Probabilistic Routing Protocol for Low Earth Orbit Satellite Networks

Hüseyin Uzunalioğlu

Network Application and Integration Lab
Georgia Tech Research Institute
Atlanta, GA 30332, USA
huseyin@ece.gatech.edu

Abstract:
Low Earth Orbit (LEO) satellite networks have dynamic, yet deterministic, topologies. The time-varying connectivity pattern would result in the re-routing of all connections passing through a link that is turned off as a result of the topology change. In this paper, a routing algorithm called Probabilistic Routing Protocol (PRP) is introduced. The PRP reduces the number of re-routing attempts due to dynamic topology of the network. During the routing phase of a newly arriving call, the PRP eliminates the links that will be turned off before the call releases the link due to call termination or connection handover. Since the algorithm has no knowledge of the call duration or exact terminal location, route usage time is only known probabilistically. The probability distribution function of the route usage time of the call is determined to realize the algorithm. Since the routing algorithm works in parallel with a handover re-routing algorithm, the application to the Footprint Handover Re-routing Protocol (FHRP) is also demonstrated in the paper. Performance of the algorithm is investigated using simulation experiments.

Keywords: Satellite communications, routing, handover.

1 Introduction

A number of *low earth orbit (LEO)* satellite systems have been proposed [3, 4, 8] to provide global wireless communications services. 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. In LEO systems, connections can be routed through inter-satellite links (ISL) without requiring any terrestrial resources. When a call arrives to the network, the connection route is determined by the gateways that track satellites continuously. As an example, assume that, in Figure 1, user A wants to communicate with user D with a handheld phone. The connection request is sent to the gateway via satellite 1 (source satellite), which covers user A. The gateway locates user D and computes the connection route. Then, the route information is sent to satellite 1, 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. The satellite serving user D is called *destination satellite.* The term *end satellite* is used for either of source or destination satellites.

The ISL connectivity between satellites changes based on the distance and viewing angle between them. Change of the ISL connectivity results in a dynamic network topology with challenging routing problems. Any connection is subject to re-routing if it is passing through a link that will be turned off before the connection is over. This event is referred to as *link handover.* If the number of connections that need to be re-routed due to link handover is large, resulting re-routing attempts cause signaling overhead in the network. Moreover, handover calls can be blocked during the re-routing process. The routing in LEO satellite networks has been addressed in [6] with an emphasis on setting up routes between pairs of satellites to minimize the re-routing attempts during link handovers, i.e., optimization was performed for the routes between satellite pairs. 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. Moreover, the network connectivity pattern is assumed to be static in the reported simulation study. This assumption is not realistic in the LEO satellite environment. In [2], a LEO satellite network is modeled as a Finite State Automaton (FSA) by dividing the system period of the satellite network into equal-length intervals, where the system period is defined as the least common multiple of the orbit period and the earth period. In this approach, two satellites are defined to be visible from each other in a state if they are within line-of-sight throughout the state. The information about intersatellite visibility within a state is encoded into a visibility matrix. In this manner, the LEO satellite network in a state can be regarded as having a fixed topology. The purpose of the FSA algorithm is to determine an optimum link assignment (e.g., topological design) to make best use of the limited number of ISL's in each satellite. The algorithm determines the optimum link assignments for each state using the visibility matrix. Optimal link assignment is defined as the one that gives the best performance when the optimal static routing is used. The FSA approach does not address reducing the number of re-routing attempts due to link handovers. In contrast, more connections would need to be re-routed during the state changes of the FSA model

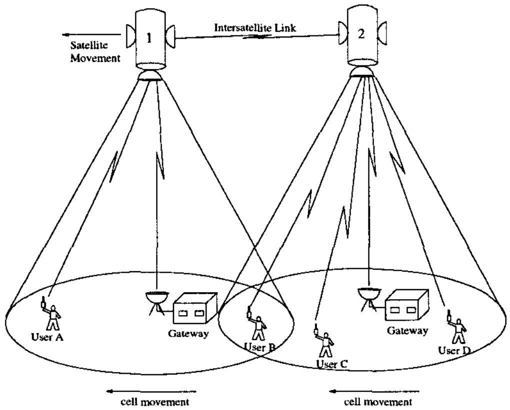

Figure 1: Wireless Communication via Satellite Network.

