



Implementation of a geographical routing scheme for low Earth orbiting satellite constellations using intersatellite links

Manuel Roth | Hartmut Brandt | Hermann Bischl

Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany

Correspondence

Manuel Roth, Institute of Communications and Navigation, German Aerospace Center, Oberpfaffenhofen, 82234 Wessling, Germany.
Email: manuel.roth@dlr.de

Summary

A major problem for low Earth orbit (LEO) constellations with intersatellite links is the efficient routing of the data packets through such a highly dynamic network. In order to achieve a worldwide coverage even in remote areas and Internet access with a limited amount of gateway stations, intersatellite links are a promising approach. Since LEO constellations represent a distinct, highly dynamic routing environment, specific strategies are needed. To this end, a suitable geographical routing scheme is proposed and investigated in two Walker Star constellations. The proposed scheme targets reliable transmissions with low latency and high data rates. The approach is based on a geographical address identifier in Layer 2 of the communication stack. The globe is thus divided into geographical areas that determine this identifier in the MAC address of the terminals. As mobile terminals are considered, the MAC addressing scheme is flexible, whereas the IP addresses of the terminals remain static. This decoupling allows for flexibility in the choice of the address resolution scheme. Moreover, the geographical identifier in the MAC address enables fast routing table lookups and switching. The proposed routing scheme also takes possible overloads of the satellites due to traffic into account and applies a rerouting procedure. When a packet arrives in the geographical area of the destination terminal, a local rerouting scheme is applied if needed. The proposed approaches take handover events that possibly occur during a transmission into account. Furthermore, the scan angles of the satellites have been adapted to the constellations to provide full coverage and high elevation angles. So a robust and adaptable routing scheme is provided for a dynamic environment where satellites and terminals are constantly moving. The proposed definitions and procedures have been implemented in a system level simulator, which allows for comparisons with adjustable parameters in various scenarios. In this work, an Iridium-like constellation and a megaconstellation are investigated and compared regarding the address resolution procedures, the average end-to-end transmission delay, and the dropping and rerouting rates. Additionally, the signaling overhead is compared with other approaches. The simulator and results of the simulations

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. International Journal of Satellite Communications and Networking published by John Wiley & Sons Ltd

provide grounds for further research w.r.t. the routing in satellite constellations using intersatellite links.

KEYWORDS

intersatellite links, low Earth orbit, routing, satellite constellation

1 | INTRODUCTION

Most of today's operational satellite systems are geostationary, but satellite constellations have experienced resurgence of interest in the recent years. Both concepts have their advantages and disadvantages. Satellite constellations make use of low Earth orbits (LEOs) thus providing lower latency in the transmission paths than geosynchronous Earth orbit (GEO) satellites. This is an important aspect, as time-critical services like voice or video conferencing are somewhat impaired if applied via GEO satellites. On the other hand, GEO satellites appear stationary from the ground, whereas LEO satellites are moving relative to the terminals on ground so that tracking antennas are required for the terminals to keep users connected with high data rates. LEO satellites can be built smaller and cheaper than GEO satellites, but a large number of LEO satellites are needed to provide permanent connectivity to the user terminals. If a worldwide coverage is envisaged with a satellite constellation, intersatellite links (ISLs) are important to keep the number of ground stations at a moderate level. With such ISLs, the traffic can be routed from satellite to satellite without the involvement of ground stations. An operational LEO satellite constellation with RF-ISLs is the Iridium system. Recently, its second generation "Iridium NEXT" has been successfully deployed. The Iridium constellation consists of 66 operational satellites in six orbital planes with 11 satellites per plane. The satellites are in a near-polar orbit at an altitude of 780 km. Each satellite can be connected to four other satellites (two satellites in the same orbital plane and two satellites on the adjacent orbital planes).¹ Iridium offers true global voice and data communications coverage in L-band. As the data rates are moderate, mobile phones with omnidirectional antennas are sufficient. While the Iridium system is a low data rate system, the proposed scheme considers high data rates, for instance, for broadband applications. If high data rates are targeted, optical ISLs are a promising approach. Optical ISLs have the advantage of high bandwidth while the size, and required power of the laser terminals is relatively small. Due to narrow beam properties, high data rates can be achieved even for large distances. Newly proposed megaconstellations consist of hundreds or thousands of satellites. With this high number of satellites, the terminals can generally operate at high elevation angles. This reduces shadowing and increases link availability, especially in urban and suburban areas. Examples of megaconstellations with optical ISLs are SpaceX's Ku- and Ka-band system with 4425 satellites in 83 orbital planes and Telesat's Ka-band system with 117 satellites in 11 orbital planes. The estimated maximum system throughput of the SpaceX constellation is 23.7 Tbps with 123 ground stations and of the Telesat constellation 2.66 Tbps with 40 ground stations.² To achieve high system throughput and to keep transmission delays and packet dropping rates low, suitable quality of service (QoS) routing and addressing QoS scheduling algorithms are of paramount importance. In this paper, we focus on one of these algorithmic problems, namely, on the routing methods for satellite constellations with ISLs. As the network topology of a satellite constellation with ISLs is highly dynamic, we propose an addressing scheme for the link layer, which is bound to the geographical position of the terminals. This alleviates the routing problem coming along with the movement of the satellites. For the terminals, the satellite network is acting like a globally distributed IP switch, but the internal routing in the satellite network will be based on these geographical addresses. The novelty of this paper is this geographical Layer 2 routing scheme for LEO satellite constellations, which is investigated in a dedicated system level simulator built from scratch.

The paper is organized as follows. Section 2 reviews related research in the subject area and positions the paper in the context of these works. In Section 3, the system model, on which this investigation is based, is discussed. This includes an overview of the proposed satellite constellations, the assumed terminals, and the communication model. Section 4 explains the general approach of the proposed geographical routing scheme. In Section 5, the proposed address resolution and routing schemes for the reference system are described. Section 6 shortly illustrates the implementation and parameters of the previously discussed system and schemes. In Section 7, the results of the simulations regarding the address resolution and routing schemes are presented for various conditions and compared with a source-based routing scheme. Additionally, the signaling overhead is compared with other approaches. Finally, Section 8 concludes the discussion and provides an outlook on further research aspects.

2 | RELATED WORKS

Several approaches for efficient routing in satellite constellations have been proposed in the literature. In a paper by Pratt et al.³ an overview of the Iridium LEO satellite system is given. The work underlines the advantages of dynamic routing algorithms on the performance of a satellite network. Since the routing in the Iridium system utilizes proprietary routing algorithms, the authors investigated two candidate algorithms. Loop-free

routing, a decreased storage and complexity overhead, and transmission overhead are mentioned as requirements when dealing with the dynamic topology of LEO satellite networks. The first approach is an Extended Bellman–Ford (EXBF) algorithm, which only considers direct paths and utilizes interneighbor coordination. The second approach is Darting, which delays network updates until they are necessary in order to reduce the signaling overhead. The results indicated that EXBF has a performance advantage over Darting. While these routing algorithms are not considered in this work, the analysis highlighted the demands of a working satellite system employing ISLs. Moreover, one of the constellations considered in this work was chosen to resemble the design of the Iridium constellation. Although, in contrast to the Iridium system, the reference system of this paper focuses on high data rates.

In general, a common routing strategy is the utilization of a virtual topology. Such schemes are based on snapshots of the constellation, which are utilized to calculate a path between two given satellites. Most schemes apply Dijkstra's algorithm⁴ or a variation of it in order to decrease the propagation delay. The paths are normally calculated by the ground stations, which means the packets are source-routed. These routing calculations can be computed in advance, since the ISLs of the constellation are predictable in the absence of unexpected failures. However, such schemes are particularly susceptible to congestions due to “hot spots”.⁵ By applying load-balancing mechanisms, the queueing delay and other QoS parameters can be improved. The explicit load balancing (ELB) routing algorithm⁶ represents a routing scheme based on a virtual topology that allows for a better traffic distribution and lower packet loss in congested scenarios. Based on explicit exchanges of their current loads, the satellites adapt their forwarding rates. This signaling is triggered when certain predetermined thresholds are reached. An important drawback of this approach is the traffic flow cascade problem, which possibly occurs during the detours of the packets. As we test our proposed scheme in scenarios with significant traffic, we applied a rerouting mechanism similar to the one used by the ELB algorithm in our system.

Such connection-oriented communication systems can introduce a significant amount of signaling overhead in LEO constellation networks. If mobile terminals are considered, handover events need to be communicated to all nodes of the network. In order to reduce the signaling overhead, connectionless communication schemes have been investigated. To this end, strategies involving virtual nodes, such as the datagram routing algorithm (DRA),⁷ have been proposed. A virtual node represents a geographical location as a communication node that is served by different satellites over time. These logical locations remain relatively static, even though the network is dynamic. The routing algorithm is distributed and routing decisions are made for each incoming packet. Thus, the required number of network-wide updates is reduced. However, due to the virtual nodes, this approach is potentially less robust to link failures. Furthermore, the scheme is based on polar, symmetrical LEO constellations and needs adaptations for more complex constellations.

