1 Topological design, routing and hand-over in satellite networks

AFONSO FERREIRA

JÉRÔME GALTIER

PAOLO PENNA†

Abstract

In this chapter, we survey communication related issues arising in the context of Low Earth Orbit (LEO) satellite constellations. In particular, we study the impact of the predictable movement of the satellites on the techniques used in topological design, routing, and hand-over strategies.

1.1 INTRODUCTION

A Low Earth Orbit (LEO) satellite constellation consists of a set of satellites orbiting the Earth with high constant speed at a relatively low altitude (a few thousands of kilometers) [1]. Each satellite is equipped with a fixed number of antennas that allow it to communicate with ground transmitters/receivers and with other satellites. One of the major advantages of LEO satellites (as opposed to geostationary – GEO – satellites) is that they are closer to the Earth's surface. This allows to reduce the communication delay and the energy required to directly connect a user with a satellite.

On the other hand, two major issues arise due to their low altitude. First, a single satellite can only cover a small geographical area (called *footprint*) at the Earth surface, many satellites being thus required to provide global coverage. Second, the footprint of each satellite moves continuously, implying a high mobility of the *whole* network, in contrast with other cellular systems.

† Partially supported by a European RTN fellowship from the ARACNE project.

In the following, we will see how the topology of LEO constellations is limited by physical constraints. Then, we will review how these factors have been taken into account in the design of routing and hand-over policies.

1.2 TOPOLOGIES

During the systems design phase, several parameters come into play, such as satellites altitude, number of satellites, number of orbits and of satellites per orbit, how to deploy the orbits, and how to inter-connect the satellites. All such factors determine the topology of the network, as shown in this section.

1.2.1 Orbits

A closer look at the feasible types of orbits shows that unless the orbits have the same altitude and inclination, their relative position change so often that inter-satellite links (ISLs) can hardly connect them for a sufficient amount of time (for more details on orbit mechanics with respect to telecommunication services, see [1, 39]). Under such constraints, different kinds of constellations can be obtained according to how the orbits are deployed.

The so-called π-*constellations* are the structure of the Iridium system [20, 22] and at the basis of the original plans for Teledesic [21, 30]. The basic structure of a π-constellation consists in a set of orbits that are deployed along a *semi-circle* when viewed from a pole, as shown in Figure 1.1(a). The satellites are placed along the orbits so as to obtain a *maximum coverage* of the Earth's surface. In Figure 1.1(c) the deployment of satellites along with their footprints is shown. We can see that in a π-constellation there are two extreme orbits which are adjacent, but whose satellites move in opposite directions. As a result, a *seam* appears, that divides the network into two parts: those satellites moving from south to north and those moving from north to south (see Figure 1.1(a)-(b)).

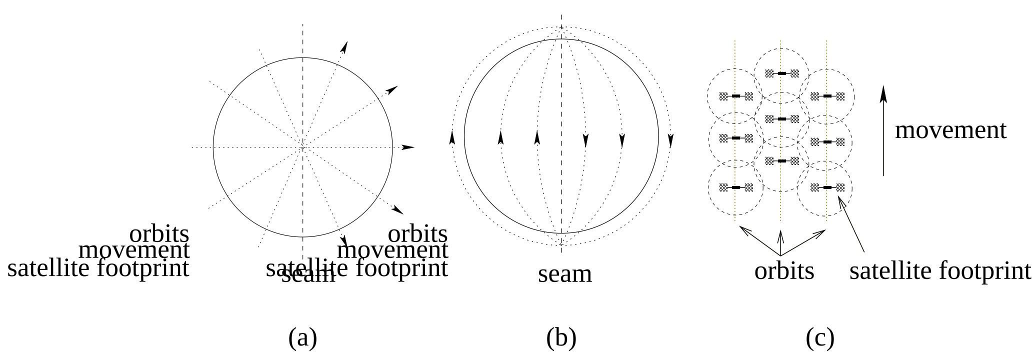

Fig. 1.1 The structure of π-constellations: (a) view from the north pole; (b) view from the equatorial plane; (c) the position of satellites on adjacent orbits and the resulting coverage.

From a communication network viewpoint, the seam is the main drawback of π-constellations, as it will be seen later in the text. Also, π-constellations suffer from excessive polar coverage. Finally, their unique coverage in many areas and, therefore, sensibility to many obstacles, like trees and buildings, does not always ensure a sufficient radio signal quality.

In order to avoid this kind of problems, 2π-*constellations* have been proposed. A 2π-constellation, consists in spacing the orbits along a complete circle as shown in Figure 1.2. The 2π-constellation is used in the Globalstar constellation [9], and has also been planned for the future Skybridge project and the now abandoned Celestri.

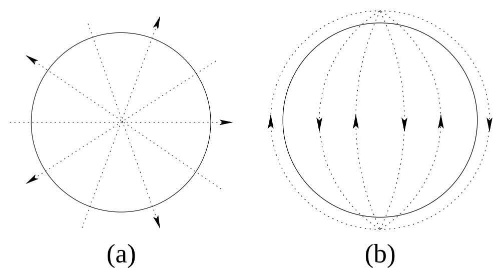

Fig. 1.2 2π-constellations: (a) view from the north pole; (b) view from the equatorial plane.

Another important aspect concerns the use of "inclined" orbits, that is, orbits whose inclination is between the equatorial inclination (0 degrees) and the polar one (90 degrees). Usually, π-constellations use polar orbits (informally, orbits that "roughly" cross the polar axis) for coverage reasons (see Section 1.2.5 below), and therefore are called "polar" constellations. On the other hand, inclined orbits allow a better optimization of 2π-constellations, hence the name "inclined" constellations. The use of inclined orbits allows to compensate for the satellites mobility with the Earth's self rotation, so as to increase the amount of time a satellite is visible from a fixed point on the Earth surface (see Section 1.2.5 below).

It is worth to observe that there is no technical reason to forbid the use of polar orbits on 2π-constellations and reversely. Moreover, the use of inclined orbits does not affect the network topology (for instance, π-constellations that use inclined orbits still result in the mesh-like topology shown in Figure 1.4(b)).

1.2.2 Inter-Satellite Links

The next step is to inter-connect the satellites through the ISLs. In particular, we distinguish between *intra-orbital* and *inter-orbital* links: The former connect consecutive satellites on the same orbits, while the latter connect two satellites that are on different orbits. In Figure 1.3 we show three possible patterns that can be obtained by using inter-orbital links between adjacent orbits: the "W" pattern and the "inclined" pattern in Figures 1.3(a)-(b) use four ISLs per satellite, while the pattern in Figure 1.3(c) uses only three ISLs.

