A routing algorithm for connection-oriented Low Earth Orbit (LEO) satellite networks with dynamic connectivity

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

[a] *Bell Laboratories, Lucent Technologies, 101 Crawfords Corner Road, Holmdel, NJ 07733, USA*

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

[c] *Department of Defense, 9800 Savage Road STE 6512, Fort Meade, MD 20755-6512, USA*

Low Earth Orbit (LEO) satellites move with respect to a fixed observer on the Earth surface. Satellites in the polar regions and the seam switch off their intersatellite links to the neighbor satellites. As a result, the connectivity pattern of the network changes. Ongoing calls passing through these links need to be rerouted. A large number of simultaneous rerouting attempts would cause excessive signaling load in the network. Moreover, the handover calls could be blocked because of the insufficient network resources in the newly established routes or large connection re-establishment delay. In this paper, a routing protocol is introduced to reduce the number of routing attempts resulting from link connectivity change. The protocol does not use the links that will be switched off before the connection is over. Since the call durations are not known a priori, the proposed protocol utilizes a probabilistic approach. The performance of the protocol is evaluated through simulation experiments. The experimental results indicate that the routing protocol reduces the number of rerouting attempts resulting from connectivity changes of the network.

1. Introduction

Low Earth Orbit (LEO) satellites are located at altitudes of 500–1500 km [2,8,9]. This low altitude results in low propagation delay, lower power requirements in the user terminals and the satellites, and efficient spectrum utilization using small coverage area for each satellite. When intersatellite links (ISLs) are present, connections can be routed through these links without requiring any terrestrial resources. LEO satellite networks could have a dynamic connectivity structure resulting from the satellite movement. The ISL connectivity between satellites may change based on the distance and viewing angle between them. Any connection is subject to rerouting if it is passing through a link that will be switched off before the connection is over. This event is referred to as *link handover*. Large numbers of rerouting attempts during the link handover would result in excessive signaling load in the network. Moreover, the handover calls could 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 [13] 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. The optimization process results in a unique route with minimum number of link handovers during a system period[1] 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 number of link handovers; however, it can 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. The performance of the algorithm should be investigated for satellite networks with dynamic connectivity.

Recently, the algorithm in [13] has been improved in [12] by introducing a *sliding window* mechanism. In the new algorithm, when optimization process is performed, the routes are determined such that the number of handovers occurring in a time window is minimized. By sliding the window, new routes are determined after each topology change. This algorithm uses a fixed window size. The performance of the algorithm is sensitive to the relative magnitudes of the call duration and the window size, which should be around the average call duration to achieve good performance. In [4,5], 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 way, 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

[1] System period is defined as the time interval in which a satellite circulates the Earth.

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 this paper, we introduce a routing algorithm that reduces the number of rerouting attempts due to link handovers by taking advantage of the LEO satellite system dynamics and call statistics. Basically, the algorithm tries not to use links that would be switched off before the connection is over. Since the algorithm has no knowledge of the exact call duration and the user location, the probability distribution function (pdf) of the time duration in which the call uses the established route is utilized by the routing algorithm. The determined pdf is used to find a route that will not experience a link handover with a certain probability during connection's lifetime. The suggested routing algorithm can be applied to any type of connection-oriented network such as circuit switched voice networks or ATM networks. In this paper, we assume that the LEO satellite network carries mostly voice calls as in the case of major LEO satellite networks such as Iridium and Globalstar [9]. The organization of this paper is as follows. In section 2, the system model is introduced. In section 3, our routing protocol that reduces the number of rerouting attempts resulting from the link handovers is introduced. In section 4, the application of the proposed routing algorithm on the Footprint Handover Rerouting Protocol (FHRP) [11] is described. In section 5, the performance of the proposed algorithm is evaluated. Finally, the paper is concluded in section 6.

2. System model

In the LEO satellite system described, satellites are moving in circular polar orbits as shown in figure 1. Similar to the Iridium system [6,7], there are six orbits in the network as shown in figure 1. 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 and assuming that the orbit is maintained, if the satellites are located at equal distances over the orbit, the system configuration as shown in figure 1 repeats itself with a period T_o, which is referred to as *intersatellite gap*, i.e., if Loc(t) is a function that gives the location of the satellites at time t, then $\mathrm{Loc}(t) = \mathrm{Loc}(t+T_o)$. In the Iridium system, the satellites are moving at a speed of 26,000 km/h (7 km/s). A satellite circulates the Earth in 100 minutes. The satellite visibility period (equivalently, the intersatellite gap) is less than 10 minutes. Given this small visibility period, the user mobility and the rotation of the Earth are negligible when designing mobility management algorithms.

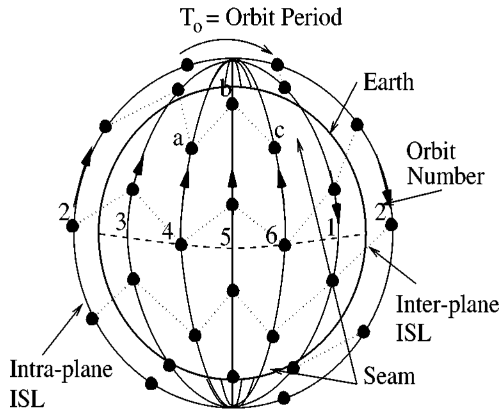

Figure 1. LEO satellite network.

