A CONNECTION HANDOVER PROTOCOL FOR LEO SATELLITE ATM NETWORKS

Hüseyin Uzunalioğlu, Wei Yen and Ian F. Akyildiz

Broadband and Wireless Networking Laboratory
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
{huseyin, wei, ian}@ece.gatech.edu

ABSTRACT

Low Earth Orbit (LEO) satellite networks require a reliable handover re-routing protocol that is critical for connections with multihop intersatellite links (ISLs). In this paper, we introduce a footprint handover re-route protocol (FHRP) that maintains the optimality of the initial route without performing a routing algorithm after satellite handovers. Furthermore, the FHRP handles the inter-orbit handover problem. The FHRP makes use of the footprints of the satellites in the initial route as the reference for re-routing. More specifically, after an optimum route has been determined during the call establishment process, the FHRP ensures that the new route due to handover is also optimum. The FHRP is applicable to any type of connection-oriented networks. The adaptation of the FHRP to the widely accepted ATM technology, which will be employed in future satellite networks, is also addressed in this paper. The FHRP demands easy processing, signaling, and storage costs. The performance results show that the FHRP performs similar to a network without any handovers in terms of call blocking probability.

Keywords: Satellite communications, handover, re-routing, ATM.

1. INTRODUCTION

Terrestrial wireless networks such as cellular and PCS networks provide mobile communication services with limited geographic coverage. In order to provide global coverage to a more diverse user population, a number of *low earth orbit (LEO)* satellite systems have been proposed [8, 11, 15]. The LEO systems can support both the areas with terrestrial wireless networks and areas which lack any wireless infrastructure. In the former case, a satellite system could interact with the terrestrial wireless network to absorb the instantaneous traffic overload of the terrestrial wireless network. In other words, mobile users would alternatively access a terrestrial or a satellite network through dual-mode handheld terminals. In the latter application area, LEO satellites would cover regions where terrestrial wireless systems are economically infeasible due to rough terrain or insufficient user population.

The term LEO is used to classify satellites with orbiting altitudes between 500 and 2000 km above the Earth's surface. This low altitude provides small end-to-end delays and low power requirements for both satellites and terminals. As Figure 1 depicts, users can access LEO satellites with their small handheld phones. Moreover, the satellites can be connected to terrestrial networks via gateways. Another feature of LEO satellites is that intersatellite links (ISL) allow the routing of a connection through the satellite network without requiring any terrestrial resources. To achieve the routing capability, the satellites have to carry packet switches on-board [9, 11, 13]

In contrast to geostationary (GEO) satellites, LEO satellites circulate the Earth at a constant speed. Because of this non-stationary characteristic, the coverage area of a LEO satellite changes continuously. The global coverage at any time is still possible if a certain number orbits and satellites are used. As an example, the IRIDIUM project uses 6 polar orbits with 11 satellites in each orbit [8]. Due to the non-stationary coverage regions of individual satellites, the source or the destination terminals on the ground may not stay in the coverage region of the initial source or destination satellites throughout the communication. Thus, the initial source and the destination satellites may need to transfer the ground source and destination terminals to other satellites whose coverage regions contain the ground source and destination terminals. This event is called as *handover*. There are two types of handovers in LEO systems: *intra-orbit* and *inter-orbit* handovers. The former refers to handovers between two adjacent satellites in the same orbit, while the latter refers to handovers between two satellites in adjacent orbits. In LEO systems, handovers occur frequently because of the relatively high speed of the satellites (19,000 to 25,000 km/h). A handover may result in the addition of new satellites in the existing connection route. During a handover, the existing connection route should be updated accordingly. In the extreme case,

MOBICOM 97 Budapest Hungary

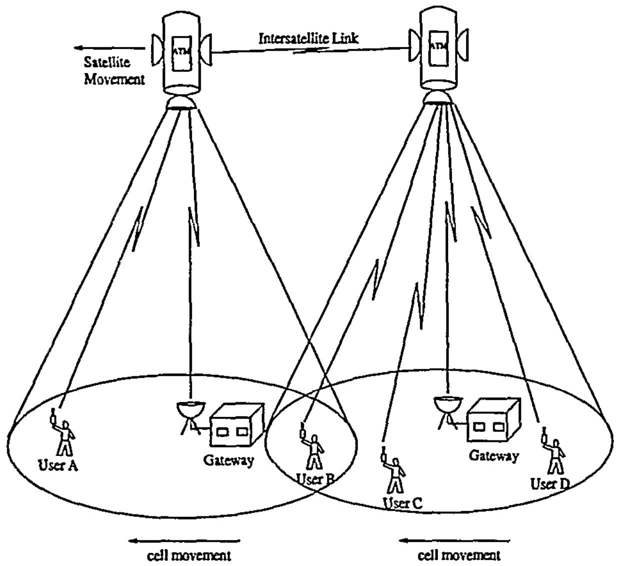


Figure 1: Wireless Communication via Satellite ATM Network.

a whole new route may have to be set up for the communication.

When a call is accepted to the network, the connection route is determined using a certain *optimality* criterion such as minimum hop or minimum cost. The route established using such criterion is called as *optimal route.* Due to their high mobility, satellites are not used to determine the optimal route by themselves. Indeed, in IRIDIUM system [9], the connection routes are determined by the gateways as shown in Figure 1. As an example, assume that user A wants to communicate with user D with a handheld phone. The connection request is sent to the gateway via the LEO satellite covering user A. The gateway locates user D and computes the multihop satellite route accordingly. Then, the route information is sent to the satellite which handles the signaling to establish the connection. The resulting path, in Figure 1, is a 2-hop route. However, in other cases, the route could have more than 2 hops. This procedure causes signaling overhead between the satellites and the gateways. Thus, minimizing the number of routing operations performed by the gateways would enhance the performance of the network. In general, appending new nodes to an optimal route may result in a sub-optimal route as the augmented route does not necessarily satisfy the initial optimality criterion. This may cause inefficient utilization of satellite resources. This problem can be alleviated by re-routing the connection whenever a handover is necessary. Re-routing introduces additional signaling and processing overhead between satellites and gateways. However, the route between the source and destination terminals remains optimal after the handover is performed. The handover re-routing problem has not been adequately addressed in the context of satellite networks. In this paper, we propose a handover scheme called *Footprint Handover Re-routing Protocol* (FHRP) which consists of two phases: route augmentation and re-routing. The motivation is to balance the optimality of re-routing with the simplicity of the path augmentation. This protocol addresses both intra- and inter-orbit handovers. Moreover, the re-routing phase in the FHRP can be handled by satellites without any intervention of gateways. Therefore, the signaling and computation overhead can be reduced in the gateways.

The FHRP is applicable to any type of connection-oriented networks. In this paper, we present two algorithms to demonstrate that the FHRP can be adapted to the ATM technology, which will be employed in the future satellite networks to support multimedia traffic [11, 13]. In ATM networks, the information packets are fragmented into 53-byte ATM cells with 48-byte of information field. The ATM cells basically do not carry any ordering information, and thus, in-sequence cell delivery is vital for the correct re-assembly of the cells into packets at the destination. Hence, the connection handover scheme should guarantee in-sequence delivery of the ATM cells. The FHRP has been extended to guarantee the in-sequence cell delivery during handover instants. Moreover, a signaling algorithm is also presented to realize the route augmentation in the ATM environment.