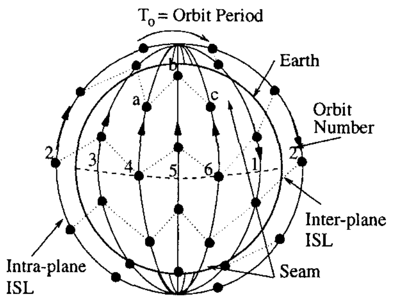

Figure 2: LEO Satellite Network.

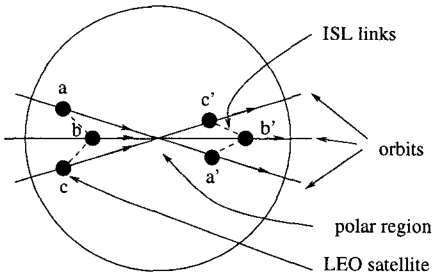

Figure 3: LEO Satellites in Polar Region (top view).

since the link assignment is optimized only with respect to the traffic pattern. In this paper, we suggest a routing algorithm that reduces the number of re-routing attempts due to link handovers by taking advantage of LEO satellite system dynamics and call statistics. Basically, the algorithm tries not to use links that would be turned off before the connection is over. Since the algorithm has no knowledge of the exact call duration and 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 developed pdf is used to find a route that will not experience a link handover with a certain probability during connection's lifetime.

The remaining of this paper is structured as follows. In the next section, system model is presented. In Section 3, the routing algorithm is described. In Section 4, the application of the presented routing algorithm to Footprint Handover Re-routing Protocol (FHRP) [5] is discussed. In Section 5, performance of the routing algorithm is investigated. Finally, Section 6 concludes the paper.

2 System Model

In the LEO satellite system described, satellites are moving in circular polar orbits as shown in Figure 2. Each satellite has up and down wireless links for communication with ground terminals and ISLs for communication between satellites. 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 due to the change in distance and viewing angle between satellites in neighbor orbits. In [7], it is concluded that only ISLs between latitudes of approximately 60^o north or south would be maintained between counter-rotating orbits in IRIDIUM system. This is labeled as *seam* in the example network model depicted in Figure 2. Satellites going into seam switch their ISLs to the neighbor orbits off temporarily. Any connection passing through these links requires re-routing.

Second type of topology change in LEO satellite network occurs due to the ISLs temporarily switched off by the satellites crossing the polar regions [6]. Figure 3 depicts the satellites passing through a pole. Drawing reflects the top view, i.e., looking at the pole from viewing position above the satellites. Satellites a, b, and c (also shown in Figure 2) 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 3. During the transition, the ISL links $a-b$ and $b-c$ are turned off. Thus, the calls passing through these links require link handover. Exact switching times of ISLs are system dependent and beyond the scope of this paper. Without loss of generalization, we assume that a satellite passing just over the pole will switch its inter-plane ISL off until its neighbors swap their positions. In Figure 3, 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.

The service area, i.e., the *footprint*, of a single satellite is assumed to be a hexagonal 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. Due to the movement of the satellites, the ground 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, existing connection route should be updated accordingly. Connection handover algorithm implemented in a system determines how often re-routing 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 re-routing attempts.

3 Probabilistic Routing Protocol

LEO satellites move around the Earth with a constant speed, i.e., the movement and ISL connectivity patterns are known a priori. The Probabilistic Routing Protocol (PRP) presented in this paper makes use of this property. The network topology is represented by an $N \times N$ cost matrix C where N is the number of satellites in the system. 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 [1] or shortest distance [1] 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 re-routed 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 re-routing operations during a link handover. A connection route is held until the call is terminated due to connection termination, connection handover, or link handover. 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. Connection handover occurs due to the moving coverage of the satellites serving the source and destination user satellites as explained in the preceding section. Connection handover may result in 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 until a connection handover that results in complete re-routing is called as *connection han-*

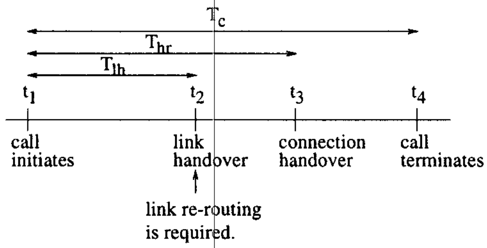

Figure 4: Timing Diagram of Call Routing Events.