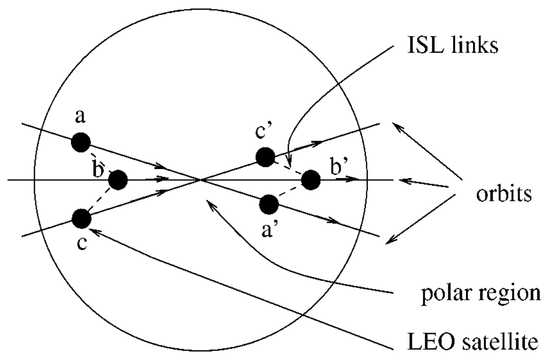

Figure 2. LEO satellites in polar region (top view).

In the LEO systems using intersatellite links and onboard processing, connections can be routed without requiring any terrestrial resources. 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 [14] 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 off their ISLs to the satellites in the neighbor orbits. Similarly, the satellites passing through polar regions switch off their ISLs to the satellites in the neighbor orbits [13]. Figure 2 shows the satellites passing through a pole. The figure 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 positions are labeled as a', b', and c' in figure 1. During

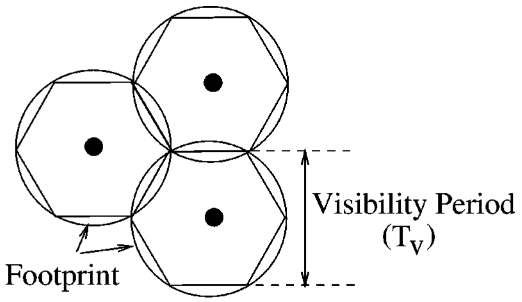

Figure 3. The footprints of the LEO satellites.

the transition, the ISL links a–b and b–c are switched 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 off its inter-plane ISL until its neighbors swap their positions. In figure 2, 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 polar crossing, results in link handovers and a dynamic network topology.

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. 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 3. 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 intersatellite gap are identical, i.e., $T_\text{o} = T_\text{v}$. However, note that, because of 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. Due to the movement of the satellites, the user terminals on the ground may not stay in the coverage region of the initial end satellites throughout the communication. To ensure that ongoing calls are not disrupted, calls should be transferred to other satellites whose coverage regions contain the ground terminals. This event is referred to as *connection handover*. During a connection handover, the existing connection route should be updated accordingly. Connection handover algorithm implemented in a system determines how often rerouting is used during connection handovers. The routing algorithm presented in the next section assumes the knowledge of the probability distribution function for the time between connection handover rerouting attempts.

3. Probabilistic Routing Protocol (PRP)

Link handovers, as explained in section 2, occur as a result of the movement of the satellites. Satellites going into the seam or the polar regions switch their ISLs to the neighbor orbits off temporarily. Any established route is subject to rerouting if it is passing through a link that will be switched off before the connection is over. If the number of connections that need to be rerouted is large, the resulting rerouting attempts cause signaling overhead in the network. Moreover, call blocking would occur because of insufficient network resources in the new route or large connection re-establishment delay. In this section, we introduce the Probabilistic Routing Protocol (PRP), which uses the knowledge of the ISL connectivity pattern to reduce the number of rerouting attempts resulting from link handovers.

LEO satellites move around the Earth with a constant speed, i.e., the movement and the ISL connectivity patterns are known a priori. The knowledge about this deterministic pattern can be used to reduce the number of rerouting attempts because of the link handovers. The Probabilistic Routing Protocol makes use of this property. Basically, the protocol does not use the links that will be switched off before the connection is over. The algorithm can be realized only if the exact call duration is known at call set-up instant, which is not a realistic assumption. However, the probability distribution function of the time duration in which the call uses the established route can be determined based on the call statistics and the dynamics of the satellite system. We use the fact that a call releases the established route due to two events: call termination and intersatellite handover that results in complete rerouting. The developed probability distribution function is used to find a route that will not experience a link handover with a certain probability when the connection is active.

The network topology is represented by an $N \times N$ cost matrix C, where N is the number of satellites in the system[3]. The entry c_{ij} represents the cost of the communication link from satellite i to satellite j. If there is no active ISL between satellites i and j, the cost is equal to infinity, i.e., $c_{ij} = c_{ji} = \infty$. The cost matrix C is time dependent since the entries change based on the dynamics of the satellite network. Any routing algorithm, such as minimum cost [3] or shortest distance [3], can be used when the connectivity matrix is defined. However, the routing algorithm has no knowledge about the topology changes of the network. A newly established connection would need to be rerouted due to link handover occurring in one of the satellites in the connection route. In PRP, a *probabilistic connectivity matrix* R is used to limit the number of rerout-

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

[3] As an example, the Iridium system has 66 satellites while Globalstar has 48 satellites [9].

