Ultra-Dense LEO-MEO Constellation Integrated 6G: A Distributed Hierarchical Mobility Management Approach

Xiaohan Qin, *Student Member, IEEE*, Ting Ma, *Member, IEEE*, Xin Zhang, *Graduate Student Member, IEEE*, Yilei Wang, *Graduate Student Member, IEEE*, Haibo Zhou, *Senior Member, IEEE*, and Lian Zhao, *Fellow, IEEE*

***Abstract*— The booming renaissance and rapid development of ultra-dense low earth orbit (LEO) satellite networks (UD-LSNs) are envisioned to realize a giant leap forward for the future sixth generation (6G) coverage expansion, bridging digital divide for remote areas and providing continuous services for user terminals worldwide. However, the inherent dual mobility, massive access scenarios and highly overlapped coverage may trigger frequent, vast and ping-pong handovers, especially with the existing limited and fixed deployment of terrestrial mobility functional entity. To this end, by exploiting the unique opportunity of UD-LSNs, we devise a medium Earth orbit (MEO) assisted distributed hierarchical mobility management architecture (HDMMA) with flexible function configuration to adapt the high dynamic and large scale network. Subsequently, the lightweight handover procedures (LHPs) are proposed for two scenarios under the HDMMA to ensure service continuity, that is on-orbit handover and off-orbit handover. Considering the user mobility attributes and satellite available resources, the on-orbit handover introduces user aggregate to share signaling overhead, while the off-orbit handover is further classified into intra-cluster, inter-cluster and inter-group handover based on the clustering and grouping. Furthermore, we conduct theoretical analysis model on the proposed LHP in terms of signaling overhead and handover latency. Simulation results verify the handover characteristics in UD-LSNs, illustrate the superiority of our HDMMA and demonstrate the handover performance improvement of the proposed LHP.**

***Index Terms*— Ultra-dense LEO satellite network, mobility management, user aggregate handover.**

Received 24 November 2023; revised 10 June 2024; accepted 25 October 2024. Date of publication 12 November 2024; date of current version 10 January 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1806104 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The associate editor coordinating the review of this article and approving it for publication was M. C. Vuran. *(Corresponding author: Haibo Zhou.)*

Xiaohan Qin, Ting Ma, Xin Zhang, Yilei Wang, and Haibo Zhou are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: nju.edu.cn; majiawan27@163.com; zanxin@smail.nju.edu.cn; yileiwang@smail.nju.edu.cn; haibozhou@nju.edu.cn).

Lian Zhao is with the Department of Electrical, Computer, Biomedical Engineering, Toronto Metropolitian University, Toronto, ON M5B 2K3, Canada (e-mail: l5zhao@torontomu.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3491794.

Digital Object Identifier 10.1109/TWC.2024.3491794

I. Introduction

THE sixth generation (6G) networks are envisioned to provide global and ubiquitous connectivity by the whole-earth coverage [1], [2], [3]. Satellite networks are the recognized solution to break through the terrain limitations for the forthcoming 6G wireless coverage extension to realize worldwide three-dimensional seamless access [4]. Accordingly, companies such as SpaceX [5], OneWeb [6] and TeleSat [7] have planned or launched thousands of low earth orbit (LEO) satellites for mega constellation construction. With relatively wider coverage, higher capacity and lower latency, future ultra-dense LEO satellite networks (UD-LSNs) represent a pillar technology to meet the ever-growing and stringent traffic demands [8], [9], [10]. As the foundation and supporting technology of UD-LSNs, mobility management is used to ensure the best connection between user terminals and satellites in a dynamic network environment, which ensures service continuity and maintains communication reachability during the user roaming. Currently, international standardization organizations including the Third Generation Partnership Project (3GPP) [11] and the Internet Engineering Task Force (IETF) [12] have actively engaged in research on mobility management in satellite networks.

Most existing works have focused on learning from the well-developed standard protocols in terrestrial networks to deal with mobility issues in satellite networks. As the cornerstone of current mobility management, mobile IP (MIP) series are widely adopted by various research [13], [14], [15], [16] to provide enhancements in satellite networks. Moreover, several novel works have proposed to improve the mobility support in satellite networks based on the core idea of the Location/Identifier separation (LISP) [17], [18], [19]. Nevertheless, the literatures above cannot be directly extended to the UD-LSNs with not adequately considering the unique characteristics of future UD-LSNs. Mobility management in UD-LSNs is still in its infancy and remains technically challenging due to the following regards.

- **Dual Mobility**: In terrestrial networks, HO is primarily caused by the movement of users, while both ground users and satellites in motion will trigger HOs in UD-LSNs, which will suffer frequent HOs under unreasonable HO strategy.

- **Network Deployment**: The UD-LSN has the characteristics of large-scale, ultra-dense, posing great challenges to network management architecture design and functional entities deployment. Existing mobility management solutions predominantly place functional entities in terrestrial networks. However, the deployment of satellite earth stations (SESs) is always uneven, limited and fixed.
- **Massive Access Scenarios**: The wide coverage of LEOs and immense connectivity demands imply a rapid expansion of potential user base, leading to a large number of simultaneous HO. The average HO rate can reach $9912/s$ with the user density of $33.36/km^2$ [11], which may increase the processing burden and HO failure.

To address the aforementioned challenges, we design a two-layer distributed mobility management mechanism based on the grouping and clustering networking architecture [20]. Specifically, LEO satellites under the coverage of a same MEO satellite are divided into one LEO group and each medium Earth orbit (MEO) satellite as the global controller oversees the satellites within its group, responsible for intra-group HO logic, and also has access to a global network view to assist with inter-group HO. Then several LEO satellites are further selected as local controllers in each group, where each local controller manages the satellites within its cluster, responsible for intra-cluster HO strategy and user registration information maintenance. Meanwhile, each access LEO satellite can also be responsible for some simple HO decisions. According to the relative motion law between users and LEO satellites, we divide the HO into two types: the HO process from the current access satellite to its adjacent satellites on the same orbital plane is called on-orbit HO; otherwise, it is off-orbit HO. In on-orbit HO, a user aggregation algorithm is proposed to simplify the HO procedure and deal with massive simultaneous handover request, which takes into account the mobile characteristics of users and the available resources of satellites to reasonably determine the aggregation scale. On this basis, we devise lightweight HO procedures in UD-LSNs to provide comprehensive assurance for seamless HO in various mobility scenarios. The major contributions of our paper are summarized as follows.

- We develop a hierarchical distributed mobility management architecture (HDMMA) with flexible function reconfiguration based on the grouping and clustering networking architecture to reduce large-scale network management complexity, where MEO with higher capacity and LEOs with lower delay are combined into the control plane to cope with the limitation of SES deployment.
- We combine the mobility characteristics of satellite networks for HO classification and devise the lightweight HO procedure (LHP) for different mobility scenarios. On-orbit HO utilizes user aggregation to deal with massive simultaneous HO considering user traveling and satellite resources, while off-orbit HO is further classified into intra-cluster HO, inter-cluster HO and inter-group HO based on the management areas the user's previous and subsequent access satellites belong to.

TABLE I
SUMMARY OF ACRONYMS

Acronym	Description
LEO, MEO	Low Earth Orbit, Medium Earth Orbit
CH, CM	Cluster Header, Cluster Member
SES	Satellite Earth Station
AS	Access Satellite
p-AS, n-AS	previous Access Satellite, new Access Satellite
ISL, GSL	Inter-Satellite Link, Ground-to-Satellite Link
UD-LSN	Ultra-dense LEO Satellite Networks
HO	Handover
HDMMA	Hierarchical Distributed Mobility Management Architecture
LHP	Lightweight Handover Procedure
MMF	Mobility Management Function
BCE	Binding Cache Entry
PBU, PBA	Proxy Binding Update, Proxy Binding Acknowledgement
HI, HAck	Handover Initialization, Handover Acknowledgment
RD, RS, RA	Route Discovery, Router Solicitation, Router Acknowledge
FlowMod	Flow Modification
HS, HA	Handover Solicitation, Handover Admission

- We conduct a comprehensive performance analysis of the proposed architecture and procedures from three aspects: signaling cost, handover delay, and buffered overhead. Furthermore, extensive simulations are conducted to verify the HO characteristics of UD-LSNs and the superiority of HDMMA.

The rest of this paper is organized as follows. In Section II, we summarize the related work. In Section III, we introduce the characteristics of UD-LSNs and design the mobility management architecture. We propose the lightweight handover procedure for various scenarios in Section IV. In Section V, we develop significant key performance indicators (KPIs) for HO performance evaluation. Simulation result is provided in Section VI and Section VII concludes the paper. The acronyms used in this paper are summarized in Table I.

II. RELATED WORKS

Most related works typically focus on HO strategy and algorithm [21], [22], [23], but the architecture design and procedure optimization are equally indispensable in practical applications, especially in the emerging UD-LSNs completely different from traditional terrestrial networks. Existing literature on mobility management architecture mainly draws on solutions from terrestrial networks, which can be summarized into three categories: MIP-based, LISP-based and distributed methods.

A. Mobile IP Series Based Approaches

MIP series are the most widely used mobility management solutions in legacy IP networks [24]. According to the

responsible entities, MIP series can be divided into host-based mobility management protocols including MIPv6, Hierarchical Mobile IPv6 (HMIPv6), Fast Mobile IPv6 (FMIPv6) and network-based mobility management protocols such as Proxy Mobile IPv6 (PMIPv6), Fast Proxy Mobile IPv6 (FPMIPv6). [25] reported a performance evaluation analysis and comparison about the existing IPv6 mobility management protocols and proposed that the HO performance of all mobility approaches is greatly affected by the network topology. Considering the differences with static terrestrial network environments, [13] conducted formulaic and numerical analysis on MIPv6 and FMIPv6 in LEO satellite networks, which found that an effective HO prediction mechanism is crucial to improve the HO performance. To support the LEO satellite motion behavior, [14], [15] proposed enhanced mechanism with the basis of MIP series in satellite networks. Specifically, two HO mechanisms were designed in [14], where flexible agent aims to reduce the burden of the fixed home satellite moving far away and aggregate HO can relieve the control signaling storm caused by continuous, frequent and global HOs. In response to the IP mobility support in LEO satellite networks, [15] proposed a network-driven mobility (NDM) management solution, which makes full use of satellite motion patterns and cooperate with terminal-driven methods for robust consideration. Reference [26] summarized the extended applications and limitations of IP-based standardized mobility management protocols in future LEO satellite constellations, and proposed that the deterministic aspects of LEO mega constellation and fixed terminal can be used to support mobility management. Reference [27] analyzed the application and challenge of the existing mobile IP protocol in satellite networks, and put forward the optimization directions for mobility management supporting satellite access. Essentially, MIP-based mobility management is a centralized method, where all control signaling and data forwarding rely on a central anchor point to deal with, leading to the single point of failure, reliability and scalability issues. Especially in future UD-LSNs with limited computational capacity, highly dynamic and large scale, it is infeasible to directly transplant the ground-based IP mobility solutions.

B. Loc/Id Seperation Based Approaches

In the MIP mechanism, IP address serves not only as the identification for users, but also indicators of location information. The dual role of IP address is regarded as the root cause hindering mobility support and limiting network scalability. To address the above issues, LISP [28], [29] methods have been put forward. Reference [17] proposed the LISP-LEO, a LISP based mobility management protocol for LEO satellite networks. In LISP-LEO, terminal address need not be changed during the HO since the location information of the new address represents the terminal partition on the Earth. Home agents in MIP may be far away from terminals, while service satellites in LISP-LEO are always located near terminals, thus avoiding triangular routing. Considering the high mobility of satellites, an indirect binding scheme based on virtual attachment point was proposed in [18] to solve the frequent binding updates. The virtual attachment point was kept fixed with respect to the ground, thus reducing the complexity and signaling cost. Reference [19] introduced the application of LISP in aeronautical communication system consisting of both satellite and terrestrial networks. The LISP protocol is used to support multi-homing mobility and reduce the transmission delay and packet loss in HO process. However, LISP-based mobility management methods require additional location resolvers. Placing the resolver on the ground will introduce extra delays in location updating and resolution, while placing on satellites may lead to inter-satellite link congestion due to the mapping records propagation. Moreover, since LISP is not fully compatible with IP-based networks, it is necessary to fully consider the backward compatibility with IP system in practical applications.