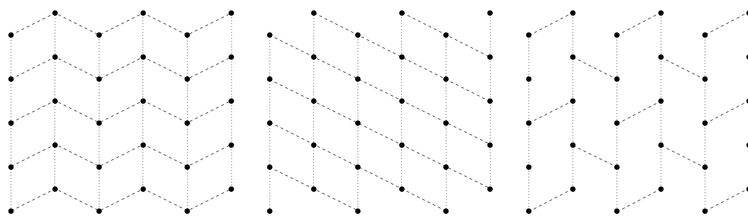

Fig. 1.3 Some inter-satellite link patterns.

Consider now the "W" pattern in Figure 1.3(a). In order to obtain the network topology we have to take into account the seam and the relative position of satellites crossing the poles, as follows.

For π-constellations, one has to consider the problem of connecting two satellites moving in opposite directions, which is too expensive or even infeasible with the existing technology (see Section 1.2.5). Hence, it is commonly assumed that two such satellites cannot be directly connected over the seam, even though they are "physically" close one to each other. Therefore, long user-to-user delay can occur even when the two parties are geographically close to each other but the covering satellites are separated by the seam. Also notice that two adjacent satellites swap their relative position whenever crossing the poles (see Figure 1.4(a)). Hence, the network topology can be represented as a two-dimensional mesh where columns are wrapped around, but rows are not (see Figure 1.4(b)).

In [15] the impact of the ISLs architecture (for instance, the use of antennas that support higher angular velocity) has been studied, and further patterns to connect the satellites of a π-constellation have been proposed. Such patterns use inter-orbital links that connect satellites in non adjacent orbits, typically the neighboring orbit of the neighboring orbit. This reduces the user-to-user delay when the communication takes place between two positions that are quite far (or when the communications have to go across the seam). Under the assumption of ISLs that support high angular velocity, the delay effects of ISLs that cross the seam have also been investigated in [15].

With respect to inter-satellite links for 2π-constellations, neither Globalstar nor Skybridge have implemented ISLs in their design, although it seems that they have been considered in the early phase of these projects, as it was the case in Celestri. At that point in time, many designers thought those projects were innovative enough to delay the introduction of this additional new feature. Nevertheless, there is a strong belief that future designs of 2π-constellations will introduce such links.

From the topology point of view, it is worth to observe that the regular torus turns into a *skewed torus* if an inclined ISL pattern such as the one of Figure 1.3(b) is adopted [17]. Notice that 2π-constellations do not present any seam. Thus, their coverage has smoother properties. On the other hand, a unique position may be covered by two satellites quite far one from another in the network topology (e.g.,

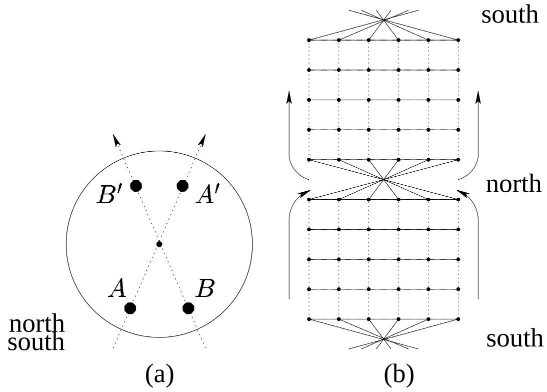

Fig. 1.4 The relative position of two adjacent satellites crossing the pole and the resulting topology of π-constellations.

two satellites that move in opposite directions), especially when the user is close to the equator.

1.2.3 ISLs versus Terrestrial Gateways

The use of ISLs is intended to implement communications that do not use any terrestrial infrastructure. However, the use of terrestrial gateways still present some advantages such as a reduced number of computing devices on-board the satellites. For instance, gateways can be used to compute the routing tables that are used by the satellites.

A more extensive use of the gateways has been adopted in the Globalstar system, where the satellites operate in a "bent-pipe" mode. Their main function is to redirect user signals to ground gateways, and vice-versa. As a result, the operator has to build many gateways, one for each area in which the service is opened. Additionally, part of the radio spectrum is used to support the communications between the satellites and the gateways. Unfortunately, radio resources are becoming a scarce resource. Currently, several systems share the same spectrum of frequencies (Globalstar, ICO, and probably Ellipso) which is the source of several interference problems.

We note that the use of ISLs presents significant advantages, like reducing the communications between the satellites and the gateways, reducing the number of gateways, balancing the load between the gateways, and preventing gateway faults.

1.2.4 Multiple Coverage

Another important issue for satellite constellations with ISLs is risen by *multi-coverage* goals. From the radio and signal propagation point of views, a single satellite may not suffice to ensure the real-time connection, especially if some obstacles exist between the user and the satellite. Systems like Globalstar [9] answer this problem using *multi-path techniques*: Instead of being received by one satellite, the signal is received by two to four satellites and merged to recover a clear signal. When a new satellite is visible to the user, its signal contribution is introduced progressively into the global merging of the signals.

We remark that routing with multi-path techniques in a satellite constellation is very challenging. A single user may be directly connected to two (or more) satellites that are very far one from another in the network topology, mainly in inclined constellations. From the algorithmic point of view, this characteristic essentially turns a basic network routing problem into a multicasting problem.

1.2.5 Physical and Technological Constraints

In this section we discuss some of the main physical and technological factors that impact on many of the above design choices.

ISLs Geometry. The main technological constraints to take into account in ISL design are the relative angular velocity of the endpoints and their visibility [17]. This is because antennas cannot afford excessive angular speed and the atmosphere is also a source of fading of the signal.

Mobility of ISLs. As a satellite moves along its orbit, the set of satellites visible from it changes continuously. This happens for those satellites that are not in adjacent orbits and, in polar constellations, whenever the satellite approaches the poles. This is due to the small distance between adjacent satellites approaching the pole which results in a higher angular velocity [1, 15]. Additionally, ISLs between adjacent orbits must be turned off when crossing the poles because of the satellites relative position switching (see Figure 1.4). As observed in [15], ISLs that support higher angular velocity allow to maintain intra-orbital links at higher latitudes. An unexpected side-effect of the angular velocity is that the tracking system may affect the stability of the satellite within its orbit and therefore result in an additional consumption of fuel, which in turn impacts the satellite's weight and time in service.