The remaining of this paper is structured as follows: In Section 2 we briefly review the state-of-the-art on satellite handover schemes. In Section 3 we introduce the footprint handover re-routing protocol (FHRP). We also describe how the FHRP can handle intra-orbit and inter-orbit handovers. In Section 4 we describe the algorithms necessary for the ATM adaptation. In Section 5 we investigate the performance of the algorithm. Finally, we discuss the future research directions and conclude the paper in Section 6.

2. RELATED WORK

Satellite handover problems have become an active research area recently [6, 10]. An analytical model has been proposed to calculate the handover rate for single-hop satellite connections in [6]. The model only considers intra-orbit handovers. Due to the single-hop nature of the connections, no re-routing scheme is proposed. In a more recent study [10], inter-orbit handovers are addressed in a single-hop network environment. After developing call blocking probabilities for new calls and handovers, authors investigated a handover prioritization strategy based on the queueing of handovers. This study lacked the support for a multihop handover scheme. Although neglected in the existing literature, multihop connection routing is necessary in mobile satellite networks, since, even in the case of a connection between two parties near to each other, the source and destination terminals would be covered by different satellites; hence, necessitating at least two satellites for the connection.

The multihop satellite routing problem has been addressed in [13] with an emphasis on setting up routes between pairs of satellites to minimize the re-routing frequency, i.e., optimization was performed for the routes between two satellites. Realistically, the optimization is needed for the route between two ground terminals. An optimal route between two satellite nodes is not necessarily optimum for a connection between two ground terminals since the handovers between the ground terminals and the satellites result in changing satellite end nodes for the connection. The

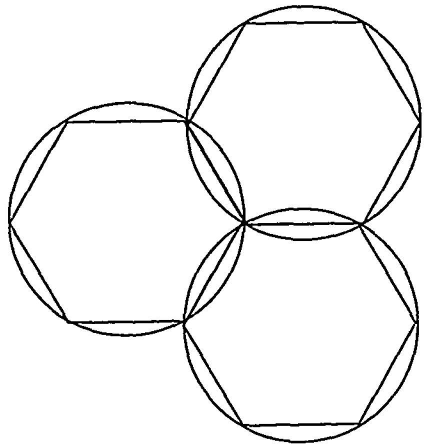


Figure 2: The Footprints of the LEO Satellites.

study did not address the handover re-routing problem.

The handover re-routing problem has been studied in the context of terrestrial wireless networks [1, 7, 12]. For example, a whole new route is established after a handover in [7]. Although an optimum route is used all the time, frequent re-routing would cause excessive signaling and computational overhead due to the optimum route calculation. Partial re-routing algorithms have been proposed in [1, 12]. These algorithms basically make use of a tree-based structure for the network. During a handover, the node which is a parent of both nodes involved in the handover are determined, and the route between the parent and the original end node is replaced with a route between the parent and the new end node. Even though partial re-routing algorithms cause much less overhead in the network compared to the new route establishment, the route after the handover is not optimum.

3. THE FOOTPRINT HANDOVER RE-ROUTING PROTOCOL (FHRP)

The service area, i.e., the *footprint*[1], 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 or greater than the minimum elevation angle determined by the link budget requirement of the system. For a complete coverage of the Earth's surface, some overlapping between the footprints of 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.

In the LEO system described, the satellites are moving in O circular polar orbits. Each orbit has L satellites and the total number of satellites is $N = O \cdot L$. 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 10 minutes. The period of an orbit T_O, on the other hand, is the minimum time interval required for the location of the satellites sharing a common orbit to repeat itself. If $Loc(t)$ is a function that gives the location of the satellites at time t, then $Loc(t) = Loc(t + T_O)$. 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., $T_O = T_V$[2]. The case in which multiple satellites are visible to a ground terminal is also possible if more than the minimum number of satellites for global coverage is used. In this case, the user can pick any of the visible satellites subject to a certain objective such as the Signal-to-Noise Ratio (SNR) maximization or maximum time to a possible handover [4]. After the selection of a particular satellite, the handover problem is identical to the handover problem in a satellite network with minimal coverage. Thus, for the sake of presentation clarity, the satellite network discussed in this paper is assumed to provide minimal coverage, and $T_O = T_V$. The analytical relations between the orbit altitude, satellite speed, total number of orbits and satellites, and the visibility period can be found in [14].

A single satellite in the i^{th} polar orbit traces the coverage region of that orbit, R_i, as it circulates the Earth. In other words, all satellites in the i-th polar orbit have exactly the same coverage region, R_i. But, at a given time, each satellite in the i^{th} 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 3. Note that the overlapping coverage regions of adjacent orbits are different than the overlapping coverage 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.

Each satellite has up and down wireless links for communication with ground terminals and I intersatellite links (ISL) for communication between satellites. The ISLs that connect the adjacent satellites in the same orbit are called as *intraorbit* ISLs while the ISLs that connect neigbor satellites in adjacent orbits are called as *inter-orbit* ISLs. Intraplane ISLs are permanent while interplane ISLs might be turned off temporarily when the satellites are crossing polar regions [14]. Moreover, left and right neighbor satellites of a satellite crossing polar regions switch their positions, i.e., the left neighbor becomes the right neighbor and vice versa. This results in a dynamic, but deterministic, network topology. Routing strategies that handle dynamic LEO satelite network topology have been investigated in [5, 13]. Note that the routing problem and the re-routing protocol described in this paper are orthogonal to each other. Hence, for the clarity of the presentation, it is assumed that the LEO satellite network described in this paper has static topology and each satellite has 2 permanent intraorbit and 2 permanent inter-orbit ISL links as shown in Figure 4. Moreover, in the IRIDIUM system, the orbits 1 and 6 are counter-rotating, i.e., the satellites in the neighbor orbits 1 and 6 rotate in reverse directions which result in a *seam*. It is very difficult to maintain the ISL links accross the seam and, thus, they are turned off [9]. The network depicted in Figure 4 is a seamless one, but the protocol described in this paper

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

[2]Due to 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.

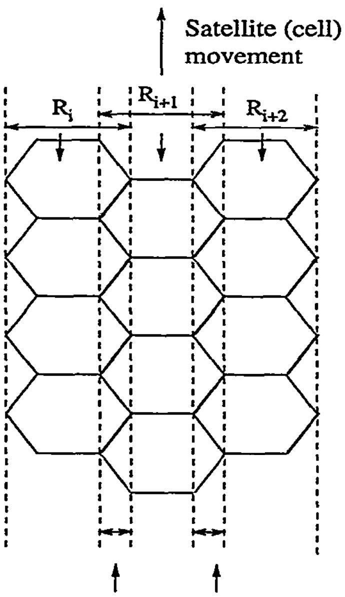


Figure 3: Overlapped Coverage Regions of Adjacent Orbits.

handles the networks with a seam as explained later in this section.