Geographical routing strategies represent a similar approach. In such schemes, packets are routed towards a geographical position, as opposed to a specific (virtual) node. Due to the irregular movement of most satellite constellations in relation to geographical locations on Earth, a division of the surface of the Earth into a grid, which corresponds exactly to the covered areas is impossible. Nevertheless, carefully designed systems using geographical routing maintain a minimal signaling overhead in comparison to other approaches.

Henderson and Katz⁸ investigated distributed, geographic-based packet routing. The authors described the practical difficulties of distributed routing based on geographic distance for LEO satellite systems. A hybrid routing scheme was proposed, which provided adequate results. However, the robustness of the scheme was questioned, as inherent issues were illustrated. The first complication is the ambiguity in the last hop of the transmission. As multiple satellites possibly cover a geographic area, geographical routing methods are not inherently able to find the correct last hop. Therefore, multiple additional hops after the geographical routing might be required to reach the destination. In addition, the mesh structure near the polar regions, and the seam caused by counter-rotating planes in polar orbits are mentioned. The presented issues were considered and resolved during the design of our routing schemes and mechanisms.

Tsunoda et al.⁹ introduced a handover-independent mobility management scheme that is based on geographical location mapping. The paper illustrates that the management cost can be reduced in comparison with Mobile IP schemes.^{10,11} In general, Mobile IP schemes are unfavorable for LEO satellite networks, as the additional mobility of the satellites causes a significant management cost. The analysis also shows that appropriate cell sizes are required for the efficiency of geographical routing approaches. Based on the results, the team proposed a mobility management scheme that focuses on solving the last-hop ambiguity of geographical routing schemes.¹² In particular, the scheme tries to resolve ambiguities that result in multiple additional hops, that is, in scenarios where interplane communication is difficult due to counter-rotating planes. In order to circumvent this case, a method is proposed, which uses both geographical and orbital information. Based on an identifier of the terminal on ground, the last-hop candidates are limited to satellites with the defined orbit. This approach can reduce the cost for local forwarding after the geographical routing and has thus been partially utilized in this work.

The proposed scheme in this paper builds on these investigations. In our approach, the addressing scheme and the topological information are decoupled. Hence, the Layer 3 identity of a terminal on the ground does not include its geographical information, which is embedded in its Layer 2 MAC address. Therefore, the constellation can be interpreted as an IP switch from the point of view of the terminals on ground. This allows for practical improvements while maintaining the advantages of geographical routing for LEO satellite systems. Moreover, the scheme is analyzed using a specifically implemented system level simulator built from scratch. The simulated results illustrate the effectiveness of the proposed scheme for different traffic scenarios and lay a foundation for further research.

3 | REFERENCE SYSTEM

For the investigation of the routing method and addressing schemes, the reference system is broken down into nodes, terminals, and local area network (LAN) segments as shown in Figure 1. In the given networking scenario, nodes represent the satellites. They have several links and contain switching functionality, which forwards traffic packets entering the node through one link to another link based on constantly updated routing and switching tables. As the nodes represent LEO satellites, they are considered to be permanently moving. For the given LEO constellations, the movement follows regular patterns, which can be precomputed with sufficient accuracy in advance. In order to allow for ad hoc routing within the satellite constellation, ISLs are used. The usage of ISLs thus reduces the number of required gateway stations for routing purposes on the ground. The ISLs form a regular but time-variable pattern for the given scenario. The ISL connections are not necessarily permanent. They may change because of failures of the link or a node or because of node movement. Some of these changes, such as the movement of the satellites, are predictable in the long term. Others, such as failures, may not be predictable at all. The amount of signaling necessary for the required handovers is different in each case. While the assumed system consists of several types of terminals, the term is used to describe all terminals on ground, which are connected to the satellites in this paper. The proposed system is loosely related to IP routing, as MAC and IP addresses are utilized in a similar manner. For the identification on Layer 3, IPv4 addresses are used. On Layer 2, modified MAC addresses are used, whose structure depends on the address type. While the nodes utilize conventional MAC addresses, the MAC addresses of the terminals include additional geographical information. This type of MAC addresses is mutable, as the geographical information can change for mobile terminals. Figure 2 shows the generic structure of a node. The central element of the node is a switch that relays data between the input and output ports. A node possesses a dedicated switch port for each of its four ISLs. Furthermore, a node terminal with a dedicated switch port is used to communicate with other nodes. In addition, this on-board terminal has interfaces to all other components to carry out its control and management tasks. The node communicates with the terminals on ground via user links, which utilize spot beams. The user beam transceivers may use a dedicated switch port per user beam or, for larger configurations, a common switch port.

3.1 | Satellite constellations

In this work, two Walker Star constellations¹³ are considered. The first constellation is roughly based on the Iridium system,¹⁴ consisting of 66 satellites in six planes in a nearly polar orbit. As mentioned, the Iridium system is a low data rate system, while the proposed scheme focuses

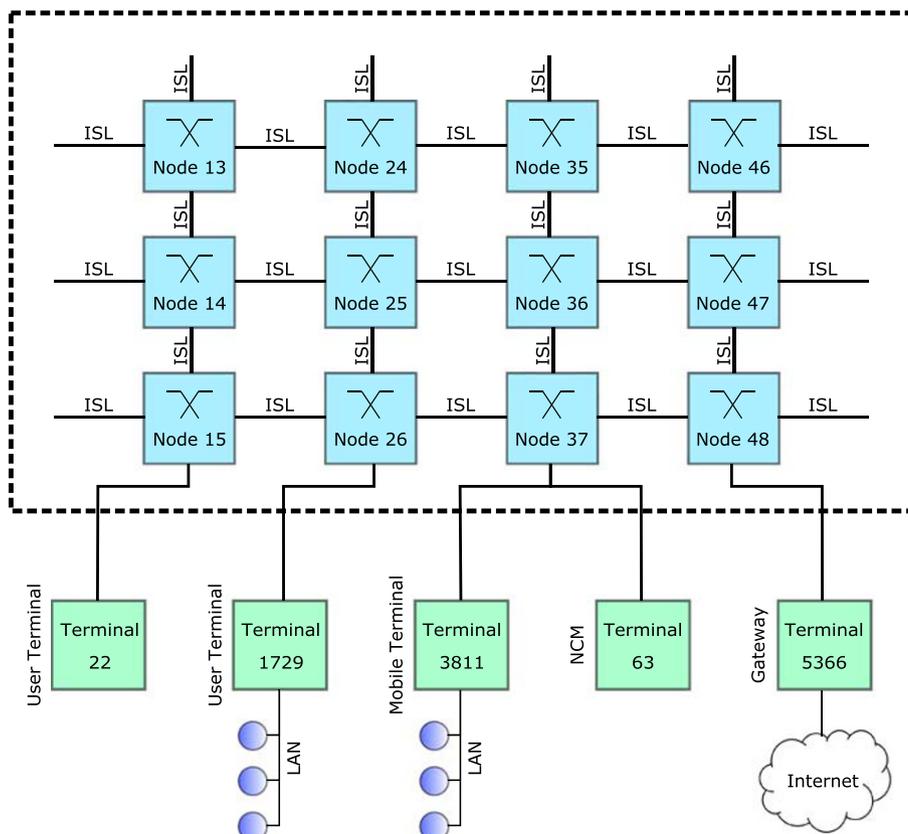
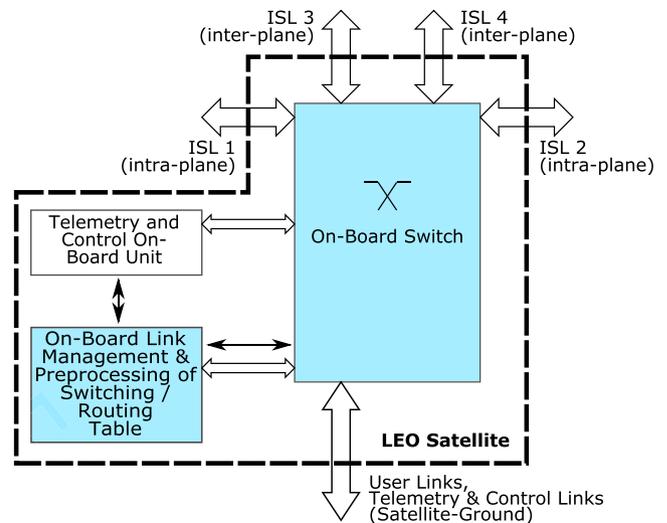


