

Stochastic Geometry-Based Low Latency Routing in Massive LEO Satellite Networks

RUIBO WANG 

King Abdullah University of Science and Technology (KAUST), CEMSE
Division, Thuwal, Saudi Arabia

MUSTAFA A. KISHK , Member, IEEE

National University of Ireland, Maynooth, Ireland

MOHAMED-SLIM ALOUINI , Fellow, IEEE

King Abdullah University of Science and Technology (KAUST), CEMSE
Division, Thuwal, Saudi Arabia

In this article, the routing in massive low earth orbit (LEO) satellite networks is studied. When the satellite-to-satellite communication distance is limited, we choose different relay satellites to minimize the latency in a constellation at a constant altitude. First, the global optimum solution is obtained in the ideal scenario when there are available satellites at all the ideal locations. Next, we propose a nearest neighbor search algorithm for realistic (nonideal) scenarios with a limited number of satellites. The proposed algorithm can approach the global optimum solution under an ideal scenario through a finite number of iterations and a tiny range of searches. Compared with other routing strategies, the proposed algorithm shows significant advantages in terms of latency. Furthermore, we provide two approximation techniques that can give tight lower and upper bounds for the latency of the proposed algorithm, respectively. Finally, the relationships between latency and constellation height, satellites' number, and communication distance are investigated.

Manuscript received 18 November 2021; revised 5 March 2022 and 4 August 2022; accepted 11 August 2022. Date of publication 18 August 2022; date of current version 11 October 2022.

DOI No. 10.1109/TAES.2022.3199682

Refereeing of this contribution was handled by K.-M Cheung.

This work was supported by KAUST Office of Sponsored Research.

Authors' addresses: R. Wang and M.-S. Alouini are with the King Abdullah University of Science and Technology (KAUST), CEMSE Division, Thuwal 23955-6900, Saudi Arabia, E-mail: (ruibo.wang@kaust.edu.sa; slim.alouini@kaust.edu.sa); M. A. Kishk is with the Department of Electronic Engineering, National University of Ireland, Maynooth W23 F2H6, Ireland, Email: (mustafa.kishk@kaust.edu.sa). (*Corresponding author: Mustafa A. Kishk.*)

0018-9251 © 2022 IEEE

I. INTRODUCTION

In recent years, we have witnessed the booming development of low earth orbit (LEO) satellite networks. Companies such as SpaceX, Amazon, and OneWeb are accelerating the formation of a network of tens of thousands of LEO satellites [1]. Since LEO satellite communication has relatively low latency and unique ability to provide seamless global coverage [2], [3], part of real-time communication services are being delivered from ground to space [4]. In terms of low latency and ultra-long distance communication, the LEO satellite network has excellent advantages over ground networks and high orbit satellite networks [5]. In ultra-long distance communication, multiple satellites are used as relays to complete multihop routing. How to select the relay satellite to achieve the minimum latency routing becomes one of the challenges [6], [7].

Different from the traditional planar routing, satellites are distributed on a closed sphere, and the maximum distance between satellites is limited due to earth blockage [8]. For a network, where the number and location of satellites are constantly changing, it is more challenging to implement routing in the time-varying topology than in the traditional static topology [9]. For small LEO satellites, both computing and storage capacity are limited [10]. In a massive LEO satellite network, frequent position changes lead to high computational costs. In addition, each satellite collects only the current state of its neighbors in most cases, which means that it is highly demanding for a single satellite to obtain and store global information such as the location of the satellite. However, using only local information can only get the approximate shortest path, which has limited improvement on the whole constellation latency performance [11].

Existing routing schemes provide strategies to address some of the challenges, but they are not suitable for dynamic large-scale satellite constellations. Stochastic geometry provides a powerful mathematical method for routing in massive constellations. The coverage probability of LEO satellite constellation and two-dimensional (2-D) plane routing have been studied based on stochastic geometry. Based on these studies, we propose an algorithm to solve the routing problem of a dynamic constellation. At the end of this section, the contributions of this article are described in more detail.

A. Related Work

Most of the existing LEO satellite routing is based on the store-and-forward mechanism [12], [13], which undoubtedly brings considerable delay. The following algorithms can achieve real-time communication in specific scenarios [11], [14], [15]. In [14], medium orbit satellites and high orbit satellites are used to collect and exchange global information to find a route with minimum latency for low-orbit satellites. However, due to the increased complexity of the algorithm, this method is only suitable for small-scale networks but not a massive dynamic network. In [15], the latency is effectively reduced according to the regular

motion of the satellite, and there is no need to collect global information and pay the great computational cost. However, the algorithm is only suitable for a specific small network composed of eight satellites, and the algorithm cannot optimize link latency. Compared to [15], the algorithm in [11] is local optimum and scalable. By dividing the sphere into many grids, the satellite is positioned by the grid. However, the algorithm reaches the square complexity and is only suitable for static topology. In addition, from the global point of view, it is difficult to guarantee the lower bound of the algorithm.

For massive dynamic satellite networks, the main reason the existing routing algorithms can not combine low complexity and global optimization is that they are designed for each satellite's specific constellations and specific behavior. As an effective mathematical tool, stochastic geometry is especially suitable for analyzing network topology from system-level [16]. So far, many methods have been developed to analyze LEO satellite systems based on stochastic geometry. Binomial point process (BPP) is used to model a closed-area network with a finite number of satellites in [17] and [18]. [8], [18], and [19] give different forms of contact distance distribution, respectively, that is, the distribution of the distance between a reference point and the nearest satellite. Contact distance distribution provides an important theoretical basis for the analysis of this article.