In the remainder of this paper, it is assumed that a connection is routed through a number of LEO satellites for the source and the destination terminals on the ground. The route is denoted by $S_1 \rightarrow S_2 \rightarrow ... \rightarrow S_K$ where K is the number of satellites in the route. S_1 and S_K are called as the *source* and *destination* satellites, respectively. For the sake of clarity, S_1 and S_K are also labeled as S_s (source satellite) and S_d (destination satellite), respectively. The ordered set of satellites involved in the route is called as the *routing set*, $\mathcal{A}$, i.e., $\mathcal{A} = \{S_1, S_2, ..., S_K\}$. Assume that the connection is set-up at $t = t_e$ using a routing algorithm such as minimum hop or minimum cost. If the route is optimum, we use the notation $\mathcal{A}_{opt}$, $S_{s_{opt}}$ and $S_{d_{opt}}$ for $\mathcal{A}$, S_s, and S_d, respectively. The optimum routing set and the current routing set are identical if no handover is performed after the optimum route is established.

Given the limited visibility period and the high speed of the satellite, which is much faster than the velocities of mobile ground terminals, it is realistic to assume that ground terminals are stationary in this specific environment [6, 10]. In other words, all handovers are caused by the mobility of the LEO satellite instead of the ground terminal. In the case of intra-orbit handovers, the takeover satellite that assumes the responsibility for the ground terminal is called the *successor* of the handover satellite. However, as shown in Figure 3, if a ground terminal is located in the overlapping coverage regions of two adjacent orbits, the inter-orbit handover may occur. In this scenario, the takeover satellite is referred as the *neighbor successor* of the handover satellite because these two satellites reside in the neighbor (adjacent) orbits. Also note that, due to the overlapping footprints of the adjacent satellites, the ground terminal would communicate with two satellites simultaneously when it is located

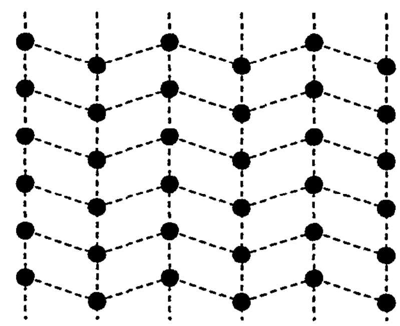


Figure 4: Intersatellite Links between Adjacent LEO Satellites.

in the overlapping region.

Definition: The re-routing of a connection passing through satellites $S_1 \rightarrow S_2 \rightarrow ... \rightarrow S_K$ to $S_1' \rightarrow S_2' \rightarrow ... \rightarrow S_K'$ where S_i' is the successor satellite of S_i is called **Footprint Re-routing** (FR).

Theorem 1: Let P be a multihop LEO satellite route established at time $t = t_e$. Also, let P' be another route determined by the footprint re-routing of P. Then;

A. If P is a minimum hop route between S_s and S_d, then P' is a minimum hop route between S_s' and S_d'.

B. If P is a shortest path route between S_s and S_d and the link cost is a function of the time-homogeneous traffic load, then P' is a shortest path route at time $t = t_e + T_O$.

C. If P is a shortest path route between S_s and S_d, then P' is a shortest path route between S_s' and S_d' under the assumption that link cost is a function of time- and location-homogeneous traffic load.

Proof: The first part of the proof is trivial since there is a one-to-one correspondence between the nodes of the original and the footprint re-route. The proof of the second part is based on an imaginary terrestrial network where a one-to-one correspondence between each terrestrial node and a satellite switch exists. The relation between the satellites and the nodes of the imaginary terrestrial network is determined by two functions, f and g. The function f maps the satellites in the network to the ground points. Each ground point, s_i, in the mapping corresponds to the center of coverage of a unique satellite at time t, i.e., $f(S_i, t) = s_i$ for $i = 1, .., N$ where S_i denotes the i^{th} satellite of the system and N is the number of satellites in the system. The function g maps a ground point s to a satellite at time t if $f(S_i, t) = s$ for any i in the set $\{1, 2, ..., N\}$. The value of $g(s, t)$ is undefined if there is no satellite i that results in $f(S_i, t) = s$.

Next, we construct a ground network which has the same topology of the satellite network at time t_e. Specifically; if there exists a route between satellites S_i and S_j shown as $\overline{S_iS_j}$, then there exists a route between ground nodes s_i and s_j shown as $\overline{s_is_j}$ where $s_i = f(S_i, t_e)$ and $s_j = f(S_j, t_e)$. Also, the cost function associated with the satellite and the ground routes are equal, i.e., $C(\overline{S_iS_j}) = C(\overline{s_is_j})$ where $C(\cdot)$ is the cost function. Suppose the traffic load in

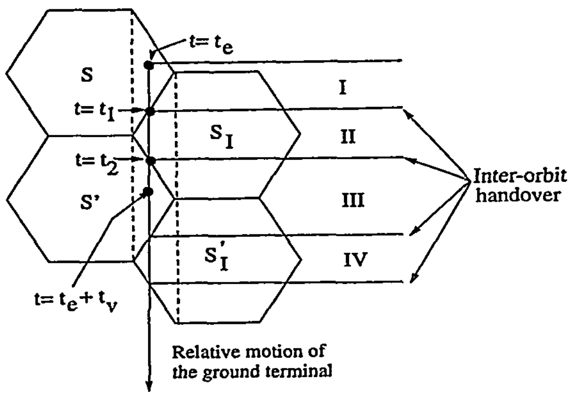


Figure 5: Successive Inter-orbit Handovers.

location (x, y) at time t_e is $L(t_e, (x, y))$. An optimal routing algorithm A can compute the optimal path P for a source and a destination pair, i.e., $A(L(t_e, (x, y))) = P$. For time-homogeneous traffic load, the time reference can be ignored. Thus, the path P is optimal along time in the ground network. On the other hand, $f(S_i, t_e) = s_i = f(S_i', t_e + T_O)$. Thus, the optimum route in the ground network corresponds to the satellite route P at time t_e and to the satellite route P' at time $t = t_e + T_O$ where P and P' correspond to the paths $S_1 \rightarrow S_2 \rightarrow ... \rightarrow S_K$ and $S_1' \rightarrow S_2' \rightarrow ... \rightarrow S_K'$, respectively.

