Handover Schemes in Satellite Networks: State-of-the-Art and Future Research Directions

Pulak K Chowdhury\(^1\), Mohammed Atiquzzaman\(^1\), and William Ivancic\(^2\)

\(^1\)Telecom and Networks Research Lab
School of Computer Science, University of Oklahoma, Norman, OK 73019-6151
Email: \{pulak, atiq\}@ou.edu

\(^2\)NASA Glenn Research Center, Satellite Networks & Architectures Branch
21000 Brookpark Rd., Cleveland, OH 44135
Email: wivancic@grc.nasa.gov

Abstract—Low Earth Orbit (LEO) satellites will work as an important component in future data communication networks. LEO satellites provide low end-to-end delays and efficient frequency spectrum utilization, making it suitable for personal communication. However, due to high satellite speed, the ongoing communication using LEO constellations experiences frequent handover. In this paper, we provide an up-to-date comprehensive literature survey on proposed handover schemes for LEO satellite systems. We also present a detailed classification of handover schemes based on a common framework. We first classify the schemes into link layer and network layer handover schemes. Link layer handover schemes are further classified into three categories: (a) spotbeam handover schemes, (b) satellite handover schemes, and (c) ISL handover schemes. Spotbeam handover schemes are categorized based on channel capacity, handover guarantee, and handover prioritization techniques. Network layer handover schemes are also classified depending on connection transfer strategies. Finally, we compare the handover schemes using different Quality of Service.

I. INTRODUCTION

The trend in designing future global communication networks is to offer fast and integrated service to ubiquitous users on demand, any time [1] [2]. In order to provide complete coverage to a diverse population, satellites will play an integral part of future global communication infrastructure [1] [2] [3]. First generation satellite based communication systems (like Iridium, Globalstar, Odyssey, ICO, Ellipso) were proposed in the early 90’s, and were primarily intended to carry only voice and low speed data traffic [2]. However, due to the competition from terrestrial broadband, meshed and WIMAX networks [4], lack of real market incentives and some inconvenient LEO satellite characteristics (like price, size and complexity of terminals), most of these projects failed. Currently, we are going through an era of high speed worldwide Internet where global information network should offer bandwidth-intensive multimedia data services. In order to fulfill these requirements, a new generation of satellite communications (SATCOM) networks, called broadband satellite networks, has been proposed. Astrolink, Cyberstar, Spaceway, SkyBridge, Teledesic and iSky (KaStar) are examples of this generation of satellite communication networks [1]. These satellite communication networks will provide a large array of services like video on demand, multimedia traffic, fast Internet access, interactive video and other existing Internet based applications along with voice services [1] [2]. National Aeronautics and Space Administration (NASA) and its enterprises are aiming to build future space communication architecture based on LEO satellite systems [5]. They have already incorporated Internet technology in one of their LEO satellites for IP based data communication [6] [7].

These satellite systems are intended to complement and extend the existing terrestrial networks to provide complete global coverage. They can interact with existing fixed networks to share instantaneous traffic overload. In general, satellite networks will extend the coverage area where terrestrial wire-line and wireless systems are infeasible both economically and geographically. Satellite UMTS (S-UMTS) [8], USRAN (UMTS Satellite Radio Access Network) [2] and 3GPP [9] are good examples of standardization and organizational bodies which are integrating satellite and terrestrial networks for future global communications; the interest on these systems in the research community will increase in future [2]. Satellite based communication networks can be broadly classified into three categories depending on the type of the satellites used: Geostationary (GEO), Medium (MEO) and Low Earth Orbit (LEO) satellite systems [10], although mixed constellations (like Spaceway design contains both GEO and MEO satellites [1]) exists.

Geostationary Earth Orbit (GEO) satellites are deployed 35,786 km above the equator line [10]. These satellites are called geostationary as, at this altitude, the satellites move synchronously with earth, i.e., a GEO satellite completes a circular movement around the earth in 24 hours. Consequently, the satellite position and coverage area are stationary relative to a fixed location or observer on earth. At this altitude, a GEO satellite covers almost 3/4 of the earth’s surface (not including the polar area), requiring only three satellites to cover the whole earth (see Fig. 1). Although, a small number of GEO satellites is needed for global coverage, GEO systems exhibit some significant disadvantages for communication networks. The user terminals and satellites consume lot of power, and the propagation delay for real time communications is very...
high in these systems.

A number of LEO satellite systems (like Iridium, Glob-alstar, SkyBridge) [2] have been proposed to overcome the disadvantages of GEO systems in high speed data and voice communications (Fig. 1). In contrast to GEO systems, LEO satellite systems have a number of advantages, such as efficient bandwidth usage, lower propagation delays, and lower power consumption in the user terminals and satellites [2]. However, in contrast to GEO satellite systems, the coverage area of a LEO satellite is not stationary. This is due to the asynchronous movement of the satellite relative to Earth, resulting in handing over of a satellite between ground stations as it passes over different areas of the Earth. The mobility management in LEO satellite systems is thus more challenging than in GEO systems.

![Mixed constellation of Iridium and GEO.](image)

Fig. 1. Mixed constellation of Iridium and GEO.

In some LEO satellite systems (for example, Iridium), satellites communicate among themselves using Inter Satellite Links (ISL). As shown in Fig. 1, ISLs are of two types: intraplane ISLs which connect satellites within the same orbit, and interplane ISLs which connect satellites in adjacent orbits [10]. The footprint of a satellite is a circular area on the earth’s surface [10]. To achieve efficient frequency reuse, a footprint is divided into smaller cells or spotbeams (Fig. 2). Two different schemes are proposed regarding cellular coverage geometry for LEO satellites: (a) Satellite Fixed Cell (SFC) systems, and (b) Earth Fixed Cell (EFC) systems [11]. As most of the research work on handover schemes in space networks are carried out on Satellite Fixed Cell (SFC) systems, in this paper, we focus mainly on SFC systems.