C. Distributed Approaches

Distributed mobility management (DMM) proposes the decentralized deployment of multiple anchors near users to overcome the limitations of centralized management [30], [31]. Aiming at dynamic topology of LEO satellite networks, [16] presented a virtual distributed mobility management mechanism called VMIPv6, where a virtual agent cluster is formulated to co-manage users in the corresponding virtual agent domain and HO only needs to update intra-domain relations, thereby reducing HO overhead and delay. Reference [32] proposed a MEO-assisted distributed mobility management architecture, where a dual location area including satellite area and user area was designed to accommodate the distinct mobility patterns of LEOs and users. Meanwhile, a low-cost location update and paging algorithm was proposed to effectively reduce the overhead and latency of the mobility management. In [33], a distributed mobility management architecture for satellite-terrestrial integrated networks was proposed including network, protocol, function implementation and HO process, which introduces geostationary Earth orbit (GEO) and MEO together with the ground mobility management functions (MMFs) to provide better HO delays and lower signaling overheads. Under the architecture, a dynamic mobility management node selection algorithm based on distributed multi-agent reinforcement learning was proposed in [34], so as to achieve high timeliness and low-cost mobility management in mega LEO satellite constellations. Reference [35] proposed a secure distributed location management (DMM) scheme for LEO-satellite integrated vehicular network, which has lower computational and communication overhead while protecting privacy. Reference [36] proposed a distributed lightweight stateless satellite core network architecture, combining LEO, GEO satellite and terrestrial core network, where an online algorithm was proposed to solve the distributed deployment problem of multiple network functions. Reference [37] proposed a distributed location management with low synchronization overhead and fast lookup response based on split identifier and network address, combining GEO and LEO to provide natural mobility support. The above DMM based methods allow the placement of mobility management functions as needed, which effectively enhances network

scalability and flexibility and is the preferred choice for future large-scale satellite networks.

Although one or another of the aforementioned approaches show some potential when applied to satellite networks, they are not the case for the next generation of UD-LSNs. Many challenges will be encountered with mobility management in UD-LSNs due to the ultra-dense deployment of mega constellation and the high dynamic of LEO satellites. Moreover, considering the explosive growth of HO demands and diverse access options, the placement location and processing capability of mobility management functional entities are under immense pressure. At the same time, it is required that the HO procedure should be completed with the minimum signaling overhead as quickly as possible to ensure seamless service transition. Hence, it is crucial to devise a flexible distributed mobility management architecture to adapt the high dynamic and large scale features for the next generation of UD-LSNs, and develop lightweight mobility management procedures so as to ensure service continuity and improve the QoS of users.

III. Architecture Design

In this section, we first analyze the unique characteristics of UD-LSNs, which bring new opportunities to the design of mobility management. Then mobility management architecture named HDMMA are introduced in UD-LSNs.

A. Mobility Management Opportunity in UD-LSNs

The differences between UD-LSNs and terrestrial network are considered as the fundamental factors that hinder the applicability of existing mobility management solutions to satellite networks. On the other hand, the unique characteristics of UD-LSNs will bring opportunities for mobility management, which are summarized here.

- **Network Motion Regularity:** Taking the Starlink Phase I satellite with an orbital altitude of 550 km as an example, its motion speed is about 27,697 km/h [5], while for ground users, pedestrian travel at 5 km/h and the speed of vehicles is about 40 - 80 km/h [14]. Compared with the high-speed movement of LEO satellite, the ground user mobility can be considered negligible, so the HO is primarily triggered by satellite motion in UD-LSNs. Unlike the intermittent, localized, and random nature of user mobility patterns, satellite movement exhibits continuity, regularity and periodicity, which can be leveraged to simplify HO model and decision.
- **Non-terrestrial Management:** Compared to SES, MEO satellite has broader coverage, unlimited deployment and higher capacity, making them more suitable as MMFs in satellite networks. Additionally, with the enhancement of satellite processing capabilities, LEO satellites can also provide partial MMF services. In our previous work [20], a dual-layer MEO/LEO grouping and clustering based network model have been proposed, which can significantly improve the management efficiency of large-scale networks. In line with this model, we design a mobility management architecture to reduce the management complexity in UD-LSNs.
- **User Aggregation Handover:** Users in close geographical proximity tend to be within the coverage range of the same satellite, thus having similar HO decisions. This means that users within the coverage range of the same satellite are about to collectively switch to the next same one. With the rapid increase of LEO satellite network access density, making individual HO decisions for massive users at the same time would incur significant overhead with a large number of repeated signal exchange. Therefore, we can aggregate geographically close users together to perform HO collectively to simplify HO process and share the decision-making overhead.

B. Handover Classification

According to the network motion regularity in the UD-LSN, satellites on the same orbit will generally pass through the same target area in sequence within a short period of time. Due to the wide coverage characteristics of LEO satellites, the duration of successive coverage by satellites on the same orbital plane over the same target area on Earth remains relatively long, even considering the influence of Earth's rotation. We define the on-orbit coverage probability p as the probability that one target area will be covered by at least one satellite on the same orbital plane during a given period. Otherwise, it is the off-orbit coverage probability, which is equal to $1-p$. The regularity of satellite passive motion and the intensity of user motion will affect the value of p. Based on the analysis above, it holds $p \gg 1-p$, which will also be verified in later simulation.

As shown in Fig. 2, LEO satellites have relatively short coverage time over a certain area of the Earth's surface due to their lower altitude. To ensure continuous coverage of a target area, multiple satellites on the same orbital plane are typically used to take turns serving as access points. When one satellite leaves the coverage area, another satellite on the same orbital plane takes over, which is called the on-orbit HO. The on-orbit HO probability refers to the probability that a target area will switch from the current satellite to another on the same orbital plane within a given period. Otherwise, it is the off-orbit HO probability, and the sum of off-orbit HO probability and on-orbit HO probability is equal to 1. In general, the on-orbit HO is a relatively simple HO logic, which only involves the information exchange between adjacent satellites on the same orbit. Given the ultra-dense deployment of mega constellations, users are usually in the coverage range of many satellites, facing the problem of multiple access point selection. To minimize the HO cost, it is better to switch between adjacent satellites in the same orbital plane, known as the on-orbit HO. Under this HO strategy, the on-orbit coverage probability is equal to the on-orbit HO probability.

In other words, when satellites on the same orbital plane can cover a certain area on Earth, the service satellites for that area are all located on the same orbital plane, indicating that on-orbit HOs accounts for a large proportion. Hence, users undergo on-orbit HOs in general circumstances. Additionally, taking into account some special cases, users can switch

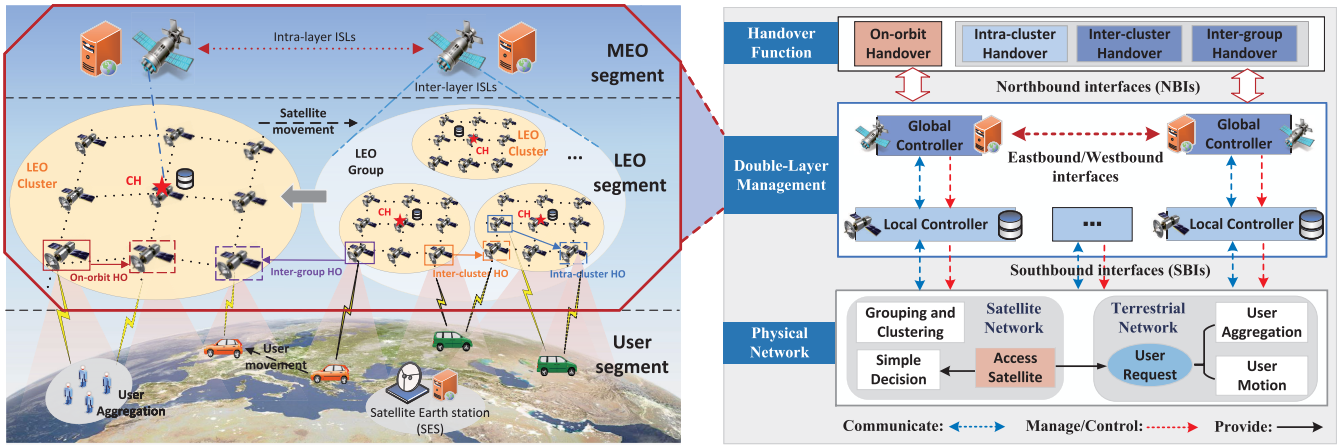


Fig. 1. HDMMA for UD-LSNs.

from the current satellite to satellites on other orbital planes, when the off-orbit HO occurs. For example, users move so fast that are out of the coverage of all satellites on the same orbital plane, like civil aircraft. And users may have higher QoS demands, such as poor experience due to an excessive number of service users on satellites within the same orbital plane. In addition, factors like the cumulative effects of Earth's rotation or on-orbit satellite failures and same orbit inter-satellite link (ISL) interruptions must be considered.

C. Management Architecture

Based on the clustering and grouping, we introduce the two-layer distributed mobility management architecture with flexible function configuration. As shown in Fig. 1, we consider a network scenario including MEO segment, LEO segment and user segment. Each network segment has partial or full mobility management function. In such a large-scale network composed of thousands of LEO satellites, LEO satellites are grouped and each group is managed by an MEO satellite. Furthermore, each group performs clustering, where cluster member (CM) LEO satellites are managed by each cluster head (CH) LEO. In UD-LSNs, several mobility-related functional entities are described as follows.

- **User Segment:** The user segment consists of cellular systems, a limited number of fixed SESs, and massive users. When encountering situations such as satellite malfunctions, insufficient computational power or the need for human intervention, the SESs will act as the global controller to deal with. Initially, a user gets access to a certain LEO satellite for Internet service. When the user leaves the coverage of its previous access satellite (p-AS), it establishes a connection with a new access satellite (n-AS) to ensure service continuity.
- **LEO Segment:** In each cluster, CM LEO serves as an access point, providing connectivity to users within its coverage. Meanwhile, CM LEO maintains the routing table and is responsible for the forwarding of arriving packets. Additionally, CM LEO can also handle certain simple mobility logic, such as on-orbit HO strategy. CH LEO serves as the local controller, responsible for off-orbit intra-cluster HO strategy, routing strategy computation for its CM LEO and user registration information track by Binding Cache Entries (BCEs).
- **MEO Segment:** As the global controller, MEO satellite takes charge of inter-cluster and inter-group routing calculation and deals with complex mobility logic and HO decision, such as off-orbit inter-cluster and inter-group HOs. Each MEO also disseminates control signaling by information exchange with CH LEOs. Moreover, MEO satellite can communicate with adjacent MEO satellites and coordinate with each other to obtain a net-wide perspective through network status and traffic information share in real time, ensuring the efficient maintenance, management and operation of the network.

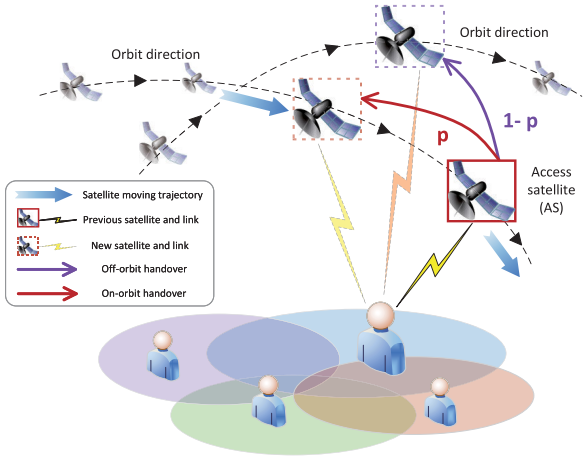


Fig. 2. Handover classification in the UD-LSNs.