FIGURE 1 System entities of the proposed system: different terminal types are connected to the nodes. The nodes of the constellation are interconnected via intersatellite links (ISLs) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Generic node structure of the proposed reference system. The on-board switch of the satellite switches the incoming packets to the corresponding output port. The on-board link management generates the routing and switching tables [Colour figure can be viewed at wileyonlinelibrary.com]



on high data rates. The second constellation represents a scaled up version of this constellation, using 1000 satellites in 25 planes. This constellation is roughly comparable with the constellation proposed by OneWeb.¹⁵ It has been chosen represent megaconstellations, which are prevalent in envisaged “NewSpace” systems.¹⁶ Table 1 summarizes the properties of the two constellations. The initial setup of the constellations is illustrated in Figure 3. The satellites of the constellations are interpreted as individual communication nodes. A node possesses a networking table that is used for address resolution, a geographical switching table that is used for fast geographical routing, and a routing table that is used for signaling between nodes and for local routing procedures. Each node has up to four ISLs as well as multiple user links from the satellite to the ground. Whether RF or optical ISLs are considered depends on the mass and power consumption requirements of the realization. The current state of the art of ISLs favors optical communication for high capacity links (links with data rates of several tens of Mbit/s or more).¹⁷ Free-space optical communication is a particularly interesting prospect for ISLs, due to the lack of interfering weather effects.¹⁸ Therefore, the ISLs are assumed to be optical links, while the links to and from the ground are assumed to be RF links. Upon arrival on an input port, the header of a packet is stored in a buffer. This header buffer is used to compensate for the table lookup time. When the entry for the given MAC address is found in the table, the packet is directed to the corresponding output port. Individual switch buffers between the input and output ports store the packets, while the output is busy. Therefore, each incoming packet follows an inherent first in, first out (FIFO) scheduling. If this queue is empty, dummy packets are sent. As the uplink and downlink are not the focus of this investigation, these links are assumed to be free of congestion in this

TABLE 1 Considered satellite constellations

Name	Inclination	Number satellites	Number planes	Relative spacing	Height	Minimum elevation angle
Iridium-like	86.4°	66	6	180°/66	1000 km	10°
Megaconstellation	86.4°	1000	25	180°/1000	1000 km	50°

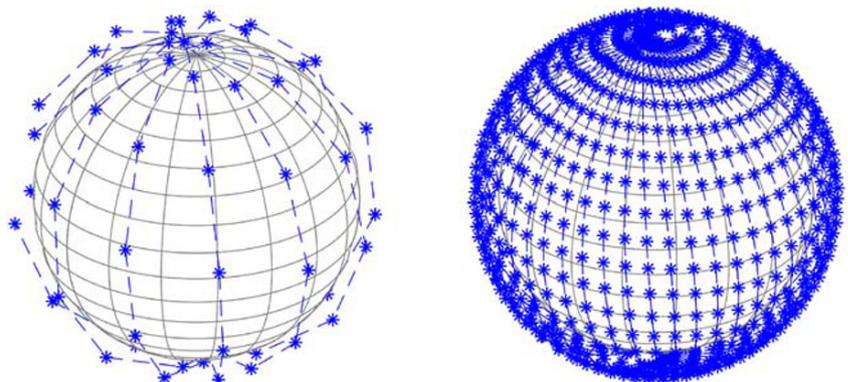


FIGURE 3 Initial distribution of the satellites of the Iridium-like (left) and megaconstellation (right). A plane is connected by a dashed line [Colour figure can be viewed at wileyonlinelibrary.com]

context. If the buffer of a satellite meets a certain threshold, an overload message is sent to its neighbors. This effects the routing procedures of its neighbors, which consider local rerouting options, as explained in Section 5.3. Due to the movement of the satellites, the nodes are not permanently attached to the terminals on the ground. Thus, handover events will occur rather frequently. There are two types of handovers: intranode and internode handovers. Intranode handovers are necessary when the node has multiple user downlinks to disjunct areas. These handovers are assumed to be handled internally inside the user-link equipment of the node. To the routing and switching functionality, they appear as a movement from one switch port to another if each user-beam has its own switch port. Internode handovers happen when a node is no longer able to communicate effectively with the terminal or a more advantageous connection to another node is possible. This generally involves some signaling between these two nodes, which may also generate more signaling traffic to other nodes. The handover procedures are detailed in Section 5.1.

In the assumed system, the coverage beams of the satellites are broken down into spot beams to use frequency reuse. In addition, it is generally advantageous to adapt the maximum scan angle of the satellite antennas according to the given constellation. On one hand, the coverage areas (or footprints) of the satellites have to overlap in order to provide permanent connectivity. On the other hand, smaller coverage areas result in higher elevation angles for the terminals on ground, allowing for improved transmission conditions. Furthermore, the interference between the satellites due to the overlapping coverage areas is decreased. Therefore, the scan angles of the satellites should be chosen as small as possible, while still guaranteeing permanent global connectivity. The scan angle of a satellite ϑ_{\max} , also called maximum nadir angle, is linked to the minimum elevation angle ε_{\min} of the terminal on ground. In this context, a terminal is an unspecified user or special purpose terminal on the ground, which are described in detail in Section 3.2. ψ_{\max} denotes the angle from the center of the Earth between the subsatellite point and a terminal at the edge of the coverage area. As the three angles describe a triangle, ϑ_{\max} can be expressed with the following formula¹⁹:

$$\vartheta_{\max} = \pi - \psi_{\max} - \left(\varepsilon_{\min} + \frac{\pi}{2} \right)$$

In order to discern a fitting scan angle for the proposed constellations, we set a minimum elevation angle and verify global coverage. Based on the minimum elevation angle, the maximum Earth central angle ψ_{\max} is calculated:

$$\psi_{\max} = \arccos\left(\frac{r_{\text{Earth}}}{r_{\text{Earth}} + h} \cos(\varepsilon_{\min})\right) - \varepsilon_{\min}$$

As the coverage areas are circular and thus need to overlap, we use spherical hexagons without overlapping to represent the coverage areas instead.²⁰ The area is calculated by splitting the spherical hexagon into six spherical triangles, which have the edge angle α :

$$\alpha = \arctan\left(\frac{\sqrt{3}}{\cos(\psi_{\max})}\right)$$

$$A_{\text{Hexagon}} = 6r_{\text{Earth}}^2 \left(2\alpha - \frac{2\pi}{3} \right)$$

The area of such a spherical hexagon A_{Hexagon} is then compared with the area of the Earth A_{Earth} . The number of spherical hexagons needed is equal to the number of satellites N_{Sat} needed to achieve global coverage.¹⁹

$$N_{\text{Sat}} = \left\lceil \frac{A_{\text{Earth}}}{A_{\text{Hexagon}}} \right\rceil = \left\lceil \frac{\pi}{3\alpha - \pi} \right\rceil$$

Specifically for polar and near-polar constellations, the number of satellites required for global coverage can be approximated by the following formula²¹:

$$N_{\text{Sat}} \approx \frac{4}{1 - \cos(\psi_{\max})}$$

For the Iridium-like constellation, a minimum elevation angle of 10° was chosen and for the megaconstellation a minimum elevation angle of 50° was chosen. For a height of 1000 km, a minimum elevation angle of 10° results in a scan angle of approximately 58.34° . A minimum elevation angle of 50° results in a scan angle of approximately 33.75° for the same height. Based on the previous formulas, these angles provide global coverage when used in the proposed satellite constellations.

3.2 | Terminals

The assumed overall system consists of regular user terminals, node terminals, and special purpose terminals. Special purpose terminals include network control and management (NCM terminals) as well as gateways to relay traffic to other networks. Links between terminals and nodes may be built using different technologies, for example, DVB-S2, DVB-RCS2, or others. In the context of this paper, minimal assumptions about the user-link technology are made. Regular user terminals are terminals used by end users of the system to send or receive traffic. In this case, an end user is not strictly a single person or computer. The terminal may have a more or less complex LAN on its user-side interface. Node terminals are special terminals that are integrated with a node. They do not have an RF-link to the node but are directly connected. They are used to handle traffic between higher protocol layers in different nodes and to the network operation terminals. Gateway terminals generally have large networks like the Internet connected on their user side. NCM terminals are connected to local or global network operation centers. As this work focuses on the routing aspects of the system, only user terminal to user terminal transmissions are considered. Therefore, it is sufficient to solely include user terminals in the simulations for this work. For the sake of simplicity, the term “terminal” is thus solely used to describe the previously defined user terminals for the remainder of this work. The various types of terminals on the ground are also shown in Figure 1.

As future communication systems with integrated satellite networks might include mobile terminals, such as ships or airplanes, stationary and mobile terminals are considered in this paper. Nevertheless, the number of stationary terminals is assumed to be significantly larger than the number of mobile terminals. Since the velocity of mobile terminals is lower in comparison with the relative velocity of the satellites, all terminals are often accepted as stationary in the literature.³ However, since longer time frames are also considered, the implementation of mobile terminals is necessary. The mobile terminals move in a random direction with a random speed limited to 360 m/s. The speed limit has been chosen in order to approximate the traveling speed of a commercial airplane. These assumptions slightly differ from real-world scenarios, in which planes and ships oftentimes travel on common routes at specific speeds. In the proposed system, a terminal can only be connected to one node at a time, whereas a node can be connected to multiple terminals on the ground. The link of a terminal to a node is called terminal link. It is assumed that no packets are dropped in these links for the proposed data rates.

3.3 | Communication sessions

In order to communicate, a terminal starts a session. A session describes the communication between two terminals over the satellite network for a certain duration. When the session is initiated, an address resolution scheme has to be applied on the node. Once the recipient has been found, the communication over the network is started. The initial request consists of the IP address of the destination terminal. The address resolution then provides the MAC address of the destination terminal, which is used for routing. Potential handover scenarios due to the mobility of the nodes and the terminals have to be respected during the address resolution phase and the routing calculations.