Transfer of an ongoing connection to a new spotbeam or satellite is called link layer handover. Three types of link layer handovers are observed in satellite systems [10]: (a) Satellite handover, (b) Spotbeam handover, and (c) ISL handover. Satellite handover refers to the switching a connection between satellites, whereas spotbeam handover involves switching of a connection between spotbeams. ISL handovers occur due to the change of connectivity patterns of satellites.

Till now, we have considered link layer (layer 2) handover in the satellite networks. However, for IP-based data communication using satellites as IP nodes, network layer (layer 3) handovers are also required. End terminals (satellites or user) which have Internet Protocol (IP) connectivity may need to change their IP address while moving, experiencing a network layer handover. When a satellite or a user needs to migrate its ongoing connections to a new IP address due to the change of coverage area of the satellite or mobility of the user, a network layer handover is also required. Due to fast satellite movement, hosts on the Earth frequently come under new satellite footprints or spotbeams. Change of satellite footprint or spotbeam requires change of IP address at the end hosts during data communication. Fu et al. [12] considered two satellite scenarios where a network layer handover has to be performed to maintain the ongoing data communications.

In LEO systems, mobility management issues like location management (registration and paging) is similar to current terrestrial networks. In contrast, handover management differs significantly from terrestrial networks, as handovers occur frequently due to the movement of satellites. Many research efforts have focused on handover management in LEO satellite networks. However, the authors are not aware of any paper which brings all the work together in a common framework for comparison purposes. In this paper, we focus on handover schemes in LEO networks and present a comparison of their performance.

The objective of this paper is to introduce the basics of handover schemes in LEO satellite networks and classify the schemes based on handover strategies. Akyildiz et al. [10] provide an overview of link layer handover problems and suggested solutions for LEO satellite networks. Papapetrou et al. [13] give a short description of different handover schemes. However, the above studies do not include all handover solutions proposed in the literature and do not consider network layer handover issues in space networks. Our study includes a detailed classification and overview of all the proposed handover solutions for both link and network layers in space networks. We compare handover performance of different schemes based on call dropping and forced termination probabilities. Our contributions in this paper are to classify all available satellite handover schemes and compare them based on a common comparison framework.

The rest of the paper is structured as follows: Section II summarizes the handover schemes in LEO satellite networks. In Section III, we present the basics and classification of spotbeam handover schemes. Next, in Section IV, we cover the fundamentals of satellite handover schemes and categorize the schemes. Section V gives an overview of ISL handovers and also discusses different ISL handover schemes. Following that, a brief introduction and classification of network layer handovers is given in Section VI. Section VII outlines areas of future research in LEO satellite handover schemes. Finally, concluding remarks are presented in Section VIII.

II. HANDOVER IN LEO SATELLITE SYSTEMS

LEO satellites will work as the core element of future data communication systems for some of its important character-
istics like lower propagation delay, lower power requirements both on satellite and user end and more efficient spectrum allocation due to frequency reuse among the satellite spotbeams [2]. However, LEO satellites are not stationary with respect to a fixed user on the Earth’s surface. The satellite ground track speed ($V_{trk}$) is much greater than Earth’s rotation speed and the user speed [11]. Due to constant rotation of the LEO satellites, the visibility period of a satellite in a cell is very small. For this reason, a user terminal can be served by a number of spotbeams and satellites during a connection.

Supporting continuous communication over a LEO satellite system may require changing of one or more links as well as the IP address of the communication endpoints. Thus, both link layer and higher layer handovers may be required for satellite networking. Mobility management of LEO satellites is therefore much more challenging than GEO or MEO systems. The mobility of LEO satellite systems is rather similar to cellular radio systems with a few differences. In both systems, the relative position between the cells and the mobile hosts changes continuously, requiring handover of the mobile hosts between adjacent cells [14]. The differences between the mobility of these two systems are as follows. In cellular systems, the mobile hosts move through the cells, while in LEO systems the cells move through the mobile hosts [14]. The cell size of LEO satellite systems is larger compared to cellular systems. Moreover, the mobile host’s speed can be ignored in LEO satellite systems since that speed is negligible compared to the LEO satellite’s rotational speed [14]. Bandwidth and power are also some constraints to be considered while designing mobility management schemes in LEO satellite systems. However, unlike terrestrial cellular mobile systems where the movement of mobile devices is not easily predictable, in LEO satellite systems, it is possible to predict the movement of satellites, and thus selection of next servicing satellite is relatively simple. At any instant, we can get an actual scenario of the satellite constellation which facilitates careful selection of the satellites in a communication path between endpoints to avoid unnecessary handovers. Handovers in satellite networks can be broadly classified as:

- **Link Layer Handover**: Link layer handover occurs when we have to change one or more links between the communication endpoints due to dynamic connectivity patterns of LEO satellites. It can be further classified as:
  - **Spotbeam Handover**: When the end point users cross the boundary between the neighboring spotbeams of a satellite, an intrasatellite or spotbeam handover occurs. Since the coverage area of a spotbeam is relatively small, spotbeam handovers are more frequent (every 1-2 minutes) [10].
  - **Satellite Handover**: When the existing connection of one satellite with the end user’s attachment point is transferred to another satellite, an intersatellite handover occurs.
  - **ISL Handover**: This type of handover happens when inter-plane ISLs would be temporarily switched off due to the change in distance and viewing angle between satellites in neighbor orbits. Then the ongoing connections using these ISL links have to be rerouted, causing ISL handovers.

The performance of different link layer handover schemes can be evaluated using two classic connection level Quality of Service (QoS) criteria [15]:
- call blocking probability ($P_b$), the probability of a new call being blocked during handover.
- forced termination probability ($P_f$), the probability of a handover call being dropped during handover.