dover re-routing time, T_{hr}. 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 to the link handover of satellite i, or *link handover time*, $T_{i,lh}$, is the amount of time until satellite i experiences a link handover. 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 $T_c = t_4 - t_1$. A connection 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 re-routed 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. Note that, after re-routing, new values of residual call holding time and connection handover time become $t_4 - t_2$ and $t_3 - t_2$, respectively. If no more link handover occurs before the call terminates, the next re-routing event occurs at $t = t_3$ due to connection handover. Call terminates at $t = t_4$ upon the request of the communicating parties.

The connection re-routing due to link handover 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 connection handover events occur randomly and are independent of the routing algorithm. In the example timing diagram depicted in Figure 4, if link handover would have occurred after the connection handover event, i.e., $t_2 > t_3$, no re-routing due to link handover would have been required. Suitable choice of connection route would delay the occurrence of the link handover until the connection releases the existing route due to either call termination or connection handover. The goal of PRP is to establish connection routes such that routes are terminated by call termination or connection handover, instead of link handover, with a *target probability* p, i.e.,

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

for each satellite i in the established route. Equation 1 can be utilized to ensure that call termination or connection handover re-routing event occurs before a satellite in the connection route experiences a link handover event. The PRP removes any satellite that violates Equation 1 from consideration for routing. As a result, connection experiences either a call termination or a connection handover re-routing 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 probabilistic connectivity matrix R, i.e., $r_{ij} = c_{ij}$ for $1 \leq i, j \leq N$.

2. Find the *target route holding time*, T_{tr}, value such that
$$P(\min(T_c, T_{hr}) < T_{tr}) = p. \quad (2)$$

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 or minimum hop using R.

The *route usage time*, $T_{ru} = \min(T_c, T_{hr})$, is equal to the time interval the connection uses a route if no link handover occurs before call termination or handover re-routing. The algorithm simply removes the ISL links, which will be turned off in a time interval shorter than T_{tr}. Connection handover re-routing time, T_{hr}, depends on the cell geometry, 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 routing algorithm to the FHRP is presented.

Since the PRP removes certain links from consideration for routing, the call blocking rate of the network increases when PRP is used. Hence, a trade-off exists between the call blocking rate and the number of re-routing attempts due to the link handover. A distinction between new calls and 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 new calls.

4 Application of PRP to FHRP

The Footprint Handover Re-routing Protocol (FHRP) [5] has been proposed to balance the simplicity of route augmentation and the optimality of complete re-routing during a connection handover. The FHRP has two phases: Augmentation and Footprint Re-routing (FR). In the augmentation phase, a route between the new end satellite and a satellite already in the route is established, and unused portion of the route is removed. FR phase is applied after both end satellites are replaced by the satellites in their respective orbits. Connection route changes completely in the FR phase. Connection handover re-routing 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 re-routing time, T_{hr}, is equal to $\max(T_{hs}, T_{hd})$ 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}, $F_{hr}(t) = P(T_{hr} < t)$ is given by:

$$F_{hr}(t) = P(T_{hs} < t)P(T_{hd} < t) = [P(T_h < t)]^2, \quad (3)$$

where *single re-routing time*, T_h, is a random variable denoting T_{hs} and T_{hd}, which are independent and identically distributed random variables.

The pdf of single re-routing 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[1] 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 that 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 $[0, T_v]$ 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.

[1] 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.

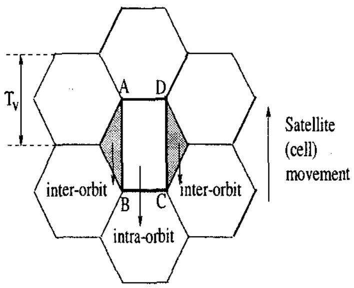

Figure 5: Inter- and Intra-orbit handover regions.