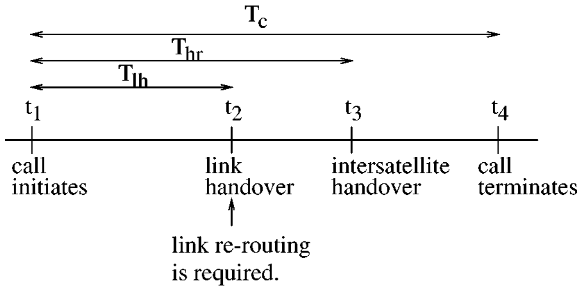

Figure 4. Timing diagram of call routing events.

ing operations during a link handover. A connection route is held until the call is terminated due to connection termination, intersatellite handover, or link handover. A call termination event occurs when the communicating parties complete their call. The time interval between the route establishment and the call termination event is called as *residual call duration*, T_c. *Intersatellite handover* occurs due to the moving coverage of the satellites serving the source and destination user satellites, as will be explained in section 4. An intersatellite handover may result in the addition of new satellites in the existing connection route. The resulting route is still expected to use a portion of the previous route. In some cases, a whole new route may have to be set up for the communication. The time from route establishment to an intersatellite handover that results in complete rerouting is called as *intersatellite handover rerouting time*, T_{hr}. The link handover event for satellite i depends on the network architecture and the position of the satellite i relative to any polar region and the seam. The time interval from the route establishment to the link handover of satellite i in the route is referred to as *link handover time*, $T_{i,lh}$, for satellite i. The relation among different route termination events is depicted in figure 4. A call arrives at time $t = t_1$. Since the call terminates at time t_4, the residual call holding time is $T_c = t_4 - t_1$. An intersatellite handover is expected to occur at time $t = t_3$. However, at least one of the satellites in the connection route experiences a link handover at time $t = t_2 < t_3$. The connection is rerouted at this instant. The link handover time, $T_{i,lh}$, is equal to $t_2 - t_1$, where i is the index of the satellite that experiences link handover. The connection is rerouted at time $t = t_2$. Note that, after rerouting, new values of residual call holding time and intersatellite handover time become $t_4 - t_2$ and $t_3 - t_2$, respectively. The value of the link handover time changes based on the new route established. If no more link handover occurs before the call terminates, the next rerouting event occurs at $t = t_3$ due to the intersatellite handover. The call terminates at $t = t_4$ upon the request of the communicating parties.

The connection rerouting resulting from link handovers can be controlled if the link handover time of each satellite in the system is utilized during the routing process. In contrast, the call termination and intersatellite handover events occur randomly and are independent of the routing algorithm. In the example timing diagram depicted in figure 4, if a link handover would have occurred after the intersatellite handover event, i.e., $t_2 > t_3$, no rerouting would be required as a result of the link handover. A suitable choice of a connection route would delay the occurrence of the link handover until the connection releases the existing route due to either call termination or intersatellite handover. The goal of PRP is to establish connection routes such that routes are terminated by call termination or intersatellite handover events, instead of by link handover, with a *target probability* p, i.e.,

$$P\big(\min(T_c, T_{hr}) < T_{i,lh}\big) > p, \tag{1}$$

for each satellite i in the established route. The value of the target probability p is chosen by the network, and represents the level of reduction in the number of rerouting attempts as will be explained in section 5. Since call termination and connection handover events occur randomly with known probability distribution functions, equation (1) can be utilized to ensure that the call termination or intersatellite handover rerouting event occurs before a satellite in the connection route experiences a link handover event. The PRP removes any ISL from consideration for routing that violates equation (1). As a result, the connection experiences either a call termination or a connection handover rerouting with probability p before a link handover event occurs for any of the satellites in the route. The proposed routing protocol works as follows:

1. Copy connectivity matrix C to the probabilistic connectivity matrix R, i.e., $r_{ij} = c_{ij}$ for $1 \leqslant i, j \leqslant N$.
2. Find the *target route holding time*, T_{tr}, value such that
$$P\big(\min(T_c, T_{hr}) < T_{tr}\big) = p. \tag{2}$$
3. Remove the ISLs of the satellites with $T_{i,lh} < T_{tr}$ from the probabilistic connectivity matrix R, i.e., $r_{ij} = r_{ji} = \infty$ for satellites i and j in neighbor orbits if $T_{i,lh} < T_{tr}$.
4. Apply a routing algorithm such as minimum cost [3] or minimum hop [3] using R.

The term $\min(T_c, T_{hr})$ is a random variable and will be referred to as *route usage time*, $T_{ru} = \min(T_c, T_{hr})$, which is the time interval the connection uses a route if no link handover occurs before the route is released as a result of the call termination or an intersatellite handover rerouting. The algorithm simply removes the ISL links that will be switched off in a time interval shorter than T_{tr}. For a newly arriving call, T_c is equal to the call holding time. Intersatellite handover rerouting time, T_{hr}, depends on the cell geometry, the initial position of the user terminal in the cell, speed of the satellites relative to the user terminal, and the handover protocol used in the system. In section 4, the application of the PRP to the Footprint Handover Rerouting Protocol (FHRP) [10,11] is presented.

The PRP removes certain links from consideration for routing to reduce the number of rerouting attempts in the future. In other words, the call blocking rate of the network is expected to increase. Hence, a trade-off exists between the call blocking rate and the number of rerouting attempts because of the link handover. A distinction between new calls and intersatellite handover calls can be made. The call blocking rate for latter type of calls should be smaller compared to that for new calls since the interruption of an ongoing call is more annoying for users than the blocking of a new call. Thus, the PRP is suggested only for newly arriving calls. For handover calls, the route should be determined using all the ISLs with available capacity.