There is a tradeoff between $P_b$ and $P_f$ in different handover schemes. The priority can be given via different treatments of new and handover calls to decrease handover call blocking [16].

- **Network Layer Handover**: When one of the communication endpoints (either satellite or user end) changes its IP address due to the change of coverage area of the satellite or mobility of the user terminal, a network or higher layer handover is needed to migrate the existing connections of higher level protocols (TCP, UDP, SCTP, etc.) to the new IP address. This is referred to as Network or higher layer Handover. Three different schemes can be used during this call transfer process [17]. They are:
  - Hard handover schemes: In these schemes, the current link is released before the next link is established.
  - Soft handover schemes: In soft handover schemes, the current link will not be released until the next connection is established.
  - Signalling Diversity schemes: Similar to soft handover. Only exception is that, in signalling diversity schemes, signalling flows through both old and new links and the user data go through the old link during handover [17].

### III. Spotbeam Handover

The service area or footprint of a satellite is a circular area on the Earth’s surface. To allow frequency reuse, the footprint of an individual satellite is divided into smaller cells or spotbeams. This results in better frequency utilization through the use of identical frequencies in non-adjacent spotbeams which are geographically separated to limit interference [18]. To ensure uninterrupted ongoing communications, a current communication link should be handed off to the next spotbeam if needed. A spotbeam handover involves the release of the communication link between the user and the current spotbeam and acquiring a new link from the next spotbeam to continue the call (Fig. 2). Since both spotbeams are served by the same satellite, no other satellite is involved in the handover process.

Due to the small area covered by spotbeams and high satellite speed, spotbeam handovers are the most common type of handovers experienced in LEO satellite systems [10]. We can consider the user mobility negligible compared to high satellite speed. As a result, the deterministic and constant movement of the satellites makes the solving of the spotbeam handover problems easier. During the handover process, if a new link or channel can not be found in the next spotbeam, the ongoing call should be dropped or blocked. From the
user viewpoint, the interruption of a call is less desirable than the blocking of a newly arrived call [10]. It will be the best for a user if handovers can be guaranteed, ensuring smooth ongoing calls. Again, the selection of a suitable policy in resource management (channel allocation) can ensure new channel availability during handover. Thus, the channel allocation strategies and the handover guarantee are the prime issues in managing handover requests.

To solve the spotbeam handover problem, several handover policies/schemes are proposed in the literature. We can classify the spotbeam handover schemes according to two different criteria:

- channel allocation strategies
- handover guarantee

Here, while classifying, we take into account of the capacity issue, i.e. the classification is based on channel quantity in the respective spotbeams. The other radio interface issues like pathloss, interference, better quality of the channels can also be used in classifying the handover schemes, but is not the focus of this paper.

A. Classification based on Channel Allocation Strategies

Various channel allocation strategies can be used to assign a channel to a call. Handover requests can also be considered a transferred call for the next cell, requiring allocation of a channel. Based on channel allocation strategies, handover schemes can be divided into three broad categories [19] [20] as follows:

- Fixed Channel Allocation (FCA) based handover schemes
- Dynamic Channel Allocation (DCA) based handover schemes
- Adaptive Dynamic Channel Allocation (ADCA) based handover schemes

The differences between those schemes are: in FCA schemes, fixed number of channels are allocated to each cell. In DCA schemes, number of allocated channels to a particular cell may vary depending on the network traffic. ADCA schemes are variations of DCA schemes where less call dropping during handover is guaranteed. Table I compares different channel allocation schemes based on several link layer QoS criteria.

1) FCA based Handover Schemes: In FCA schemes, a set of channels is permanently assigned to each cell, according to the frequency reuse distance [20] [19]. A handover call can only be given a channel if any channel belonging to the set of the cell is available. If no channel is available, the call is blocked or, in the worst case, dropped.

Fixed channel allocation schemes have a very simple implementation due to the fixed predefined channel distribution [19]. However, in nonuniform traffic conditions, the implementation becomes complex as a sophisticated network planning is required to assign more capacity to cells when a high traffic rate is expected [20]. In LEO constellations, this traffic planning is almost meaningless as it is not easy to predict the traffic conditions in a given cell. Statistical methods coupled with user behavior model and precise predictions of satellite tracks relative to the earth surface allow general characterization of the traffic load for a particular satellite or spotbeam. In LEO satellite systems with fixed channel allocation schemes expected traffic load varies from time to time and place to place while fixed channel allocation does not, resulting in poor resource utilization [19]. Thus, a number of schemes have been proposed to provide a more suitable solution for resource management in handover schemes.

An interesting variation of the FCA based handover scheme is Channel Sharing Handover [21]. Channel Sharing Handover uses a channel allocation scheme called channel sharing scheme [21], where channels can be shared between adjacent cells. A pair of adjacent cells is called a meta-cell. Two adjacent cells that form a meta-cell are called the component cells [21]. Fig. 3 shows two adjacent meta-cells with three component cells for a linear cellular system.

![Fig. 3. Linear cellular system.](image)

We can now describe this FCA scheme using channel sharing between component cells. Here, we assume the movement of users is towards higher numbered cells, i.e., users move from cell 1 to cell 2, and so on. When there is a new call in cell 1, it is given a channel if there is any idle channel in the meta cell \((i, i+1)\), otherwise, the call is blocked [21]. Consequently, the call can “carry” the channel to cell \((i+1)\) during handover. Handover calls arriving at cell \(i\) are assigned a channel from the meta-cell \((i, i+1)\) if a channel is idle. If the call has already a channel from the meta-cell \((i-1, i)\), it
is allowed to carry the same channel in cell \( i \), and is queued in a FIFO queue for acquiring channels belonging to meta-cell \((i, i + 1)\) [21]. However, during handover, a call is dropped if it is using a channel from the meta-cell \((i - 2, i - 1)\). Each time a channel becomes free in the meta-cell \((i, i + 1)\), the channel is assigned to the first call waiting in the queue of that meta-cell. In case of an empty queue, the channel is idle and can be used for future new or handover calls. This scheme offers a significantly lower call blocking probability \(P_b\) for the same handover dropping probability \(P_f\) when compared to FCA based schemes [21].