Shortest (Delay) Path. It is worth to observe that the distance between two adjacent polar orbits decreases as they get closer to the poles. Hence, for π-constellations using the "W" ISLs pattern, for instance, the minimum delay path is the one that uses a minimal number of ISLs and inter-orbital links whose latitude is the maximum latitude between the two satellites to be connected.

Notice that routing algorithms on mesh-like topologies may return sub-optimal time/delay paths, since such models do not consider that the orbits distance varies with the latitude. In [10] a model that takes into account this issue has been investigated.

1.3 NETWORK MOBILITY AND TRAFFIC MODELING

There are two main factors that should be taken into account when designing routing algorithms for LEO satellite constellations:

- *Users' distribution:* the fact that the position of the users and the duration of the communications are not known in advance.
- *Network mobility:* the fact that satellites move, constantly changing the network topology.

Although the first aspect has been extensively studied for classical cellular networks, such networks use wired connections in order to connect two base stations. Hence, the main issue in these "terrestrial networks" is to provide enough resources for the user's connection to last. There is a lot of flexibility in the size of the cells, but the users may move from one to another, and at different speeds. Conversely, LEO cells are big enough to consider the users immobile. However, routing problems occur since on-board resources – in particular the maximum number of connections using ISLs – are scarce.

The second aspect, namely the network mobility, is a distinguishing factor of LEO constellations. Indeed, even if we assume a static set of communications (i.e., pairs of users that want to communicate one with the other), the problem of maintaining active these connections over time is not a trivial task: The satellites movement triggers both hand-overs and connections updates (re-routing), when a topology change occurs.

In both cases, mobility is the main cause of call blocking, call dropping, and unbounded delay in communications. However, there is a fundamental difference between the users' mobility and the network's mobility: The users' behaviour is *not deterministic*, while changes to the network topology are *predictable*. Hence, two different approaches are generally adopted:

- The network's behaviour is deterministic and can be "predicted" quite accurately (see Section 1.3.1).
- The users' behaviour is usually modeled by means of a probability distribution (see Section 1.3.2).

It is worth to observe that if we consider the relative movement between a user and the satellites, then the major part of such movement is due to the satellites speed. Hence, the probability distributions used to model users' mobility mainly focus on the issue of managing requests whose position and duration are not known prior to their arrival.

1.3.1 Satellites Mobility

One of the main differences between "classical" cellular networks and LEO constellations is the high mobility of the system. Complicating factors such as the satellites movement and the Earth's self rotation make the problem of connecting "immobile" users non trivial. In the following we describe the interplay between these two factors and the previously mentioned aspects, and also how network mobility can be modeled.

1.3.1.1 Satellites Movement The satellites movement is the main cause of hand-overs. Two types of hand-over may occur:

- *A satellite hand-over* is the transfer of a user from a satellite to another during a communication.
- *A cell hand-over* is the transfer of a user from a spot-beam to another, within the same satellite. A satellite antenna directed to terminals is composed by a series of beams. Such a decomposition of the satellite footprint allows to reuse the radio frequencies several times in its coverage area. Those hand-overs have no consequence on the inter-satellite routing, but impact seriously on-board computations.

If a user is just on the border of the coverage area of a satellite, his/her connection time to an individual satellite can be extremely small. Hence, in general, constellations are designed in such a way that the footprints overlap and extremely small connection times to an individual satellite never happen. Nevertheless, the *maximum* connection time is still limited. A user's trajectory, viewed from the satellite, will resemble a straight line crossing the center of the coverage area. The apparent (or relative) speed of the user is then the speed of the satellite. This causes the following undesirable phenomena: Visibility changes, varying topologies (ISLs changes), footprint hand-over, and need for re-routing.

1.3.1.2 Earth's Self Rotation The Earth's self-rotation introduces some more complication in the system. In Figure 1.5, we plot the maximum time between two satellite hand-overs against the altitude h and the elevation angle ε of a constellation, in two cases:

- The Earth's self rotation is not taken into consideration and the satellite's inclination can be arbitrary.
- The Earth's self rotation is taken into account and the orbit of the satellite is equatorial.

1.3.1.3 Modeling the Network Mobility Notice that the maximum hand-over time, shown in Figure 1.5, can vary from some minutes up to several hours. Also, inclined orbits can be used to exploit the Earth's self rotation to increase the visibility period. Hence, the mobility of the network can also vary a lot. Roughly speaking, they can be distinguished between low and high mobility, depending on the maximum hand-over time.

Low Mobility (periodic). In [5], the mobility of a satellite constellation is described as a Finite State Automata (FSA) by a series of *states* described along the time in

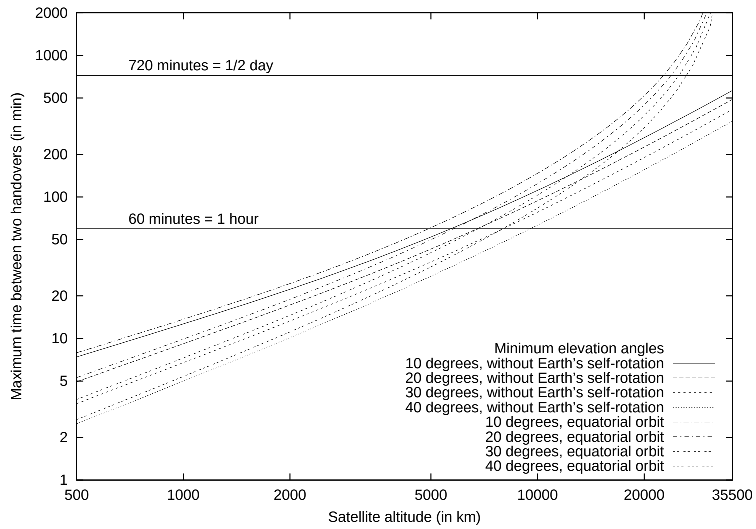

Fig. 1.5 Maximum time between two satellite hand-overs.

round-robin fashion. The main advantage of this model is that we have to consider only a finite set of configurations of the satellite constellation (where the satellites are assumed to be immobile), and provide efficient routing solutions for each of them, inspired from classical telecommunication problems.

Low Mobility (aperiodic). It is worth to observe that the "periodicity" assumption of the FSA model may be, in some cases, too strong. This is essentially due to the combination of "physical" factors, such as the Earth's self rotation, the satellites speed, the use of inclined orbits, etc. They make the system aperiodic for all practical usages, i.e. a satellite will find again the same position only after such a long time that too many intermediate states would be necessary. In this case, a possible approach consists in taking a series of snapshots or fixed constellation topologies, method sometimes referred to as *discretization* [11, 37, 38]. Then, the routing problem is solved with respect to that fixed "constellation".