IV. Handover Procedure

When a user leaves the coverage of its current access satellite (p-AS), HO process is triggered to ensure service continuity. Under the proposed HDMMA, we design the lightweight handover procedure named LHP. To support seamless HO in various mobility scenarios, the procedures for on-orbit HO and off-orbit HO are designed respectively. In order to facilitate the implementation of our procedures,

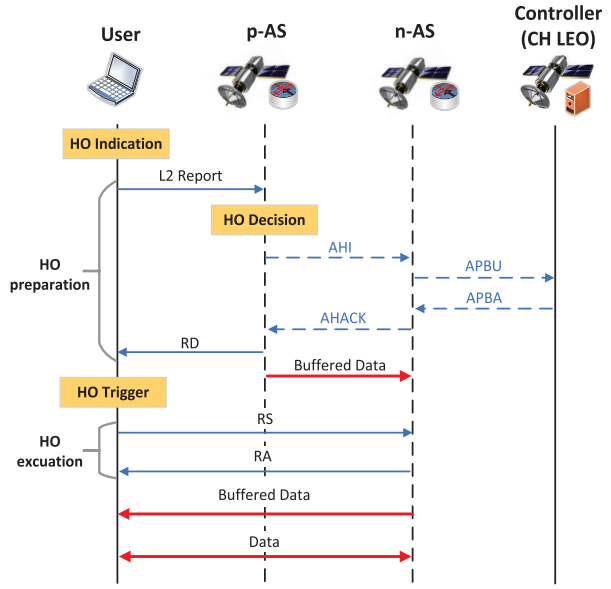


Fig. 3. On-orbit HO Procedure.

LHPs are designed drawing on the classical MIPv6 [25] and some SDN-based distributed HO procedures [38], aiming to ensure compatibility between future UD-LSNs and terrestrial IP networks and support the satellite-terrestrial integrated network.

A. On-Orbit Handover

It is worth noting that since the on-orbit HO is a relatively simple mobility management logic, the AS is allowed to perform HO decisions without uploading to the controller, which significantly simplifies the procedure. In case of a HO indication, such as the received signal strength from the p-AS is much lower than that from on-orbit LEOs (n-AS), the on-orbit HO is triggered. The whole HO process consists of two parts. The first part is HO preparation, where HO decision is made by the p-AS, user binding update is performed by the CH LEO and buffered data is transferred from the p-AS to n-AS. The second part is HO execution, where user connects to the n-AS and maintains its ongoing sessions. The following procedures are executed as shown in Fig. 3.

- First, user sends an L2 report to the p-AS including its identifier and received satellite signal strength.
- Upon receiving the report containing the information of its neighboring LEO in the same orbit, p-AS will make a HO decision for on-orbit HO and send a HO Initialization (HI) message to the adjacent satellite, i.e. n-AS.
- After receiving the HI, n-AS allocates resources for the upcoming user and sends a Proxy Binding Update (PBU) message to its CH LEO for binding information update. When two neighboring LEOs in the same orbit belong to different clusters, the p-AS also sends PBU to the CH LEO within its cluster upon detecting user separation.
- Upon receiving the PBU, CH LEO updates the user's binding information in the BCE and responds to the n-AS with a Proxy Binding Acknowledgement (PBA) message.
- The n-AS replies a HO Acknowledgment (HAck) message to the p-AS when receiving the PBA message. At the same time, the p-AS sends the buffered data of the user to the n-AS, thus avoiding the packet loss.
- Upon receiving HAck, p-AS sends the user a Route Discovery (RD) message to inform of the upcoming AS.
- In HO execution, the user sends the n-AS a Router Solicitation (RS) message to request connection.
- The n-AS responds with an Router Acknowledge (RA) message to user upon receiving the RS. The previously requested packets can continue to be forwarded to the user, thus restarting the communication service.

Algorithm 1 User Aggregation Algorithm

Input: User set $\mathcal{U}$, Maximum available resources $\boldsymbol{R}_{max}$, Maximum distance threshold θ_{max}.
Output: Aggregation results $\mathcal{A}$.

Initialize the set $\boldsymbol{c} = \mathcal{U}$, $\mathcal{A} = \emptyset$, Flag=TRUE;
while Flag **do**
  Calculates the distance between every two user classes in the user classes set $\boldsymbol{c}$;
  Select the two user classes c_m, c_n that have the smallest distance $c_m, c_n = \arg\min_{i,j} Dist_{i,j}$ and the same target satellite;
  if $Dist_{m,n} \leq \theta_{max}$ **then**
    if $R_{g_m} + R_{g_n} < R_{max}$ **then**
      $c_m + c_n \rightarrow \boldsymbol{c}$;
      Calculate the average value as the attributes of the new class after c_m and c_n are merged;
    else if $R_{c_m} + R_{c_n} = R_{max}$ **then**
      $c_m + c_n \rightarrow \mathcal{A}$;
      Move c_m and c_n out of $\boldsymbol{c}$;
    else
      $c_m + c_n - \lambda \rightarrow \mathcal{A}$, $\lambda \rightarrow \boldsymbol{c}$;
      Calculate the average value as the new class attributes after c_m and c_n are partly merged;
    end if
  else
    $\boldsymbol{c} \rightarrow \mathcal{A}$, Flag=FALSE;
  end if
end while

What calls for special attention is that in the on-orbit HO, user aggregation HO is employed to distribute the decision-making cost and handle massive HOs. A user aggregation algorithm named UAA is described in Algorithm 1 [39]. Taking into account the mobility characteristics of users and the available resources of satellites, the algorithm can appropriately determine the scale of user aggregation to avoid the additional signaling overhead and HO failure. Initially, each user is considered a class, i.e., $\boldsymbol{c} = \mathcal{U}$, where $\boldsymbol{c}$ stands for the set of classes. In each iteration, the correlation between each two classes in $\boldsymbol{c}$ is measured based on their mobility attribute. Generally, users in close proximity with the same speed tend to trigger HO to the same satellite at the same time. Therefore, user's location vector $\boldsymbol{p}$ and velocity vector $\boldsymbol{v}$ are adopted as the aggregation attributes. The correlation

between the user classes c_i and c_j can be calculated using Euclidean distance as follows.

$$Dist_{i,j} = \sqrt{\left(\boldsymbol{p}_{c_i} - \boldsymbol{p}_{c_j}\right)^2 + \left(\boldsymbol{v}_{c_i} - \boldsymbol{v}_{c_j}\right)^2},\ \forall c_i, c_j \in \boldsymbol{c}. \quad (1)$$

Next, the two user classes are selected with the smallest distance and the same target satellite in the set of user classes. If the smallest distance is still larger than the maximum distance threshold θ_{max}, the user class set is added to the aggregation result $\boldsymbol{c} \rightarrow \mathcal{A}$ and the algorithm terminates. If the strongest related user classes do not have the same target satellite, they cannot be aggregated either. In aggregation process, when the sum of resources required by the two classes is more than the maximum resources $\boldsymbol{R}_{max}$, users whose demands are closest to but not greater than $\boldsymbol{R}_{max}$ are regrouped and aggregated to $\mathcal{A}$, while the others λ are added to $\boldsymbol{c}$. Typical resources include the bandwidth, storage and power.

For each user class, the representative user utilizes Aggregated HI and Aggregated HAck (AHI/AHAck) instead of HI/HAck. Similarly, Aggregated PBU and Aggregated PBA (APBU/APBA) are used in place of PBU/PBA. When other users in the class need to handover, there is no need to repeat the HI/HAck and PBU/PBA processes. It is assumed that all users within the crowd are bound to handover to the same LEO, while other exceptional cases will be considered in off-orbit HO in the next part. Since the p-AS is aware of all users' ID currently connected, all users about to undergo HO can be inferred, which is included in the AHI/AHAck and APBU/APBA messages. Upon receiving AHI/AHAck, the n-AS needs to allocate resources for all upcoming users. Since the UAA has taken into account the relationship between user requirements and available satellite resources, the n-AS can always accommodate all users in the same class. In this case, only one AHI/AHAck and APBU/APBA is required to complete the HO initialization, confirmation, and binding update processes for all users within the class, thus greatly simplifying the HO process and reducing the HO overhead.

B. Off-Orbit Handover

In general, the off-orbit HO decision is relatively complicated, especially in certain special cases that require targeted discussions. Therefore, the AS, usually a CM LEO, cannot undertake the responsibility for HO decision, which needs to be handed over to the controller for processing. For off-orbit HO, it further includes intra-cluster HO, inter-cluster HO and inter-group HO according to the different locations of p-AS and n-AS, so as to design tailored HO procedure.

1) Intra-Cluster Handover: If p-AS and n-AS are located in the same cluster, intra-cluster HO occurs. Similarly, the HO process contains two parts: HO preparation with HO decision and route update and HO execution with user connection and session maintaining. The procedures are shown in Fig. 4.

- In the HO preparation, user sends an L2 report to p-AS.
- If the L2 report does not contain information about on-orbit neighboring LEOs or the user with high priority requests higher QoS, the p-AS sends a Dereg PBU message to the CH LEO in its cluster for HO decision.

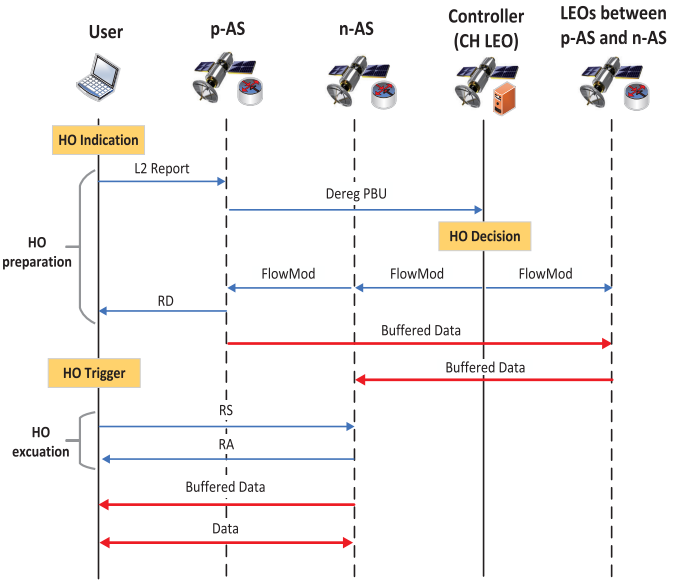


Fig. 4. Off-orbit intra-cluster HO Procedure.

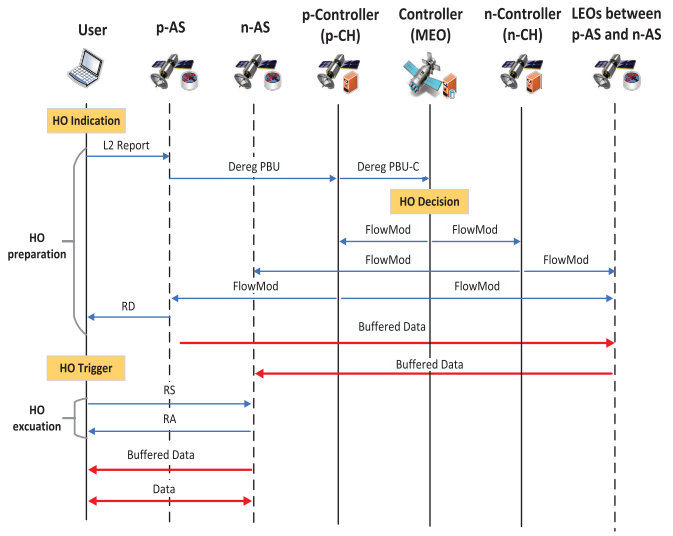


Fig. 5. Off-orbit inter-cluster HO Procedure.