2) DCA based Handover Schemes: DCA based handover schemes use dynamic channel allocation, where channels are grouped together in a central pool. Any cell requiring a channel uses a channel from the pool satisfying the channel reuse distance [20] [19]. Allocated channels are removed from the common channel pool during call time. When the call is terminated, the channel is transferred to the central pool for future reuse. DCA based schemes provide the important advantage of coping with traffic variations and overload conditions in different cells. This adaptability of DCA schemes makes it a fundamental channel allocation strategy in third generation cellular networks. It is concluded that there is a reduction of \(P_b\) and \(P_f\) in DCA compared to FCA based schemes under same conditions. A number of DCA based resource management schemes (DCA1, DCA2) for handover strategies have been discussed in [22], [23].

3) ADCA based Handover Schemes: Adaptive Dynamic Channel Allocation (ADCA) is an extension of DCA scheme (Sec. III-A.2). It uses guard channel during handover (Handover with Guard Channel (HG), described in Sec. III-B.2.a). A handover scheme with guard channel technique has to deal with the tradeoff between the number of guard channels and the number of normal channels. Excessive guard channels will create new call blocking, and fewer guard channels may block handover calls. Hence, ADCA keeps track of the current traffic load, and dynamically adapts the optimal number of guard channels according to user location information [18]. ADCA thus tries to make appropriate use of the guard channels.

Cho et al. [18] proposed a new connection admission control scheme based on ADCA, called Geographical Connection Admission Control (GCAC), for LEO satellites to limit the handover blocking probability. Based on user location information, GCAC estimates the future handover blocking probability \(P_b\) of a new call and existing calls [18]. From the estimated \(P_b\), the GCAC technique either accepts or rejects a call. The GCAC algorithm guarantees that the “handover blocking probability \(P_f\)” is less than a target handover blocking probability \(P_{QoS}\)” [18].

B. Classification based on Handover Guarantee

A number of handover schemes provide guaranteed handover to prevent calls from being blocked or dropped during handover. Other schemes try to ensure best service by prioritizing handover over the new calls, but do not ensure any handover guarantee. Based on handover guarantee, handover schemes can be classified as:

- Guaranteed Handover (GH) schemes
- Prioritized Handover schemes

1) Guaranteed Handover Schemes: In a guaranteed handover (GH) scheme, a new call is assigned a channel only if there is an available channel simultaneously in the current cell and the next transit cell. If such channels cannot be found immediately, the call is blocked. As the name indicates, this scheme guarantees each handover to be successful. Maral et al. [24] proposed a guaranteed handover scheme. In that scheme, when the first handover occurs, a new channel reservation request will be issued to the next candidate transit cell. If all the channels in the candidate transit cell are busy, the handover request is queued in a FIFO queue until the next handover. Thus, this scheme provides almost zero \(P_f\) while the value of \(P_b\) is unacceptably high. This is due to the early channel reservation (also known as channel locking in GH) for a call which is still not transferred to the cell, exhibiting bad resource management. To improve resource allocation, a few modified GH schemes are proposed: (a) Elastic Handover Scheme [25], (b) TCRA Handover Scheme, and (c) DDBHP Scheme. All of them provide techniques to delay the channel allocation for the next cell by a calculated time, and trade off the handover guarantee to a certain extent. The main difference of these schemes is in the determination of the time instant when the channel reservation request should be sent to the next cell so that call during handover is not dropped. In Table II, we compare different guaranteed handover schemes based on several link layer QoS criteria.

a) Elastic Handover Scheme: The elastic handover scheme is based on the Elastic Channel Locking (ECL) scheme [25]. The idea behind the ECL scheme is that an entering call does not issue a channel locking request to the next cell immediately; instead it postpones the request for a period of time until \(T_a(0 \leq T_a \leq T_c)\) (Fig. 4) [25]. The time \(T_a\) is decided by the QoS requirement for handover failure probability.

In Fig. 4, if a call which originated in cell \(i\) is entering cell \((i + 1)\), the channel reservation request for cell \((i + 2)\) is postponed till \(T_a\). If a free channel in \((i + 2)\) exists after the request is made, it is reserved for the call. Otherwise, the request is placed in a queue at cell \((i + 2)\) [25]. Anytime a channel is available in cell \((i + 2)\), it is given to the first request in the queue. If a channel can be locked in \((i + 2)\) before the call enters \((i + 2)\), the call continues; otherwise it is forced
to terminate. Whenever a call ends (either forced or natural), all the channel locking requests for the call are cleared. This scheme does not guarantee that a request can lock a channel eventually in the next transit cell, thus reducing the degree of handover guarantee [25].

b) TCRA based Handover Scheme: Boukhatem et al. [26] [27] [28] proposed a Time based Channel Reservation Algorithm (TCRA) to improve GH performance and resource utilization. TCRA locks a channel in the next candidate cell with the cell movement. Considering deterministic and constant satellite movement [11], TCRA can evaluate the expected crossing time of the user in the next candidate cell from the current one. This time interval is used to reserve a channel in the next cell which will be used during handover [26] [27] [28]. TCRA is a variation of ECL (Sec. III-B.1.a) except that the time instant to send the channel reservation request ($T_a$ in ECL) is calculated using the estimated user location in the current cell, instead of the QoS parameters in ECL.

c) DDBHP Scheme: Dynamic Doppler Based Handover Prioritization Technique (DDBHP) is yet another variation of GH scheme proposed by Papapetrou et al. [13]. This method uses Doppler effect in order to determine the terminal location, and to reserve channels at the estimated time in the next servicing cell. The system must reserve a channel in the next cell during the corresponding time interval known as handover threshold ($t_{HH}$) [13]. Clearly, different values of $t_{HH}$ will provide different level of service. DDBHP is comprised of three activities: (a) station monitoring, (b) channel reservation, and (c) reservation cancellation. During station monitoring, a satellite determines the time to handover occurrence ($t_H$), and schedules the channel reservation phase at time ($t_H - t_{HH}$).