High Mobility. The above two models are interesting when the mobility of the satellite network is negligible with respect to the mobility of the users requests, e.g. if most of the requests have very low duration, let us say a few minutes, while the hand-over time would be of one hour or more. In this case, before the network configuration changes (significantly) several (many) requests will have been satisfied.

On the other hand, these models do not take into account the dependence between consecutive states of the network. Thus, between two states the *complete routing scheme* of the constellation should be changed. Clearly, in the case of highly dynamic constellations and/or long call durations, almost all requests may pass through several states and thus may be re-routed several times.

1.3.2 Users Distribution: Common Traffic Assumptions

Depending on the application, three major scenarios can be identified for satellite markets. The first one, and the most natural one, states that satellites will serve countries where the telecommunication infrastructure is insufficient or inexisting. The second one, which appears as more and more probable, is that the satellites will provide additional capacities to countries that already have good telecommunication infrastructures, but which suffer from an overload of the resources. A third market concerns people who require a seamless connection in their international activities. Of course, depending on the scenario, the traffic may have different characteristics, as summarized in Table 1.1.

Type	**Developing**	**Overload**	**International**
Location	poor countries/oceans	rich countries	international
Time distribution	Poisson-like	bursty	nearly deterministic
User concentration	sparse	huge	irregular
Call duration	short	exponential	long
Call distance	average	short	long

Table 1.1 Characteristics of foreseeable usages of satellite constellations.

Little is known on the two first classes of applications. The last one has been investigated in [35], where an analysis of the international activities led to a map of different zones worldwide. In this model, the planisphere is divided into 288 cells, with 24 bands along the longitudes and 12 along the latitudes. The intensity levels from 0 to 8 shown in Table 1.2 correspond to traffic expectations for the year 2005, of 0, 1.6, 6.4, 16, 32, 95, 191, 239, and 318 millions of addressable minutes/year, respectively. In [15] the traffic requirement matrix is obtained from trading statistics, namely the imports/exports between any two regions. Further market studies on satellites can be found in the quarterly report [23].

In the following we describe how the users mobility can be modeled by means of some traffic assumptions. In particular, we group traffic assumptions into three categories:

- Geographical distribution: On which satellite the user requests are expected to arrive.

- Time distribution: How long they are expected to be active.

- Rate distribution: How much of the resources they will require.

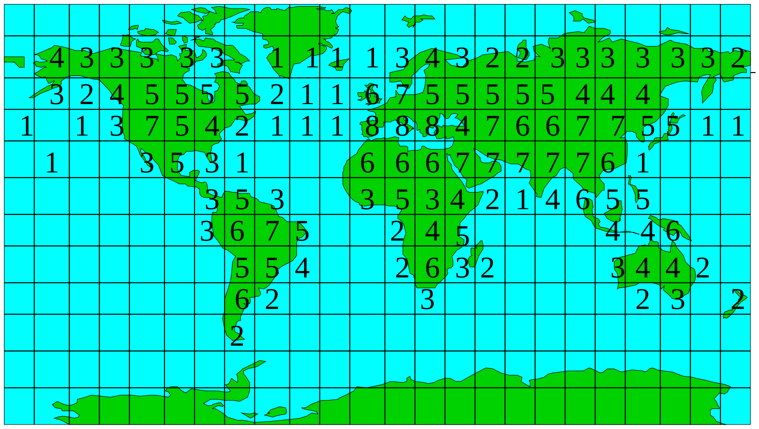

Table 1.2 **Intensity levels on the planisphere for the distribution of users.**

1.3.2.1 Geographical distribution Statistical models have been developed to represent the load all over the Earth. A structure that appears promising is the notion of *point process* over the 2-dimensional Euclidean space [8, 3, 16, 18]. A point process is a family $X = \{X(B),\ B \in \mathcal{B}\}$ of non-negative integer-valued variables, where $X(B)$ denotes the random number of points that lie in the set B. A *homogeneous Poisson process* with parameter $\lambda > 0$ is a point process X as follows.

- The number of points $X(B)$ is Poisson distributed with parameter $\lambda\mu(B)$ for each bounded Borel set B.
- The random variables $X(B_1), \ldots, X(B_n)$ are independent for each sequence $B_1, \ldots, B_n$ of disjoint Borel sets.

In fact, the homogeneous Poisson process reflects quite well the traffic load within a country with uniform development. More generally, when trying to map a point process to the entire world it would be interesting to either choose a measure μ that reflects the economic development of each region (i.e., for instance, the map of Table 1.2), or try a model with different properties, such as MMPP or multi-fractal models [2] (for more details, we refer the reader to the survey in [13]).

1.3.2.2 Time Distribution The traditional way to model the distribution of the call durations consists in using a Poisson law. In fact, the behavior of the traffic is then very close to that obtained on phone systems. However, new broadband applications, made possible by the Internet, generate other types of traffic. In [24] a comparative study between self-similar and Poisson traffics is done in the satellite constellation context.

1.3.2.3 Rate Distribution It is quite natural to relate the rate distribution to the locations of the different parties of a communication. In [36], the load of the intercontinental traffic is evaluated. It is estimated that between 81% and 85% of the traffic is within continents, the remaining traffic being shared with the closest and/or most populated areas. Another method of generating traffic is suggested in [6]: Once a pair (u,v) of locations is selected (u and v are viewed here as points on the unit sphere), associated with *potential requirement densities* w_u and w_v, the traffic requirement between the two nodes is given by

$$T(u,v) = \frac{(w_u w_v)^\alpha}{d(u,v)^\beta},$$

where α and β are two parameters set by the user. In [6], it is assumed that $\alpha = 0.6$ and $\beta = 0.5$.

1.4 ROUTING AND HAND-OVER

A good routing strategy should mainly prevent from (1) the congestion of ISLs due to too many routes passing trough them; (2) routing requests along paths containing many links since this results into a poor resources utilization and in a higher delay in the communication; (3) dropping an ongoing call or blocking a new one.

1.4.1 Problems and Optimization Criteria

Here, we describe more in detail the main goals concerning the design of efficient routing algorithms for LEO constellations.

Maximum Throughput and ISLs Usage. Maximizing the throughput under limited ISL capacity appears as one of the main objectives of the constellation designers. Clearly, because of the limited ISL capacity, a good routing strategy should minimize the maximum link usage (e.g., the load due to the overall traffic passing through such link) among all the ISLs.