4. Application of the PRP to the FHRP

The Footprint Handover Rerouting Protocol (FHRP) [10,11] has been proposed to balance the simplicity of route augmentation and the optimality of complete rerouting during a connection handover. The FHRP has two phases: augmentation and footprint rerouting (FR). In the augmentation phase, a route between the new end satellite and a satellite already in the route is established, and the unused portion of the route is removed. The FR phase is applied after both end satellites are replaced with the satellites in their respective orbits. The connection route changes completely in the FR phase. Connection handover rerouting time, which is used in PRP, is equal to the time interval between the route establishment time and the time instant where both end satellites are replaced with satellites in their respective orbits. Thus, the connection handover rerouting time, T_{hr}, is equal to $\max(T_{\mathrm{hs}}, T_{\mathrm{hd}})$, where T_{hs} and T_{hd} are time intervals between the call establishment and time instants when the original source and destination satellites are replaced with satellites in their respective orbits. The pdf of T_{hr}, $F_{\mathrm{hr}}(t) = P(T_{\mathrm{hr}} < t)$, is given by

$$F_{\mathrm{hr}}(t) = P(T_{\mathrm{hs}} < t)P(T_{\mathrm{hd}} < t) = \left[P(T_{\mathrm{h}} < t)\right]^2, \qquad (3)$$

where *single rerouting time* T_{h} is a random variable denoting T_{hs} and T_{hd}, which are independent and identically distributed random variables. For the clarity of the presentation, T_{h} is called as the *single rerouting time*.

The pdf of the single rerouting time depends on the location of the user terminal inside the footprint and the size of the satellite footprint. The location of the user terminal is uniformly distributed in the hexagonal area. A terminal located in the rectangular area shown in figure 5 experiences an intra-orbit handover[4] and is ready for the FR after the first handover. The probability of a user terminal being located in the rectangular area of the cell is equal to $2/3$, which is the ratio of the area of the rectangle and the area of the hexagon. Hence, a user terminal experiences intra-orbit handover with probability $2/3$. The distance traveled by such a user terminal is distributed uniformly in $[0, T_{\mathrm{v}}]$, where T_{v} is the *visibility period*, which is defined as the longest time interval in which a satellite is visible to a ground terminal as shown in figure 5. Note that the visibility period T_{v} is assumed to be equal to the intersatellite gap T_{o} in this paper. A terminal located in one of the shaded triangles in figure 5 experiences an inter-orbit handover. Figure 6 illustrates a timing diagram for a call located in the right triangle region. The call arrives at the network at time $t = t_{\mathrm{e}}$. 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 $t = t_1 > t_{\mathrm{e}}$, 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'. The user terminal is ready for the FR phase at $t = t_2$. Single rerouting time, T_{h}, for this user terminal is equal to $t_2 - t_{\mathrm{e}}$, which has a pdf given as

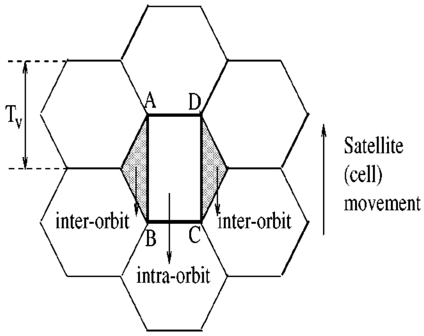

Figure 5. Inter- and intra-orbit handover regions.

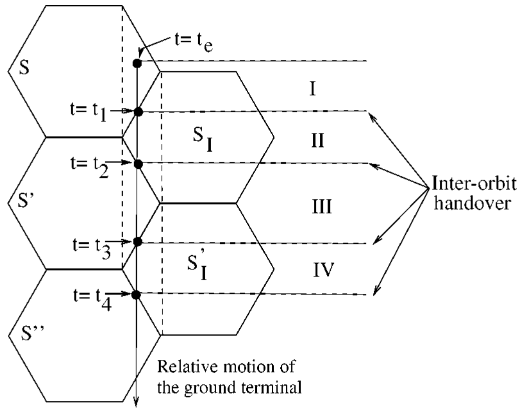

Figure 6. Timing diagram for FR.

$$F_{\mathrm{h}}(t \mid \text{interorbit handover}) = \begin{cases} \dfrac{t^2}{T_{\mathrm{v}}^2} & \text{for } t \in [0, T_{\mathrm{v}}], \\ 1.0 & \text{for } t \geqslant T_{\mathrm{v}}. \end{cases} \qquad (4)$$

The second line of equation (4) is intuitive since a call has to use FR in a time interval shorter than T_{v}. Combining the distribution functions for square and triangle regions, the pdf of single rerouting time, $F_{\mathrm{h}}(t) = P(T_{\mathrm{h}} < t)$, is determined as

$$F_{\mathrm{h}}(t) = \begin{cases} \dfrac{2t}{3T_{\mathrm{v}}} + \dfrac{t^2}{3T_{\mathrm{v}}^2} & \text{for } t \in [0, T_{\mathrm{v}}], \\ 1.0 & \text{for } t \geqslant T_{\mathrm{v}}. \end{cases} \qquad (5)$$

[4] Intra-orbit handovers are the ones between adjacent satellites in the same orbit, while inter-orbit handovers are the ones between satellites in adjacent orbits.