The proof of part **C** is trivial since the shortest path route is also minimum hop route if the traffic is time- and location-homogeneous, i.e., the cost of each ISL link is identical at any time. Thus, the route P' is also a shortest path route all the time. □

The goal of the Footprint Re-routing (FR) is to find an optimum route without performing the optimum route finding algorithm after a handover. Theorem 1 guarantees the optimality at all times if the minimum cost algorithm is used to obtain the original route. On the other hand, the optimality is guaranteed at certain time instants if the original route is obtained using location dependent traffic load. Even in the case of an inter-orbit satellite, FR is possible after the second inter-orbit handover. Figure 5 demonstrates this situation for a connection established at $t = t_e$. For simplicity, only one of the end-satellites (either source or destination) is shown. Also, for the sake of clarity, the footprints of the satellites are stationary, but the terminal moves with a speed relative to the satellites. The ground terminal is served by the original end-satellite S initially (region I). At $t = t_1 > t_e$, the first inter-orbit handover occurs. The ground terminal is served by S_I (region II) until $t = t_2 > t_1$ when the second inter-orbit handover occurs. After $t = t_2$, the ground terminal is served by S' that is the successor of the original end satellite S. Parts A and C of Theorem 1 become applicable when the ground terminal enters the cell served by S', and FR can be applied after the second handover. Part C of the Theorem 1 is also applicable when $t = t_e + t_V$. This example demonstrates that for a connection experiencing multiple inter-orbit handovers, the FR can be applied after even-numbered handovers.

A handover is necessary when one of the end satellites, either the source or the destination satellite, goes out of the visibility region of the ground terminals involved in the communication. The connection should be transferred to either the successor or the neighbor successor satellite that is visible to the ground terminal that requires the handover. It is not possible to use the footprint re-routing at this instant since FR replaces both end satellites with their respective successors. In other words, FR is only possible when the new end satellites are the successors of the end satellites in the original route. Hence, there is a need for a mechanism to handle the routing problem until the FR becomes applicable. The mechanism that we propose for this task called as the *Footprint Handover Re-routing Protocol* (FHRP) consists of two phases. In the first phase which is called *augmentation*, a direct link is set up between the new end-satellite and the original route. If no such link exists, the connection is re-routed using the original routing algorithm. FR algorithm is applied after both of the end-nodes become the successor of the original end nodes. During a handover process, the ground terminals decide whether the augmentation or the FR should be used. The decision depends on the current time, the set-up time of the most recent optimum route, t_e, the routing set $\mathcal{A}_{opt}$ of the optimum route and current routing set $\mathcal{A}$. The mobile terminals keep this information during the lifetime of their connections. The storage overhead of FHRP is discussed in Section 3.3.

3.1. Augmentation Algorithm:

In this section, the augmentation algorithm for the source node of a connection is described. It is shown later that the same algorithm is also applicable to the destination node. Assume that the most recent optimum route establishment has been performed at $t = t_e$. At $t = t_1 > t_e$, the ground terminal goes out of the coverage region of the source satellite, S_s. Since the global coverage is guaranteed, a new satellite which is denoted as S' covers the mobile terminal. S' is either the successor (intra-orbit handover) or the neighbor successor (inter-orbit handover) of S_s as shown in Figure 6. Assume that the necessary conditions as explained in the next section for FR is not held at this moment, and, thus, the source terminal decides to start the augmentation algorithm. A service request message including the current routing set, $\mathcal{A}$, is sent to S' to initiate the augmentation algorithm. The rest of the algorithm is handled by S' as follows:

1. The satellite S' checks whether the handovering connection can be supported by its up-/downlinks. The connection would be blocked due to insufficient capacity in the uplink channels. Moreover, the connection blocking is also possible due to insufficient capacity in the downlink the destination terminal is also in the coverage region of S'. If there is no sufficient bandwidth in the up and/or downlink(s), the connection is blocked and the source terminal releases the previous route.
2. If there is sufficient capacity in up- and/or downlinks, the new source satellite, S', first checks whether it is already serving the connection, i.e., it is checked

whether S' is already in the routing set $\mathcal{A}$. If the result is positive, say $S' = S_i$ where $i = 2, 3, ..., K$, the portion of the route up to S' is deleted and the new route becomes $S_i \to S_{i+1} \to ... \to S_K$. The new routing set, $\mathcal{A}$, is sent to the ground terminals.

3. If S' is not in $\mathcal{A}$, a direct link to one of the satellites in $\mathcal{A}$ is searched starting with the last member (satellite with the largest index) of $\mathcal{A}$. This is because a link to a satellite with high index number results in a route with shortest length. If a direct link with sufficient capacity to support the connection is found, the link is augmented to the original route. As an example, assume that a link between S' and S_i is found. Then, the new route is $S' \to S_i \to S_{i+1} \to ... \to S_K$. The unused portion of the previous route, $S_1 \to S_2 \to ... \to S_{i-1}$, is removed. The handover process is completed after the ground terminals are informed about the route changes.

4. If a direct link between S' and the nodes in $\mathcal{A}$ with required capacity is not found, the original routing algorithm is performed. If a route with required capacity is found, the resulting routing set, $\mathcal{A}_{opt}$ and route establishment time are sent to the ground terminals.

Note that Step 4 handles the handovers between satellites in counter-rotating orbits that was defined as the seam earlier in this section. Since no ISL is maintained between the satellites in counter-rotating orbits, Step 4 enforces the connection to be completely re-routed.

The described augmentation algorithm can be applied to a handover involving the destination ground terminal with minor changes. To avoid simultaneous handovers of the source and the destination satellites of a connection, the destination terminal sends a handover request message to the source terminal. If the source terminal is in the process of a source handover, the destination's request is held until the source handover is completed. The source sends a handover permission message back to the destination that follows a similar augmentation process. Only difference is that, when a direct link is searched between S' and $\mathcal{A}$ in Step 3, the satellite with smallest index is checked first. We omit the details of the destination augmentation algorithm because of the space limitations. Upon completion of the augmentation process, the new routing set is sent to the source terminal.

3.2. Footprint Re-routing Phase:

A disadvantage of the augmentation algorithm is that, when the route is augmented several times, the resulting route will not hold the routing optimality criterion of the original route. Thus, there is a need to update the route at certain time intervals. The selection of the update interval is important since the frequent re-routing attempts waste the network resources while a large re-routing interval results in the use of a non-optimal route for the connection most of the time. Here, we use Theorem 1 to solve the routing update interval problem. Theorem 1 states that the optimality of the original route is preserved in the Footprint

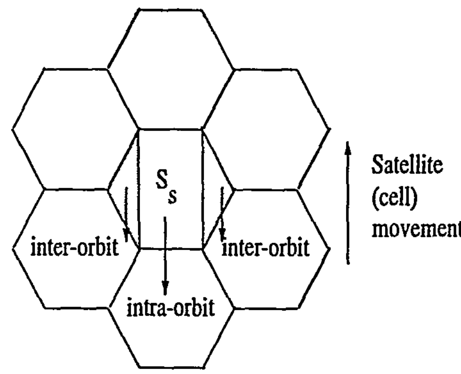


Figure 6: Inter- and Intra-handover in Augmentation Algorithm.

Re-route at certain times. The time when the Footprint Re-route is optimum would only be known by the end terminals and, thus, FR should be initiated by one of the end terminals. Here, we assume that the FR is initiated by the source terminal. The route update time is based on the establishment time of the most recent optimum route t_e, optimum routing set $\mathcal{A}_{opt}$ and the current routing set $\mathcal{A}$.