The traffic model used for the communication sessions is the following: two randomly chosen terminals commence a session at a random begin time with a random duration. For simplicity, simplex communication is assumed. The time of the session beginnings during the simulation is uniformly distributed, while the durations of the sessions follow a Poisson distribution. A terminal can participate in more than one active session at a time.

4 | GEOGRAPHICAL ROUTING SCHEME

As the system is highly dynamic and the connection between a terminal and a node is relatively short (in the order of dozens of minutes), a specialized packet-based routing scheme is considered. Traditional mobile communication systems, such as Mobile IP,¹⁰ handle this situation by using a central location register, which is updated with every reattachment and that source terminals consult to map the name of the destination to a routable address. Unfortunately, this approach scales poorly due to the highly dynamic topology of the system: the nodes as well as some terminals are constantly moving. The frequent handover events between nodes and terminals cause significant signaling overhead. To overcome this problem, we propose to use a routing scheme based on the geographic location of the terminals. To this end, the complete surface of the Earth is split into a grid (which is not necessarily homogeneous). Each cell in the grid has a unique area identifier. This identifier is used to route the packets to the destination area and is independent of a specific destination node. In general, such schemes are more robust against node failures, due to their inherent decentralization. In contrast to conventional geographical routing schemes, the location of the terminal is embedded in its MAC address in the proposed approach. Consequently, the topological information is decoupled from the public identity of the terminal, that is, its IP address. Due to this separation, the movement of a terminal does not necessarily require network-wide updates. Naturally, this implies that the MAC address of a terminal is mutable and has to change when the terminal moves to a new cell of the grid. Moreover, the decoupling implies a dedicated address resolution scheme. In principle, the geographical routing scheme only requires signaling when a terminal of an active transmission changes its MAC address or when a node failure arises. As the geographic

location can be considered as private information, the decoupling inherently increases the privacy of the terminals in comparison with schemes that embed the area identifier in the IP addresses.

A node periodically calculates the next hop for any possible geographic destination, that is, each cell of the grid, based on its own location. Thus, a switching table is generated by the node. When a packet arrives at the node, the area identifier is extracted from the MAC address of the destination terminal. Using the switching table, the next hop is then identified. From the perspective of a terminal, the packets traverse a distributed switch until arrival in the destination area. Then, the terminal identifier in the MAC address of the destination terminal is used to find the corresponding user-link port, if this identifier is known by the node. Beyond the generation of the switching table, the nodes possess the following Layer 3 capabilities: the source node of the transmission has to obtain the MAC address of the destination terminal by applying an address resolution scheme, overload messages of other nodes have to be taken into account, and possible ambiguities once the destination area is reached have to be resolved. The address resolution as an independent mechanism allows for additional degrees of freedom in terms of its implementation. Address resolution updates are only required when a terminal enters a new area. Thus, the signaling is reduced if there are significantly more fixed terminals than mobile terminals, which is the case for the assumed scenario. In order to minimize the dropping of packets, a rerouting mechanism can be added in the nodes. If the buffer allocation of a node reaches a certain threshold, an overload signal is sent to the neighboring nodes. The surrounding nodes then regenerate their switching tables to account for the overloaded node. When the node is no longer overloaded, the switching tables have to be regenerated again. While this functionality is not required for successful routing, it is useful to mitigate congestions. As multiple satellites might be covering the same grid cell, the node reached by the geographical routing scheme might not be connected to the destination terminal and is thus not familiar with the terminal identifier. This ambiguity in the last hop of the routing has been investigated previously.¹² In this work, it is solved by a local broadcasting of the connection in an extended area. The broadcasted updates are triggered by inter-node handover events. The procedure is explained in detail for the reference system in Section 5.3. There are also certain trade-offs concerning the cell size of the grid. If the size is too small, there are many cells and the area identifier needs more bits. If the size is too large, a single cell may be covered by a multitude of satellites and the local routing is more complex. A small cell size also requires more location updates because of mobile users. The cell size in the simulations have been chosen to fit the requirements of the reference system.

5 | ROUTING METHODS AND PROCEDURES

5.1 | Internode handover procedures

Internode handovers are needed when establishing a new connection between a terminal and a node. When multiple nodes are within the accepted proximity, a simple approach is to use the satellite that is closest to the terminal. Additionally, it is useful to include the future trajectory of the node in relation to the satellite in this decision. Furthermore, in particularly dense regions, this might not be the best choice due to a significant workload on the chosen node. However, these schemes require that a terminal probes multiple nodes. Since we assume that the terminals have no such prior information available and an inspection of neighboring nodes takes additional time, this is not taken into account in the handovers. If the connected node experiences a significant workload, it proposes another node to the connected terminal, which then initiates a handover. Since the nodes are informed about the overload status of their neighboring nodes, no additional traffic is generated. In order to facilitate the handover between two nodes, a proactive procedure has been implemented. As soon as the elevation of the current node in relation to the terminal is below a specified angle, the terminal initiates the previously described handover procedure. When a new node with better connectivity is found, the original connection is cut and a connection to the new node is established. The session information is handed over from the previous node to the currently connected node.

5.2 | Address resolution procedures

In order to compare the efficiency of the proposed approaches, two basic address resolution procedures have been implemented. The first approach can be described as request flooding. Each node is oblivious to the feeder links of the others. The address resolution process begins with the flooding of the complete network in order to find the destination node. Only the destination node retains the address resolution information and subsequently sends a response to the source node. When the response is received, the transmission commences. In order to avoid loops, a cool down for the specific session identifier is set, after the request has been transmitted to all neighboring nodes. Naturally, other loop-free network flooding techniques²² can be considered as well. As this approach has to find the destination node and establish a connection after the session is initialized, an initial connection delay is expected. The second approach is to broadcast every mapping in the networking table to every satellite. Consequently, the networking table of each satellite includes all mappings between active terminals and their MAC address. This effectively provides an immediate address resolution. Updates are only required in three cases: when a terminal joins or leaves the network, when a terminal changes its IP address, or when a terminal moves to a new area, which causes the area identifier in the MAC address to change. The

signaling overhead of this scheme depends thus on the number of terminals and the changes in their connectivities. Therefore, the choice of the addressing scheme depends on the use case. The broadcasting scheme provides an immediate address resolution and is preferable if the number of terminals is does not cause overly large networking tables, and required updates are relatively scarce. On the other hand, the request flooding scheme only generates traffic once a session is initiated, which is advantageous if the number of sessions is low. A local clustering of certain sub-groups of satellites can also be used to combine the two approaches. Within a cluster, the second approach is applied to broadcast the mapping to all satellites of the cluster. An adjusted request flooding scheme is used to handle the address resolution between clusters. Furthermore, the schemes can be improved by adaptive caching.

5.3 | Geographical and local routing procedures

The routing and switching procedures are relatively simple thanks to the geographical addressing scheme proposed in Section 4. As stated before, a node possesses a networking table for address resolution purposes, a switching table, which allows for a streamlined geographical routing, and a routing table, which is used by the local routing scheme and needed for signaling between nodes. When the MAC address of the destination terminal has been identified by the address resolution scheme, it is added to the packet headers. Based on the geographic area identifier of this MAC address, the packets are then sent to the grid cell with the corresponding area identifier. A mask is applied by the nodes to filter out the relevant geographic information. Each node decides which ISL should be used based on the relevant geographic area identifier and its knowledge of the constellation. To this end, a switching table is generated. A mask can even be applied to the area identifier in order to decide the next hop more quickly, if the number of area identifiers is high. Once a node that covers the grid cell of the destination terminal has been reached, the local routing scheme is used. The local procedure is based on connectivity updates, which are triggered by internode handover events. The connectivity update is sent to all nodes within an extended area of the grid cell of the terminal. Due to the movement of the satellites, it is insufficient to only update the nodes that are covering the grid cell at the beginning of the transmission. Otherwise, the transmission could reach a node that is covering the grid cell but does not have information about which node is actually connected to the terminal. To circumvent this case, the area is extended to include satellites that cover the cell at a later instant. The size of the extended area is based on the maximum duration of a fly by and depends on the size of a grid cell and the constellation. In order to avoid deprecated entries in the routing tables of the nodes, the entries of the connectivity updates have a lifetime equal to this duration. The ambiguity in finding the destination node can introduce multiple additional hops if the destination node is on the other side of a seam of the constellation. No short path to the destination node is available in this scenario due to the counter-rotating planes. This issue can be resolved by encoding the orbital information of the connected node as well as the area identifier of the terminal.¹² In the reference system, this approach introduces additional signaling depending on which address resolution scheme is used. If the broadcasting scheme described above is utilized, an update is required each time the terminal connects to a satellite in a different orbit, since its MAC address changes. If a node that has been reached by the geographical routing does not have an entry in its routing table clarifying which node is connected to the destination terminal, the packet is dropped. Due to the local updating scheme, this case only arises when a satellite fails after address resolution.