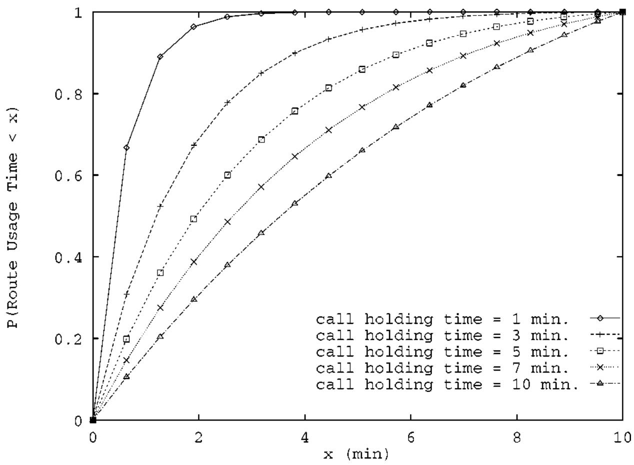

Figure 7. Route usage time pdf for various call holding times.

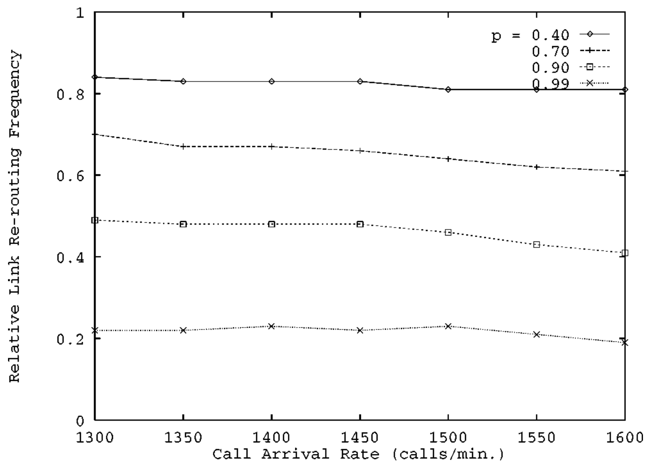

Figure 8. Relative rerouting frequency as a function of call arrival rate.

The distribution function for connection handover rerouting time, $F_{\mathrm{hr}}(t)$, is determined using equations (3) and (5). The route usage time is equal to the minimum of the residual call holding time and the intersatellite handover rerouting time, i.e., $T_{\mathrm{ru}} = \min(T_{\mathrm{c}}, T_{\mathrm{hr}})$. Using exponential call holding time, the distribution function of the route usage time, $F_{\mathrm{ru}}(t) = P(T_{\mathrm{ru}} < t)$, is found as

$$F_{\mathrm{ru}}(t) = \begin{cases} 1 + \mathrm{e}^{-\mu t}\big(F_{\mathrm{hr}}(t) - 1\big) & \text{for } t \in [0, T_{\mathrm{v}}], \\ 1.0 & \text{for } t \geqslant T_{\mathrm{v}}, \end{cases} \tag{6}$$

where μ is the inverse of the call holding time. Figure 7 shows the pdf of the route usage time for various values of call holding time with a visibility period of 10 min. When the call holding time is small compared with the visibility period, such as when the mean call holding time is equal to 1 min, route usage time is almost exponentially distributed with a parameter equal to that of call holding time. The visibility period becomes more effective on the route usage distribution when the connections stay in the network for longer time periods, as in the case of calls with mean holding times equal to 10 min.

5. Performance evaluation

The performance of the PRP has been evaluated to investigate the trade-off between the number of rerouting attempts during the link handovers and the call blocking probabilities. The performance of the PRP is compared for different values of the target probability and mean call holding times. When the target probability is equal to zero, route usage time information is not used at all for routing, i.e., PRP is identical to the direct application of the Dijkstra algorithm [3]. The connections are assumed to be voice calls. Both the call interarrival and call holding times are exponentially distributed. No traffic is generated in polar regions. The simulated LEO satellite network has 6 orbits and each orbit has 11 satellites. The simulation time for each experiment is 300 min. First 60 min of the experimental data are discarded to remove the transient behavior of the simulation experiments. The number of ISL channels between neighbor satellites is equal to 150. The Dijkstra algorithm [3] is used in combination with the PRP to determine routes for new calls. The cost of each ISL is equal to one, and, thus, the resulting route corresponds to the minimum hop 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.

Figure 8 shows the performance of the PRP in terms of relative link rerouting frequency, which is defined as the ratio of the number of link rerouting attempts for a given target probability to that of for a target probability of zero. The effects of the PRP become noticeable when the target probability p, as defined in equation (1), increases. The relative frequency decreases as larger target probabilities are used. As an example, a target probability of 0.99 results in 80% decrease in the number of link rerouting operations. The reduction in the number of rerouting attempts is less for smaller target probabilities. As seen in figure 8, relative link rerouting frequency is almost independent of traffic load since the PRP operates independently for each arriving call.