The first requirement to apply the FR is that the source and the destination satellites of the current route are the successors of the source and the destination satellites of the original route, respectively, i.e., $S_d = S'_{d_{opt}}$ and $S_s = S'_{s_{opt}}$ where $S'_{d_{opt}}$ and $S'_{s_{opt}}$ are the successors of the source and the destination satellites of the optimum route. Second requirement depends on the nature of the traffic load of the system and the optimum routing criterion based on Theorem 1 as follows:

1. If the traffic load is time- and location-homogeneous or the original route is a minimum hop route, FR can be applied anytime when the successor end nodes are serving the connection, $S_d = S'_{d_{opt}}$ and $S_s = S'_{s_{opt}}$.
2. If the traffic load is only time-homogeneous and the routing criterion depends on the traffic load (minimum cost routing), the FR should be performed at $t = t_e + t_O$.

One of the above conditions is chosen based on the traffic and the routing criterion of the network and is applied for all connections. FR is initiated with a re-routing request sent from the source terminal to the source satellite. The re-routing request includes the optimum route set $\mathcal{A}_{opt}$. The source satellite tries to establish a connection traversing the optimal route. If the connection re-routing is performed successfully, i.e, no blocking occurs, the current route is removed. Upon the completion of the re-routing process, the source and the destination terminals update their routing information.

3.3. Storage Requirements

The storage requirement of the route information does not introduce major overhead for the ground terminals, since the longest possible loop-free route in a LEO network is bounded by $N-1$ where N is the number of satellites in the network. This result is trivial since the longest loop-free route from the source to the destination passes through all the nodes in the network only once. The length of such a route is equal to $N-1$. When the minimum hop routing algorithm is used, the bound for the length of the route is smaller as proven in the following theorem.

Theorem 2: Assume that P is a loop-free route in a LEO satellite network and $Length(P)$ is the length of the route given in number of links. If P is a minimum hop route, then $Length(P) \leq \lfloor \frac{O}{2} \rfloor + \lfloor \frac{L}{2} \rfloor + 2$ where O is the number of the orbits and L is the number of satellites per orbit.

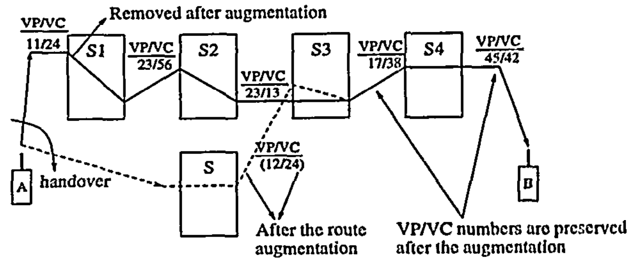
Proof: The proof is based on the connectivity structure of the network and the properties of the minimum hop routing algorithm. Assume that the orbits are indexed as R_i for $i = 1, ..., O$ where O is the number of orbits. The orbits R_i and R_{i+1} are adjacent to each other. Moreover R_1 and R_O are also neighbors due to the circular symmetry of the system. Thus, the maximum length of a route between two orbits is equal to $\lfloor \frac{O}{2} \rfloor$. The satellites in each orbit are indexed similarly as S_i for $i = 1, ..., L$ where L is the number of satellites per orbit. Each satellite has direct links with its up/down and left/right neighbors. Thus, S_i can communicate with S_{i+1} with a direct link. Similar to the orbits, S_1 and S_L have a direct link between them due to the circular symmetry of the system. The maximum length of a route between two satellites sharing the same orbit is equal to $\lfloor \frac{L}{2} \rfloor$. The length of the minimum hop route between two satellites in different orbits is equal to the sum of the maximum distance between the orbits and the maximum distance between the satellites sharing the same orbit, i.e., the length of the minimum hop routes is bounded by $\lfloor \frac{O}{2} \rfloor + \lfloor \frac{L}{2} \rfloor$. The constant term 2 in the Theorem is due to the augmentation algorithm. As explained in Section 3.1, the augmentation algorithm would extend the route at most 1 link. In the worst case, the augmentation algorithm is applied twice (one for each end terminal for inter-orbit handover). Thus, the length of the worst-case minimum hop route is bounded by $\lfloor \frac{O}{2} \rfloor + \lfloor \frac{L}{2} \rfloor + 2$. □

4. ADAPTATION OF THE FHRP TO ATM SWITCHING

The FHRP is applicable to any type of connection-oriented networks. Since the ATM platform is a promising multimedia networking technology, it will be deployed in future LEO satellite networks [13, 11]. Thus, in this paper, we address the issue of adapting the FHRP to ATM switching. First problem addressed in this section involves the modification of the routing tables in the ATM switches to realize the connection augmentation. The second problem deals with the in-sequence delivery of the ATM cells to the destination during a route change.

4.1. Routing Table Update

In ATM networks, the routing is performed using the Virtual Path (VP) and Virtual Circuit (VC) numbers inserted into the cell header. Each switch, upon receiving a cell from an input port, determines the VP/VC numbers in the cell header and finds the corresponding entry in the routing table. The entries in the routing table match the input port and the incoming VP/VC numbers to the output port and the outgoing VP/VC numbers. Thus, the cell is forwarded to the output port after its VP/VC information is replaced with the new values determined from the routing table. Based on this algorithm, ATM switches have no global knowledge about the connections, i.e., the switch can distinguish different connections only by their incoming port and VP/VC numbers. No information regarding the source or destination identification is known by the switch. The lack of global information about connections poses a problem for the augmentation phase of the FHRP. Assume the scenario depicted in Figure 7. Connection traversing the route $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$ is being re-routed through $S \rightarrow S_3 \rightarrow S_4$. The FHRP requires that S communicates with S_3 and a path between S and S_3 is set-up for the connection. The problem is how S_3 would distinguish this connection from the other connections. With the existing routing scheme, only way is to send a signaling message to S_3 through S_1 and S_2 with the initial VP/VC numbers (11/24 for this example). S_1 and S_2 could translate the VP/VC numbers to the related routing information according to their routing table. Upon receiving this signaling message, S_3 would determine the connection, specified by the local VP/VC numbers and the incoming port number, is involved in the augmentation process. Although it can be implemented in ATM networks, this algorithm introduces a signaling overhead into the network. This overhead can be alleviated using an identification scheme in which each switch has global information besides the local VP/VC information about the connections.



GIT: AB1

Routing Table of S3

Incoming		Outgoing	
Port	VP/VC	Port	VP/VC
4	23/13	4	17/38
3	12/24	4	17/38

Figure 7: Routing Table Update Algorithm.

During the connection set-up, each switch negotiates with their upstream and downstream switches[3]. During this phase, a unique connection identification number, or **connection_id**, is also negotiated among the switches on the route. The negotiation of the **connection_id** would be time-consuming if a random number is chosen since a randomly chosen **connection_id** would be in use in one of the switches in the connection path. A simple rule to determine the **connection_id** is needed to avoid this problem.

[3]The upstream and the downstream switches are the switches preceding and subsequent to the switch in discussion on the route of the connection.