In addition, there are several of 2-D planar routing strategies based on stochastic geometry [20], [21], [22]. Among them, [23] and [24] provide the concept of a reliable region, which ensures the routing can always follow the established direction. The concept of routing efficiency is used to measure the maximum gap between the proposed routing strategy and the optimal one [24]. By sacrificing the optimality of the algorithm, a suboptimal routing strategy is given on the premise that only local information is available [25]. According to this idea, the optimal routing is derived in an ideal scenario. Then, the suboptimal routing strategy is proposed when only local information is available.

B. Contribution

So far, this is the first study of satellite routing based on stochastic geometry. The contributions can be summarized as follows:

- 1) Three propositions are given in the ideal scenario, where there are available satellites at any location. Based on these propositions, we provide a solution for the ideal scenario and use it as an upper bound for the proposed algorithm.
- 2) Equal interval, minimum deflection angle, and maximum step size relay strategies are derivatives of propositions in the ideal scenario. We obtain the proposed algorithm by improving the equal-interval relay strategy. The remaining two are used as the baselines.
- 3) We provide two approximations to estimate the gap between the algorithm and the best possible solution.

Numerical results show that these two approximations can give tight upper and lower bounds for the algorithm delay.

- 4) According to three LEO satellite constellations, algorithm complexity, average, and maximum search area required for finding at least one satellite are analyzed.
- 5) We study the influence of parameters such as communication distance, constellation height, and the number of satellites on latency.

II. OPTIMAL ROUTING SCHEME

Let us consider a scenario, where two satellites are too far apart to communicate directly. Several satellites act as relays to complete multihop satellite to satellite link communication. The notations and corresponding descriptions sections are shown in Table I.

A. Problem Formulation

To formalize the problem, this section introduces 1) satellite distribution, 2) link routing model, 3) coordinate system, and 4) optimization problem in order.

Consider a massive constellation composed of N_{Sat} satellites, which are independently distributed on a spherical surface according to a homogeneous BPP [18]. The radius of the sphere is denoted as $r = r_{\oplus} + r_{\text{Sat}}$, where $r_{\oplus} = 6371$ Km is the radius of the Earth, and r_{Sat} is the height of the satellite orbits.

The latency required for transmission is often measured in milliseconds, which is much smaller than the orbital period of LEO satellites. The change of satellite position with time in single routing is negligible. A transmission from one satellite to another is called a hop. A link with n hops can be expressed as $\mathcal{H} = \{h_0, h_1, \dots, h_n\}$. h_i is the ID of the i th satellite, which is a positive integer less than N_{Sat} . x_{h_0} and x_{h_n} are the positions of the starting point and the ending point, respectively.

Since the distribution of satellites forms a homogeneous BPP, the rotation of the coordinate system do not affect the distribution. Set the center of the Earth as the origin. All satellites have the same radial distance r . We establish the coordinate system by the coordinates of the starting satellite x_{h_0} and the ending satellite x_{h_n} of the multihop link. As is shown in Fig. 1, the x -axis is parallel to the line segment between x_{h_0} and x_{h_n} , and the z -axis is the midperpendicular of this segment, so the y -coordinates of x_{h_0} and x_{h_n} are 0. Since satellites are distributed on a sphere, spherical coordinates are more practical than rectangular coordinates. Coordinate (r, θ_i, φ_i) is used to represent the location of i th satellite x_i . θ_i and φ_i are the polar and azimuth angles, respectively. Furthermore, the homogeneous BPP is denoted as $\Phi = \{x_1, x_2, \dots, x_{N_{\text{Sat}}}\}$. d_i is used to describe the distance of the i th hop, that is, the spatial distance from $x_{h_{i-1}}$ to x_{h_i}

$$d_i = r \left[2 \left(1 - \cos \theta_{h_{i-1}} \cos \theta_{h_i} - \sin \theta_{h_{i-1}} \sin \theta_{h_i} \cos (\varphi_{h_{i-1}} - \varphi_{h_i}) \right) \right]^{\frac{1}{2}} \quad (1)$$

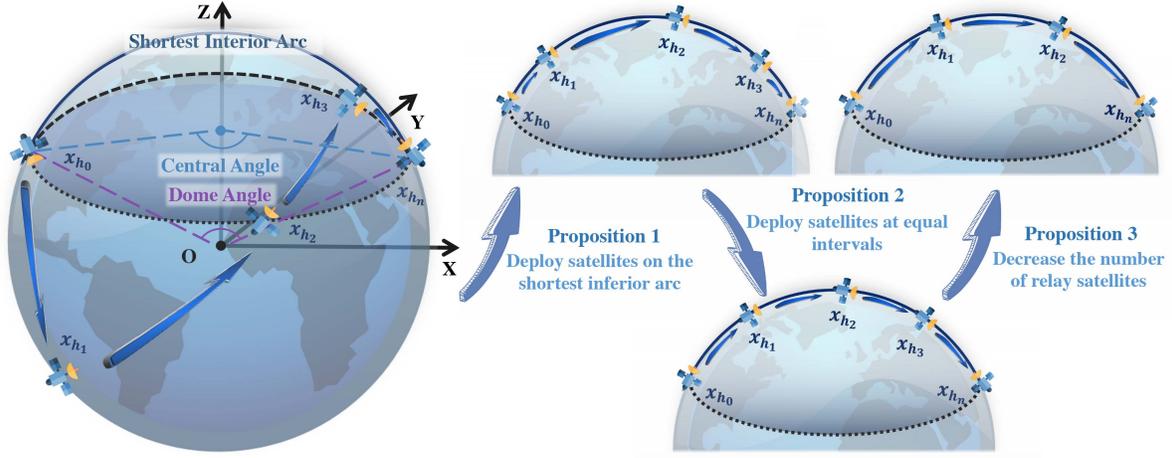


Fig. 1. Explanatory figure of the three propositions.

where $i = 1, 2, \dots, n$.