- The CH LEO, which has the network information of the entire cluster, can select a more appropriate AS for the user within the cluster. After making HO decision, the CH LEO will update the BCE. Meanwhile, the CH LEO calculates the optimal route between p-AS and n-AS, and sends Flow Modification (FlowMod) messages to the satellites involved in the route to update their forwarding rules. As a result, the buffered packets on the p-AS can be forwarded to the n-AS.
- Upon receiving the FlowMod, the p-AS is informed about the n-AS and then replies the user with an RD message. Meanwhile, the n-AS allocates resources for the user.
- The user sends an RS message to the n-AS for access request, and the n-AS promptly responds with an RA message. Consequently, the data packets previously transmitted from p-AS to n-AS can continue to be forwarded to the user, thus guaranteeing seamless service.

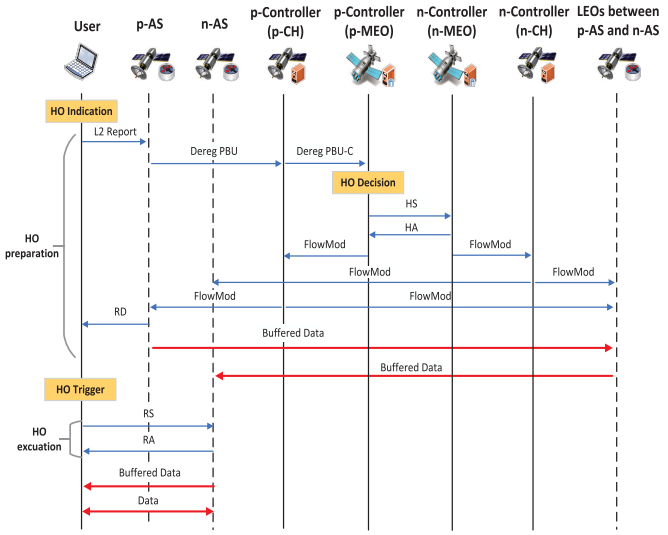


Fig. 6. Off-orbit inter-group HO Procedure.

2) Inter-Cluster Handover: Inter-cluster intra-group HO occurs if p-AS and n-AS are in different clusters but in the same group. As shown in Fig. 5, we describe the HO process with HO preparation and HO execution in the following procedures.

- Firstly, the user sends its L2 report to the p-AS.
- If all satellites in the cluster where the p-AS belongs to are not included in the report or do not meet user's QoS demands, the p-AS and its CH LEO are unable to make suitable HO decisions for the user. At this time, the p-AS sends a Dereg PBU message to its CH LEO (p-CH), and then the p-CH will send a Dereg PBU-C message to the MEO in its group to request HO decisions.
- The MEO holds the complete network view of the group, so it can select the appropriate n-AS within the group. Upon the decision-making, the MEO simultaneously sends FlowMod to the CH LEOs in the cluster of both the p-AS and n-AS to update the user binding information as well as the intra-cluster routing information.
- Afterwards, the p-CH and n-CH send FlowMod to the relevant satellites in their respective clusters according to the optimal routes between p-AS and n-AS calculated by the MEO. This allows the buffered data packets on the p-AS to be forwarded to the n-AS.
- The p-AS will respond an RD message to user about the HO decision, while the n-AS reserves resources.
- The connection between the user and n-AS will be established through RS/RA messages; Thereafter, the n-AS forwards the received buffered packets to the user.

3) Inter-Group Handover: There occurs inter-group HO if p-AS and n-AS are in different groups. In Fig. 6, we describe the HO process with HO preparation and HO execution in the following procedures.

- Firstly, user sends its L2 report to the p-AS.
- When there is no on-orbit adjacent satellite information in L2 report, p-AS will send Dereg PBU message to its CH LEO (p-CH); Subsequently, if p-CH finds that none of satellites in its cluster are included in the report, it sends a Dereg PBU-C message to the controller of its group (p-MEO); Then, when p-MEO discovers that all the satellites in its group cannot satisfy user's QoS demands, it will negotiate with other MEOs to get global view and make HO strategies. Afterward, the p-MEO sends a HO Solicitation (HS) message to the n-MEO.
- Upon receiving the HS message, n-MEO replies with a HO Admission (HA) message to the p-MEO.
- The p-MEO and n-MEO send FlowMod to the p-CH and n-CH respectively to update their BCEs.
- Subsequently, p-CH and n-CH send FlowMod to the satellites involved in their respective clusters. Therefore, the buffered packets can be forwarded to the n-AS.
- On receiving the FlowMod message, the p-AS sends an RD to user to notify the n-AS, which allocates resources for user at the same time.
- Therefore, the user establishes access to the n-AS via RS/RA messages, allowing service.

V. Performance Analysis

In this section, we propose a performance analysis model to theoretically evaluate the HO performance of the LHP under the HDMMA from aspects such as signaling overhead, HO delay, HO failure probability, and buffered overhead.

A. Signaling Overhead

It is evident that HO procedure involves signaling exchange between users and satellites, which inevitably incurs signaling overhead. The more the data amount of signaling message, and the greater the number of hops involved, the higher the signaling overhead. Therefore, we define the signaling overhead as the product of the size of control message and the number of transmission hops. In on-orbit HO, the HO procedure of the representative user is different from the others in the class, so signaling overhead needs to be calculated separately. In addition, off-orbit HOs with three cases correspond to different procedures and we discuss separately here.

1) On-Orbit Handover: For the on-orbit HO, both p-AS and n-AS are within the visual range of user. Meanwhile, since p-AS and n-AS are adjacent satellites, the distances between them are all one hop. According to Fig. 3, for the representative user in the on-orbit HO, the signaling overhead can be calculated as follows.

$$\begin{aligned} S_{R-IC} &= \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}) + \frac{\rho_c}{1-q_2} \\ &\quad \cdot \left[L_{AHI} + L_{AHACK} + H_{n-CH} \cdot (L_{APBU} + L_{APBA})\right], \end{aligned} \tag{2}$$

where L_{l2}, L_{RD}, L_{RS}, L_{RA}, L_{AHI}, L_{AHACK}, L_{APBU}, L_{APBA} are the size of L2 report, RD, RS, RA, AHI, AHACK, APBU, APBA message respectively; q_1, q_2 represent the packet loss rate of ground-to-satellite links (GSLs) and LEO ISLs respectively; ρ_c is the link control weight; H_{n-CH} indicates the number of hops between n-AS and its CH LEO.

However, when the p-AS and n-AS belongs to different clusters, they both need to send APBU/APBA messages to their

respective CH LEO, at which point the signaling overhead for the representative user is calculated as

$$\begin{aligned} &S_{R-IE} \\ &= \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}) + \frac{\rho_c}{1-q_2} \\ &\quad \cdot [L_{AHI} + L_{AHACK} + (H_{n-CH} + H_{p-CH}) \\ &\quad \cdot (L_{APBU} + L_{APBA})] \end{aligned} \tag{3}$$

where H_{p-CH} indicates the number of hops between p-AS and its CH LEO.

For other users in the same class, they do not need the signaling interaction between p-AS and n-AS as well as the binding update, so their signaling overhead is

$$S'_{R-IC} = S'_{R-IE} = \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}). \tag{4}$$

2) Off-Orbit Handover: The off-orbit HO contains three different signaling interaction flows, which are denoted as follows.

a) Intra-cluster handover: According to Fig. 4, the signaling overhead of the off-orbit intra-cluster HO is

$$\begin{aligned} &S_{D-IC} \\ &= \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}) \\ &\quad + \frac{\rho_c}{1-q_2} \cdot \left[H_{p-CH} \cdot L_{D-PBU} + \sum_{x=1}^{H_{p-n}} h^x_{p-n} \cdot L_{FMod} \right], \end{aligned} \tag{5}$$

where L_{FMod} and L_{D-PBU} indicates the size of FlowMod and Dereg PBU message; H_{p-n} is the number of hops between p-AS and n-AS; h^x_{p-n} is the hop number between the CH LEO and the x-th LEO along the path from p-AS to n-AS.

b) Inter-cluster handover: In Fig. 5, the signaling overhead of the off-orbit inter-cluster HO can be obtained as

$$\begin{aligned} &S_{D-IE} \\ &= \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}) \\ &\quad + \frac{\rho_c}{1-q_2} \cdot \left[H_{p-CH} \cdot L_{D-PBU} + \sum_{x=1}^{H_{p-n}} h^x_{p-n} \cdot L_{FMod} \right] \\ &\quad + \frac{\rho_c}{1-q_3} \cdot (2L_{FMod} + L_{D-PBU-C}), \end{aligned} \tag{6}$$

where $L_{D-PBU-C}$ indicates the size of the Dereg PBU-C message and q_3 is the packet loss rate of cross-layer ISLs from LEOs to MEOs.

c) Inter-group handover: For Fig. 6, signaling overhead of the off-orbit inter-group HO can be obtained as

$$\begin{aligned} &S_{D-IG} \\ &= \frac{\rho_c}{1-q_1} \cdot (L_{l2} + L_{RD} + L_{RS} + L_{RA}) \\ &\quad + \frac{\rho_c}{1-q_2} \cdot \left[H_{p-CH} \cdot L_{D-PBU} + \sum_{x=1}^{H_{p-n}} h^x_{p-n} \cdot L_{FMod} \right] \\ &\quad + \frac{\rho_c}{1-q_3} \cdot (2L_{FMod} + L_{D-PBU\_C}) \\ &\quad + \frac{\rho_c}{1-q_4} \cdot H_{M-M} \cdot (L_{HS} + L_{HA}), \end{aligned} \tag{7}$$

where L_{HS} and L_{HA} indicates the size of HS and HA message; q_4 is the packet loss rate of ISLs in MEO segment.

B. Handover Latency

It is the time interval during which a user cannot send or receive any packets while it performs its handover between different access satellites. We define the HO delay as the time interval from the L2 disconnection (Link Down) to the resumption of data communication (Link Up), which includes the propagation delay, transmission delay and processing delay of all signaling interaction during the HO process. In fact, the interactions of signaling messages may be synchronous. For example, when p-AS sends buffered data to n-AS after HO preparation, the user may have already received RA from the n-AS, thus completing the HO excuation. Therefore, the HO latency calculation is not simply the sum of all signaling transmission delays, where time axis is used to illustrate various cases.