6 | SIMULATION SCENARIO

The simulator has been created from scratch in C++17. An object oriented approach has been applied using template programming techniques providing a generic, easily extendable system simulator. The characteristics of the constellation, the terminals and the transmissions are set in a configuration file. The general system settings of the simulations can be found in Table 2. If not stated differently, these are the parameters used in the simulations. Again, the two constellations of interest are an Iridium-like constellation and a megaconstellation, which is loosely based on planned constellations. All system entities are managed by the events of the chronologically ordered event loop. When the defined time limit of the simulation is reached or no events are queued in the event queue, the simulation is ended. During the simulation, parameters of interest are constantly monitored and periodically written to output files. The results of the simulation have been analyzed using MATLAB. The movement of the satellites is given as an input by a prior script written from scratch in MATLAB. The current position of the satellite is a linear interpolation

TABLE 2 General parameter settings of the simulation, if not specified differently

System	Constellation		Population		Sessions		
Simulation duration	43 200 s	Number satellites	66/1000	Number terminals	1000	Number sessions	10000
Movement update	1 s	ISL data rate	100 Mbit/s	Number mobile	50	Average duration	30 s / 300s
Network update	30 s	Node buffer size	25 Mbit	Maximum speed	360 m/s	Data rate	10 Mbit/s

between the two closest instants in the input file. The given input positions for the simulation have a step size of 60 s. The movement of a satellite (in steps of 60 s) is illustrated in Figure 4. Due to the rotation of the Earth, the satellites drift relative to the fixed terminals with each fly by.

In order to illustrate various network loads, the average duration of a session varied between 30 and 300 s. The size of a packet is 100 kbit. During a session, a packet is sent every 10 ms. This results in a data rate of 10 Mbit/s for a session. The geographical locations of the terminals on ground are randomized, following the distribution of a population density map of the Earth. A session is between two random terminals, which is of interest when investigating routing procedures. This scenario slightly differs from real-world scenarios, in which sessions tend to be locally correlated (e.g., more sessions within a country are expected).

The proposed system is also compared quantitatively with a source-based routing scheme. This approach uses Dijkstra's algorithm at the sending terminal to compute the optimal route in terms of propagation delay. It does not utilize any load-balancing mechanisms. By comparing the proposed scheme to this approach, we can evaluate the average transmission delay and dropping rates. In terms of the average transmission delay, the source-based scheme is a good benchmark in scenarios with little congestion. However, in more demanding traffic scenarios, some packets might be delayed due to the queueing delays in the nodes. In terms of the dropping rates, it represents a system without a rerouting mechanism. The comparison is more interesting when evaluating these quantities than a comparison to other geographical routing schemes, which would yield results similar to those of the proposed system depending on their implementation. In general, the principal disadvantage of the source-based routing is its significant signaling overhead.

7 | SIMULATION RESULTS

7.1 | Analysis of address resolution schemes

The first investigation concerns the address resolution delay introduced by the request flooding procedure described in Section 5.2. The address resolution procedure in case of the request flooding scheme consists of a flooding, followed by a response. Thus, the duration of the procedure is approximately equal to the round-trip-time (RTT) of the transmission. Therefore, the duration of this address resolution scheme varies depending on the satellites that are involved. The proposed scheme has been investigated for the Iridium-like constellation and the megaconstellation. Figure 5 is a histogram of the address resolution delays of each session for the Iridium-like constellation. Figure 6 is the histogram for the megaconstellation. The results of the histogram are normalized: the number of occurrences of each delay has been divided by the overall number of occurrences. The periodical delay spikes in the histogram are due to sessions, in which two satellites of the same plane communicate. Since the distance of the nodes in a plane does not vary, the RTT of the address resolution is equal in these cases. Figure 5 illustrates this circumstance: six spikes are visible with equidistant spacing. As this constellation has eleven satellites in a plane, a maximum of five hops is needed to reach every satellite in the plane. Of course, if the ingress satellite has the requested IP address in its networking table, the address resolution is instant—the address resolution delay is assumed to be zero in this case. The results of the simulation using the megaconstellation are similar. However, as shown in Figure 6, due to the size of the constellation, not all intraplane delays are as clearly visible.

The scheme applying a broadcasting of the mapping of all terminals to every node provides a direct address resolution response. The networking table of each node contains the information of all terminals. In the simulated scenario, no terminals enter or leave the network. Thus, updates are only required when a mobile user moves to a new grid cell or when the orbital information of a newly connected satellite is different (as described in Section 5.3). As the number of mobile terminals is low in comparison with the number of fixed terminals, few updates are required. If orbital information is used, the corresponding updates are only required in areas that are below counter-rotating planes. While the choice of the address resolution scheme generally depends on the use case, this approach introduces significantly less signaling overhead for the given scenarios. Therefore, for the remainder of the simulations, the broadcasting scheme is utilized, if not specified differently.

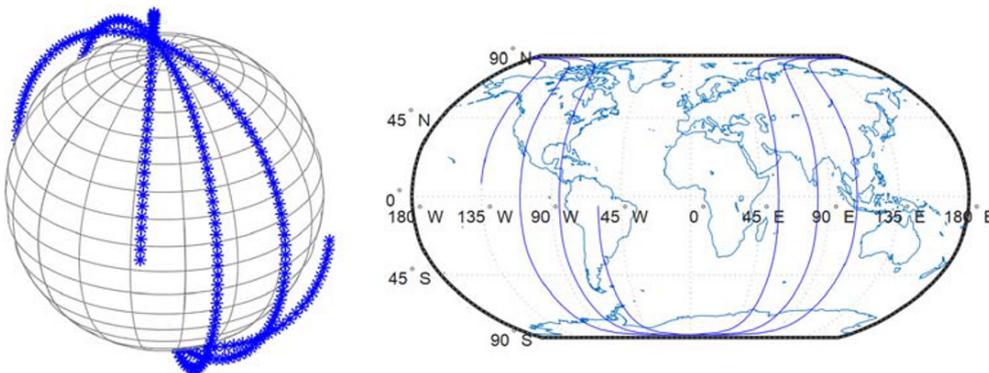


FIGURE 4 Movement of a satellite over a duration of 5 h (18 000 s). Position of satellite is updated every 60 s [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 Histogram of address resolution delay of on demand scheme for Iridium-like constellation

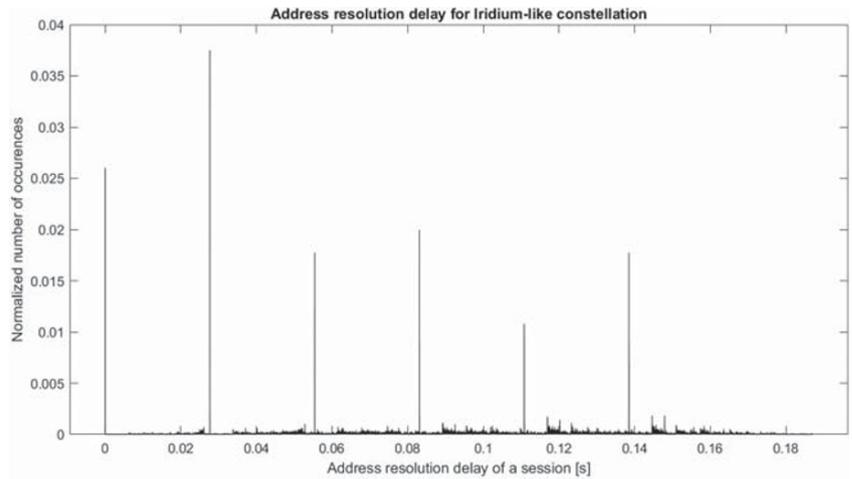
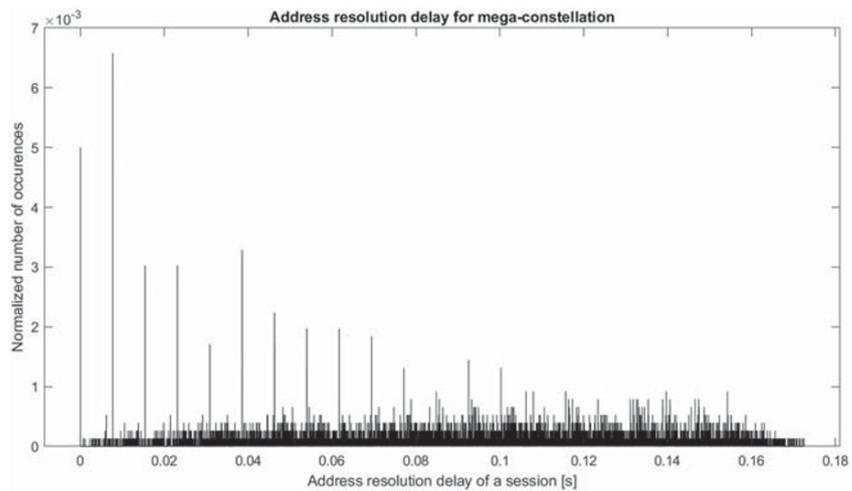


FIGURE 6 Histogram of address resolution delay of on demand scheme for megaconstellation



7.2 | Analysis of routing procedures

In order to evaluate the proposed system, we first take a look at the average transmission delay during a session for the Iridium-like constellation. So, for each session, the end-to-end transmission delay of each transmitted packet is taken into account and then averaged over the number of packets, which were transmitted. As this approach does not take dropped packets into account, the dropping rate is also of interest when interpreting the results. Additionally, the proposed scheme is compared with a source-based routing scheme using Dijkstra's algorithm in order to evaluate its performance. Generally, the network load can be scaled by varying the data rate of a transmission, the number of sessions, and their duration for the given simulation duration, as well as the ISL data rate. The chosen traffic scenarios consist of 10 000 sessions uniformly distributed over the simulation time of 43 200 s (12 h). In the first scenario, an average session duration of 30 s is assumed. Therefore, on average, approximately 6.94 sessions are active at the same time. This results in an average overall network load of about 69.4 Mbit/s, as the data rate of a transmission is 10 Mbit/s. This scenario represents a relatively low load on the network with the given parameters. The second scenario assumes an average session duration of 300 s. Thus, approximately 69.4 sessions are active at a time, and the overall load on the network is about 694 Mbit/s. The average transmission delay of the sessions is compared for these two traffic scenarios in Figure 7. As before, the results of the histogram are normalized by dividing the number of occurrences of each delay by the overall number of occurrences, which results in the relative probability. Due to the increased load in the network, packets are delayed in the buffers of the traversed nodes, and additional rerouting is possibly necessary. Furthermore, drops occur more frequently.