To minimize the latency by selecting the number of satellites and their positions, we consider the following optimization problem:

$$\mathcal{P}_0 : \underset{n, \mathcal{H}}{\text{minimize}} \quad T = \frac{1}{c} \sum_{i=1}^n d_i \quad (2a)$$

$$\text{subject to:} \quad d_i \leq 2\sqrt{r^2 - r_{\oplus}^2}, \quad \forall i \quad (2b)$$

$$d_i \leq d_{\max}, \quad \forall i. \quad (2c)$$

In (2a), the optimization objective is the latency of the multihop link, where $c = 3 \times 10^2$ Km/ms is the speed of laser propagation. Constraint (2b) guarantees that the satellites are within line-of-sight of each other [18], and constraint (2c) limits the maximum communication distance d_{\max} between satellites. Note that we omit the power constraint issues in \mathcal{P}_0 . Since the objective function is related to the position and number of satellites, the problem is not convex.

B. Ideal Scenario Solution

To make the problem more manageable, we start with an ideal scenario, which assumes satellites are available anywhere on the sphere. Before solving the optimization problem \mathcal{P}_0 , the following definitions are required.

DEFINITION 1 (CENTRAL ANGLE): For a circle passing satellites A and B, the central angle of the circle is the angle between the line connecting A and the center of the circle and the line connecting B and the center of the circle.

DEFINITION 2 (DOME ANGLE): For a circle centered at the origin, passing satellites A and B, the central angle for this specific circle is called the dome angle.

DEFINITION 3 (SHORTEST INFERIOR ARC): The circle centered at the origin, with radius r , passing the starting point x_{h_0} and the ending point x_{h_n} , are divided into two arcs by

x_{h_0} and x_{h_n} . The arc with a shorter arc length is called the shortest inferior arc.

An ideal solution of problem \mathcal{P}_0 is derived through the following three propositions.

PROPOSITION 1: In the ideal scenario, optimal positions $x_{h_i}^*$ in \mathcal{P}_0 are located on the shortest inferior arc.

PROOF: See Appendix A. ■

Based on proposition 1, all satellites are assumed to locate on the shortest inferior arc. Therefore, an equivalence problem for \mathcal{P}_0 is given by

$$\mathcal{P}_1 : \underset{n, \mathcal{H}}{\text{minimize}} \quad T = \frac{1}{c} \sum_{i=1}^n 2r \sin\left(\frac{\theta_i^h}{2}\right) \quad (3a)$$

$$\text{subject to:} \quad \theta_i^h \leq 2 \arccos\left(\frac{r_{\oplus}}{r}\right), \quad \forall i \quad (3b)$$

$$\theta_i^h \leq 2 \arcsin\left(\frac{d_{\max}}{r}\right), \quad \forall i \quad (3c)$$

$$\sum_{i=1}^n \theta_i^h = \theta_{02n}^h \quad (3d)$$

where θ_i^h is the dome angle between satellites $x_{h_{i-1}}$ and x_{h_i} . As is shown in Fig. 2 θ_{02n}^h is the dome angle between starting satellite x_{h_0} and ending satellite x_{h_n} , which is given as

$$\theta_{02n}^h = \arcsin\left(\frac{\sqrt{2}}{2} \left(1 - \cos \theta_{h_0} \cos \theta_{h_n} - \sin \theta_{h_0} \sin \theta_{h_n} \cos(\varphi_{h_0} - \varphi_{h_n})\right)^{\frac{1}{2}}\right). \quad (4)$$

θ_{02n}^h is also defined as the dome angle of the multihop link. It can be derived intuitively by the formula (1) with the aid of simple geometric relations. The following proposition will further give a more specific distribution of relay satellite positions.

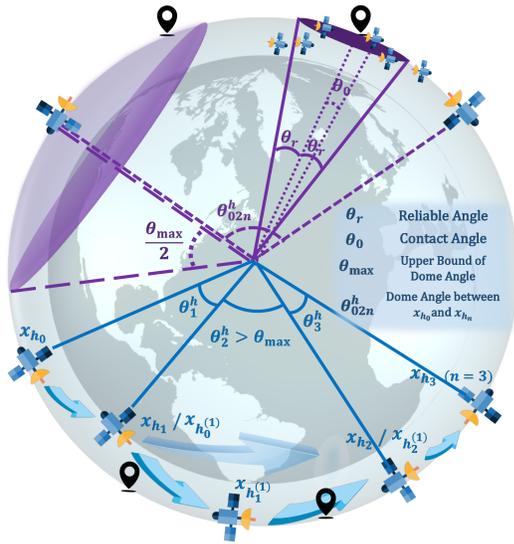


Fig. 2. Schematic diagram of equal interval routing strategy.