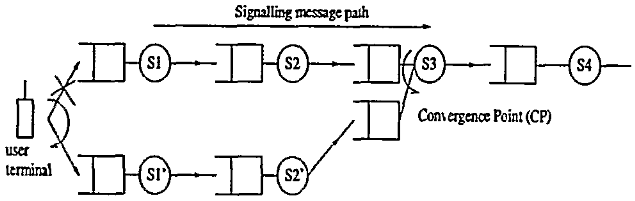


Figure 8: In-sequence Cell Delivery Algorithm.

Note that each user terminal has a unique user identification number, or **user_id** for addressing and billing purposes in commercial systems [9]. A unique **connection_id** would be formed by concatenating the source and the destination **user_id**'s. If two communicating user terminals would have more than one connection simultaneously, an index number would be added to the **connection_id**. The index number would be chosen as the index of the connections to the same destination for the source terminal. As an example, assume that the source and the destination **user_id**'s are A and B, respectively. If there is only one connection between A and B initiated by A, the **connection_id** is $AB1$. If A initiates another call to B, this connection will have a **connection_id** of $AB2$. It is obvious that the **connection_id**'s determined by this algorithm are unique.

When a connection is set-up, each connection is given a connection identification number as explained. The **connection_id**'s of the calls served by the switch is kept in a *Global Information Table* (GIT). Each entry in the GIT has a pointer to the routing table that relates the incoming port and the VP/VC numbers to the outgoing port and the VP/VC numbers. Upon arrival of a re-routing request as shown in Figure 7, the switch checks the **connection_id** in the incoming signaling message and a new entry is placed in the routing table with outgoing VP/VC numbers identical to those of the original routing entry. The old entry is kept until the *terminate connection* signal is received from S_1 as explained in the following subsection. The introduction of the GIT requires additional functionality in the ATM switch fabric compared to the ATM switch fabrics used in B-ISDN networks. Given the high cost of building a satellite network [2], the cost increase in the ATM switch fabric due to the use of GIT would be a good trade-off to achieve a lower signaling overhead.

4.2. In-sequence Cell Delivery

The ATM technology requires that the transmitted cells arrive to the destination in the order that they have been sent by the source, i.e., in-sequence cell delivery should be guaranteed. When a handover occurs, the order of the ATM cells would change due to the route augmentation or complete re-routing. Thus, a mechanism is needed to preserve the cell order during the connection handover. In the following, we suggest an algorithm to achieve this goal. The algorithm, first, identifies the convergence point of the original and the augmented routes. The location of the convergence point depends on whether an augmentation or complete re-routing is used during the handover. In the augmentation phase, the convergence point is one of the satellites already in the route. If the connection is re-routed, the convergence occurs in the destination terminal. After the identification of the convergence point, the rest of the algorithm is generic. Thus, below, we outline the algorithm for a general convergence point.

Figure 8 depicts a case where two routes converge at switch S_3 that is also designated as CP. The path between S_1 and CP is being replaced by the path between S_1' and CP. During the handover process, the ground link between the user terminal and S_1 is removed by the user terminal's request. Thus, the user terminal starts sending its cells to the new source satellite S_1'. The in-sequence cell delivery can be guaranteed if the last cell sent to S_1 is served by CP before the first cell sent to S_1' by the user terminal. This service order can be achieved if CP buffers the cells arriving from S_1' until the last cell sent by $S1$ is served as shown in Figure 8. Upon forwarding the last cell of the connection to S_2, the switch S_1 sends a signaling message to S_2 to inform S_2 that the last cell of the connection has been forwarded, and the link between S_1 and S_2 is torn down. The signaling message is forwarded to S_3, similarly, by S_2 when all the cells of the connection is sent to S_3 and the link between S_2 and S_3 is torn down. The process is repeated by other switches preceding CP. Upon receiving a similar message, CP identifies the last cell of the removed path, and, hence, cell order is preserved.

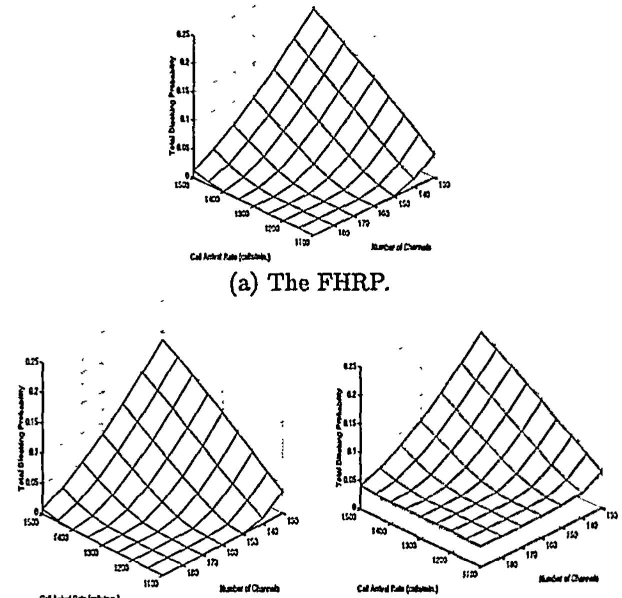
5. PERFORMANCE EVALUATION

An event-driven LEO satellite network simulator has been written to evaluate the performance of the FHRP. The connections are voice calls that can be served by the constant bit rate (CBR) service in ATM environment. The simulation variables are the call arrival rate, call holding time, number of ground channels, and number of ISL channels. Both the call interarrival and call holding times are exponentially distributed. In particular, the average call holding time is set to 3 minutes for all experiments. The simulated LEO satellite network has 6 orbits and each orbit has 6 satellites as depicted in Figure 4. The simulation time for each experiment is 200 minutes.

The main performance metric is the total blocking probability that is computed by dividing the number of blocked calls by the total number of calls during the simulation. The total blocking probability is the sum of the probabilities of the new call blocking and the handover blocking. The handover blocking is more important than the new call blocking, because the interruption of a conversation is more annoying than blocking a new call. Blocking would occur due to insufficient ground or ISL channels. Since the scope of this paper is related to the multihop ISL routing, we intentionally increase the number of ground channels in the simulations so that most of the results presented below do not have any blocking due to insufficient ground channels.

The Dijkstra algorithm [3] is used to find routes for new calls. The cost of a path is equal to the sum of the costs of all links in that path. Two different cost functions are tried in the simulations. In the first, the cost of each ISL link is equal to the number of busy channels. In the second, the cost of each ISL link is equal to one, and, thus, the resulting route corresponds to the minimum hop (minimum delay) route. Note that even in the minimum hop routing, the load on the ISL channel is considered so that the Dijkstra algorithm finds the minimum hop route that does not contain any congested ISL link. However, we found that the simulation results for different cost functions are very close in all cases. We present only the results of the minimum hop routing due to the space limitation.