Figure 7 shows that many sessions are subject to increased transmission delays. For the proposed scheme, shown in plots (A) and (B), the additional delay is due the queueing delay in the buffers as well as the rerouting mechanism. As the source-based scheme, shown in (C) and (D), does not implement any form of load balancing, so the increased transmission delay in Scenario 2 is due to the queueing delay. The results show that for most of the sessions, the proposed scheme is able to provide transmission delays that are close to the delays of a

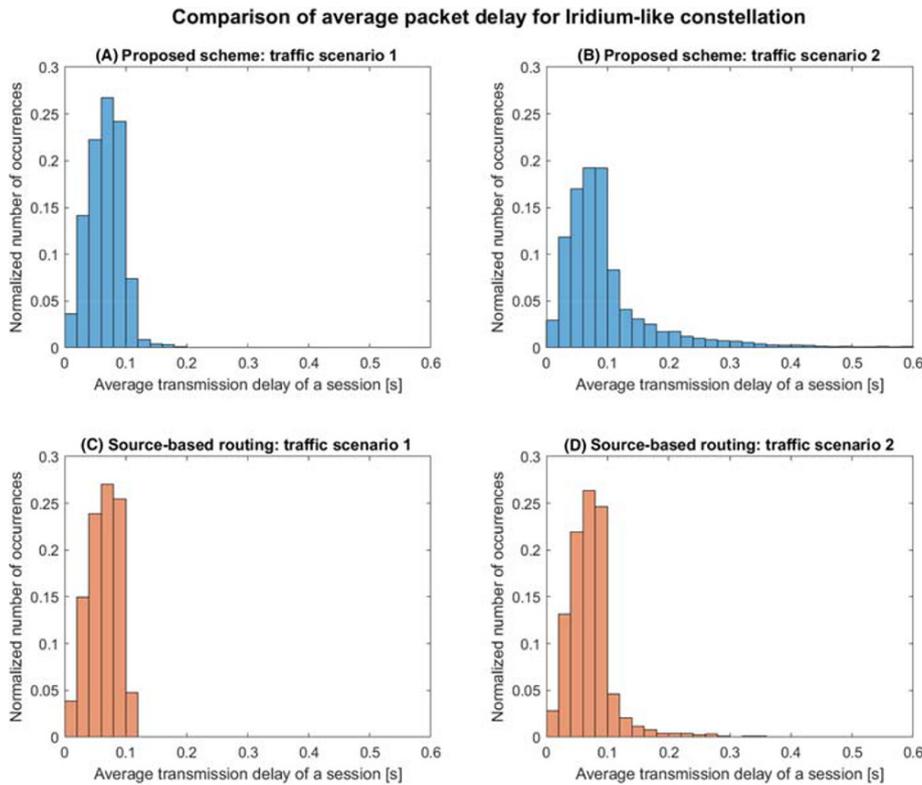


FIGURE 7 Histograms of average delay during sessions: Iridium-like constellation. Histogram (A), proposed scheme with an average network load of 69.4 Mbit/s; histogram (B), proposed scheme with an average network load of 694 Mbit/s; histogram (C), source-routed scheme with an average network load of 69.4 Mbit/s; histogram (D), source-routed scheme with an average network load of 694 Mbit/s [Colour figure can be viewed at wileyonlinelibrary.com]

source-based system that calculates the best routes in terms of propagation delay. Since the proposed scheme significantly decreases the signaling overhead, as described in Section 7.3, this is a satisfying result.

Due to the implemented load balancing in the proposed system, the dropping rates are also significantly reduced in comparison with the source-based scheme. In the first traffic scenario, the dropping rate is negligible when using the proposed scheme. An average of less than 0.01% of the incoming packets were rerouted on a node. Thus, the network was able to handle the load reliably without resorting to rerouting procedures. When using the source-based routing scheme, the packet dropping rate was 4.0×10^{-4} . In the second scenario, the average overall network load of 694 Mbit/s resulted in a dropping rate of 3.6×10^{-3} for the proposed scheme and a dropping rate of 7.5×10^{-3} for the source-routed scheme. In this scenario, each node rerouted 2.34% of the incoming packets on average in the proposed scheme. So, utilizing a simplistic rerouting mechanism, the dropping rate was more than halved in Scenario 2. If the network load is increased further, a rerouting mechanism becomes increasingly important in order to circumvent local congestions.

The load-balancing strategy of the proposed system can be illustrated by the bar diagram of the buffer allocations in Figure 8. For each node, a snapshot of its buffer allocation has been made every 60 s of the simulation time. Each snapshot represents the maximum allocation of the buffer in this time frame. The number of occurrences of the level is then represented as a bar. As a packet has a size of 100 kbit, and the overall internal buffer of a node is 25 Mbit, a maximum of 250 packets can be present in the buffer. In Figure 8, the sum of the occurrences of all nodes is used. In order to visualize the higher levels of buffer allocation, which are of interest, the plot is cut short at 525 occurrences. The overall buffer allocation for the second traffic scenario shows significantly more occurrences around the levels of 12.5 Mbit (125 packets in the buffer) and 25 Mbit (250 packets in the buffer). The nodes send an overload signal to their neighboring nodes when their overall buffer exceeds a size of 12.5 Mbit. As this induces the rerouting scheme, this buffer level indicates that more rerouting procedures were triggered in this scenario. Furthermore, a higher dropping rate is indicated, since the nodes drop incoming packets when the limit of the overall buffer of 25 Mbit is reached. These findings are coherent with the overall statistics of the simulation.

A possible approach to decrease the routing delays in the system due to rerouting procedures is the increase of the data rate of the ISLs. Simulations using an ISL data rate of 1 Gbit/s instead of 100 Mbit/s resulted in decreased end-to-end transmission delays and fewer drops. All other parameters of the simulation were kept identical. The results were comparable with the simulations with less traffic, traffic scenario 1. Therefore, depending on the use case, the application of ISLs with higher data rates should be considered in order to minimize end-to-end transmission delay and dropping rates, if high network loads are expected.

Consisting of significantly more satellites, the megaconstellation has more possible routes and buffer space. Thus, the simulations generally yield fewer delays and dropped packets when applying the same traffic loads as before. This is illustrated in Figure 9, where the previous traffic loads are tested on the megaconstellation. Again, the results are normalized using their relative probability: the number of occurrences of each delay is divided by the overall number of occurrences. As before, the proposed system is compared with a source-based routing scheme using

FIGURE 8 Bar diagram of overall buffer allocation on nodes: Iridium-like. Scenario (A) applies an average session duration of 30 s, resulting in an average network load of 69.4 Mbit/s; scenario (B) has an average session duration of 300 s, resulting in an average network load of 694 Mbit/s [Colour figure can be viewed at wileyonlinelibrary.com]