PROPOSITION 2: In the ideal scenario, for an n -hop link, if the satellites are located on the shortest inferior arc, the optimal dome angle θ_i^{h*} in \mathcal{P}_1 is equal to θ_{02n}^h/n .

PROOF: See Appendix B. ■

Proposition 2 decreases delay by the equally spaced distribution of relay satellites, while proposition 3 minimizes the latency by determining the optimal number of satellites. Both propositions are shown in Fig. 1.

PROPOSITION 3: In an ideal scenario, assume the satellites are equally spaced distributed on the shortest inferior arc, the optimal number of hops is

$$N_{\min} = \left\lceil \frac{\theta_{02n}^h}{\theta_{\max}^h} \right\rceil + 1 \quad (5)$$

where $\lceil \cdot \rceil$ means rounding up to an integer, and

$$\theta_{\max} = \min \left\{ 2 \arccos \left(\frac{r_{\oplus}}{r} \right), 2 \arcsin \left(\frac{d_{\max}}{r} \right) \right\}. \quad (6)$$

PROOF: See Appendix C. ■

In proposition 3, θ_{\max} is the upper bound of the dome angle between satellites that have established communication links. θ_{\max} ensures that two satellites are within the LoS region of each other and the maximum communication distance d_{\max} . By combining the above propositions, the global optimum solution to the problem \mathcal{P}_0 under ideal conditions is given by the following theorem.

THEOREM 1: In the ideal scenario, the global optimal multihop link in \mathcal{P}_0 has N_{\min} hops, and each hop is located on the inferior arc with equal interval distribution, and the dome angle between each hop is θ_{02n}^h/N_{\min} .

C. Practical Strategies Discussion

Although the optimal solution is derived in Section II-B, it cannot be applied in practice because the satellites are not going to be available exactly at the ideal positions. Based on

propositions 1–3, we designed three strategies to transition multihop routing from the ideal scenario to the practical situation. Fig. 3 is a top view along the direction of the negative z -axis. It gives an example of these strategies.

In *minimum deflection angle strategy*, each satellite should look for the satellite with the least deflection from the shortest inferior arc as its next hop. Only satellites satisfying the distance constraints are eligible to be relay satellites. The next-hop satellite also needs to be shorter from the ending satellite than the previous one to ensure that each hop keeps approaching the destination satellite. These requirements also need to be met in the two subsequent strategies. From the algorithm's perspective, since $\theta = 0$ for the shortest inferior arc, the strategy finds the satellite with the minimum value of θ_i that meets the requirements.

Equal interval strategy finds the nearest satellite as the relay in every optimal position obtained under the ideal scenario. As an intuitive extension of the ideal scenario solution, this strategy can bring extremely low latency. The cost of low delay is the poor reliability since it is highly likely that relay satellites do not meet the constraints (2b) and (2c).

In *maximum stepsize strategy*, the satellite chooses the farthest satellite within communication range as its next hop. It reduces the number of hops as much as possible on the premise of ensuring successful communication. In order to avoid the relay satellite being too far away from the shortest inferior arc, we set up a reliable region, which is the dark area in Fig. 3.

As a result, minimum deflection angle strategy and maximum stepsize strategy are set as baselines. The proposed algorithm is designed on the basis of the equal interval strategy, and it is proved to have the lowest latency and high reliability.

III. ALGORITHM DESIGN AND PERFORMANCE ANALYSIS

In this section, we first determine the number of hops of the multihop link by introducing contact angle and reliable angle. After that, a complete nearest neighbor search algorithm is given, and its reliability is analyzed. Finally, we define link efficiency to measure the maximum gap between algorithm delay and possible optimal solution.

A. Contact Angle and Reliable Angle

Since the interval θ_{02n}^h/n decreases as the number of hops n increases, one way to improve the reliability of equal interval strategy is to increase n . However, proposition 3 shows that the latency is also increased with n . To choose a proper n which can balance the latency and reliability, the concepts of contact angle θ_0 and reliable angle θ_r need to be introduced first, which is shown at the top of Fig. 2.

DEFINITION 4 (CONTACT ANGLE): The contact angle is the dome angle between a randomly placed reference and the closest point from the process (the nearest satellite in this article).

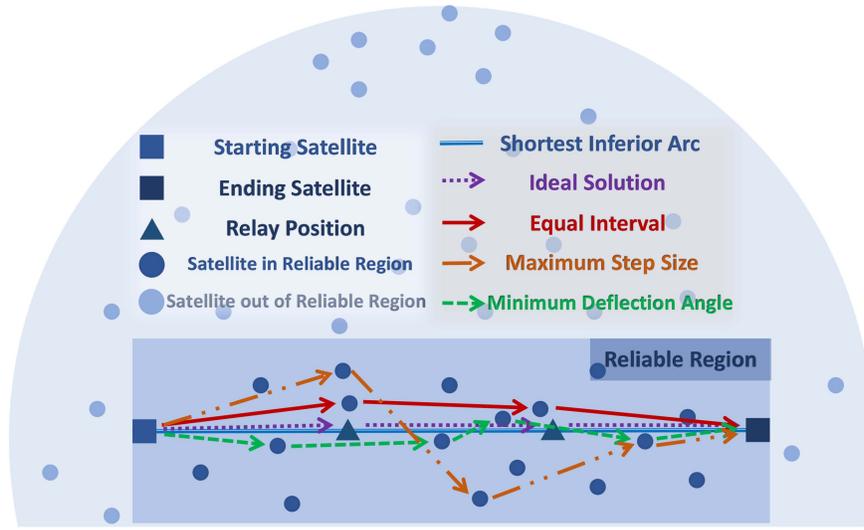


Fig. 3. Example of three strategies.