1) On-Orbit Handover: In the time axis, the upper part represents the HO preparation and the buffered data handoff. Here we denote T'_{R-IC} as the time interval between user sending the L2 report and n-AS receiving the buffered data. The lower part mainly represents the HO execution, using T''_{R-IC} on behalf of the time interval between user sending the L2 report and receiving the RA message. For the representative user in the class, there are

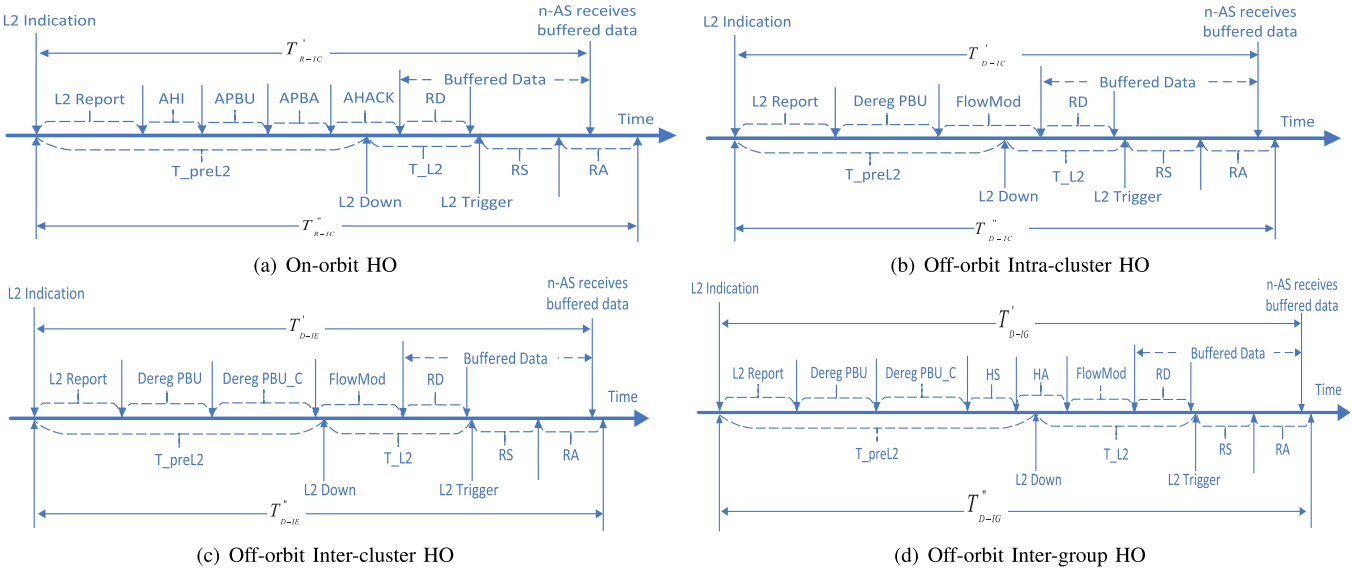
$$\begin{aligned} &T'_{R-IC} \\ &= \frac{L_{l2}}{(1-q_1)B_1} + \omega_1 + \gamma_{LEO} + \frac{L_{AHI} + L_{AHACK}}{(1-q_2)B_2} \\ &\quad + 2\omega_2 + max\left\{ \frac{L_{RD}}{(1-q_1)B_1} + \omega_1, \frac{L_{BUF\_DATA}}{(1-q_2)B_2} + \omega_2 \right\} \\ &\quad + H_{n-CH}\left[\frac{L_{APBU} + L_{APBA}}{(1-q_2)B_2} + 2\omega_2 \right], \end{aligned} \tag{8}$$

$$\begin{aligned} &T''_{R-IC} \\ &= T_{preL2} + T_{L2} + \frac{L_{RS} + L_{RA}}{(1-q_1)B_1} + 2\omega_1, \end{aligned} \tag{9}$$

where L_{BUF_DATA} indicates the size of buffered data in the p-AS; B_1, B_2 is the bandwidth of GSLs and ISLs in LEO segment respectively; ω_1, ω_2 is the propagation delay of GSLs and LEO ISLs respectively; γ_{LEO} represents the processing delay of LEO; T_{preL2} is the interval between L2 report and L2 Down, which is usually difficult to predict and calculate; T_{L2} indicates the interval between L2 Down and L2 Trigger.

By contrast, for other users in the same class, there is no AHI/AHACK and APBU/APBA process, which are both in HO preparation phase. Hence, the lower part T''_{R-IC} is same, while the upper part should be calculated as

$$\begin{aligned} T'_{R-IC} = &\frac{L_{l2}}{(1-q_1)B_1} + \omega_1 + \gamma_{LEO} \\ &+ max\left\{ \frac{L_{RD}}{(1-q_1)B_1} + \omega_1, \frac{L_{BUF\_DATA}}{(1-q_2)B_2} + \omega_2 \right\}. \end{aligned} \tag{10}$$



(a) On-orbit HO

(b) Off-orbit Intra-cluster HO

(c) Off-orbit Inter-cluster HO

(d) Off-orbit Inter-group HO

Fig. 7. Timing diagram under different HO.

According to Fig. 7(a), user will not send RS message until it receives the RD, that is, the inequality should be satisfied as $T_{preL2}+T_{L2} \geq \frac{L_{l2}}{(1-q_1)B_1}+\omega_1+\gamma_{LEO}+\frac{L_{GHI}+L_{GACK}}{(1-q_2)B_2}+2\omega_2+\frac{L_{RD}}{(1-q_1)B_1}+\omega_1$.

The HO latency is the time interval from the L2 down to the re-establishment of data communication, which can be represented in the time axis as the time difference between the larger of T'_{R-IC} and T''_{R-IC} and T_{preL2}. However, the timing of when n-AS receives the buffered data and when user completes the HO cannot be determined precisely, which is related to the buffered data amount. If $T'_{R-IC} > T''_{R-IC}$, it indicates that when user receives the RA message and connects to n-AS successfully, the buffered packets along the original path have not been transferred from p-AS to n-AS. At this point, user needs to wait until the buffered data reaches n-AS before resuming communication. Otherwise, it means that buffered packets reaches n-AS before user connects to n-AS, so that user can re-establish the previous communication with the n-AS immediately after receiving the RA. Hence, the on-orbit intra-cluster HO latency can be represented as:

$$T_{R-IC} = \begin{cases} T_{L2}+\dfrac{L_{RS}+L_{RA}}{(1-q_1)B_1}+2\omega_1, T'_{R-IC} \leq T''_{R-IC} \\ T_{L2}+\dfrac{L_{RS}+L_{RA}}{(1-q_1)B_1}+2\omega_1+T'_{R-IC}-T''_{R-IC}, others. \end{cases} \tag{11}$$

However, when the p-AS and n-AS belongs to different clusters, in addition to n-AS sending APBU/APBA to its CH, p-AS also needs to send its CH LEO APBU/APBA messages. T'_{R-IE} represents the time interval between user sending the L2 report and n-AS receiving the buffered data, while T''_{R-IE} is the time interval between user sending the L2 report and receiving the RA. In this case, only the upper latency of the representative user in the class is different, which can be calculated as follows.

$$\begin{aligned} T'_{R-IE} &= \frac{L_{l2}}{(1-q_1)B_1}+\omega_1+\gamma_{LEO}+\frac{L_{AHI}+L_{AHACK}}{(1-q_2)B_2}+2\omega_2 \\ &+ max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1, \frac{L_{BUF\_DATA}}{(1-q_2)B_2}+\omega_2\right\} \\ &+ max\left\{H_{n-CH}, H_{p-CH}\right\}\left[\frac{L_{APBU}+L_{APBA}}{(1-q_2)B_2}+2\omega_2\right]. \end{aligned} \tag{12}$$

The representation of other latency, such as the lower latency for the first user T''_{R-IE}, the two-part latency for other users and the total latency T_{R-IE}, are the same, so there is no need to elaborate further.

2) Off-Orbit Handover: Similarly, HO latency also need to be calculated in three cases. Indicating the upper part of the time axis, $T'_{[\cdot]}$ denotes the time interval between user sending the L2 report and n-AS receiving the buffered data. As the lower part of the time axis, $T''_{[\cdot]}$ is the time interval between user sending the L2 report and receiving the RA message, where $[\cdot]$ stands for different HO types, including intra-cluster HO $[D-IC]$, inter-cluster HO $[D-IE]$, and inter-group HO $[D-IG]$.

a) Intra-cluster handover: Comparing to Fig. 7(a), the lower part in Fig. 7(b) is same, i.e., $T''_{D-IC}=T''_{R-IC}$ and the upper part can be calculated that

$$\begin{aligned} T'_{D-IC} &= \frac{L_{l2}}{(1-q_1)B_1}+\omega_1+H_{p-CH}\left[\frac{L_{Deg\_PBU}}{(1-q_2)B_2}+\omega_2\right] \\ &+ \gamma_{CH}+\max\left(h^x_{p-n}\right)\left[\frac{L_{FMod}}{(1-q_2)B_2}+\omega_2\right] \\ &+ max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1, H_{p-n}\left[\frac{L_{BUF\_DATA}}{(1-q_2)B_2}+\omega_2\right]\right\}, \end{aligned} \tag{13}$$

where γ_{CH} represents the processing delay of CH LEO; $\max\left(h_{p-n}^{x}\right)$ is the maximum hop number between CH LEO and the x-th LEO along the route from p-AS to n-AS. Similarly, there is $T_{preL2}+T_{L2} \geq \frac{L_{l2}}{(1-q_1)B_1}+\omega_1+H_{p-CH}\left[\frac{L_{Deg\_PBU}}{(1-q_2)B_2}+\omega_2\right]+\gamma_{CH}+\max\left(h_{p-n}^{x}\right)\left[\frac{L_{FMod}}{(1-q_2)B_2}+\omega_2\right]+\frac{L_{RD}}{(1-q_1)B_1}+\omega_1$.

b) Inter-cluster handover: The lower part of the time axis satisfies that $T''_{D-IE}=T''_{D-IC}=T''_{R-IC}$ in Fig. 7(c), while for the upper part of the time axis T'_{D-IE}, there is

$$\begin{aligned}T'_{D-IE}&=\frac{L_{l2}}{(1-q_1)B_1}+\omega_1+H_{p-CH}\left[\frac{L_{Deg\_PBU}}{(1-q_2)B_2}+\omega_2\right]\\&+\frac{L_{Deg\_PBU\_C}}{(1-q_3)B_3}+\omega_3+\gamma_{MEO}+\frac{L_{FMod}}{(1-q_3)B_3}+\omega_3\\&+\max\left(h_{p-n}^{x}\right)\left[\frac{L_{FMod}}{(1-q_2)B_2}+\omega_2\right]\\&+max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1,H_{p-n}\left[\frac{L_{BUF\_DATA}}{(1-q_2)B_2}+\omega_2\right]\right\}.\end{aligned}\tag{14}$$

where γ_{MEO} represents the processing delay of MEO; B_3, ω_3 is the bandwidth and propagation delay of LEO-to-MEO cross-layer ISLs respectively. Likewise, T_{L2} satisfies the inequality: $T_{preL2}+T_{L2}>T'_{D-IE}+\frac{L_{RD}}{(1-q_1)B_1}+\omega_1-\max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1,H_{p-n}\left[\frac{L_{BUF_{DATA}}}{(1-q_2)B_2}+\omega_2\right]\right\}$.

c) Inter-group handover: In the same way, we have $T''_{D-IG}=T''_{D-IE}=T''_{D-IC}=T''_{R-IC}$ in Fig. 7 and

$$\begin{aligned}T'_{D-IG}&=\frac{L_{l2}}{(1-q_1)B_1}+\omega_1+H_{p-CH}\left[\frac{L_{Deg\_PBU}}{(1-q_2)B_2}+\omega_2\right]\\&+\frac{L_{Deg\_PBU\_C}}{(1-q_3)B_3}+\omega_3+\gamma_{MEO}+\frac{L_{HReq}+L_{HAck}}{(1-q_4)B_4}+2\omega_4\\&+\frac{L_{FMod}}{(1-q_3)B_3}+\omega_3+\max\left(h_{p-n}^{x}\right)\left[\frac{L_{FMod}}{(1-q_2)B_2}+\omega_2\right]\\&+max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1,H_{p-n}\left[\frac{L_{BUF\_DATA}}{(1-q_2)B_2}+\omega_2\right]\right\}.\end{aligned}\tag{15}$$

where B_4 and ω_4 represents the bandwidth and propagation delay of ISLs in MEO segment respectively. The inequality here is satisfied, i.e., $T_{preL2}+T_{L2}>T'_{D-IG}+\frac{L_{RD}}{(1-q_1)B_1}+\omega_1-max\left\{\frac{L_{RD}}{(1-q_1)B_1}+\omega_1,H_{p-n}\left[\frac{L_{BUF\_DATA}}{(1-q_2)B_2}+\omega_2\right]\right\}$.