The channel reservation phase tries to reserve a channel in the next cell. Reservation cancellation is used to cancel any reservation corresponding to a dropped or ended call.

Using station monitoring, a satellite can calculate the position of its neighboring satellites. Consequently, the serving satellite is able to determine if the destination cell corresponds to a different satellite. Thus this technique can be used in spotbeam handover as well as in satellite handover [13].

2) Prioritized Handover Schemes: Probability of handover failure is a common criteria for performance evaluation of handovers in satellite networks. In non-prioritized schemes, handover requests are treated equally as new calls, thereby increasing the probability of call dropping during handover [19]. As discussed in Section III, ongoing call dropping is less desirable than new call blocking from a user’s viewpoint. Thus, handover prioritization schemes have been proposed to decrease handover failure at the expense of increased call blocking [19]. These prioritized handover techniques can be used along with the channel allocation strategies defined in Sec. III-A to increase handover performance. Table III compares different prioritized handover schemes based on $P_b$ and $P_f$. The differences between these schemes are: handover with guard channel prioritizes handover by reserving a set of guard channels for handover calls. Handover with queuing queues the handover requests for a certain time period before servicing. Channel rearrangement based handover uses rearrangement of channels in the adjacent cells for prioritizing handover. In the following, we discuss different handover prioritization categories:

a) Handover with Guard Channel (HG): HG scheme [29] [30] provides successful handover by reserving a set of channels (either fixed or dynamically adjustable) exclusively for handovers [19]. The remaining channels can be used for handover or normal calls. This reduces the probability of forced termination of calls during handover, while increasing new call blocking probability as fewer channels are available for new calls. Therefore, an important design issue is carefully choosing the number of guard channels [19].

b) Handover with Queueing (HQ): HQ scheme takes advantage of the overlapping area between adjacent cells [20]. While in the overlapping area, a mobile host can be served by any of the cells. This makes provision of queuing the handover requests for a certain time period equal to the time of mobile host’s existence in the overlapping area [19]. When a new channel becomes available, the cell checks the queue for waiting requests and grants the channel to the longest waiting request. Several schemes, depending on the strategy to order the handover requests in the queue, have been proposed. First in first out (FIFO) scheme [29] [31] is the most common queuing discipline where handover requests are ordered according to their arrival times.

A more complex scheme called MBPS (Measurement Based Priority Scheme), is based on dynamic priority, where the handover priorities are defined by the power levels of the corresponding calls (received from the satellite) from their current spotbeam [32]. The objective is to first serve the call with the most degraded link. Another alternative priority scheme is called LUI (Last Useful Instant) scheme [20] where a handover request with the shortest residual time (time remaining until the handover must occur for preserving the ongoing call) is queued ahead of other requests. In such a way,

<table>
<thead>
<tr>
<th>Criteria</th>
<th>FCA</th>
<th>DCA</th>
<th>ADCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>For uniform traffic conditions, complexity is low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>$P_b$</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>$P_f$</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Non-uniform Traffic conditions</td>
<td>To handle non-uniform traffic conditions, complex network planning is required</td>
<td>Network planning always same</td>
<td>Network planning always same</td>
</tr>
<tr>
<td>Frequency reuse/Resource management</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
the system tries to serve the most urgent handover request.

c) Channel Rearrangement based Handover: This scheme is only used with dynamic channel allocation schemes [33] and manages handover requests in exactly the same manner as new call attempts. Whenever a call termination occurs in a cell, the scheme performs a channel rearrangement to de-allocate the channel which becomes available in the greatest number of cells.

d) HQ+HG Handover: HQ+HG scheme takes advantages of both guard channel and queueing schemes.

IV. SATELLITE HANDOVER

Satellite handover occurs when a satellite involved in the connection between two users can not provide service to a user (one reason may be going out of sight from the user). In that case, the connection has to be transferred to a new satellite.

Let us consider the scenario in Fig. 5(a). User 1 is in communication with user 2 using satellites A and B. Since the satellites are moving left, user 2 will soon come under the footprint of satellite C. Thus, satellite C should be involved in the connection from user 1 to user 2 to keep the connection alive. The connection of user 2 to satellite B should be handed off to satellite C, and the new communication path from user 1 to user 2 will be through satellites A, B and C (Fig. 5(b)).

From the discussions in Section III, it can be concluded that the spotbeam handover issue and its solutions are well investigated in the literature. However, there is a lack of thorough studies for satellite handover techniques [15]. This is due to the fact that spotbeam handovers are more frequent than satellite handovers. Satellite handover is very important in LEO satellite based diversity systems. In a spotbeam handover, a user is constrained to choosing only one possible next cell. In contrast, for satellite handover, the user can select among different satellites. Moreover, the user has to first select the servicing satellite, and then will be served by the cell covering the user. Satellite handover schemes should aim to select the most suitable satellite depending on $P_b$, $P_f$, and the quality of communication from the satellite. Consequently, a well investigated satellite handover scheme can reduce bandwidth wastage and also fulfill the QoS requirements of $P_b$ and $P_f$ [15]. In Table IV, we compare different satellite handover schemes based on several link layer QoS criteria.

Gkizeli et al. [34] [35] [36] proposed two handover schemes for systems with satellite diversity: (a) hard handover scheme, and (b) hybrid Channel Adaptive Selective (CASD) scheme. Hard handover scheme uses two thresholds during handover while hybrid CASD scheme uses dual satellite diversity coupled with two thresholds under critical channel conditions [36].