Since the satellites form a uniform BPP, any randomly selected reference points have the same contact angle distribution.

LEMMA 1: The Cumulative Distribution Function (CDF) of the contact angle is obtained as

$$F_{\theta_0}(\theta) = 1 - \left(\frac{1 + \cos \theta}{2} \right)^{N_{\text{Sat}}}, \quad 0 \leq \theta \leq \theta_{\max} \quad (7)$$

where θ_{\max} is defined in (6).

PROOF: See Appendix D. ■

Based on Lemma 1, the probability density functions (PDF) of the contact angle can be obtained by taking the derivative of CDF with respect to θ .

LEMMA 2: The PDF of the contact angle is obtained as

$$f_{\theta_0}(\theta) = \frac{N_{\text{Sat}}}{2} \sin \theta \left(\frac{1 + \cos \theta}{2} \right)^{N_{\text{Sat}}-1}, \quad 0 \leq \theta \leq \theta_{\max} \quad (8)$$

where θ_{\max} is defined in (6).

DEFINITION 5 (RELIABLE ANGLE): Reliable angle is the minimum dome angle that ensures that at least one satellite can be found within a specified range.

However, even given a large region for search, no satellite may be available because of the randomness. Therefore, we can only guarantee that the probability of not finding any satellite is lower than an acceptable threshold. The value of reliable angle is related to this predefined threshold.

DEFINITION 6 (LINK TOLERABLE PROBABILITY OF INTERRUPTION): Link tolerable probability of interruption ε is the upper bound of the probability that no satellite is available within the reliable angle range in at least one hop.

The absence of a satellite available within a reliable angle range is not a sufficient condition for the interruption. Therefore, ε is not equivalent to the average link interruption probability but an upper bound. In addition, ε can be regarded as a system parameter determined by the requirements rather than an optimization variable. For a fixed ε , the more hops the link has, the higher the reliability required for a single hop. Therefore, reliable angle θ_r is a monotonically increasing function of n . The following lemma will give the relationship among reliable angle θ_r , link tolerable probability of interruption ε , and the number of hops n .

LEMMA 3: For an n -hop link with link tolerable probability of interruption ε , the reliable angle θ_r is given by

$$\theta_r(n) = \arccos \left(2 \left(1 - (1 - \varepsilon)^{\frac{1}{n}} \right)^{\frac{1}{N_{\text{Sat}}}} - 1 \right). \quad (9)$$

PROOF: See Appendix E. ■

B. Type-I Interruption Analysis

Through the above analysis, the following results about the number of hops n can be summarized. The latency increases monotonically with the increase of n . The relationship between interruption probability and the number of hops is not intuitive. Increasing n requires a lower interruption probability for a single hop but brings a larger area for finding a satellite. If n is too large and the single-hop interval is too small, two relay locations of the one-hop may choose the same satellite, which leads to severe errors. To satisfy the distance constraints, the dome angle of each hop θ_i^h should satisfy

$$\theta_i^h + 2\theta_r(n) \leq \theta_{\max}. \quad (10)$$

To ensure that the multihop communication can be completed within n hops, we have

$$n \cdot \theta_i^h \geq \theta_{02n}^h. \quad (11)$$

Algorithm 1: Iterative Method for Deriving the Number of Hops.

- 1: **Input:** Dome angle θ_{02n}^h , number of satellites N_{Sat} and link tolerable probability of interruption ε .
 - 2: $n \leftarrow N_{\text{min}}$.
 - 3: $\theta_r \leftarrow \arccos \left(2 \left(1 - (1 - \varepsilon)^{\frac{1}{n}} \right)^{\frac{1}{N_{\text{Sat}}}} - 1 \right)$.
 - 4: **while** $\frac{1}{2} \left(\theta_{\text{max}} - \frac{\theta_{02n}^h}{n} \right) \leq \theta_r \leq \frac{1}{2} \theta_{\text{max}}$ **do**
 - 5: $n \leftarrow n + 1$.
 - 6: $\theta_r \leftarrow \arccos \left(2 \left(1 - (1 - \varepsilon)^{\frac{1}{n}} \right)^{\frac{1}{N_{\text{Sat}}}} - 1 \right)$.
 - 7: **end while**
 - 8: **Output:** Minimum number of hops n and reliable angle θ_r .
-

By combining the above two inequalities, a loose lower bound on $\theta_r(n)$ can be obtained

$$n(\theta_{\text{max}} - 2\theta_r(n)) \geq \theta_{02n}^h \quad (12)$$

To avoid the possibility of selecting the same satellite for two relay positions of a single hop, an upper bound of $\theta_r(n)$ is given as

$$\theta_r(n) \leq \frac{\theta_{\text{max}}}{2} \quad (13)$$

The following algorithm can give the minimum number of hops between the upper and lower bounds through iteration.