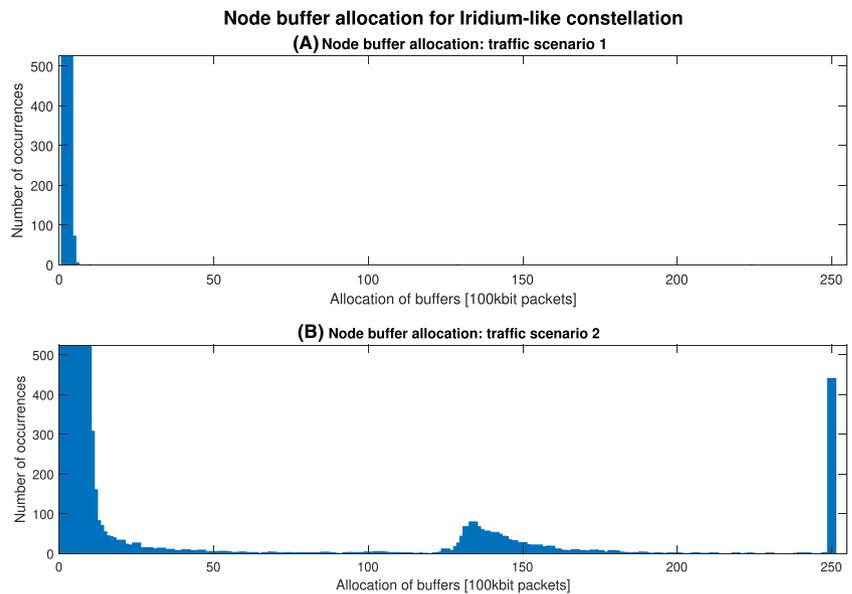
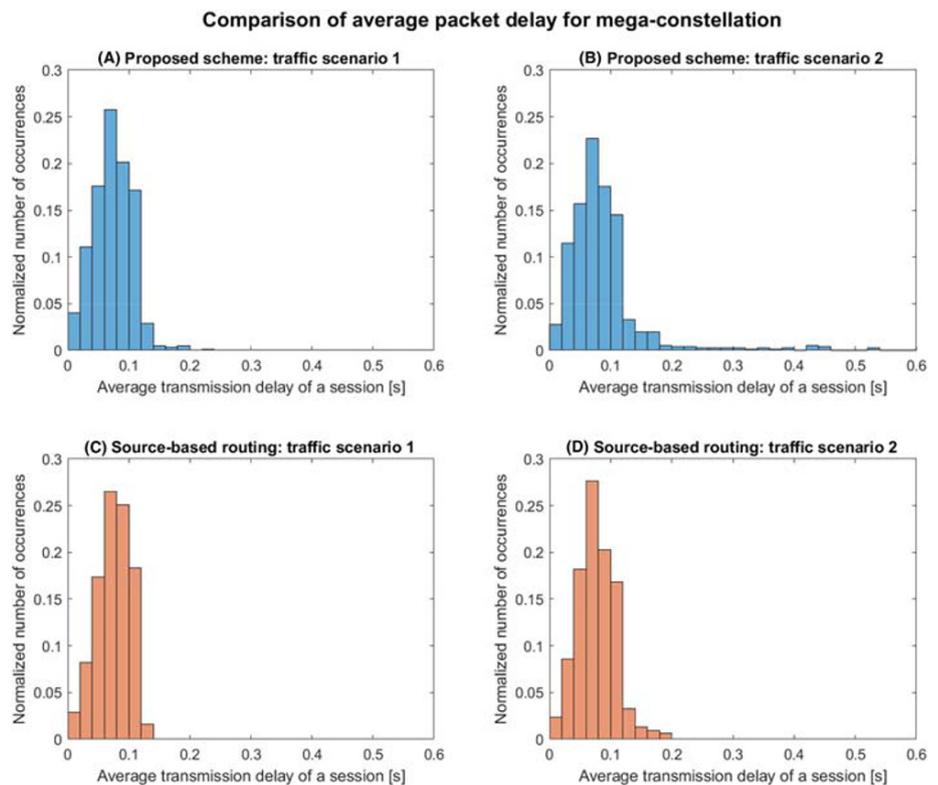


FIGURE 9 Histograms of average delay during sessions: megaconstellation. Histogram (A), proposed scheme with an average network load of 69.4 Mbit/s; histogram (B), proposed scheme with an average network load of 694 Mbit/s; histogram (C), source-routed scheme with an average network load of 69.4 Mbit/s; histogram (D), source-routed scheme with an average network load of 694 Mbit/s [Colour figure can be viewed at wileyonlinelibrary.com]



Dijkstra's algorithm to find routes with minimal propagation delay. For the megaconstellation, we have the following overall statistics: in the first traffic scenario, no packets were dropped and no significant rerouting was needed using the proposed scheme. The dropping rate of the source-based routing approach was also negligible. The slight "tail" in histogram (A) are caused by rerouting decisions due to the low rerouting threshold, which were effectively not required to mitigate drops. In the second scenario, the proposed system had a dropping rate of $9.3e-4$. In this scenario, each node rerouted 0.64% of the incoming packets on average. Using the source-based approach resulted in a dropping rate of $1.7e-3$.

While the megaconstellation is therefore more resilient and able to handle higher traffic volume, longer delays can occur due to the aggregation of buffer delays. An average transmission traverses approximately 20 satellites, allowing for significant delays. As the average transmission path of the Iridium-like constellation consists of approximately five satellites, the delays introduced by the buffers are not as apparent. Therefore, the scaling of the buffers on the satellites in conjunction with other parameters of the constellation is of importance. For the megaconstellation, the results concerning different ISL data rates are similar to the previous results of the Iridium-like constellation.

7.3 | Comparison of signaling overhead

An important aspect of routing schemes for LEO satellite systems is the generated signaling overhead. To this end, we compare the mathematical signaling cost of Mobile IP, Hierarchical Mobile IP (HMIP), a previously proposed system based on geographical routing and the proposed scheme. Fundamentally, the geographically based approaches are similar in their functionality. However, as the geographical information is embedded in the MAC addresses of the terminal in our approach, the necessities for updates differ. In an analysis of Mobile IP with and without paging extensions,²³ the management cost has been calculated as the product of the message size M of the update and the number of hops H , which are required to reach the destination:

$$\text{Cost} = M \cdot H$$

We use this measure of management cost in order to evaluate the signaling overhead introduced in the proposed architecture. For the sake of simplicity, all update messages of all schemes are assumed to be of the same size M . The costs of required control updates are neglected in this analysis. In the Mobile IP system,¹⁰ a location directory is required. In the context of the reference system, an update has to be sent to said location directory each time an internode handover event occurs. For Mobile IP, the number of hops between the terminal and the location directory, where the update is processed by a Home Agent (HA), is assumed to be $H_{T,HA}$. As an improvement to the scheme, the usage of local location directories has been proposed (HMIP).²⁴ The number of hops needed to reach the local location directory in a HMIP system is denoted as $H_{T,LD}$. We consider R_H to be the rate of inter-node handovers. The update as well as a response are sent. Thus, the cost at time t for these schemes can be estimated by the following formulas¹²:

$$C_{MIP}(t) = 2 \cdot M \cdot H_{T,HA} \cdot R_H(t)$$

$$C_{HMIP}(t) = 2 \cdot M \cdot H_{T,LD} \cdot R_H(t)$$

In general, sending an update for each internode handover introduces significant overhead. Even though the number of hops is decreased for HMIP, signaling for every handover is costly. In the Iridium-like constellation, around 7600 internode handover events occur per hour. In the simulations of the megaconstellation, around 25028 internode handover events occur per hour. For some approaches, like the previously described source-based routing scheme, these updates are required. When geographical routing is utilized, these updates are unnecessary. The signaling that is required to solve the ambiguity of the last hop in the destination area is comparably lower, as the problem can be resolved locally. Geographical routing schemes, such as the approach by Tsunoda et al.¹² cause demonstrably less signaling than Mobile IP. In this scheme, an update is sent to the location directory when a mobile terminal crosses the border between two grid cells. The rate of such a cell crossing event is denoted as R_C . This approach embeds the geographical information in the IP address of the terminal, which might raise privacy concerns. A local forwarding scheme is used for internode handovers and paging areas are used to resolve local forwarding issues and ambiguities. So there is an additional local forwarding cost C_{Local} as well as a paging cost C_{Paging} . The overall cost at time t is the following:

$$C_{Conv.Geo}(t) = 2 \cdot M \cdot H_{T,LD} \cdot R_C(t) + C_{Local}(t) + C_{Paging}(t)$$

For our approach, the address resolution mechanism can be chosen to fit the requirements of the expected scenarios. As we used the address resolution scheme based on broadcasting changes in the networking tables, it is analyzed in this context. $H_{Broadcast}$ describes the number of hops due to a network-wide broadcasting of messages. The rate of terminals entering or leaving the network is denoted as $R_{E/L}$. As before, R_C is the rate of mobile users moving to a new cell. In contrast to the previously discussed approach, we use orbital information to mitigate additional hops caused by the inability to establish ISLs between counter-rotating planes. The rate of such changes in the direction of the orbit of the corresponding satellite, after an internode handover, is denoted as R_{Orbit} . While this causes signaling overhead, the transmission delay is reduced in these cases. This results in a cost at time t of

$$C_{Proposed.AR}(t) = M \cdot H_{Broadcast} \cdot (R_{E/L}(t) + R_C(t) + R_{Orbit}(t)).$$

The precomputed switching tables of our approach do not require additional signaling. Generally speaking, the geographical routing scheme only requires updates in the case of node failures. However, the local routing scheme, which is sometimes required to find the actual destination node, introduces signaling overhead. Depending on the applied scheme, the resulting overhead varies. In the simulations, a local broadcasting of internode handover events was used. The number of hops of these updates is denoted as H_{Area} . Furthermore, the rerouting procedure, which is used to reduce dropping rates, is based on overload updates. $H_{Overload}$ and $R_{Overload}$ describe the number of hops of such updates and their rate,

respectively. If we omit signaling due to node failures (because of their rarity and unpredictability), the cost of the signaling at time t by routing mechanisms becomes

$$C_{Proposed, Routing}(t) = M \cdot H_{Area} \cdot R_H(t) + M \cdot H_{Overload} \cdot R_{Overload}(t).$$

In our simulations, sending the overload signal to the neighboring nodes was sufficient. Hence, $H_{Overload}$ is equal to 1. For the reference system, the local routing scheme only needs to contact the neighboring nodes as well, so H_{Area} is also equal to 1. Therefore, the overall signaling cost at time t of the proposed scheme using the previously described address resolution mechanism is

$$C_{Proposed}(t) = M \cdot H_{Broadcast} \cdot (R_{E/L}(t) + R_C(t) + R_{Orbit}(t)) + M \cdot (R_H(t) + R_{Overload}(t)).$$

Based on our reference system and scenarios, R_C is low in comparison to the handover rate R_H . Depending on the constellation and distribution of the terminals, R_{Orbit} is also low in comparison to the handover rate. The rate of terminals which are entering or leaving the network is assumed to be low as well. As the handover rate causes only local signaling, our scheme generally generates less overhead in comparison to Mobile IP. The comparison with other geographical routing approaches is difficult. From a qualitative point of view, the signaling costs of the other approach are comparable or even lower than the overall cost of our scheme. However, we applied certain adjustments in the proposed system, which decrease the average transmission delay and the dropping rates. Therefore, the proposed approach represents a combination of fast and reliable transmissions, decreased signaling overhead, scalability, and flexible implementation options.