Second set of experiments focuses on the call blocking performance for different values of the target probability. The new call blocking probability, which is defined as the ratio of the number of blocked new call arrivals and the number of new call arrivals, is shown in figure 9. The results confirm that the use of PRP increases the new call blocking probability. Especially, the blocking probability for a target probability of 0.99 is much larger than those of lower target probabilities. This can be explained using figure 7. The call holding time for this experiment is equal to 3 min. The probability distribution function of the route usage time reaches 0.9 at the end of the third minute. To achieve a probability of 0.99, the target route hold time should be as much as 7 min, i.e., the links that will be switched off within 7 min are not considered for routing for a newly arriving call. This clearly results in a high call blocking rate compared to the blocking rates achieved

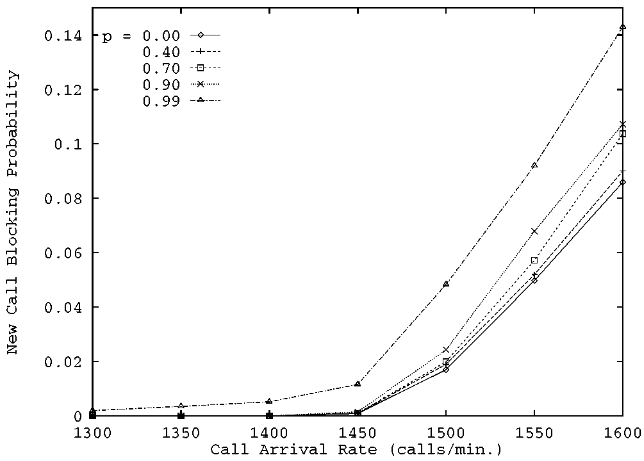

Figure 9. New call blocking probability as a function of call arrival rate and target probability.

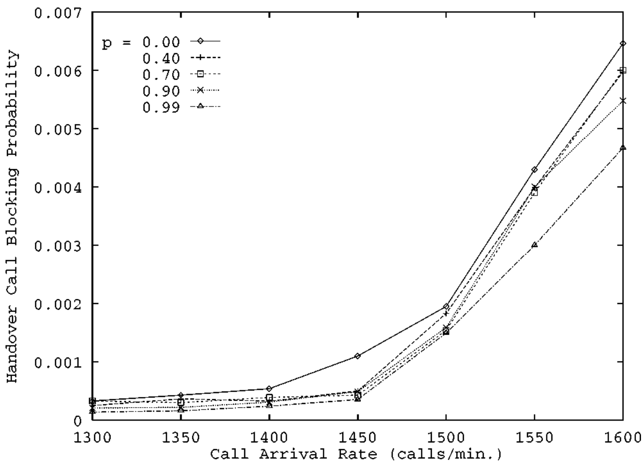

Figure 10. Rerouted call blocking probability as a function of call arrival rate and target probability.

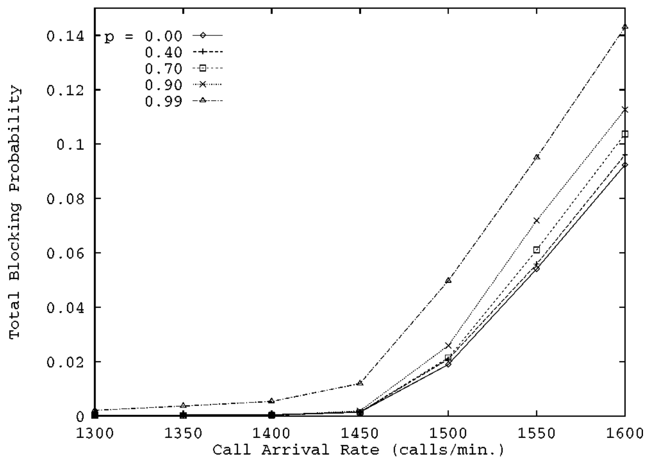

Figure 11. Total blocking probability as a function of call arrival rate and target probability.

Table 1
Relative rerouting and blocking probabilities as a function of average call blocking time.

Avg. call holding time	Relative rerouting frequency	Handover blocking probability	
		$p = 0.0$	$p = 0.9$
1 min	0.33	0.0001 (0.009)	0.00009 (0.01)
3 min	0.47	0.001 (0.008)	0.0009 (0.011)
5 min	0.45	0.0015 (0.003)	0.001 (0.009)
7 min	0.43	0.0026 (0.003)	0.0016 (0.01)

using smaller target probabilities. For calls with holding times of 3 min and visibility period of 10 min, target probability would be set to 0.90 to decrease the relative link rerouting frequency to 0.5. Thus, an empirical choice of the target probability based on the traffic characteristics and the system geometry would solve the trade-off between the normalized rerouting frequency and the call blocking probability. The blocking probability in this case is very similar to that of target probability of zero. Figure 10 shows the blocking probability for handover calls. The blocking probability for handover calls is slightly better for high target probabilities since the capacity kept by denying service to new calls is utilized partially by the rerouted calls. Not surprisingly, the rerouted call blocking for a target probability of 0.99 is smaller for every call arrival rate simulated. Total call blocking probability, which is the ratio of number of blocked calls to number of call arrivals, is very similar to the new call blocking probability as shown in figure 11. Finally, the algorithm performance for calls with different average call holding times is presented in table 1. For this set of experiments, the call arrival rates are adjusted to achieve a total blocking probability of around 0.01 for the PRP algorithm with a target probability of 0.9. The relative rerouting frequency is smallest when the average call holding time is 1 min. This is because, even for a target probability of zero, most of the calls are terminated by the communicating parties before a link handover occurs. When the average call holding time is 3 min, the relative rerouting frequency becomes 0.47; however, for further increases in call holding times, the relative rerouting time decreases slightly. This is because, for calls with very long holding times, the connections experience a route change due to intersatellite handover. Note that the PRP is not applied when a connection is rerouted as a result of link or intersatellite handover. The third and fourth columns of table 1 show the handover call blocking probabilities for the target probabilities of 0.0 and 0.9. The total blocking probabilites are given in the parentheses. Handover blocking probability is increasing with increasing average call holding times. Total blocking probabilities are comparable for small average call holding times. For average call holding times of 5 and 7 min, the total call blocking probabilities for a target probability of 0.0 is much smaller than that for a target probability of 0.9. This is because, for large holding times, the PRP removes an excessive number of links from consideration when a new call arrives and, as a result, the new call blocking probability increases.