Note that the minimum number of hops is not related to the positions of satellites and it can be expressed as

$$\hat{N}_{\text{min}} = \min \left\{ n : (\theta_{\text{max}} - 2\theta_r(n-1))(\theta_{\text{max}} - 2\theta_r(n)) < 0 \text{ or } \left(\theta_{\text{max}} - \frac{\theta_{02n}^h}{n-1} - 2\theta_r(n-1) \right) \left(\theta_{\text{max}} - \frac{\theta_{02n}^h}{n} - 2\theta_r(n) \right) < 0 \right\} \quad (14)$$

which is another representation of step (4) of the algorithm. Both $\theta_r(n)$ and $\frac{1}{2}(\theta_{\text{max}} - \theta_{02n}^h/n)$ increase with n . When the algorithm ends the loop as $\frac{1}{2}(\theta_{\text{max}} - \theta_{02n}^h/n) > \theta_r(n)$ satisfied, the output n is the required minimum number of hops. Otherwise, when the algorithm ends the loop as $2\theta_r(n) > \theta_{\text{max}}$, no value of n guarantees tolerable probability of interruption ε . For a constellation with a small number of satellites, it is not realistic to guarantee a low tolerable probability of interruption. Such problems due to poor system design are defined as type-I interruption.

DEFINITION 7 (TYPE-I INTERRUPTION): Type-I interruption is a qualitative indicator to describe the rationality of multihop communication system design.

In addition to running an algorithm to determine whether the type-I interruption occurred, The proposition also provides a sufficient condition for the type-I interruption not to occur.

PROPOSITION 4: If there exists a θ_r , which is smaller than $\frac{1}{2}\theta_{\text{max}}$, satisfying the following inequality, then there must be a routing scheme such that the probability of no satellite being available in the range of reliable angle is lower than ε

$$N_{\text{Sat}} \geq \frac{1}{\ln \left(\frac{1+\cos \theta_r}{2} \right)} \ln \left(1 - (1 - \varepsilon)^{1 / \left(\left\lceil \frac{\theta_{02n}^h}{\theta_{\text{max}} - 2\theta_r} \right\rceil + 1 \right)} \right) \quad (15)$$

where θ_{02n}^h and θ_{max} are defined in (4) and (6), respectively.

PROOF: See Appendix F. ■

C. Type-II Interruption Analysis and Nearest Neighbor Search Algorithm

Considering that even if the constellation is suitable for multihop transmission, communication interruption may still happen due to the randomness of the satellite position. Such interruptions are defined as type-II interruption.

DEFINITION 8 (TYPE-II INTERRUPTION): Type-II interruption is an event that happens when the distance in any hop does not satisfy at least one constraint in \mathcal{P}_0 .

Although the two types of interruptions happen for different reasons, the occurrences of these two types of interruptions are not independent. The occurrence of type-I interruption often leads to type-II interruption. Because type-II interruption cannot be avoided by the parametric design of the satellite constellation, so we deal with the interruption after it occurs. Suppose the distance between each satellite is too far to communicate. In that case, the satellite at the starting point of the hop is expected to looking for one or several satellites closest to the shortest inferior arc as relays. As is shown at the bottom of Fig. 2, it can be regarded as using minimum deflection angle strategy within a single hop.

As mentioned, the algorithm proposed in this article is based on the equal interval strategy. If two types of interruptions are resolved, the probability of errors occurring in the equal-interval strategy is significantly reduced, thus ensuring low latency and high reliability. The practical nearest neighbor search algorithm is divided into four stages: 1) calculate the minimum number of hops through iteration, 2) find the relay position according to equal interval strategy, 3) find nearest satellite in the neighborhood of the relay position to establish the link, and 4) adopting minimum deflection angle strategy in the single hop when the two satellites of the hop cannot satisfy the distance constraints. The last three steps of the algorithm are as follows.

To simplify the description of the algorithm, the distance between two points is defined as

$$d(\theta_1, \varphi_1, \theta_2, \varphi_2) = r \left(2(1 - \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \cos(\varphi_1 - \varphi_2)) \right)^{\frac{1}{2}}. \quad (16)$$

In addition, the start ID $h_0^{(i)} = h_{i-1}$ and the end ID $h_n^{(i)} = h_i$ in set $\mathcal{H}^{(i)}$. The nearest neighbor search algorithm cannot guarantee finding the optimal solution even when the two

Algorithm 2: Nearest Neighbor Search Algorithm.

- 1: **Input:** Locations of point process Φ , the number of hops \hat{N}_{\min} , starting point ID h_0 and ending point ID $h_{\hat{N}_{\min}}$.
- 2: **Initialize:** $T \leftarrow 0$.
- 3: **for** $i = 1 : \hat{N}_{\min} - 1$ **do**
- 4: $\theta_i^h \leftarrow \theta_{h_0} \left| \frac{2i}{\hat{N}_{\min}} - 1 \right|$.
- 5: **if** $i < \frac{\hat{N}_{\min}}{2} + 1$ **then**
- 6: $\varphi_i^h \leftarrow 0$.
- 7: **else**
- 8: $\varphi_i^h \leftarrow \pi$.
- 9: **end if**
- 10: $h_i \leftarrow \arg \min_j d(\theta_i^h, \varphi_i^h, \theta_j, \varphi_j)$.
- 11: **end for**
- 12: $\mathcal{H} \leftarrow \{h_0, h_1, \dots, h_{\hat{N}_{\min}-1}, h_{\hat{N}_{\min}}\}$.
- 13: **for** $i = 1 : \hat{N}_{\min}$ **do**
- 14: **if** $d(\theta_{h_{i-1}}, \varphi_{h_{i-1}}, \theta_{h_i}, \varphi_{h_i}) > d_{\max}$ **then**
- 15: Use minimum deflection angle strategy to find the relay satellite IDs $\mathcal{H}^{(i)}\{h_0^{(i)}, h_1^{(i)}, \dots, h_n^{(i)}\}$ in i^{th} -hop.
- 16: $T \leftarrow T + \sum_{j=1}^n d(\theta_{h_{j-1}^{(i)}}, \varphi_{h_{j-1}^{(i)}}, \theta_{h_j^{(i)}}, \varphi_{h_j^{(i)}})$.
- 17: **else**
- 18: $T \leftarrow T + d(\theta_{h_{i-1}}, \varphi_{h_{i-1}}, \theta_{h_i}, \varphi_{h_i})$.
- 19: **end if**
- 20: **end for**
- 21: **Output:** IDs of the multi-hop link \mathcal{H} and Latency T .