In summary, the HO latency of the off-orbit HO is given by Eq.(16).

$$T_{[\cdot]}=\begin{cases}T_{L2}+\dfrac{L_{RS}+L_{RA}}{(1-q_1)B_1}+2\omega_1, T'_{[\cdot]}\leq T''_{[\cdot]}\\ T_{L2}+\dfrac{L_{RS}+L_{RA}}{(1-q_1)B_1}+2\omega_1+T'_{[\cdot]}-T''_{[\cdot]}, other.\end{cases}\tag{16}$$

C. Handover Failure Probability

When the time a user stays within the coverage area of the new access satellite is less than its HO time, the HO for the user is failed. There are several reasons for HO failure, such as poor channel conditions due to signal-to-noise deterioration, HO request rejection due to unavailable resource. Suppose that T_R denotes the user residence time with its probability density function $f_R(t)$. Here T_{HO} is used to represent HO delay uniformly, whose specific form under different cases is shown in the previous section. For convenience in analysis, we assume that the HO delay T_{HO} follows an exponential distribution with its cumulative distribution function $F_{HO}(t)$ and its mean value $E\left[T_{HO}\right]$. Then the handover failure probability [25], [40] can be denoted as

$$\begin{aligned}\mu_f=\Pr\left(T_{HO}>T_R\right)&=\int_0^{\infty}\left(1-F_{HO}(t)\right)\cdot f_R(t)dt\\&=\frac{\mu_r E\left[T_{HO}\right]}{1+\mu_r E\left[T_{HO}\right]},\end{aligned}\tag{17}$$

where μ_r represents the coverage cross rate for the user. Similar to [24] and [41], μ_r can be expressed as follows, considering the circular coverage area of the satellite.

$$\mu_r=\frac{2\boldsymbol{v}}{\pi R_{LEO}},\tag{18}$$

where R_{LEO} is the coverage radius of the LEO satellite. For ease of calculation, $\boldsymbol{v}$ is the vector sum of user's active speed, the Earth's rotation speed and the LEO satellite's orbital speed assuming that LEO satellite is static.

D. Buffered Overhead

It's necessary to buffer data at either the new or old access satellite during the HO process. The required buffering space for one user is directly proportional to the buffered data amount and the HO delay, so the buffered overhead is defined as the product of the data amount buffered at access satellites and the HO delay.

$$C_b=\rho_b\cdot\lambda\cdot {T_{HO}}^2\cdot L_{DATA},\tag{19}$$

where λ is the average packet arrival rate for the user, L_{DATA} represents the size of data packet and ρ_b is the weight factor for buffered cost.

Buffered data need to be transferred between the previous and current access satellites, so the transmission cost of buffered packets is denoted as the product of the buffered data amount and the number of hops between the previous and current access satellites. Here ρ_d is the weight factor of data plane.

$$C_d=\rho_d\cdot\lambda\cdot T_{HO}\cdot L_{DATA}\cdot H_{p-n}.\tag{20}$$

VI. Simulation Results

In this section, we conduct numerical simulations to evaluate the HO performance of the proposed LHP under the HDMMA based on the KPIs developed in the previous section.

For the MEO satellite segment, 12 MEO satellites are used for network management, where the orbit configuration is designed by 4×3 with an altitude of 12,000 km and an inclination of 60°. Taking the SpaceX's first-generation Starlink project as an example, the LEO satellite segment comprises 11,927 LEO satellites to provide ubiquitous network

TABLE II
SIMULATION PARAMETER SETTINGS

Notation	Description	Value
L_{l2}, L_{RD}	The size of L2, RD	52 Bytes
L_{RS}, L_{RA}	The size of RS, RA	52 Bytes
L_{AHI}, L_{AHACK}	The size of AHI, AHACK	72 Bytes
L_{APBU}, L_{APBA}	The size of APBU, APBA	72 Bytes
L_{FMod}	The size of FlowMod	172 Bytes
L_{D-PBU}, $L_{D-PBU-C}$	The size of Dereg PBU, Dereg PBU-C	72 Bytes
L_{HS}, L_{HA}	The size of HS, HA	52 Bytes
q_1, q_2, q_3, q_4	The packet loss rate of links	0.05
B_1, B_4	The bandwidth of GSLs, cross-layer ISLs in control plane	[3, 5] Mbps
B_2, B_3	The bandwidth of ISLs in control plane	[2, 4] Mbps
T_{pre}	The time interval from L2 link down to L2 link trigger	30 ms
γ_{LEO}, γ_{CH}, γ_{MEO}	The processing latency of LEO, CH LEO, MEO satellites	[50, 100] ms
N_{max}	The maximum number of satellites in one cluster	150
L_{DATA}	The size of data packet of users	98 Bytes
λ	The average packet arrival rate of the flow	10 packets/s
v_{user}	The average velocity of ground users	[0, 200] km/h

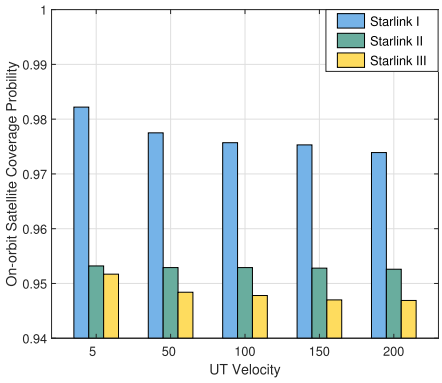


Fig. 8. On-orbit satellite coverage probability.

services globally [42], [43]. The satellite network topology is generated from Satellite Tool Kit (STK). For the ground segment, 232 ground stations are distributed worldwide with real location data and we consider the HO situation that users located in the Harbin, China [$45^o76'N$, $126^o64'E$] may experience when using satellite Internet. The size of various signaling messages involved in the HO procedures are set according to [25], [38], and [41]. Detailed simulation parameters and system settings are listed in Table II.

A. Handover Characteristics in UD-LSNs

Firstly, we validate the handover characteristics in the UD-LSNs through simulation experiments. Here,

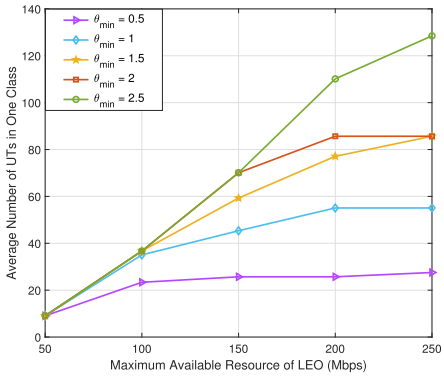


Fig. 9. User aggregation scale.

Starlink I, II and III are employed as access points, with 1584, 4409 and 11927 LEO satellites. The ground user moves within the geographical region of [10^oN, 13^oN], [0^oE, 5^oE], whose traveling pattern obeys the Gaussian Markov Mobility Model. Initially, each user selects the access satellite according to the Shortest Distance Principle, and the positions of users and satellite constellations are updated in each time interval $T = 1s$ based on their respective motion models. The statistical probability of users being covered by on-orbit satellites over a period of $T = 2000s$ is shown in Fig. 8. Representative users are selected for observation with speed of [5, 50, 100, 150, 200] km/h and the coverage probability of on-orbit satellites is all above 0.94. The greater user's active motion intensity with higher speed, the smaller the on-orbit satellite coverage probability.

B. Handover Procedure Evaluation

Subsequently, we evaluate the performance of the UAA. There are 1000 users moving at the speed of [0, 40] km/h, where the users' bandwidth requirements are randomized within [0, 3]$Mbps$ with the maximum available bandwidth of the satellite $R_{max} = 120Mbps$ and minimum correlation threshold $\theta_{min} = 0.8$. Fig. 9 shows the average user aggregation scale under different available bandwidth and distance thresholds. The average user number in one class is constrained by both the available bandwidth R_{max} and the distance threshold θ_{max}. As one of them grows, the average user aggregation size increases, but is still constrained by the other.

We then evaluate the HO performance of the LHP. Fig. 10 illustrates the signaling overhead of on-orbit HO under different numbers of HO users. Here, Hop represents the number of hops between the AS and the CH LEO satellite. Under the same user amount and hops, user aggregation HO can significantly reduce signaling overhead. This is because the representative user completes the HO confirmation and binding update process for all the others in the class only by using the original workflow. Particularly, there are signaling interactions between AS and CH LEO during the binding update, which usually involves message delivery among multiple hops and occupies a major portion of the signaling overhead. Through user aggregation HO, users within the class can share the signaling overhead together, resulting in less impact from the

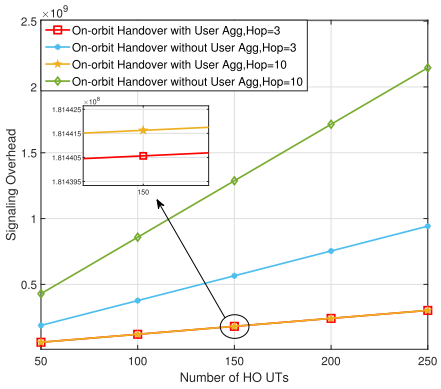


Fig. 10. Signaling overhead vs. users.

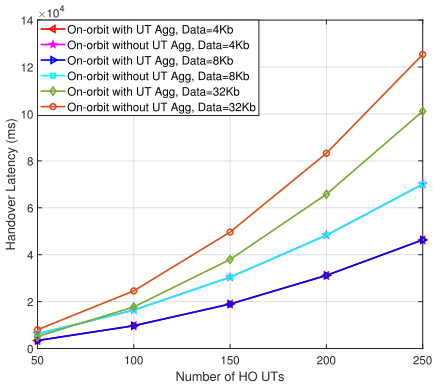


Fig. 12. Handover latency vs. users.

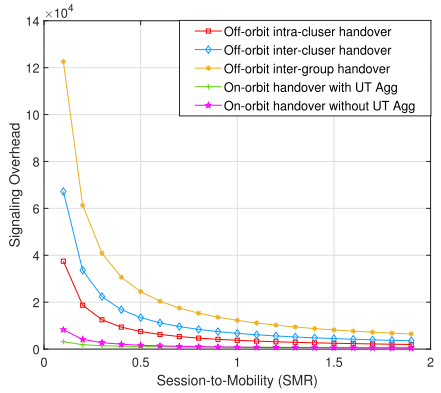


Fig. 11. Signaling overhead vs. SMRs.

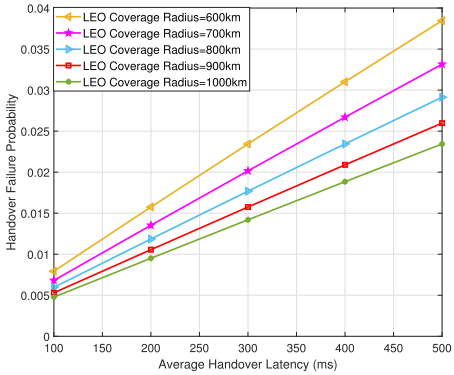


Fig. 13. On-orbit HO failure probability.

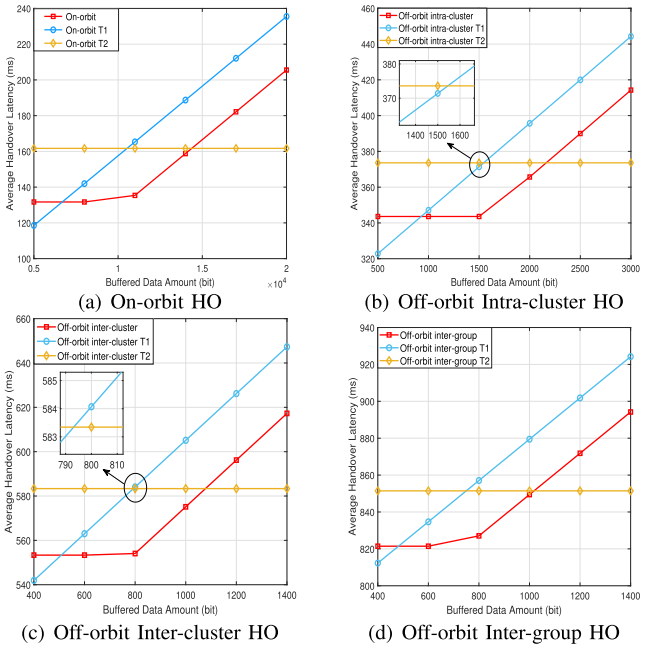
number of users and hops. Furthermore, it can be seen that the more HO users within the same class, the greater the signaling overhead saved by adopting user aggregation HO.