The performance of the FHRP is compared with two other scenarios. In the first scenario, we have a *static* network with the same topology of the simulated LEO satellite network. The static network consists of switches that do not move. Thus, no handover is necessary. In a static network, the connection routes preserve their optimality at all times. In other words, it represents the best achievable blocking performance with the given switch capacity. In the second scenario, *pure augmentation* approach is used during the satellite handover. The pure augmentation approach results in a handover blocking if the handover satellite is not on the existing route or an augmented link cannot be found between the handover satellite and the existing route. Thus, the difference between the performance of the pure augmentation and the FHRP shows the performance gain achieved by the re-routing, especially by the footprint re-routing.



(a) The FHRP.

(b) Static Network.

(c) Pure Augmentation.

Figure 9: Total Blocking Probability for (a) the FHRP, (b) Static Network, and (c) Pure Augmentation Scenario.

5.1. The Case of Homogeneous Traffic

For homogeneous traffic, the new call arrivals are distributed uniformly in the coverage regions of the satellites. Specifically, the source and the destination satellites are generated uniformly among the satellites in the network. Also, within the footprints of these two satellites, the locations of the source and the destination mobile terminals are generated uniformly. In Figure 9, the performance of the FHRP, the static network, and the pure augmentation scenarios are shown. The number of ground channels in these three figures are equal to the number of ISL channels.

The static network scenario performs better than the FHRP and the pure augmentation. The blocking performance of the FHRP is very similar to that of the static network. The pure augmentation, on the other hand, performs poorly. Especially, for each call arrival rate, the total blocking probability of the pure augmentation approach decreases to around 5% as the number of ISL channels increases to 190. However, the FHRP and the static network scenario have no blocking at all as the number of ISL channels increases. This result shows that pure augmentation is not sufficient by itself to achieve low blocking probabilities. In all scenarios, the total blocking probability for small call arrival rates or large number of channels is almost constant. When the call arrival rate increases (or, the number of channels decreases), the performance starts degrading quickly. We also observed that the blocking due to ISL channel congestion contributes to the total blocking probability much more than the blocking due to ground channel congestion, although the numbers of ground channels and ISL channels are equal. In most cases, the ISL channel blocking contributes more than 90% of the total blocking. The reason for this behavior is because each connection has about 4.8 hops on average while only one uplink and one downlink are required for the ground channels. It is more probable for a multihop route being congested compared to a single hop route.

In the second set of experiments, the number of ground channels are kept constant at 190 so that no blocking due to ground channel congestion occurs. The number of ISL channels is equal to 150. Figure 10 shows the new call blocking and handover blocking performances of the FHRP and the pure augmentation as a function of the call arrival rate. The static network case has not been simulated, since no handover occurs in the static network. According to Figure 10, the FHRP has a much smaller handover blocking than the pure augmentation. In the pure augmentation almost 4% of the handover calls are blocked. The new call blocking for the FHRP is higher than that of the pure augmentation, since the pure augmention blocks more handover calls and, hence, can accept more new calls. On the other hand, the sum of the handover and the new call blocking probabilities is lower in the FHRP case.

In the last experiment for the homogeneous call arrival case, we kept the call arrival rate at 1300 calls/min and the number of ground channels at 190. The effect of increasing the number of ISL channels is investigated. As shown in Figure 11, increasing the number of ISL channels helps to decrease the blocking probability. The FHRP and the static networks perform very similar and are superior to the pure augmentation. The results of the first three sets of experiments show that the FHRP performs very similar to a network without any mobility of the switches or the user terminals, i.e., the handovers do not degrade the performance of the system. On the other hand, pure aug-

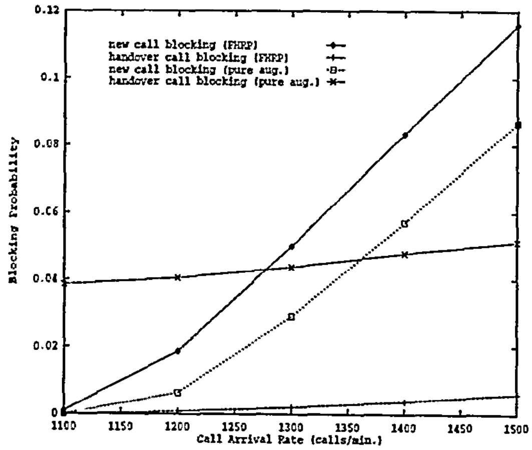


Figure 10: New Call and Handover Blocking Probabilities vs. Call Arrival Rate for Static, FHRP and Pure Augmentation Networks.

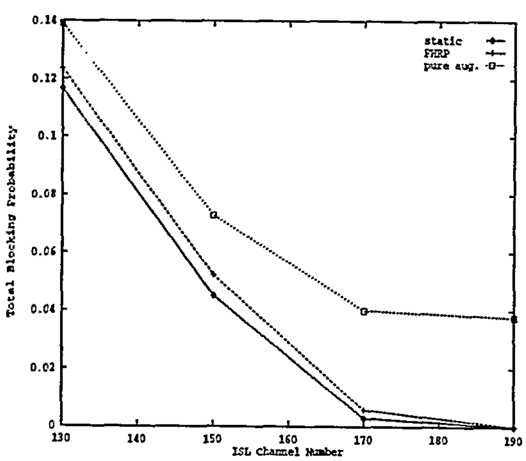


Figure 11: ISL Blocking Probabilities vs. Number of ISL Channels.

mentation results in higher call blocking compared to the other two. Especially, the handover call blocking is comparably high. Thus, it is clear that a handover algorithm that solely consists of route augmentation causes very high blocking probability. We repeated the experiments to see whether re-routing would help when the calls are blocked in pure augmentation. Thus, if a handover call is blocked, it is re-routed. The performance results for this modified pure augmentation algorithm were very similar to that of the FHRP. Thus, the FR phase of the FHRP can be replaced with a re-routing function as used in call establishment phase. On the other hand, as explained in Section 1, FHRP can be performed by the satellites while re-routing without FR needs to be done by the gateways. Furthermore, performing FR in the satellite can reduce the signaling and computation overhead in the gateways. According to our experiments, more than 90% of handovers only require a FR instead of a re-routing.

5.2. The Case of Heterogeneous Traffic

In the heterogeneous traffic, the source satellite is still generated uniformly. However, we use a different approach to select destination satellite. First, two random numbers uniformly distributed between 0 and 1 are generated to determine the magnitude of the movement in the x and y axes from the source to the destination satellite. In the meantime, a probability vector P is used where $P = (p_0, p_1, p_2, p_3, p_4, p_5, p_6) = (0, 0.1, 0.35, 0.6, 0.8, 0.9, 1)$. If the random numbers fall between p_i and p_{i+1}, then i is assigned to their associated magnitude. Since we are more interested on the multihop connection, p_1 is assigned to 0.1 so that the source and destination satellites will not be the same unless both random numbers are less than 0.1. After the magnitude is decided, two random numbers are generated to decide the direction of the movement. Once we know the direction and magnitude of the movement, the destination satellite can be obtained. Note that the movement in the y direction circulates on the orbit. For example, if the source satellite is 9 and the y movement is -4, then the resulting destination satellite is $9 \rightarrow 8 \rightarrow 7 \rightarrow 12 \rightarrow 11 = 11$ because satellites 7-12 are in the same orbit in the simulated network. However, the x movement circulates on different orbits. If the obtained destination satellite is not in the correct range[4], then the same process repeats until a correct one is selected. Since no circulation in the x axis, the satellites located in the central orbits are subject to more load in their ground channels, especially, in the downlinks.