8 | CONCLUSIONS AND FURTHER RESEARCH

In this discussion, the implementation of a Layer 2 geographical routing scheme for two LEO constellations using ISLs has been investigated. The comparison with a source-based routing scheme showed that the proposed scheme provides fast and reliable transmissions in different traffic scenarios. Furthermore, it was demonstrated that the proposed approach based on a geographical identifier in the MAC addresses of the terminals offers an efficient solution in terms of switching and routing overhead in comparison to conventional approaches. The comparison between an Iridium-like and a megaconstellation underlined the scalability of the proposed system. It was also demonstrated that the megaconstellation is more resilient in scenarios with increased network load, as more buffer space and routes are available. In general, the parameters of the system have to be chosen carefully. The application of a simple local rerouting mechanism facilitated the exploration of alternative paths and minimized the number of drops for higher network loads. Further research should take a closer look at different rerouting procedures in order to further reduce the average transmission delay and dropping rates. The choice of the address resolution scheme depends on the use case of the system. However, an approach applying local caching can provide trade-offs in terms of additional signaling overhead and address resolution delay. Such approaches are more efficient for recurrent sessions and are thus a topic for additional research in the context of the proposed system. As the nadir angle of the satellites has been calculated to provide high elevations, adapting the size of the geographical areas accordingly facilitates the local routing mechanism, which resolves last-hop ambiguities. Another possible approach would be the utilization of machine learning techniques to minimize local congestions.²⁵ It is also of interest to investigate the performance of the scheme for other constellation types, for instance, a Walker Delta constellation.

REFERENCES

1. Thales SA. Ten iridium (R) next satellites to be launched soon. Thales Space News, <https://www.thalesgroup.com/en/worldwide/space/news/ten-iridiumr-next-satellites-be-launched-soon>. Accessed July 15, 2019; 2018.
2. Portillo I, Cameron BG, Crawley EF. A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. *Acta Astronaut*. 2019;159:123-135.
3. Pratt SR, Raines RA, Fossa CE, Temple MA. An operational and performance overview of the iridium low earth orbit satellite system. *IEEE Commun Surv*. 1999;2(2):2-10.
4. Dijkstra EW. A note on two problems in connexion with graphs. *Numer Math*. 1959;1:269-271.
5. Gvozdiev N, Vissicchio S, Karp B, Handley M. Low-latency routing on mesh-like backbones. In: Proceedings of the 16th acm workshop on hot topics in networks, HotNets-XVI. Association for Computing Machinery; 2017; New York, NY, USA:136-142.
6. Taleb T, Mashimo D, Jamalipour A, Hashimoto K, Nemoto Y, Kato N. Sat04-3: Elb: an explicit load balancing routing protocol for multi-hop ngeo satellite constellations. In: IEEE Globecom 2006 Grad. Sch. of Inf. Sci., Tohoku Univ., Sendai; 2006:1-5.
7. Ekici E, Akyildiz IF, Bender MD. A distributed routing algorithm for datagram traffic in LEO satellite networks. *IEEE/ACM Trans Netw*. 2001;9(2): 137-147.
8. Henderson TR, Katz RH. On distributed, geographic-based packet routing for LEO satellite networks. In: Globecom '00 - IEEE. Global Telecommunications Conference. Conference Record (Cat. No.00CH37137), Vol. 2; 2000:1119-1123.
9. Tsunoda H, Ohta K, Kato N, Nemoto Y. Supporting IP/LEO satellite networks by handover-independent ip mobility management. *IEEE J Sel Areas Commun*. 2004;22(2):300-307.

10. Perkins CE. Mobile ip. *IEEE Commun Mag.* 1997;35(5):84-99.
11. Perkins CE. IP mobility support for IPv4, revised. RFC5944; 2010.
12. Tsunoda H, Ohta K, Kato N, Nemoto Y. Geographical and orbital information based mobility management to overcome last-hop ambiguity over IP/-LEO satellite networks. *IEEE Int Conf Commun.* 2006;4:1849-1854.
13. Walker JG. Continuous whole-earth coverage by circular-orbit satellite patterns. Technical Report 77044, Farnborough, Royal Aircraft Establishment; 1977.
14. Fossa CE, Raines RA, Gunsch GH, Temple MA. An overview of the iridium (R) low earth orbit satellite system. In: IEEE National Aerospace and Electronics Conference Proceedings; 1998:152-159.
15. Hanson WA. In their own words: Oneweb's internet constellation as described in their fcc form 312 application. *New Space.* 2016;4(3):153-167.
16. Sweeting MN. Modern small satellites-changing the economics of space. *Proc IEEE.* 2018;106(3):343-361.
17. Maral G, Bousquet M. *Satellite communications systems: systems, techniques and technology, 5th edition.* Chichester: Wiley; 2011. ISBN 978-0-470-71458-4.
18. Kaushal H, Kaddoum G. Optical communication in space: challenges and mitigation techniques. *IEEE Commun Surv Tutor.* 2017;19(1):57-96.
19. Lutz E, Werner M, Jahn A. *Satellite systems for personal and broadband communications.* Berlin, Heidelberg: Springer; 2000. ISBN 978-3-64-264101-5.
20. Werner M, Jahn A, Lutz E, Bttcher A. Analysis of system parameters for LEO/ICO-satellite communications networks. *IEEE J Sel Areas Commun.* 1995; 13:371-381.
21. Beste DC. Design of satellite constellations for optimal continuous coverage. *IEEE Trans Aerosp Electron Syst.* 1978;14:466-473.
22. Stojmenovic I, Lin X. Loop-free hybrid single-path/flooding routing algorithms with guaranteed delivery for wireless networks. *IEEE Trans Parallel Distrib Syst.* 2001;12(10):1023-1032.
23. Zhang X, Castellanos JG, Campbell TA. P-MIP: paging extensions for Mobile IP. *Mob Netw Appl.* 2002;7(2):127-141.
24. Soliman H, Castelluccia C, El-Malki K, Bellier L. Hierarchical Mobile IPv6 (HMIPv6) mobility management. RFC5380; 2008.
25. Valadarsky A, Schapira M, Shahaf D, Tamar A. A machine learning approach to routing. arXiv preprint arXiv:1708.03074(2017); 2017.

AUTHORS BIOGRAPHIES

Manuel Roth received MSc degrees in Electrical Engineering and Information Technologies from the Karlsruhe Institute of Technology and the Grenoble Institute of Technology in 2018 in the context of an international double degree program. Since 2018, he has been with the Satellite Networks Department of the Institute of Communications and Navigation of the German Aerospace Center (DLR) as a research associate. His interests include satellite communication protocols, satellite constellation networks, resource management, and deep learning techniques.

Hartmut Brandt received a Dipl.-Ing. degree in Systemtechnik from the Novosibirsk Electrotechnical Institute in 1988. He has been working on financial expert systems, asynchronous transfer mode (ATM), and multimedia applications before joining the German Aerospace Center (DLR) in 2004 where he works on system simulation, channel modeling, traffic modeling, protocol design, implementation and testing, and resource management.

Hermann Bischl received the Dipl.-Ing. degree in electrical engineering from Technische Universität München in 1989 and the Dr.-Ing. degree from the University of the Federal Armed Forces, Munich, in 1994. Since 1989, he has been with the Satellite Networks (former Digital Networks) Department of the Institute of Communications and Navigation of the German Aerospace Center (DLR). He is head of the Group "Satellite Communications Techniques" and project leader of many national and international projects dealing with novel satellite communications techniques and applications. In 2014, he has been honored with the senior scientist award of DLR. His research interests are satellite communications, in particular protocols and standards (e.g., DVB-S2X/RCS2), PHY and MAC layer, resource management, network coding, error control, channel modeling, and system design.

How to cite this article: Roth M, Brandt H, Bischl H. Implementation of a geographical routing scheme for low Earth orbiting satellite constellations using intersatellite links. *Int J Satell Commun Network.* 2020;1-16. <https://doi.org/10.1002/sat.1361>