Fig. 11 illustrates the variation of signaling overhead under different Session Mobility Ratios (SMRs) for on-orbit HO with and without user aggregation, as well as for off-orbit HO under three scenarios. SMR represents the ratio of session rate to network mobility. The smaller SMR indicates higher satellite mobility, corresponding to increased frequency of HOs, which leads to larger signaling overhead in all cases. Under the same p-AS and unit user SMR, the signaling overhead of on-orbit HO is much less than off-orbit HO. In on-orbit HO, the signaling overhead with user aggregation HO is less than non-aggregation HO. In off-orbit HO, signaling overhead for intra-cluster HO is smaller than inter-cluster HO, which is less than inter-group HO. It is because when the difference between the p-AS and n-AS is larger, the corresponding signaling interactions and transmission distances increase, resulting in larger signaling overhead.

Fig. 12 shows how HO delay varies with the number of HO users in the on-orbit HO. As the number of HO users grows, the total HO latency also increases, but not in direct proportion. This is because the link bandwidth resources allocated to each user decrease with more users, leading to an increase in average HO latency. Although more signaling overhead can be saved by increasing the number of HO users, as shown in Fig. 10, the average HO latency for each user will increase. Thus, it may not be wise to blindly increase the number of HO users, and a tradeoff needs to be made between the HO delay and signaling overhead. Clearly, aggregate HO can bring improvements in HO latency. Under the smaller buffered data amount, for example, 4 Kb and 8 Kb, different data amounts may result in the same HO latency. However, when the buffered data amount exceeds the critical point, such as 8 Kb and 32 Kb, the larger the data amount, the longer the HO latency, the reason for which will be analyzed in Fig. 14.

When a user has moved out of the coverage area of the satellite to be accessed within the HO latency, the handover fails. Fig. 13 shows the relationship between the HO failure probability and the average HO latency when on-orbit HO occurs, where stationary users at the equator of the Earth's surface are selected as the research object. It can be observed that the larger the HO latency, the greater the probability of HO failure. At the same time, the larger the satellite's coverage radius, the smaller the probability of HO failure.

Fig. 14 displays the variation of HO latency of a single user with the size of buffered data under different HO scenarios. Here T1 represents the upper part in the time axis, including HO preparation and buffered data delivery, thus being proportional to the buffered data amount. On the other hand, T2 indicates the lower part, which contains the HO execution and is independent of buffered data amount. The total HO latency is determined by the larger of the two. When buffered data amount is small, T1 is relatively short and the total HO latency mainly depends on T2. As the buffered data



(a) On-orbit HO (b) Off-orbit Intra-cluster HO

(c) Off-orbit Inter-cluster HO (d) Off-orbit Inter-group HO

Fig. 14. Handover latency vs. different buffered data.

amount increases, T1 gradually begins to dominate the total HO latency. As a result, for all the HO scenarios, the total HO latency remains constant initially and then increases with the grow of buffered data amount. We define the buffered data amount at which the total HO latency starts to be influenced as the critical data amount value. The critical point is the maximum amount of buffered data that has the minimum impact on HO latency. By comparison, the critical data amount value for on-orbit HO is significantly larger than that for off-orbit HOs.

C. Management Architecture Evaluation

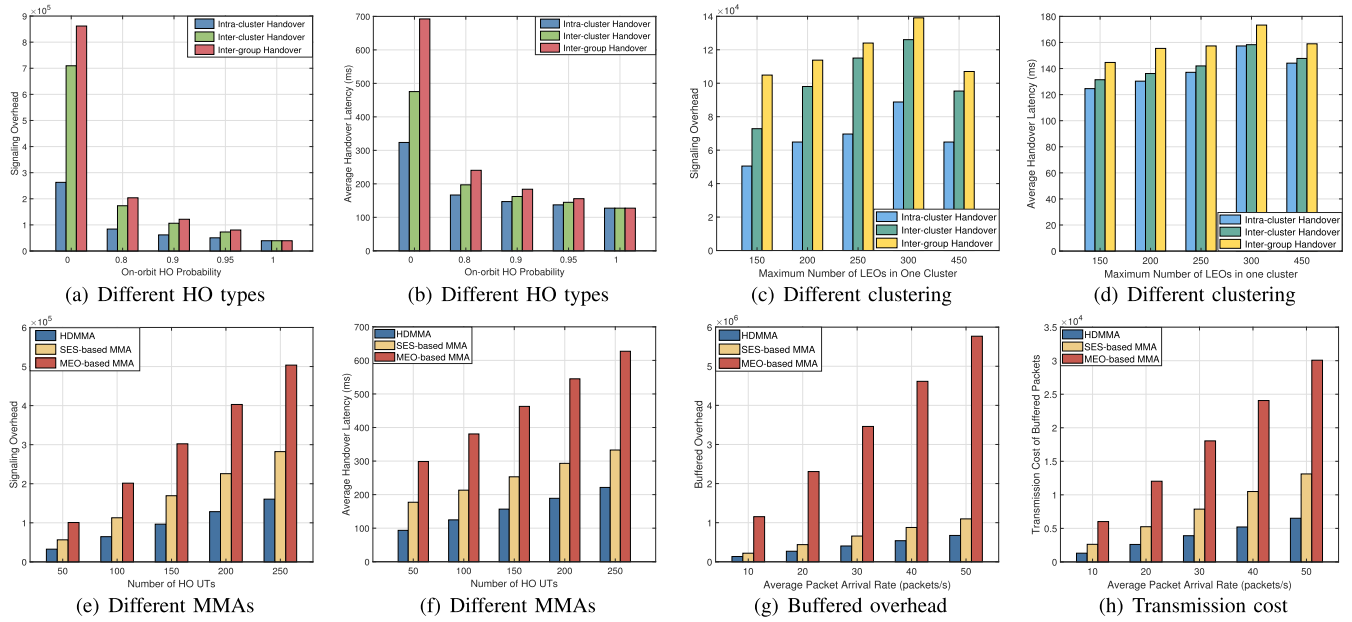
According to Fig. 11, the signaling cost of on-orbit HO are much smaller than off-orbit HO. As can be seen from Fig. 8, users always remain within the coverage range of satellites on the same orbital plane. To minimize the handover cost, the proposed architecture takes the on-orbit coverage probability p as the proportion of on-orbit HO. In the off-orbit HO, the probability of intra-cluster HO, inter-cluster HO and inter-group HO depends on users' demands and speeds, which is set to the same.

Fig. 15(a) and Fig. 15(b) show the variation of signaling overhead and HO latency with respect to the on-orbit HO probability. There are 150 users, and the probability of intra-cluster, inter-cluster and inter-group HO accounts for all the off-orbit HO probability respectively. In this case, the signaling overhead for inter-group HO is the highest, followed by inter-cluster HO, and intra-cluster HO has the smallest signaling overhead. On-orbit HO is more likely to occur with higher p, which can greatly reduce the signaling overhead. When p equals 0, there is only off-orbit HO, corresponding to the maximum overhead. On the other hand, $p = 1$ indicates all HOs are on-orbit with the least overhead as demonstrated in Fig. 11, and the signaling overhead for the three types HOs tends to be the same. The same trend applies to HO latency.

Fig. 15(c) and Fig. 15(d) depict the signaling overhead and handover delay under different satellite clustering modes respectively. Here, the on-orbit HO probability is set to $p = 0.95$, and the number of HO users is 150. The front-to-back satellites are the same for on-orbit HO, intra-cluster HO and inter-group HO. Due to the different clustering modes, the target HO satellites in inter-cluster HO will be different. It can be seen that even for the same HO type with the same previous and target satellites, the signaling overhead under different clustering modes is different. It is because the clustering mode affects the selection of CH LEOs and the network topology and routing within the cluster will change. When the maximum number of satellites in one cluster increases, the control signaling interaction may experience more hops, leading to larger overhead. However, signaling transmission has more chance to select fewer hops due to the larger cluster scale and more complete network topology, so the signaling overhead will be reduced under $N_{max} = 450$. The analysis is the same for HO latency.

The signaling overhead and HO latency under different mobility management architectures are shown in Fig. 15(c) and Fig. 15(d). In SES-based MMA, the SES serves as the controller for the UD-LSNs, responsible for handling all types of mobility logics for all satellites within its management scope, such as HO decisions and binding updates, while MMFs are located on MEO satellites in the MEO-based MMA, with each MEO acting as a controller to make mobility strategies for LEO satellites within its coverage area. The HO procedure in the both MMA adopt the SDN-MM [38]. It can be seen that our HDMMA exhibits the lowest signaling overhead. Meanwhile, the SES-based MMA incurs higher overhead since all control messages need to be transmitted to the fixed BSs on the ground for processing. In the MEO-based MMA, all HO decisions and signaling processing are handled by the MEOs. However, considering the limited number of satellite antennas, only CH LEO within each cluster can establish cross-layer ISLs with the MEO, which inevitably leads to more redundant and complex message delivery and the highest signaling overhead. The analysis on HO latency is the same.

Fig. 15(g) shows how the buffered overhead for a single user varies with user's average packet arrival rate. As packets arrive faster and HO latency becomes longer, the buffered overhead increases rapidly. Comparatively, our HDMMA has the smallest buffered overhead among the three MMAs due to the fewer signaling interactions and lightweight HO procedure. Moreover, as the HO delay increases, the buffered data amount also increases for the same data arrival rate, which cause the performance difference between the MMAs is greater here than that in terms of HO delay. Fig. 15(h) shows the transmission cost of buffered packets per user varies with the average packet arrival rate. The results here are proportional to HO delay because the number of hops between the previous and subsequent ASs is kept consistent across different schemes. Therefore, the buffered data transmission cost here is only related to the HO delay under the same packet arrival rate.



(a) Different HO types (b) Different HO types (c) Different clustering (d) Different clustering

(e) Different MMAs (f) Different MMAs (g) Buffered overhead (h) Transmission cost

Fig. 15. HO performance comparison.

VII. Conclusion

In this paper, considering the unique challenges and opportunities posed by future UD-LSNs, we have designed a hierarchical distributed mobility management architecture HDMMA, supporting MEO and LEO satellites respectively responsible for certain management functions. Under the HDMMA, handover is classified into on-orbit HO and off-orbit HO, capturing the satellite highly dynamic patterns. For on-orbit HO, taking into account the attributes of users and the available resources of satellites, users in close proximity with the same target satellite are aggregated to share the processing cost of massive simultaneous HO. For off-orbit HO, we have differentiated between intra-cluster HO, inter-cluster HO, and inter-group HO, depending on the regions that p-AS and n-AS belong to. Subsequently, lightweight handover procedures LHPs are designed separately for the four HO scenarios. We have conducted theoretical analysis and simulation experiments on the proposed LHP. Simulation results demonstrate the HO characteristics of UD-LSNs, providing insights for HO decision-making. Compared to BS-based and MEO-based management architecture, HDMMA significantly reduces signaling overhead and HO latency, which sheds a light on the superiority of MEO-assisted mobility management for UD-LSNs.

References

[1] X. Qin, T. Ma, X. Zhang, Y. Wang, H. Zhou, and L. Zhao, "A lightweight hierarchical mobility management architecture for ultra-dense LEO satellite network," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 679–684.