Shortest/Bounded Delay and Jitter. One of the main motivations of LEO systems is the reduction of the communication delay. Indeed, the minimum delay to open a connection through a geostationary satellite is around 240 ms, while a LEO could connect two users in around 20 ms. However, while the connection is roughly independent of the parties' location in a GEO system, the delay significantly increases for LEO systems when the parties get further from one another. However, since the ISLs offer straight free-space propagation, the delay between the satellites is governed by the speed of the light.

The multi-path techniques and in-the-air merging of the signal (see Section 1.2.4) should raise a new delay problem. Indeed, merging two signals that are far in the network topology takes time comparable to the one required to reach a geostationary satellite. The same problem occurs with π-constellations when the communications have to go across the seam. In each case, the communication delay could take, in the worst case, an additional 100 ms to be completed (the time to reach the furthest satellite in the constellation).

The jitter (in other words, the delay variation) is relatively important in LEO constellations, since the distance between the user and the satellite (and also between the satellite and the gateway, and even between two satellite of different orbits) changes continuously during the lifetime of a communication. This behavior cannot be avoided, but can lead to storage optimization issues in the satellite (in case the terminal is not able to handle it) for instance.

Guaranteed Hand-Over. The next optimization problem concerns the quality of service and more precisely the guarantee that is given to the users that a communication will not be dropped because of a hand-over. This can be done either by fixing an acceptable rate of call dropping, or by forcing the system to avoid call dropping at any price.

Call Admission and Routing. The guaranteed hand-over feature greatly impacts the call admission procedure and can lead to additional call blocking. Blocked calls may also be a consequence of scarce link resource availability. This may happen either because a user cannot be connected to the visible satellite(s) or no route between the two satellites used in the communication can be found without overloading the ISLs capacity.

1.4.2 Algorithmic Solutions

1.4.2.1 Call Admission for Hand-Overs The call admission procedure decides whether an incoming communication request will be handled or not. A user will be refused the entry in the system when there is not enough available capacity to take the request in consideration. A user may be also rejected because the system is not able to guarantee the duration of the service sufficiently enough to meet quality of service goals. For instance, a user could be immediately served, but after one minute the communication will have to be dropped because it interferes with other users, more privileged or who were already in the system before the latter user's arrival.

Two main schemes can be used to control the resources of the LEO satellite system:

- *Earth-fixed* cells can be drawn directly on the ground. One cell has a fixed capacity, and is served by the same (set of) satellite(s). Therefore, the hand-overs occur simultaneously for all the users of the cell. This scheme reduces the amount of usable capacity by the satellite, but simplifies the management of the users on-board the satellites. This idea has been used for the plans of the Teledesic constellation. The Earth's surface was divided into stripes parallel to the equator of 160 km from the South to the North, each stripe being redivided along the longitudes into squares of approximatively 160 km per 160 km, for a total of around 20 000 cells. Those "super-cells" were then further subdivided into elementary cells of 53.3 km per 53.3 km [21].

It is then to the responsibility of the constellation's designer to make sure that at each instant of time each cell will be served by at least one satellite.

- *Satellite-fixed* cells is the most common scheme, applied in particular for Iridium. The users are handled individually by the satellite and several satellites may serve a user at the same time.

As mentioned in Section 1.3, the satellite speed is responsible for the greatest part of the system mobility. After that comes the Earth's self-rotation, and only then the terminal's mobility.

The admission control and resource allocation work has been mainly focused on satellite-fixed cells. Usually, system designers try to include an *overlapping area* between any two satellite-fixed cells. Therefore, some researchers have proposed that, when a terminal comes into the overlapping area, and no channel is available in its new covering cell, it issues a *hand-over request* in a queue (the *Hand-over Queue* - *HQ*) that has priority over incoming calls [28]. Although this idea may enhance the quality of service, the final result depends on many factors, like the size of the overlapping area, and the distribution of the length of the calls. Although this procedure certainly enhances the system, there is no mean to reach a *target* quality of service.

Another idea consists in systematically reserving some channels to the hand-over requests [19]. In this *Hand-over Gard* (HG) system, if the number of busy channels exceeds a given threshold, then no incoming call is accepted and only hand-over requests are handled. This system becomes more efficient when the threshold goes down. However, if the threshold is too low, the system will be under-loaded in many cases, as incoming calls that could have been accepted are rejected. Hence the need of a tradeoff between quality of service and system capacity. In [7], additional concepts of geographical position are integrated into this concept, so that the authors can evaluate a call blocking probability depending on the remaining time the user has in his/her cell, and the expected duration of a call. A new user is accepted into the system if his/her hand-over call blocking probability will meet QoS requirements, and his/her inclusion does not degrade the existing calls under the QoS target. This results in a more accurate acceptance/rejection of users.

In [29], the authors consider, for each terminal, the servicing time of a satellite cell. In particular, they analyze its movement and deduce the instants t_i and t_e when the terminal enters the cell and when it leaves it, as shown in Figure 1.6. As a result, when a terminal is introduced in the system, not only the present covering cell is reserved, but also a sufficient number of future cells, associated with future utilization, so that the service duration meets the quality of service requirements (this algorithm is called *Guaranteed Hand-Over* – HG). If the communication continues, then additional reservations are done in real time. In the worst case, when this life-time reservations fail, the terminal can be notified that the connection will end within a certain amount of time.

1.4.2.2 ISL Dimensioning In [4, 15], the impact of the ISLs design phase on the overall performances of the network have been pointed out (see Section 1.2.2). It

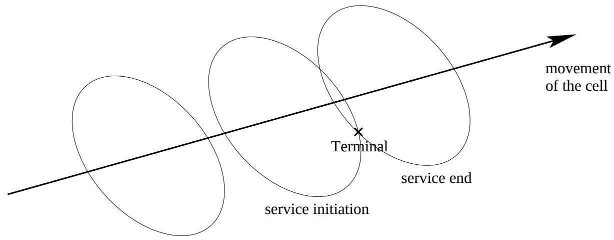

Fig. 1.6 Instants of service initiation and service end for a satellite-fixed cell.

is worth to observe that, in these works, a *unique* pattern is chosen *once and for all* during the design phase.

A different approach has been adopted in [37, 38] where the authors consider different discrete time steps (each time step corresponds to a "snapshot" of the constellation, that is, the relative positions of the satellites). For each time step, an ISLs dimensioning problem is solved so as to minimize the maximum ISL load which, in turns, yields the sufficient ISL capacity to route the requests.