types of interruptions do not occur. For example, a link with many hops may meet the distance constraints even after two links are merged. Since the sum of the two sides of the triangle is greater than the third, the combined link has a lower latency. Therefore, it is necessary to analyze the latency performance of the algorithm.

D. Efficiency Analysis

For an optimization problem that is difficult to find the optimal solution, the most concerning issue is the gap between the found solution and the optimal solution. Unfortunately, according to the available data, no algorithm can find the optimal solution to the problem. The latency of the optimal solution in the ideal scenario is an unattainable lower bound. It is also an upper bound of the difference between the proposed method and the optimal solution. Therefore, efficiency is defined to quantify the difference.

DEFINITION 9 (EFFICIENCY): Efficiency is the ratio of minimum latency in the ideal scenario to the latency of the proposed method.

Since satellites are uniformly and independently distributed on the sphere and the intervals of relay positions on multihop links are equal, the distance between single hops is independent and identically distributed. Therefore, analyzing the efficiency of multihop links can be equivalent to studying that of single-hop. The increase in the distance

caused by random distribution can be equivalent to the increase of the dome angle. In other words, it offsets the random distribution of satellites by moving their relay positions. Thus, the following two approximations are given.

THEOREM 2: For an \hat{N}_{\min} -hop link with dome angle θ_{02n}^h , the contour integral approximation of the efficiency is given as

$$\tilde{E}_1 = \frac{N_{\min} \cdot \sin\left(\frac{\theta_{02n}^h}{2N_{\min}}\right)}{\hat{N}_{\min} \cdot \sin\left(\frac{\theta_{02n}^h}{2\hat{N}_{\min}}\right) \left(2\bar{\alpha}\left(\frac{\theta_{02n}^h}{2N_{\min}}\right) - 1\right)} \quad (17)$$

where N_{\min} is defined in (5), and $\bar{\alpha}(\theta^h)$ is defined as

$$\bar{\alpha}(\theta^h) = \frac{\sqrt{2}}{2\pi} \int_0^{\theta_{\max}} \int_0^{\pi} \frac{f_{\theta_0}(\theta)}{\sin\left(\frac{\theta^h}{2}\right)} (-\cos(\theta_0)\cos(\theta^h) - \sin(\theta_0)\sin(\theta^h)\cos\varphi + 1)^{\frac{1}{2}} d\varphi d\theta. \quad (18)$$

PROOF: See Appendix G. ■

THEOREM 3: For an \hat{N}_{\min} -hop link with dome angle θ_{02n}^h , the binomial approximation of the efficiency is given as

$$\tilde{E}_2 = \frac{N_{\min} \cdot \sin\left(\frac{\theta_{02n}^h}{2N_{\min}}\right)}{\hat{N}_{\min} \cdot \eta\left(\frac{\theta_{02n}^h}{2N_{\min}}\right)} \quad (19)$$

where N_{\min} is defined in (5), and $\eta(\theta^h)$ is defined as

$$\eta(\theta^h) = \int_0^{\theta_{\max}} \int_0^{\theta_{\max}} \frac{1}{4} f_{\theta_0}(\theta_1) f_{\theta_0}(\theta_2) \times (\sin(\theta^h - \theta_1 - \theta_2) + \sin(\theta^h + \theta_1 - \theta_2) + \sin(\theta^h - \theta_1 + \theta_2) + \sin(\theta^h + \theta_1 + \theta_2)) d\theta_1 d\theta_2. \quad (20)$$

PROOF: Assuming that the contact angle between the relay position and its nearest satellite is θ_0 , the satellites are uniformly distributed on a circle with radius $r \sin \theta_0$. By approximating this distribution as a binomial distribution, satellites are distributed at the nearest or farthest from the adjacent relay position with equal probability. Take the expectation of contact angles, and the above result can be obtained. ■

IV. NUMERICAL RESULTS

This section analyzes the performance of the algorithm based on the results of numerical simulation. For the existing constellations, we analyze the feasibility of the algorithm. Then, different approximation methods and routing strategies are compared from the perspective of latency.