### A. Hard Handover Scheme

This scheme reduces handover signalling overhead and has better performance in terms of call dropping rate [36]. It has a reduced handover rate compared to “pure” satellite handover which switches satellites whenever the current signal drops below the fade margin. It uses two thresholds during handover; selection of the thresholds is based on the fading variation in hostile environments while the satellite is moving. This algorithm tries to delay the handover for as long as possible. It uses two different power thresholds to decide whether to handover to satellite 1 or satellite 2 whenever satellite 2’s or 1’s signal level goes below the fade margin.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Elastic</th>
<th>TCRA</th>
<th>DDBHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of guarantee</td>
<td>Varies with $T_a$</td>
<td>Varies with $T_a$</td>
<td>Varies with $T_a$</td>
</tr>
<tr>
<td>$P_b$</td>
<td>Increases if $T_a$ decreases</td>
<td>Depends on number of users in a predefined area</td>
<td>Depends on $T_a$</td>
</tr>
<tr>
<td>$P_f$</td>
<td>Decreases if $T_a$ increases</td>
<td>Null</td>
<td>Practically zero</td>
</tr>
<tr>
<td>$T_a$ selection criteria</td>
<td>QoS requirement of handover</td>
<td>Expected crossing time of the user in the next cell</td>
<td>Doppler effect</td>
</tr>
</tbody>
</table>

![Fig. 5](image-url)
B. Hybrid CASD based Handover Scheme

Based on the two threshold hard handover scheme, Gkizeli et al. [34] [36] proposed a Hybrid CASD handover scheme which uses dual satellite diversity (two contagious satellites sharing common coverage areas on earth surface) only under critical channel conditions (when the fading level of the signals in the channel is high). Thus, this scheme uses the two threshold concept (Sec. IV-A) of hard handover under normal conditions and during critical channel conditions; it is flexible enough to take advantage of satellite diversity for soft handover.

C. DDBHP Handover Scheme

Both the schemes mentioned in Sections IV-A and IV-B are based on the case in which a call is dropped due to power limitations. However, a call can be dropped if there is no available channel in the forthcoming satellite. Furthermore, the algorithm should maintain good QoS parameter values under heavy traffic conditions. Papapetrou et al. [15] cited one scheme based on DDBHP [Sec. III-B.1.c] which takes into account of all these issues.

As in [13], DDBHP uses the doppler effect to avoid early reservation of channels and has low blocking probability. By measuring the doppler effect at two different time instants, it is possible to determine the user location and the time of handover (station monitoring). Also, the service satellite will be able to select the possible forthcoming satellite (not in the same orbit plane) by knowing the position of other satellites.

D. Satellite Selection Criteria

As the satellites in different orbital planes share a common area on Earth, a user can select between multiple satellites during handover. Based on selection criteria of the next satellite, we can classify handover schemes into different categories. Three criteria for selection of the next servicing satellite have been proposed in [13]:

- **Maximum service time**: Select the satellite that offers maximum service period, thus minimizing the number of handovers and therefore achieving low $P_f$.
- **Maximum number of free channels**: Select the satellite with maximum number of free channels, thus achieving uniform distribution of calls among the satellites.
- **Minimum distance**: Select the closest satellite to avoid link failure.

Since the criteria can be applied to both new and handover calls, nine (each criterion applies to new and handover calls, $3 \times 3$) different satellite selection schemes result in [13].

Boedhihartono et al. [37] propose a different set of satellite selection criteria:

- **Visibility time (VT)**: Select the satellite with the longest remaining mutual visibility time.
- **Capacity (C)**: Select the least loaded satellite with mutual visibility.
- **Visibility time subject to capacity availability (VT/CA)**: Select the satellite with the longest remaining mutual visibility time.
- **Elevation angle (EA)**: Select the satellite with the highest elevation angle for the user terminal.
- **Visibility time with early channel release (VT/ECR)**: Select the satellite with the longest remaining mutual visibility time.

The satellite selection criteria proposed in [13] and [37] use link quality, system geometry and local blockage of channels as criteria for satellite selection. However, LEO satellite systems which strictly depends on these issues for satellite selection may often fail to choose the correct satellite due to local obstructions. Thus, different set of satellite selection criteria can be considered while selecting next servicing satellite in LEO satellite systems.

V. ISL Handover

Due to the change of the connectivity patterns among the satellites, satellites have to temporarily shut down their ISLs [38]. As a result, ongoing communications using those ISLs have to be rerouted. This handover, referred to as ISL handover, may create a large number of rerouting attempts and call blocking [38] due to resource scarcity in the new satellite. This type of handover is specific to satellite constellations which use ISLs among neighboring satellites for communication. It is important to note that many LEO constellation concepts (like SkyBridge) do not use ISLs [1], and thus do not require ISL handover.

In satellite constellations (like Iridium) which use polar orbits, when satellites go into the polar area, the connectivity pattern of the satellites changes [39]. As seen in Fig. 6, the ISLs between satellites A and its neighboring satellites B & C have to be turned off for a certain time as B and C change their positions relative to A. Other LEO concepts (like Globalstar, Odyssey, ICO) which do not use polar orbits have different ISL handover issues, and ISL handovers occur at different locations in the orbit. The basic question still remains the same, i.e., determining where the ISLs have to be switched off between neighboring satellites and ongoing connections handed over to different satellites. Here, we focus on ISL handovers in polar orbiting satellite constellations.