1.4.2.3 Precomputed Routes and Snapshot/FSA Techniques: Deterministic Routing In this section we survey how the deterministic behaviour of the network can be exploited to enhance the overall performance of the network. In particular, the following three aspects will be considered: (a) design choices: how satellites are interconnected through ISLs; (b) reservation strategies: how to guarantee that a communication will not be dropped; and (c) "ad-hoc" routing algorithms: algorithms that depend on the current state of the network.

Consider a *discrete* model in which each state of the network corresponds to a *visibility state* representing the set of satellites that are visible to each other (i.e., those pairs of satellites that, potentially, can be directly connected to each other). Then, two opposite policies (*off-line* or *on-line* design) can be adopted: Either provide a unique strategy which *does not depend* on the visibility state (i.e., does not vary over time), or provide a different strategy for *every* visibility state.

In the sequel we first describe some basic routing heuristics that do not exploit the knowledge of the visibility states. Then, we will see how these solutions can benefit from the deterministic behaviour of the network.

Basic Routing Strategies. In [31] different routing and reservation strategies for mesh-like topologies (a 6×11 mesh model is adopted) have been compared. As for the routing strategies, to each link is associated a cost function and the routing algorithm chooses a path with minimum cost. According to different definitions of the cost function the following four strategies have been compared[1]:

[1] In [31] such strategies are named Minimum Hops Algorithm, Minimum Cost Algorithm, Mesh Algorithm and Revised Mesh Algorithm, respectively.

- Min Hops: the cost of each ISL is 1, hence the connection is always established by choosing *any* path that minimizes the number of ISLs in it (i.e. the minimum number of hops).
- Min Load: to each link is associated a cost given by $1/vacancy$, where $vacancy$ is the number of free channels in the link; the chosen path minimizes the sum of the cost of the ISLs in such a path.
- Min Hops Min Load: among the set of paths with minimal number of hops, consider the one with lowest maximum link usage. Then, the call is accepted if and only if such route contains no link that is overloaded.
- Revised Min Hops Min Load: it is defined as the previous one, with the only difference that, if a request cannot be routed along a min-hops route, then a sub-optimal delay path is chosen.

Notice that the first strategy arbitrarily chooses any min-hop (delay) path, while the other strategies aim to keep the link usage as small as possible. Interestingly, the experimental comparison in [31] shows that, at least for some probability distribution, it is fundamental to avoid the use of highly loaded links *and* the use of paths with sub-optimal length. Indeed, the worse strategy turns out to be the Min Hops one, while the Min Hop Min Load strategy performs better than the Revised one.

From On-line to Off-line: Precomputed Routes. In [6], the deterministic behaviour of the network has been also exploited to enhance the performance (in terms of call blocking probability) of the routing strategies. In particular, the authors compare a min-cost strategy (similar to those mentioned above) to a *static routing*, in which the routes are computed *off-line* in advance. The network is modeled as a FSA (see Section 1.3.1.3) and, for each state (i.e., network topology), a set of routes is precomputed. Clearly, this reduces the communication overhead due to the periodic broadcast operations required to update the link state (i.e., the load) information. Somehow surprisingly, this approach has also a better call blocking probability. This is mainly due to the fact that, after a network change, the solution provided by the min-cost algorithm is rather “unstable”. So, several calls are rerouted.

This idea of moving the complexity of routing from an on-line problem to an *off-line* one is also the basis of [38, 37]: There, for each possible demand pair (i.e., for each possible communication), k different shortest paths are computed. Those paths are then used during the on-line phase in which the routing algorithm chooses a route between such k candidates. Based on this idea, an upper bound on the ISLs capacity is given by the number of such paths containing a link. Informally speaking, it is assumed that *there is a request for each possible pair of satellites* and the goal is to compute the minimum l such that, if every ISL capacity is l, then every request can choose between k different paths of minimal length. This problem is formulated by means of a linear programming system. In [37] the authors also consider the problem of *relating* subsequent states of the network, so that re-routing is performed only when necessary. More precisely, a routing state will be computed and optimized while keeping some remaining routes of the previous states.

1.4.2.4 Predicting Link Hand-Overs Another approach consists in taking advantage of the link hand-overs to manage the mobility of the network.

The Probabilistic Routing Protocol. Upon a call arrival, a first route through the satellites has to be assigned to it. Such a route can be chosen based on any criterium, like minimum number of hops, least congestion, minimum cost, etc. The deterministic movement of the satellites allow the prediction of the time frame where a link hand-over is to occur. Hence, the *probabilistic routing protocol* – PRP–, proposed in [33], tries to establish an arriving communication through a route which has minimum probability to be cut by a link hand-over. For this, it supposes the existence of a probability distribution function (PDF) of the call time duration over a route.

The protocol applies Dijkstra's algorithm to find the routes. The cost of each ISL is set to one, implying that the route will be minimum hop. The PDF is used to remove from consideration of Dijkstra's algorithm those ISLs that are likely to hand-over during the communication.

It is easy to see that the PRP works for very short calls, since a direct consequence of its implementation is that the route just set will not be cut (with a given target probability) during the call. For instance, a target probability of 0.99 reduces link rerouting operations by 80%, when compared to the "pure" Dijkstra's algorithm. Unfortunately, the probability of blocking new calls increases due to the forbidding of many ISLs, being almost 15% for 3-minute calls for the same target probability, over 1600 calls.

The Footprint Hand-Over Rerouting Protocol. A rerouting strategy, called *footprint hand-over rerouting* – FHR–, was proposed in [32, 34], which uses informations about the network predictability to replace a current route by a new one, based on the successor satellites in the original route, as follows.

Let $R = S_1, S_2, \ldots, S_k$ be a route connecting satellite S_1 to satellite S_k. In case *both* route-ends undergo a link hand-over, then a *footprint rerouting* can take place and R is rerouted through $R' = S'_1, S'_2, \ldots, S'_k$, where S'_i denotes the successor of satellite S_i in the same orbit. In case only the route-end S_1 (respectively, S_k) has to hand-over the communication, but S'_k (respectively, S'_1) is not yet visible to the ground-user, then the original route R is simply *augmented* to $R' = S'_1, S_1, S_2, \ldots, S_k$ (respectively, $R' = S_1, S_2, \ldots, S_k, S'_k$), until S'_k (respectively, S'_1) becomes visible to the ground-user. When the second route-end undergoes its link hand-over, then a footprint rerouting is used and the new route will be $R' = S'_1, S'_2, \ldots, S'_k$.