[2] J. Xue, K. Yu, T. Zhang, H. Zhou, L. Zhao, and X. Shen, "Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 8, pp. 8143–8157, Aug. 2024.

[3] T. Zhang et al., "Towards handover-free mobility management in FD-RAN: Architecture, challenges, and solutions," *IEEE Netw.*, 2024.

[4] N.-N. Dao et al., "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1193–1225, 2nd Quart., 2021.

[5] *SpaceX Non-Geostationary Satellite System*, Washington, DC, USA, 2018. [Online]. Available: https://www.fcc.report/IBFS/SAT-MOD-20181108-00083/1569860.pdf

[6] C. Zou, H. Wang, J. Chang, F. Shao, L. Shang, and G. Li, "Optimal progressive pitch for OneWeb constellation with seamless coverage," *Sensors*, vol. 22, no. 16, p. 6302, Aug. 2022.

[7] G. Jansson, "Telesat LightspeedT-enabling mesh network solutions for managed data service flexibility across the globe," in *Proc. IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)*, Mar. 2022, pp. 232–235.

[8] X. Qin, T. Ma, Z. Tang, X. Zhang, H. Zhou, and L. Zhao, "Service-aware resource orchestration in ultra-dense LEO satellite-terrestrial integrated 6G: A service function chain approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6003–6017, 2023.

[9] X. Zhang et al., "Cybertwin-assisted mode selection in ultra-dense LEO integrated satellite-terrestrial network," *J. Commun. Inf. Netw.*, vol. 7, no. 4, pp. 360–374, Dec. 2022.

[10] Z. Tang, K. Yu, G. Yang, L. X. Cai, and H. Zhou, "New bridge to cloud: An ultra-dense LEO assisted green computation offloading approach," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 552–564, 2022.

[11] *Study on Architecture Aspects for Using Satellite Access in 5G*, document 3GPP TR23.737 Version 17.2.0, 2021.

[12] *Network Coding for Satellite Systems*, document IETF RFC 8975, 2021.

[13] G. Su, P. You, and S. Yong, "Comparative handover performance analysis of MIPv6 and FMIPv6 in LEO satellite networks," in *Proc. Int. Conf. Netw. Inf. Syst. Comput. (ICNISC)*, Apr. 2017, pp. 30–36.

[14] W. Dai, H. Li, Q. Wu, and X. Wang, "Flexible and aggregated mobility management in integrated satellite-terrestrial networks," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 982–987.

[15] W. Dai, H. Li, Q. Wu, and X. Wang, "NDM: Network driving IP mobility support in large scale LEO satellite network," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2020, pp. 1–7.

[16] X. Zhang, K. Shi, S. Zhang, D. Li, and R. Xia, "Virtual agent clustering based mobility management over the satellite networks," *IEEE Access*, vol. 7, pp. 89544–89555, 2019.

[17] J. Hu, T. Pan, Y. Chen, X. Zhang, T. Huang, and Y. Liu, "LISP-LEO: Location/identity separation-based mobility management for LEO satellite networks," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2022, pp. 1558–1563.

[18] Z. Zhang, B. Zhao, W. Yu, and C. Wu, "Supporting location/identity separation in mobility-enhanced satellite networks by virtual attachment point," *Pervasive Mob. Comput.*, vol. 42, pp. 1–14, Sep. 2017.

[19] D. K. Luong, Y.-F. Hu, J.-P. Li, M. Ali, F. Benamrane, and K. Abdo, "Seamless handover in SDN-based future avionics networks with network coding and LISP mobility protocol," in *Proc. IEEE/AIAA 38th Digit. Avionics Syst. Conf. (DASC)*, Sep. 2019, pp. 1–7.

[20] T. Ma, B. Qian, X. Qin, X. Liu, H. Zhou, and L. Zhao, "Satellite-terrestrial integrated 6G: An ultra-dense LEO networking management architecture," *IEEE Wireless Commun.*, vol. 31, no. 1, pp. 62–69, 2022.

[21] Y. Wang, X. Qin, Z. Tang, T. Ma, X. Zhang, and H. Zhou, "QoS-centric handover for civil aviation aircraft access in ultra-dense LEO satellite networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2022, pp. 1085–1089.

[22] H. Zhou, J. Li, K. Yang, H. Zhou, J. An, and Z. Han, "Handover analysis in ultra-dense LEO satellite networks with beamforming methods," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3676–3690, Mar. 2023.

[23] F. Wang, D. Jiang, Z. Wang, J. Chen, and T. Q. S. Quek, "Seamless handover in LEO based non-terrestrial networks: Service continuity and optimization," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1008–1023, Feb. 2023.

[24] C. Makaya and S. Pierre, "An analytical framework for performance evaluation of IPv6-based mobility management protocols," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 972–983, Mar. 2008.

[25] J.-H. Lee, J.-M. Bonnin, I. You, and T.-M. Chung, "Comparative handover performance analysis of IPv6 mobility management protocols," *IEEE Trans. Ind. Electron.*, vol. 60, no. 3, pp. 1077–1088, Mar. 2013.

[26] T. Darwish, G. K. Kurt, H. Yanikomeroglu, G. Lamontagne, and M. Bellemare, "Location management in internet protocol-based future leo satellite networks: A review," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1035–1062, 2022.

[27] P. Xie, Q. Wang, and J. Chen, "A survey of mobility management for mobile networks supporting LEO satellite access," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, Aug. 2022, pp. 59–64.

[28] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "The locator/ID separation protocol (LISP)," 2013.

[29] E. M. O. Fafolahan and S. Pierre, "A seamless mobility management protocol in 5G locator identificator split dense small cells," *IEEE Trans. Mobile Comput.*, vol. 19, no. 8, pp. 1745–1759, Aug. 2020.

[30] H. Ko, I. Jang, J. Lee, S. Pack, and G. Lee, "SDN-based distributed mobility management for 5G," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2017, pp. 116–117.

[31] J. C. Zuniga, C. J. Bernardos, A. de la Oliva, T. Melia, R. Costa, and A. Reznik, "Distributed mobility management: A standards landscape," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 80–87, Mar. 2013.

[32] P. Du, J. Li, W. Bai, M. Sheng, and D. Zhou, "Dual location area based distributed location management for hybrid LEO/MEO mega satellite networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2307–2321, Feb. 2023.

[33] S. Ji, M. Sheng, D. Zhou, W. Bai, Q. Cao, and J. Li, "Flexible and distributed mobility management for integrated terrestrial-satellite networks: Challenges, architectures, and approaches," *IEEE Netw.*, vol. 35, no. 4, pp. 73–81, Jul. 2021.

[34] S. Ji, D. Zhou, M. Sheng, J. Li, and Z. Han, "Dynamic space-ground integrated mobility management strategy for mega LEO satellite constellations," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11043–11060, Sep. 2024.

[35] Q. Kong, X. Qu, F. Yin, R. Lu, S. Zhang, and M. Ma, "Achieving privacy-preserving location management in LEO-satellite integrated vehicular network with dense ground stations," *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 5616–5629, Apr. 2024.

[36] Y. Liu, L. Wang, Z. Lu, K. Du, and G. Shou, "A stateless design of satellite–terrestrial integrated core network and its deployment strategy," *IEEE Trans. Netw. Service Manage.*, vol. 21, no. 1, pp. 953–966, Feb. 2024.

[37] H. He, C. Qin, S. Chen, X. Jiang, J. Yang, and L. Hanzo, "Sparse bandit learning based location management for space-ground integrated networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10314–10329, 2023.

[38] Y. Bi, G. Han, C. Lin, M. Guizani, and X. Wang, "Mobility management for intro/inter domain handover in software-defined networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1739–1754, Aug. 2019.

[39] L. Yang, X. Yang, and Z. Bu, "A group handover strategy for massive user terminals in LEO satellite networks," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2022, pp. 1–6.

[40] H. Zhou, H. Zhou, J. Li, K. Yang, J. An, and X. Shen, "Heterogeneous ultradense networks with traffic hotspots: A unified handover analysis," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 8825–8838, May 2023.

[41] A. Ahmed, S. Jabbar, M. M. Iqbal, M. Ibrar, A. Erbad, and H. Song, "An efficient hierarchical mobile IPv6 group-based BU scheme for mobile nodes in IoT network," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 8684–8695, May 2023.

[42] X. Liu, T. Ma, Z. Tang, X. Qin, H. Zhou, and X. S. Shen, "UltraStar: A lightweight simulator of ultra-dense LEO satellite constellation networking for 6G," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 632–645, Mar. 2023.

[43] N. Cheng et al., "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 178–185, Feb. 2020.

Xiaohan Qin (Student Member, IEEE) received the B.S. degree in communication engineering from Central South University, Changsha, China, in 2020. She is currently pursuing the Ph.D. degree in communications and information systems with Nanjing University, Nanjing, China. Her research interests include space-air-ground integrated networks, network resource management, and game theory.

Ting Ma (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in statistics from Sichuan University, Chengdu, China, in 2013, 2016, and 2020, respectively. She is currently a Post-Doctoral Fellow with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. Her current research interests include robust hypothesis testing, space-air-ground integrated networks, convex optimization theory, and game theory.

Xin Zhang (Graduate Student Member, IEEE) received the B.S. degree in mathematics and physics basic science from the University of Electronic Science and Technology of China, Chengdu, China, in 2020. She is currently pursuing the Ph.D. degree in communications and information systems with Nanjing University, Nanjing, China. Her research interests include space-air-ground integrated networks, resource allocation, and convex optimization theory.

Yilei Wang (Graduate Student Member, IEEE) received the B.S. degree in electronic information science and technology from Central South University, Changsha, China, in 2021. She is currently pursuing the Ph.D. degree in communications and information systems with Nanjing University, Nanjing, China. Her current research interests include space-air-ground integrated networks, airborne internet, handover management, and resource allocation.

Haibo Zhou (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a Post-Doctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo. He is currently a Full Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks. He was a recipient of the 2019 IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award. He served as the Track/Symposium Co-Chair for IEEE/CIC ICCC 2019, IEEE VTC-Fall 2020, IEEE VTC-Fall 2021, and IEEE GLOBECOM 2022. He is currently an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, *IEEE Network Magazine*, and IEEE WIRELESS COMMUNICATIONS LETTER.

Lian Zhao (Fellow, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering (ELCE), University of Waterloo, Canada, in 2002.

She joined the Department of Electrical and Computer Engineering, Toronto Metroplitan University (formerly Ryerson University), Canada, in 2003. Her research interests include wireless communications, resource management, mobile edge computing, caching and communications, and the IoV networks.

Dr. Zhao has been a Board of Governors (BoG) Committee Member since 2023. She has served as a Panel Expert in various federal, provincial, and international evaluation committees. She is a Licensed Professional Engineer in the Province of Ontario and a Senior Member of the IEEE Communication Society and the Vehicular Technology Society. She has been an IEEE Communication Society (ComSoc) and IEEE Vehicular Technology (VTS) Distinguished Lecturer (DL). She received the Best Land Transportation Paper Award from the IEEE Vehicular Technology Society in 2016, the Top 15 Editor Award in 2016 for IEEE TRANSACTION ON VEHICULAR TECHNOLOGY, the Best Paper Award from the 2013 International Conference on Wireless Communications and Signal Processing (WCSP), and Canada Foundation for Innovation (CFI) New Opportunity Research Award in 2005. She served as the Co-Chair of the Wireless Communication Symposium for IEEE Globecom 2020 and IEEE ICC 2018, the Finance Co-Chair for 2021 ICASSP, the Local Arrangement Co-Chair for IEEE VTC Fall 2017 and IEEE Infocom 2014, and the Co-Chair of Communication Theory Symposium for IEEE Globecom 2013. From 2013 to 2021, she served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.