Werner et al. [39] investigate this rerouting problem during ISL handover. They optimized their algorithm to find a unique route with minimum ISL handovers between satellite pairs.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>HQ</th>
<th>HG</th>
<th>Channel Rearrangement</th>
<th>HQ+HG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_h$</td>
<td>Good queueing strategy decreases $P_f$</td>
<td>Depends on guard channel management</td>
<td>Depends on efficient channel rearrangement</td>
<td>Efficient uses of HQ and HG decrease $P_h$</td>
</tr>
<tr>
<td>$P_f$</td>
<td>Depends on queueing strategy</td>
<td>Depends on guard channel management</td>
<td>Depends on efficient channel rearrangement</td>
<td>Depends on efficient use of HQ and HG</td>
</tr>
</tbody>
</table>

**TABLE III**

Comparison among Prioritized Handover schemes
### TABLE IV
**Comparison among Satellite Handover schemes**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Hard CASD</th>
<th>Soft CASD</th>
<th>DDBHP based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handover Strategy</td>
<td>Hard</td>
<td>Soft</td>
<td>Guaranteed</td>
</tr>
<tr>
<td>( P_h )</td>
<td>Depends on available channels</td>
<td>Depends on available channels</td>
<td>Depends on degree of guarantee</td>
</tr>
<tr>
<td>( P_f )</td>
<td>High</td>
<td>Low</td>
<td>Zero</td>
</tr>
<tr>
<td>Traffic Conditions</td>
<td>Performance degrades in critical channel conditions</td>
<td>Can work on critical channel conditions</td>
<td>Does not matter</td>
</tr>
</tbody>
</table>

All end user connections with a satellite pair use the unique route for that pair. This algorithm minimizes ISL handover, but can be unfair in the usage of the links [39]. It also assumes static ISL links which is unrealistic in LEO satellite networks. Werner et al. further improved the performance of their rerouting algorithm using a sliding window mechanism [40].

Uzunalioglu [41] proposes a routing protocol called the Probabilistic Routing Protocol (PRP) to reduce the number of rerouting attempts during ISL handover. This protocol removes all the ISLs from the connection route of a call which may expect link handover during the estimated life time of the call. Although the call duration can not be determined accurately, it determines the call time using a certain probability (target probability). The protocol trades off between the target probability and the new call blocking rate.

### VI. NETWORK LAYER HANDOVER

As mentioned earlier in Section II, due to the movement of the satellites and the mobile users, the communication endpoints (user or satellites) may have to change their IP address, requiring a network layer handover. Fu et al. [12] identify two scenarios requiring network layer handover as follows:

- **Satellite as a Router:** As satellites move, communicating fixed/mobile hosts come under new satellite footprints or spotbeams. Different satellites or even different spotbeams can be assigned with different IP network addresses. This requires a network layer handover during the change of communication links from one satellite or spotbeam to another.

- **Satellite as a Mobile Host:** When a satellite works as an end point of a communication by generating and receiving data, it can be regarded as a mobile host. Thus, like a mobile host it always changes its communication attachment point requiring a network layer handover.

In the first scenario (Fig. 7), satellites do not have any onboard equipment to produce or consume data. They merely act as routers in the Internet. Each satellite, or even a spotbeam, can be assigned an IP address. In such cases, handover between satellites (Intersatellite handover) or spotbeams (spotbeam handover) may also require network layer handover [12]. Hosts are handed over between satellites or spotbeams as they come under the footprint of a new satellite or spotbeam.

![Fig. 6. ISL handover between the satellites in the north polar area.](image)

![Fig. 7. User handover between the satellites.](image)

In the second scenario, satellites can act as communication endpoints with all the onboard equipments which exchange data with ground stations. As in Fig. 8, the satellite’s footprint is moving from ground station A to B, while the satellite is bound with an IP address from ground station A. During movement, the satellite should maintain continuous connection with ground stations. Thus, the IP address of the satellite has to be changed when a network layer handover to ground station B takes place.

#### A. Inter Segment Handover

Future data communication systems will integrate satellites and terrestrial networks. The focal point of this integration is to provide complete global coverage, enabling mobile users to roam globally. In such an environment, a dual mode terminal can allow uninterrupted service by handing over from one segment of the network to another. This introduces a new
type of handover, called Inter Segment Handover (ISHO) [42] [43] [17]. During the handover, three different phases are considered: (a) Initiation, (b) Decision, and (c) Execution. The decision phase is realized by the handover controlling schemes. Depending on whether the mobile user or the network monitors the link quality and makes the decision, the handover initiation and decision phases can be classified into four different handover controlling schemes [17]:

- Network Controlled Handover (NCHO).
- Mobile Controlled Handover (MCHO).
- Network Assisted Handover (NAHO).
- Mobile Assisted Handover (MAHO).

The differences between these handover schemes are as follows: In NCHO, the network monitors the link quality and decides whether to initiate handover. In MCHO, the Mobile Host (MH) monitors the link quality information to the MH, and MH decides initiation of handover. On the other hand, in MAHO, the MH sends the information about the link quality to the network and the network takes the handover decision.

The execution phase of handover is a combination of connection establishment and connection transfer scheme. Based on connection establishment, the handover can be classified as: (a) Backward Handover, (b) Forward Handover [42]. In connection transfer process, all calls have to be transferred from the old connection to the new one to keep the ongoing communications alive. Three different handover strategies can be used for the connection transfer process [17]: (a) Hard handover schemes (b) Soft handover schemes (c) Signalling Diversity schemes. In this paper, we focus on these three handover schemes. The difference among those schemes can be depicted as follows: In hard handover schemes, the current link is released before establishing the new link, whereas in soft handover schemes, current link will not be released before establishing the new link. Signalling diversity schemes are similar to soft handover schemes. The only exception is, in signalling diversity schemes, during handover, signal flows through both old and new links and data flow using the old link [17]. Table V compares these network layer handover schemes based on several QoS criteria.

B. Hard Handover Schemes

In hard handover schemes, the current link is released before the next link is established [17], which may result in connection blocking during handover. NASA [6] is using Mobile IP [44], which uses hard handover, to build future space communication networks.