The third set of experiments addresses the performance of the PRP with the FHRP in terms of the change in the

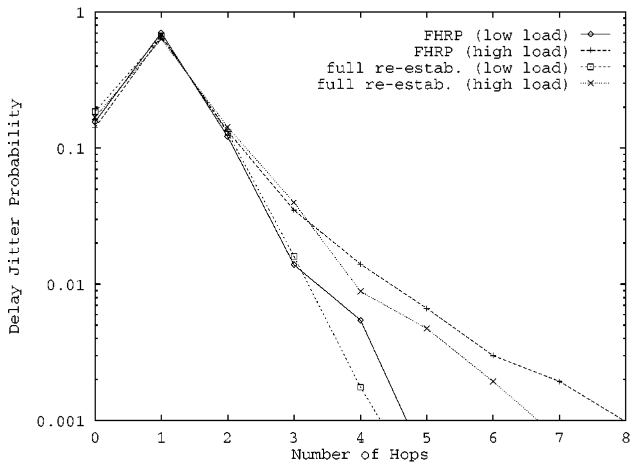

Figure 12. Probability density function of delay jitter for FHRP and full connection re-establishment.

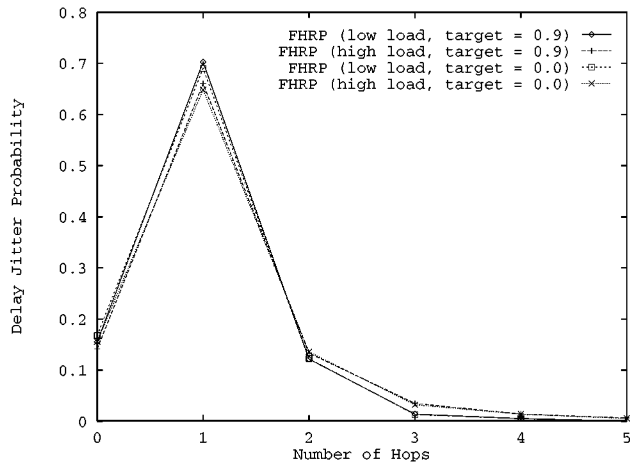

Figure 13. Probability density function of delay jitter for FHRP with target probabilities of 0.9 and 0.0.

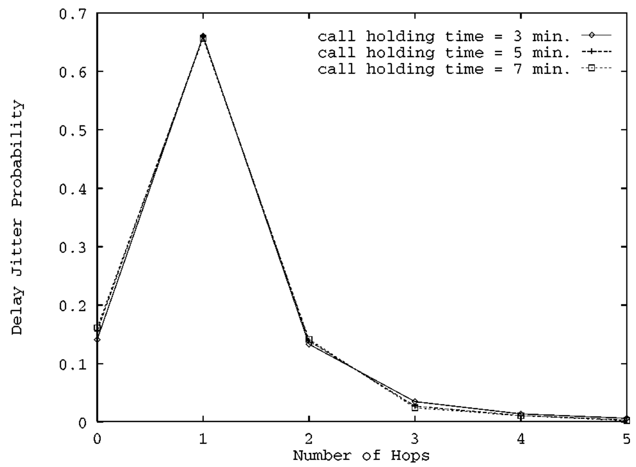

Figure 14. Probability density function of delay jitter for FHRP with different call holding times.

route length, or delay jitter, during a rerouting operation. The performance of the PRP is compared with that of the full route re-establishment algorithm in which an optimal route is determined during intersatellite and link handovers. The routing algorithm used in the simulations finds a minimum hop [3] route using the links with available capacity. Thus, when the full connection re-establishment algorithm is used, the call is always routed through the minimum number of satellites. On the other hand, the FHRP [11] performs local route updates during the augmentation phase, while it replaces the augmented route with the FR route when it is applicable. Figure 12 shows the delay jitter probability for the PRP and the full connection re-establishment for calls with holding times of 3 min. The target probability for the PRP is set to 0.9. The performance is investigated for two different traffic loads. The traffic arrival rate for the "low load" scenario is 1200 calls/min, while it is 1450 calls/min for the "high load" case. The total call blocking rates for the low and high load scenarios are in the order of 10^{-3} and 10^{-2}, respectively. For both traffic loads, the change in the route length is less than three hops for most of the rerouting attempts. In almost 70% of the rerouting operations, the route length changes by only one hop. Moreover, the FHRP performs similar to the full connection re-establishment algorithm. When the traffic load is high, the change in the route length would go up to 8 hops. However, the probability of this event is very small. Specifically, 99% of the rerouting attempts in the FHRP result in delay jitter values less than 3 and 5 hops for low and high traffic load cases, respectively. When the full connection re-establishment algorithm is used, the 99% of the rerouting attempts result in delay jitter values less than 3 and 4 hops for low and high traffic load cases, respectively. These results show that the delay jitter performance of the FHRP is indistinguishable from that of full connection rerouting algorithm.