Evidently, if R is a minimum-hop route, then FHR implements a rerouting where R' is also a minimum-hop route. Further, in the case where the link-cost is a function of traffic load, and this latter is time-homogeneous, then R being a minimum-cost route implies that R' is also a minimum-cost route.

The performance of FHR has been evaluated through simulation, particularly against a pure augmentation approach, where a call is dropped if an augmented link cannot be found between the hand-over satellite and the current route. New calls first routes are found through the implementation of Dijkstra algorithm. Two cost functions are studied, namely minimum hops, and congestion. In the case of homogeneous traffic with high call arrival rate, FHR blocks more new calls than pure augmentation, but drops much less calls because of hand-overs, than pure augmentation. This shows the interest of taking the deterministic network topology into account.

1.4.2.5 Constellations Viewed as Dynamic Networks Another idea consists in managing more specifically the mobility of the constellation. The more natural strategy consists in taking *shortest-path* algorithms, which minimize the resources taken by an individual route using a cost based on the links. Each link receives a cost (proportional to its expected load), and the route minimizes the sum of the costs of the links it uses.

An elementary model is inspired from *ad-hoc* networks [26]. The two main strategies are:

- *Proactive:* any change of topology is immediately notified to all the living nodes;
- *Reactive:* a request is likely to initiate a flood that will discover the actual state of the network.

The choice between the two methods depends on the dominating activity of the network. Proactive protocols perform better when the traffic throughput is high and the topology changes seldomly, whereas the reactive protocols perform well when the network topology often changes and the traffic is low. An example of proactive protocol is the Extended Bellman-Ford (EXBF), proposed for satellite constellations in [25], where routing tables are dynamically updated using a shortest paths policy. A reactive protocol, named Darting, was also proposed in [25]. The algorithm broadcasts network information only if it becomes absolutely necessary. In the meantime, the algorithm takes advantage of data packets to update the network topology. Experiments have shown that the Darting algorithm required as much as 72% percent more overhead when compared to the EXBF algorithm in an Iridium-like, lightly loaded network [27].

An immediate simplification occurs when one considers that the weights of the links are uniform, and the topology is regular. In the mesh or the torus/mesh case, one shortest path in terms of number of hops is obtained by a XY routing (that is a route consisting of a series of inter-orbital links first, and then a series of intra-orbital links). When one considers the minimum *delay*, instead of the minimum number of hops, it is clear that the cost of inter-orbital links are smaller when they are closer to the poles (while the intra-orbital link cost remains constant). Then, some geodesic considerations may give the advantage to routes that take more polar inter-orbital links, as explained in [10].

1.4.2.6 Reservation Strategies A second step towards reducing the call blocking probability is the choice of the *link reservation* strategy. Indeed all these algorithms attempt only to *minimize* the load of the links, without implementing congestion control mechanisms, which would give some guarantee that a communication will not be dropped when one of the links it uses experiences an hand-over.

Notice that most of the hand-over techniques we described, such as the HQ, HG, GCAC, and GH, could be applied to enhance the routing. A *basic reservation* strategy was proposed in [31]. It essentially consists in accepting a call if and only if it is possible to reserve ISLs in such a way that the communication can be maintained after any one link hand-over event (this can be seen as an extension of the GH concept). Here the deterministic behaviour of the network plays a key role: It is assumed that, for each link, we can determine its next hand-over and the overall network topology. So, if during the call *only one* of the links in the path has a hand-over, it is guaranteed that such call will not be dropped.

Another related strategy is the *slack reservation* policy, which consists in always accepting the new call and, if possible, also reserving the links for a “next hand-over” event. Intuitively, we are relaxing the basic reservation strategy, since those calls whose duration is sufficiently short will terminate *before* the link hand-over event occurs. However, the experimental results in [31] show that the slack reservation strategy performs worse than the basic one. This is mainly due to the fact that, in the slack reservation, a certain number of calls are accepted with no reserved route. Clearly, such calls are vulnerable to link hand-overs. On the other hand, calls with reserved routes “consume” more resources than they would without reservation. Hence, with less system resources available, the “vulnerable” calls are more likely to be dropped.

Going a step ahead, an algorithm is presented in [12] that forwards the reservation once it is used. It is shown that the reservation *permanently* guarantees the communication provided that only South hand-overs occur. All kinds of hand-overs are taken into account in [14], where the call admission control problem for regular satellite networks is turned into a problem of max-load of a family of rectangles in a two/three dimensional space. Communication requests are represented as a series of rectangles. Required capacity is then equivalent to the maximum number of rectangles that intersect on a given point. Therefore, capacity control for call admission policies can be done through geometric algorithms.

1.5 CONCLUSIONS

In this chapter we reviewed the literature concerning routing and hand-over techniques for LEO constellations. We showed that they are particularly complex in such a context because of both physical constraints and the movement of the satellites. Notwithstanding, several interesting results exist and, at the time of this writing, two satellite constellations are operational (Iridium and Globalstar). These facts certainly speak for more research to be undertaken in the near future.

1.6 ACKNOWLEDGMENTS

This work was partly supported by the European RTN ARACNE, the INRIA/CNPq and the RNRT/Constellation projects.

References

1. E. Altman, A. Ferreira, and J. Galtier. *Les réseaux satellitaires de télécommunication*. Dunod, 1999.
2. S. Appleby. Multifractal characterization of the distribution pattern of the human population. *Geographic analysis*, 28(2):147–160, 1996.
3. F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev. Géométrie aléatoire et architecture de réseaux. *Ann. Télécommun.*, 51:158–179, 1996.
4. P. Bergé, A. Ferreira, J. Galtier, and J.-N . Petit. A Probabilistic Study of Inter-Satellite Links Lo ad in Polar Orbit Satellite Constellations. *International Journal on Telecommunication Systems*, To appear.
5. H.S. Chang, B.W. Kim, C.G. Lee, Y. Choi, H.S. Yang, and C.S. Kim. Topological design and routing for low-earth orbit satellite networks. In *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, pages 529–535, 1995.
6. H.S. Chang, B.W. Kim, C.G. Lee, S.L. Min, Y.C., H.S. Yang, D.N. Kim, and C.S. Kim. Performance comparison of optimal routing and dynamic routing in low-earth orbit satellite networks. In *Proc. of IEEE Vehicular Technology Conference (VTC)*, pages 1240–1243, 1996.
7. S. Cho, I.F. Akyildiz, M.D. Bender, and H. Uzunalioğlu. A New Connection Admission Control for Spotbeam Handover in LEO Satellite Networks. In *Proc. of IEEE Global Telecommunications Conference (GLOBECOM), San Francisco, USA,*, pages 1156–1160, 2000.
8. D. R. Cox and V. Isham. *Point Processes*. Chapman and Hall, 1980.
9. F.J. Dietrich. The globalstar satellite cellular communication system design and status. In *Proc. of the 5th International Mobile Satellite Conference*, pages 139–144, Pasadena, CA, June 1997.
10. E. Ekici, I. F. Akyildiz, and M. Bender. Datagram Routing Algorithm for LEO Satellite Networks . In *Proc. of IEEE INFOCOM*, pages 500–508, 2000.
11. A. Ferreira, J. Galtier, P. Mahey, G. Mat eus, and A. Oliveira. An Optimization Model for Routing in Low Earth Or bit Satellite Constellations. In *Proceedings of ISPAN*, Dallas, USA, December 2000. IEEE CS Press.