Mobile IP (MIP) [44] manages mobility of Internet hosts at the network layer while keeping the upper layer connections alive. Mobile IP is based on the concept of Home Agent (HA) and Foreign Agent (FA) (which requires modification to existing routers in the Internet) for routing packets from the previous point of attachment to the new one [44]. Mobile IPv6 does not need a FA as it uses the IPv6 address autoconfiguration mechanism. Fig. 9 shows a Mobile IP based handover scenario where the satellite is acting as a Mobile Host (MH). When the satellite/MH determines that it is on a foreign network, it obtains a new Care of Address (CoA) from the new Foreign Agent (FA) (Ground Station B in Fig. 9). It registers the CoA address with the gateway router acting as Home Agent (HA) [45] (Fig. 9). The registration process begins when the satellite disconnects from the old point of attachment (Ground Station A) and starts to obtain a new CoA. After the registration process completes, data can be sent to the satellite using the new CoA. Datagrams destined
for the MH are intercepted by the home agent. Then, the HA tunnels the data to the FA, FA decapsulates and delivers them to the satellite. During the registration period (at time $h$), the MH is unable to send or receive packets through its previous or new point of attachment [45], giving rise to a large handover latency and high packet loss rate. Several schemes have been proposed in the literature to reduce the above mentioned drawbacks of Mobile IP based handover [44].

C. Soft Handover Schemes

During soft handover, the current connection is not released until the next connection is firmly established. Thus, both links can be used simultaneously for handover traffic management [17]. Many soft handover schemes have been proposed in the literature for terrestrial networks; for example [46] [47] etc. The issue of adapting them into space networks can be investigated in future research.

D. Signalling Diversity Schemes

The signalling diversity based scheme is similar to soft handover, with the difference being that the signalling procedures in signalling diversity schemes are performed through both the new and old links, while user data is sent through the old link [17]. Here no synchronization between links is needed as the old link is used for data and the new link is used for signalling.

Seamless IP diversity based Generalized Mobility Architecture (SIGMA) (previously named TRASH) [12] [48] is a signalling diversity based scheme. It is a complete transport layer mobility management scheme, and can be used with any IP diversity-based transport protocol. Fig. 10 depicts a scenario where the satellite is acting as a Mobile Host (MH). When the satellite moves into the overlapping area of two neighboring ground stations, it obtains a new IP address from the new communication agent (next visible ground station) while maintaining the old connection (via the old ground station) alive. In Fig. 10, the MH/satellite is moving from the coverage area of ground station A to ground station B. In the overlapping region, it obtains a new IP address (IP2) from ground station B while maintaining the connection through the old IP (IP1). The new address is used to carry all the signalling procedure to set up a new connection; during this time the mobile host can receive data via the old IP address (IP1). Whenever the received signal from ground station A drops below a certain threshold, the mobile host changes its primary address to the new one (IP2). When the mobile host leaves the overlapping area, it releases the old IP address (IP1) and continues communicating with the new address (IP2), thus achieving a smooth handover across ground stations. SIGMA reduces handover latency and data loss during handover.

VII. Future Research

Most of the current research work on IRIDIUM [1] [2] type LEO constellations consider only voice traffic. But future satellite networks will serve all kinds of multimedia traffic including voice, video and data. QoS requirements of multimedia traffic are different from those of voice. Consequently, multimedia traffic is more difficult to serve compared to voice. As an example, video traffic is sensitive to end to end delay but can tolerate packet losses; in contrast, data traffic expects low packet losses and is insensitive to end to end delay. Consequently, handover algorithms should provide different QoS to serve various kinds of multimedia traffic [2]. Consideration of QoS in handover management of space networks can be an active research area.

In existing handover schemes, the user mobility and earth’s rotation speed are ignored based on the assumption of short call holding times of voice traffic. Multimedia traffic has, however, longer connection holding times than that of circuit switched voice traffic [10]. The Earth’s rotation speed and user mobility in the cells have to be taken into account when designing handover schemes for connections involving multimedia traffic.

Some research efforts were directed to finding minimum number of satellites for global coverage. Thus, the overlapping coverage areas between neighboring satellites do not constitute a major portion of satellite coverage. However, in densely populated areas, for better resource management, the overlapping area between the neighboring satellites can be increased [10]. This can simplify spotbeam handover management problems, since increased overlapping areas can ensure better handover performance. As example, Globalstar was designed to provide multiple satellite coverage over the mid-latitudes. This is not a optimal satellite design, and it does not try to provide coverage to polar or central ocean. Some later constellations proposed for Teledesic also sacrificed global coverage to provide more capacity in the mid-latitudes. Thus, to improve resource and handover management, new satellite constellations in densely populated areas needs further investigation.

In contrast to spotbeam handover, satellite and ISL handover issues have not been covered in detail in the existing work.
(discussed in Sections IV and V). Developing efficient satellite and ISL handover algorithms can reduce delay during ISL and satellite handovers.

Network layer handover issues in space networks have been recently addressed in a few research works. Adapting current mobility management schemes for terrestrial wireless networks into space networks is a growing area of future research, and demands more research efforts. New efficient network layer handover schemes for space networks also need to be developed.

VIII. CONCLUSION

In this paper, we provide a comprehensive survey of handover management schemes, and propose a detailed classification of handover schemes in space networks. As far as the authors are concerned, this is the first paper which attempts to classify and compare the performance of both link layer and network layer based handover schemes for LEO satellites. We conclude that while the link layer handover schemes have been investigated in depth in the literature, further research on higher (network and above) layer handover schemes in LEO satellite systems is required. SIGMA, an IP diversity based transport layer seamless handover scheme, is suitable for LEO satellite networks.

REFERENCES


### TABLE V

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Hard</th>
<th>Soft</th>
<th>Diversity based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fault Tolerant</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Loss</td>
<td>On the fly packets are lost</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Connection Delay</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>IP Diversity</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>


