Architecture of the TELEDESIC satellite system, in: *Proc. of Int. Mobile Satellite Conf.* (1995) pp. 212–218.

[29] C.-K. Toh, The design and implementation of a hybrid handover protocol for multimedia wireless LANs, in: *Proc. Mobicom '95* (1995) pp. 49–61.

[30] H. Uzunalioğlu, Probabilistic routing protocol for low earth orbit satellite networks, in: *Proc. of IEEE ICC '98* (1998) pp. 89–93.

[31] H. Uzunalioğlu, I.F. Akyildiz, Y. Yesha and W. Yen, Footprint handover rerouting protocol for low earth orbit satellite networks, To appear in Wireless Networks Journal. A preliminary version appeared in the *Proc. of ACM/IEEE MOBICOM '97*.

[32] M. Werner, G. Berndl and B. Edmaier, Performance of optimized routing in LEO intersatellite link networks, in: *IEEE VTC '97* (1997) pp. 246–250.

[33] M. Werner, C. Delucchi, H.-J. Vogel, G. Maral and J.-J. De Ridder, ATM-based routing in LEO/MEO satellite networks with intersatellite links, IEEE J. on Selected Areas in Communications 15(1) (January 1997) 69–82.

[34] M. Werner, A. Jahn, E. Lutz and A. Bottcher, Analysis of system parameters for LEO/ICO-satellite communication networks, IEEE J. on Selected Areas in Communications 13(2) (February 1995) 371–381.

[35] R.A. Wiedeman and A.J. Viterbi, The globalstar mobile satellite system for worldwide personal communications, in: *Proc. of Int. Mobile Satellite Conf.* (1993) pp. 46–49.

Ian F. Akyildiz received his B.S., M.S., and Ph.D. degrees in computer engineering from the University of Erlangen-Nuernberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is a Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, and Director of Broadband and Wireless Networking Laboratory and serves as the elected Chair of Telecommunicatons Area within the School of Electrical and Computer Engineering. He has held visiting professorships at the Universidad Tecnica Federico Santa Maria, Chile, Universite Pierre et Marie Curie (Paris VI), Ecole Nationale Superieure Telecommunications in Paris, France, Universidad Politecnico de Cataluna in Barcelona, Spain, and Universidad Illes Baleares, Palma de Mallorca, Spain. He has published over two-hundred technical papers in journals and conference proceedings. He is an editor for "IEEE/ACM Transactions on Networking", "Computer Networks and ISDN Systems Journal", "ACM-Springer Multimedia Systems", "ACM/Baltzer/URSI Wireless Networks" and "Cluster Computing". He is a past editor for IEEE Transactions on Computers (1992–1996). He guest-edited several special issues, such as on "Networks in the Metropolitan Area" for "IEEE Journal of Selected Areas in Communications". He was the program chair of the "9th IEEE Computer Communications" workshop held in Florida in October 1994. He also served as the program chair for ACM/IEEE MOBICOM '96 (Mobile Computing and Networking) conference as well as for IEEE INFOCOM '98 conference. Dr. Akyildiz is an IEEE fellow and an ACM fellow. He received the "Don Federico Santa Maria Medal" for his services to the Universidad of Federico Santa Maria in Chile. He served as a National Lecturer for ACM from 1989 until 1998 and received the ACM Outstanding Distinguished Lecturer Award for 1994. Dr. Akyildiz received the 1997 IEEE Leonard G. Abraham Prize award for his paper entitled "Multimedia group synchronization protocols for integrated services architectures", published in the IEEE Journal of Selected Areas in Communications (JSAC) in January 1996. His current research interests are in wireless networks, satellite networks, ATM networks, multimedia communication systems.

E-mail: ian@ece.gatech.edu

WWW: http://www.ee.gatech.edu/research/labs/bwn

Hüseyin Uzunalioğlu received his B.S. and M.S. degrees from the Middle East Technical University, Ankara, Turkey, in 1990 and 1992, respectively, and the Ph.D. degree from the Georgia Institute of Technology in 1998, all in electrical engineering. From 1992 to 1996 he was a recipient of a research scholarship from Middle East Technical University. He worked as a researcher at Georgia Tech Research Institute from 1995 to 1998. Dr. Uzunalioğlu is currently with Network Planning Solutions group at Bell Laboratories, Lucent Technologies. His current research interests are in traffic and mobility management for broadband networks. He is a member of IEEE.

E-mail: huseyin@lucent.com

Michael D. Bender is the Chief of Telecommunications Science at the Laboratory of Telecommunications Science. He conducts research into emerging wireless communications systems. Mr. Bender holds a B.S. degree in electrical engineering from Johns Hopkins and is completing M.S. degrees in electrical engineering and computer engineering at Loyola.

E-mail: mdbende@afterlife.ncsc.mil