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. Call arrives to the network at time $t = t_e$. 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 $t = t_1 > t_e$, 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 re-routing time, T_h, for this user terminal is equal to $t_2 - t_e$, which has a pdf given as:

$$F_h(t \mid interorbit\ handover) = \begin{cases} \frac{t^2}{T_v^2} & for\ t \in [0, T_v] \\ 1.0 & for\ t \geq T_v \end{cases} . \quad (4)$$

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 re-routing time, $F_h(t) = P(T_h < t)$, is determined as:

$$F_h(t) = \begin{cases} \frac{2t}{3T_v} + \frac{t^2}{3T_v^2} & for\ t \in [0, T_v] \\ 1.0 & for\ t \geq T_v \end{cases} . \quad (5)$$

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

$$F_{ru}(t) = \begin{cases} 1 + e^{-\mu t}(F_{hr}(t) - 1) & for\ t \in [0, T_v] \\ 1.0 & for\ t \geq T_v \end{cases} , \quad (6)$$

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 minutes. When the call holding time is small compared to visibility period, such as when it is equal to 1 minute, route usage time is almost exponentially distributed with parameter equal to that of call holding time. Visibility period becomes more effective on the route usage distribution when connections stay in the network for longer time periods as in the case of calls with holding times equal to 10 minutes.

5 Performance Evaluation

The performance of the PRP has been evaluated to investigate the trade-off between the number of re-routing attempts

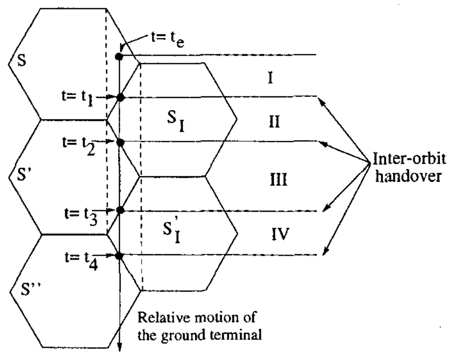

Figure 6: Timing Diagram for FR.

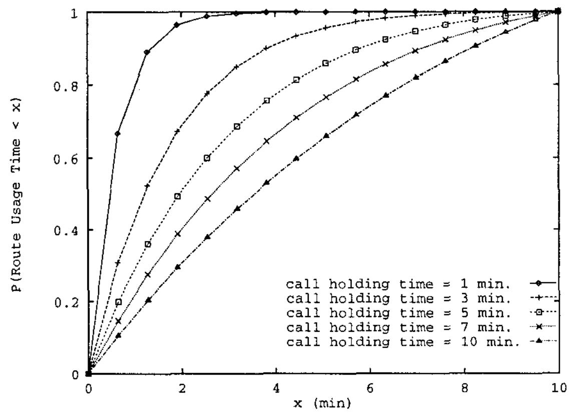

Figure 7: Route Usage Time pdf for Various Call Holding Times.

during the link handovers and 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 direct use of Dijkstra algorithm. The connections are 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 minutes. First 60 minutes of the experimental data is 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 [1] is used in combination with the PRP to find routes for new calls. The cost of each ISL 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.

Figure 8 shows the performance of PRP in terms of relative link re-routing frequency, which is defined as the ratio of the number of link re-routing attempts for a given target probability to that of for a target probability of 0. The effects of 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 re-routing operations. The reduction in the number of re-routing attempts is less for smaller target probabilities. As seen in Figure 8, relative link re-routing frequency is almost independent from traffic load since the PRP operates independently for each arriving call.

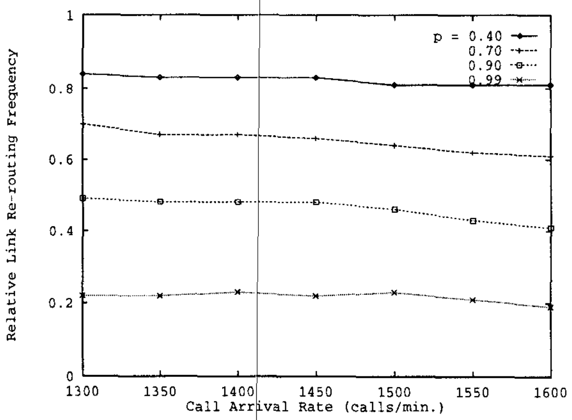

Figure 8: Relative Re-routing Frequency as a Function of Call Arrival Rate.