In the second part of the delay jitter experiments, the effects of the value of the target blocking probability is investigated. Figure 13 shows that the target probability has almost no effect over the performance. This is because the intersatellite handovers occur much more frequently than the link handovers. Thus, the delay jitter performance is determined by the FHRP. Similarly, the mean holding times of the calls have no effect over the delay jitter density as shown in figure 14, where the density functions for calls with holding times of 3, 5, and 7 min are given. When a call stays in the network for a long time period, it experiences a high number of intersatellite handovers. However, since the FHRP provides low delay jitter during the intersatellite handovers, the performance is independent of the number of rerouting attempts experienced by user connections.

6. Conclusions

In this paper, a routing algorithm, which is referred to as the Probabilistic Routing Protocol (PRP), has been introduced to handle the link handovers occurring in the LEO satellite networks with dynamic topology. The underlying network architecture is a LEO satellite network with circular polar orbits. The routing algorithm targeted connection-oriented networks and voice calls in particular. The Probabilistic Routing Protocol reduces the number of rerouting attempts during a link handover, which occurs because of the dynamic topology of the LEO satellite network. The algorithm removes all the ISLs that will experience a link handover during the lifetime of a connection from consideration for routing during the route establishment phase of a new call. However, since the call holding time is a random variable, the connection lifetime cannot 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. Based on the simulation results, the PRP reduces the number of link rerouting attempts by 80% for calls with mean holding times of 3 min when the target probability is set to 0.99. However, since a number of ISLs are removed from consideration for routing, the new call blocking probability becomes large for the target probability of 0.99. Thus, there is a trade-off between the gain in the number of rerouting attempts and the new call blocking probability. Based on the simulation results, a target probability of 0.9 provides 50% gain in the number of rerouting attempts with a tolerable increase in the new call blocking probability. Finally, the change in the route length, or delay jitter, has been investigated for the FHRP and the PRP. The delay jitter performance of these algorithms is very similar to that of the full route re-establishment algorithm. Based on the experimental results, most of the route update operations result in a route length change of only 1 hop. More specifically, in 99% of the route change events, the delay jitter is less than 5 hops. Furthermore, the performance is independent of the target probability used in the PRP and the average call durations.

A number of future improvements are possible for the PRP. The PRP assumes that the users are distributed uniformly over the satellite coverage areas. Nonuniform distributions should be studied. In addition, instead of removing a link from consideration for routing, the link would be given a high link cost value and can be kept in the routing set. This way, the new call blocking probability would be decreased. Finally, the routing problem should be studied when a connectionless network protocol such as IP is used over the satellite network.

Acknowledgements

The authors thank the anonymous referees for their insightful and thorough reviews.

References

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

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

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

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

[5] H.S. Chang, B.W. Kim, C.G. Lee, S.L. Min, Y.H. Choi and H.S. Yang, Performance comparison of static routing and dynamic routing in Low Earth Orbit satellite networks, in: *Proc. of IEEE VTC* (1996) pp. 1240–1243.

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

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

[8] E. Lutz, Issues in satellite personal communication systems, Wireless Networks 4(2) (1998) 109–124.

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

[10] H. Uzunalioğlu, I.F. Akyildiz, Y. Yesha and W. Yen, Footprint handover rerouting protocol for Low Earth Orbit satellite networks, Wireless Networks 5(5) (1999) 327–337.

[11] H. Uzunalioğlu, W. Yen and I.F. Akyildiz, A connection handover protocol for LEO satellite ATM networks, in: *Proc. of ACM/IEEE MOBICOM '97* (1997) pp. 204–214.

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

[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) (January 1997) 69–82.

[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) (February 1995) 371–381.

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 in 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.
E-mail: huseyin@lucent.com

Ian F. Akyildiz received his BS, MS, and PhD degrees in computer engineering from the University of Erlangen-Nürnberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology and Director of Broadband and Wireless Networking Laboratory. 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 is the Editor-in-Chief of "Computer Networks" (Elsevier). He is an editor for "IEEE/ACM Transactions on Networking", "ACM-Springer Multimedia Systems", "ACM/Baltzer/URSI Wireless Networks" and "Cluster Computing". He is a past editor for "IEEE Transactions on Computers" (1992–1996) and for

"Computer Networks" (1989–1999). He was the program chair of the "9th IEEE Computer Communications" workshop in 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, and Next Generation Internet.
E-mail: ian@ee.gatech.edu
WWW: http://www.ee.gatech.edu/research/labs/bwn

Michael D. Bender is the Chief of Telecommunications Science at the Laboratory of Telecommunications Science. He conducts research in emerging wireless communications systems. Mr. Bender holds a BS degree in electrical engineering from Johns Hopkins and is completing MS degrees in electrical engineering and computer engineering at Loyola.
E-mail: mdbende@afterlife.ncsc.mil