In Figures 12 and 13, the number of the ground channels is chosen as 600 to avoid any ground blocking. Figure 12 shows that the static network still has the best blocking performance. The performance of the FHRP is almost identical to that of the static network. So, this experiment confirms that the relative performance of the FHRP is not affected by the heterogeneous traffic pattern. In fact, the relation among the performances of the FHRP and the static network are similar to those in the homogeneous traffic case. However, the blocking probability for the heterogeneous traffic is higher than that for the homogeneous traffic because the call arrival rate is 1100 calls/min in Figure 12 while it is 1300 calls/min in Figure 11. In Figure 13, the ratio between the new call and handover blocking is similar to that in the case of homogeneous traffic. The number of ISL channels is 170 for this experiment. In conclusion, Figures 12 and 13 confirm that the FHRP performs consistently in the cases of homogeneous and heterogeneous traffic.

6. CONCLUSIONS

A handover re-routing algorithm called Footprint Handover Re-routing Protocol (FHRP) has been proposed for LEO satellite networks. The FHRP is a hybrid algorithm that consists of the augmentation and the footprint re-routing phases. In the augmentation phase, a direct link from the new end satellite to the existing route is found. In case, there is no such link with required capacity exists, a new route is found using the optimum route finding algorithm.

[4]For example, if the source satellite locates in the first orbit and the x movement is -2, then the resulting satellite will not be in the correct range.

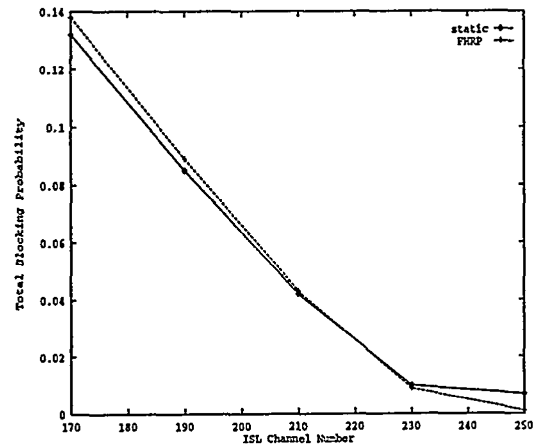


Figure 12: ISL Blocking Probabilities vs. Number of ISL Channels for Heterogeneous Case.

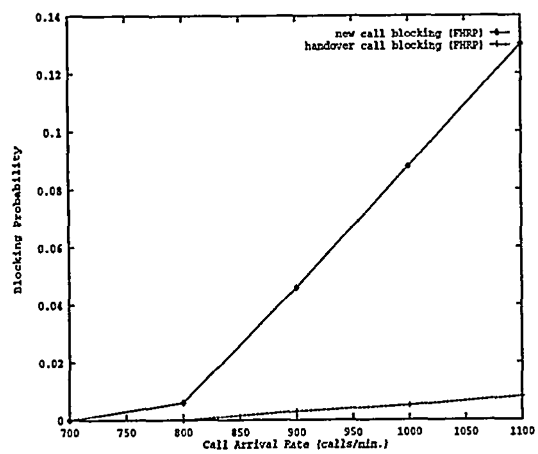


Figure 13: New Call and Handover Blocking Probabilities vs. Call Arrival Rate for Heterogeneous Case.

In the footprint re-routing phase of the FHRP, the connection is routed through footprint re-route determined by the original optimum path. The goal of the re-routing is to establish an optimum route without applying the optimum route finding 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. As proven in Section 3, the optimality of the original route is maintained during the communication. The FHRP is applicable to any type of connection-oriented networks. The FHRP has been adapted to ATM environment to guarantee the in-sequence cell delivery during handover instants. Moreover, a signaling algorithm was presented to easily achieve the route augmentation. The performance of the FHRP is compared with a static network and pure augmentation. The results show that 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.

7. REFERENCES

[1] B.A. Akyol and D.C. Cox. Rerouting for Handoff in a Wireless ATM Networks. *IEEE Personal Communications*, 3(5):26–33, Oct. 1996.

[2] F. Ananasso. System, Market and Regulatory Aspects for Satellite Personal Communications. In *Proc. of the First European Workshop on Mobile/Personal Satcoms*, pages 3–14, 1994.

[3] D. Bertsekas and R. Gallager. *Data Networks.* Prentice Hall, 1992.

[4] P. Carter and M.A. Beach. Evaluation of Handover Mechanisms in Shadowed Low Earth Orbit Land Mobile Satellite Systtems. *Int. Journal of Satellite Communications*, 13:177–190, 1995.

[5] 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*, pages 529–535, 1995.

[6] F. Dosiere, T. Zein, G. Maral, and J.P. Boutes. A Model for the Handover Traffic in Low Earth-Orbiting (LEO) Satellite Networks for Personal Communications. *Int. Journal of Satellite Communications*, 11:145–149, 1993.

[7] K.Y. Eng, M.J. Karol, M. Veeraraghavan, E. Ayanoglu, C.B. Woodworth, P. Pancha, and R.A. Valenzuela. A Wireless Broadband Ad-Hoc ATM Local - Area Network. *Wireless Networks*, 1(2):161–174, 1995.

[8] J.L. Grubb. IRIDIUM Overview. *IEEE Communications Magazine*, 29(11), 1991.

[9] Y.C. Hubbel. A Comparison of the IRIDIUM and AMPS Systems. *IEEE Networking Magazine*, 11(2):52–59, March/April 1997.

[10] E. Del Re, R. Fantacci, and G. Giambene. Handover Requests Queueing in Low Earth Orbit Mobile Satellite Systems. In *Proc. of the Second European Workshop on Mobile/Personal Satcoms*, pages 213–232, 1996.

[11] M.A. Sturza. Architecture of the TELEDESIC Satellite System. In *Proc. of Int. Mobile Satellite Conference*, pages 212–218, 1995.

[12] C. Toh. The Design and Implementation of a Hybrid Handover Protocol for Multimedia Wireless LANs. In *Proc. of MOBICOM '95*, pages 49–61, 1995.

[13] 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 Journal on Selected Areas in Communications*, 15(1):69–82, Jan, 1997.

[14] M. Werner, A. Jahn, E. Lutz, and A. Bottcher. Analysis of System Parameters for LEO/ICO-Satellite Communication Networks. *IEEE Journal on Selected Areas in Communications*, 13(2):371–381, February 1995.

[15] R.A. Wiedeman and A.J. Viterbi. The Globalstar Mobile Satellite System for Worldwide Personal Communications. In *Proc. of Int. Mobile Satellite Conference*, pages 46–49, 1993.