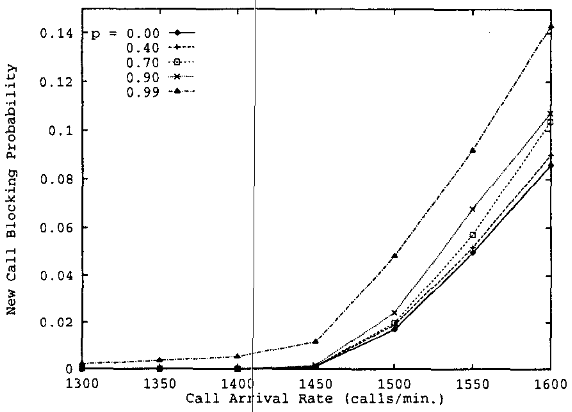

Figure 9: New Call Blocking Probability as a Function of Call Arrival Rate and Target Probability.

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 total 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 minutes. 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 minutes, i.e., the links that will be turned off within 7 minutes 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 using smaller target probabilities. For calls with holding times of 3 minutes and visibility period of 10 minutes, target probability would be set to 0.90 to decrease the relative link re-routing frequency to 0.5. Thus, an empirical choice of the target probability based on traffic characteristics and system geometry would solve the trade-off between the normalized re-routing frequency and the call blocking probability. The blocking probability in this case is very similar to that of target probability of 0.0. 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 re-routed calls. Not surprisingly, the re-routed call blocking for target probability of 0.99 is smaller for every call arrival rate simulated. Total call blocking probability, which is the ratio of

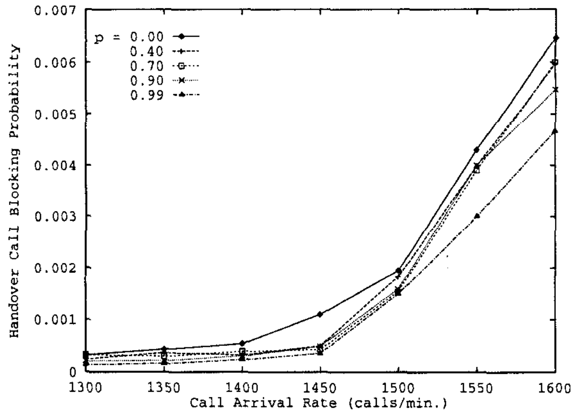

Figure 10: Re-routed Call Blocking Probability as a Function of Call Arrival Rate and Target Probability.

number of blocked calls to number of call arrivals, is very similar to new call blocking probability.

6 Conclusions

A routing protocol called Probabilistic Routing Protocol (PRP) has been proposed for Low Earth Orbit (LEO) satellite networks. The PRP reduces the number of re-routing attempts due to link handovers that occur due to the dynamic topology of the LEO satellite network. Basically, the algorithm tries not to use links that would be turned off before the connection is over. The probability distribution function of the time duration in which calls use the established route is determined. The developed probability distribution function is employed to find a route that will not experience a link handover with a certain probability when the connection is active. The simulation experiments indicate that the number of re-routing operations due to link handover can be decreased using large values of target probabilities. However, high target probability values result in high call blocking rates. Experimental results suggest that a suitable target probability value can be determined to achieve a trade off between the call blocking rate and the number of re-routing operations due to link handovers.

Acknowledgments: The author thanks Prof. Ian F. Akyildiz for helpful discussions, and Dr. Halûk Aydınoğlu and anonymous reviewers for their comments on the presentation of this paper.

References

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

[2] H.S. Chang et al. Topological Design and Routing for Low-Earth Orbit Satellite Networks. In *Proc. of IEEE GLOBECOM*, pages 529–535, 1995.

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

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

[5] H. Uzunalioglu, W. Yen, and I.F. Akyildiz. A Connection Handover Protocol for LEO Satellite ATM Networks. In *Proc. of ACM/IEEE MOBICOM'97*, pages 204–214, 1997.

[6] M. Werner et al. ATM-Based Routing in LEO/MEO Satellite Networks with Intersatellite Links. *IEEE Journal on Selected Areas in Communications*, 15(1):69–82, Jan, 1997.

[7] M. Werner et al. Analysis of System Parameters for LEO/ICO-Satellite Communication Networks. *IEEE Journal on Selected Areas in Comm.*, 13(2):371–381, Feb. 1995.

[8] 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.