12. A. Ferreira, J. Galtier, J.-N. Petit, and H. Rivano. Re-routing algorithms in a meshed satellite constellation. *Annales des Télécommunications*, 2001. To appear.

13. A. Frey and V. Schmidt. Marked point processes in the plane I - a survey with applications to spatial modeling of communication networks. *Advances in Performance Analysis*, 1(1):65–110, 1998.

14. J. Galtier. Geographical reservation for guaranteed handover and routing in low earth orbit constellations. *Telecommunication Systems*, 2001. To appear.

15. B. Gavish and J. Kalvenes. The impact of intersatellite communication links on LEOS performance. *Telecommunication Systems*, 8:159–190, 1997.

16. N. Gerlich. On the spatial multiplexing gain of SDMA for wireless local loop access. Technical Report 161, Universität Würzburg, January 1997.

17. P. Gvozdjak. *Modeling communications in Low-Earth-Orbit satellite networks*. PhD thesis, School of Computing Science, Simon Fraser University, 2000.

18. M. Hellebrandt and R. Mathar. Cumulated interference power and bit-error-rates in mobile packet radio. *Wireless Networks*, (3):169–172, 1997.

19. D. Hong and S. Rappaport. Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology*, 36(3):77–92, August 1986.

20. Y.C. Hubbel and L.M. Sanders. A comparison of the IRIDIUM and AMPS systems. *IEEE Network*, 12(2):52–59, March/April 1997.

21. D.M. Kohn. Providing global broadband internet access using low-earth-orbit satellites. *Computer networks and ISDN systems*, 29:1763–1768, 1997.

22. R.J. Leopold. Low-earth orbit global cellular communications network. In *Proc. of IEEE International Conference on Communications (ICC)*, pages 1108–1111, 1991.

23. J. A. Musey, W. Kidd, and P. Fuhrmann. The satellite report. http://www.unterberg.com.

24. E. Papapetrou, I. Gragopoulos, and F.-N. Pavlidou. Performance evaluation of LEO satellite constellations with inter-satellite links under self-similar and Poisson traffic. *International Journal of Satellite Communications*, 17:51–64, 1999.

25. S.R. Pratt, C.E. Fossa Jr. R.A. Raines, and M.A. Temple. An operational and performance overview of the IRIDIUM Low Earth Orbit satellite system. *IEEE Communications Surveys*, pages 2–10, second quarter 1999.

26. A. Qayyum. *Analysis and evaluation of channel access schemes and routing protocols in wireless LANs*. PhD thesis, University of Paris-Sud, November 2000.

27. R. Raines et al. Simulation of routing protocols operating in a low earth orbit satellite communication network environment. In *1997 IEEE Military Commun. Conf.*, volume 1, pages 429–433, Monterey, CA, November 1997.

28. E. D. Re, R. Fantacci, and G. Giambene. Efficient dynamic allocation techniques with handover queuing for mobile satellite networks. *IEEE Journal on Selected Areas in Communications*, 13(2):397–405, February 1995.

29. J. Restrepo and G. Maral. Guaranteed handover (GH) service in a non-Geo constellation with “satellite-fixed cell” (SFC) systems. In *Proc. 5th International Mobile Satellite Conference (IMSC)*, pages 19–24, Pasadena, CA, USA, June 1997.

30. M.A. Sturza. Architecture of the Teledesic satellite system. In *Proc. of the International Mobile Satellite Conference (IMSC)*, pages 214–218, 1995.

31. P.T.S. Tam, J.C.S. Lui, H.W. Chan, Cliff C.N. Sze, and C.N. Sze. An Optimized Routing Scheme and a Channel Reservation Strategy for a Low Earth Orbit Satellite System. In *Proc. of IEEE Vehicular Technology Conference (VTC 1999 - Fall). IEEE VTS 50th, Volume 5*, pages 2870–2874, 1999.

32. H. Uzunalioğlu, I.F. Akyildiz, Y. Yesha, and W. Yen. Footprint handover rerouting protocol for low Earth orbit satellite networks. *ACM Journal on Wireless Networks*, 5(5):327–337, October 1999.

33. H. Uzunalioğlu, M.D. Bender, and I.F. Akyildiz. A Routing Algorithm for Low Earth Orbit (LEO) Satellite Networks with Dynamic Connectivity. *ACM-Baltzer Journal of Wireless Networks (WINET)*, 6(3):181–190, June 2000.

34. H. Uzunalioğlu, W.Yen, and I.F. Akyildiz. A connection handover protocol for LEO satellite ATM networks. In *Proc. of the 3rd ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, pages 204–214, Budapest, Hungary, September 1997.

35. M. D. Violet. The development and application of a cost per minute metric for the evaluation of mobile satellite systems in a limited-growth voice communication market. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1995.

36. M. Werner and G. Maral. Traffic flows and dynamic routing in LEO intersatellite link networks. In *Proc. of the 5th International Mobile Satellite Conference (IMSC)*, pages 283–288, 1997.

37. M. Werner and P. Révillon. Optimization issues in capacity dimensioning of LEO intersatellite links networks. In *Proc. of ECSC 5*, November 1999.

38. M. Werner, F. Wauquiez, J. Frings, and G. Maral. Capacity dimensioning of ISL networks in broadband LEO satellite systems. In *Proc. of the 6th International Mobile Satellite Conference (IMSC)*, pages 334–341, Ottawa, 1999.

39. W.W. Wu, E.F. Miller, W.L. Pritchard, and R.L. Pickholtz. Mobile Satellite Communications. *Proceedings of the IEEE*, 82(9):1431–1448, Sept 1994.