A. Reliability Analysis of Constellations

Table II shows the simulation results of three LEO satellite constellations [26]. Set the maximum distance at which the satellite can maintain stable communication as $d_{\max} = 3000$ Km. Within this distance, the satellites in all three constellations are in the LoS region. Suppose two satellites on opposite sides of the earth need to communicate. Since Kuiper's satellites will be distributed in three

TABLE I
Summary of Notations

Notation	Description
$N_{\text{Sat}}; n; N_{\text{min}}$	Number of satellites; number of hops that one link contains; minimum number of hops
$r_{\oplus}; r_{\text{Sat}}; r$	Radius of the Earth; height of the satellite orbits; radius of the sphere where satellites locate
$\mathcal{H}; h_i; d_i$	The set which contains the IDs of a link; ID of the i^{th} satellite; distance of the i^{th} hop
$x_i, \theta_i, \varphi_i; \Phi$	The location, polar angle, azimuth angle of the i^{th} satellite; the homogeneous BPP
$T; \varepsilon$	Latency of the multi-hop link; link tolerable probability of interruption
$\theta_i^h, \theta_{02n}^h$	Dome angle between satellites $x_{h_{i-1}}$ and x_{h_i} ; starting satellite x_{h_0} and ending satellite x_{h_n}
$d_{\text{max}}; \theta_{\text{max}}$	Maximum communication distance; upper bound of dome angle between satellites
$\theta_0; \theta_r$	Contact angle; reliable angle
$\bar{E}_1; \bar{E}_2$	Contour integral approximation of the efficiency; binomial approximation of the efficiency

TABLE II
Reliability Analysis of LEO Satellite Constellations

	Starlink	OneWeb	Kuiper
Constellation altitude [Km]	550	1200	590, 610, 630
Number of (planned) satellites	11927	650	3236
Expectation of contact angle	0.0162	0.0695	0.0312
Number of hops	9 / 10	69 / 8	12 / 13
Reliable angle [rad]	0.0386 / 0.0481	0.1996 / 0.2026	≈ 0.0765 / ≈ 0.0941
Minimum number of satellites required	710 / 2535	2053 / 7889	≈ 777 / ≈ 2520
Type-I interruption occurs or not	No / No	Yes / Yes	No / No
Probability of Type-II interruption occurs	$< 0.01\%$ / $< 0.01\%$	9.41% / 100%	$< 0.01\%$ / $< 0.01\%$
Efficiency	99.44% / 99.17%	97.80% / 96.27%	97.91% / 97.56%

different altitude orbits, we approximate that all satellites are distributed in the 610 Km orbit. For the last five parameters, the left and right sides of the slash correspond to $\varepsilon = 0.1/0.01$, respectively.

Since the latency of satellite communication is usually tens to hundreds of milliseconds, it is necessary to consider the calculation delay of the algorithm and the delay of search. The complexity of iterative method for deriving number of hops is linear. Iterations can end in a finite number of steps, and the number of hops should satisfy

$$n \leq \frac{\ln(1 - \varepsilon)}{\ln\left(1 - \frac{1}{2}\left(1 - \cos\frac{\theta_{\text{max}}}{2}\right)^{N_{\text{Sat}}}\right)}. \quad (21)$$

It can be seen that the number of iterations mostly ranges from 1 to 6. The expected contact angle and reliable angle are used to analyze the area of the search region. According to the description of the nearest neighbor search algorithm, traversing all satellites can only stay at the theoretical level. In practice, since satellite systems are massive and moving, it is difficult for a single satellite to get global information. Therefore, it is more meaningful to analyze the required area for finding a satellite than the algorithm complexity. The expectation of contact angle can be derived from the following simple derivation

$$\begin{aligned} \mathbb{E}[\theta_0] &= \int_0^{\theta_{\text{max}}} 1 - F_{\theta_0}(\theta) d\theta \\ &= \int_0^{\theta_{\text{max}}} \left(\frac{1 + \cos\theta}{2}\right)^{N_{\text{Sat}}} d\theta \\ &= 2 \int_0^{\frac{\theta_{\text{max}}}{2}} (\cos\tilde{\theta})^{2N_{\text{Sat}}} d\tilde{\theta} \end{aligned}$$

$$\stackrel{(a)}{\approx} \pi \prod_{i=1}^{N_{\text{Sat}}} \frac{2i-1}{2i} \quad (22)$$

where (a) follows Wallis' integrals, since $1 - F_{\theta_0}(\theta)$ is very close to 0 when $\theta > \theta_{\text{max}}$ [8], the result can be approximated by continuation of the domain. Assume the spherical caps with radius of the expected contact angle and reliable angle as the average search area and maximum search area required for finding a satellite. This region is chosen as a spherical cap for computational convenience. Taking Starlink as an example, for a ten-hop link, the average search area is 0.066% of the entire spherical area. The maximum search area is no more than 0.58% of the spherical area. The minimum deflection angle strategy needs to search along the belt region near the shortest inferior arc. The maximum step size strategy needs to search in the whole communication range. When the reliable region is not set, the search area of the maximum step size strategy is 10.2% of the entire spherical area for a ten-hop link. In conclusion, the proposed algorithm can end in a linear number of iterations and generally only takes a few iterations. It requires a tiny search area and has a huge advantage over other strategies. At last, note that shape of the search area is not necessary to a spherical cap, as well as surface of arbitrary shape. Since satellites are uniformly distributed on the sphere, the probability of finding a satellite is a constant for a given surface area for search.

The minimum number of satellites required is obtained by testing several sets of θ_r according to proposition 4 and taking the smallest of them. The probability of type-II interruption is obtained by Monte Carlo method: 1) running the algorithm for 10^6 rounds and regenerating the satellites'

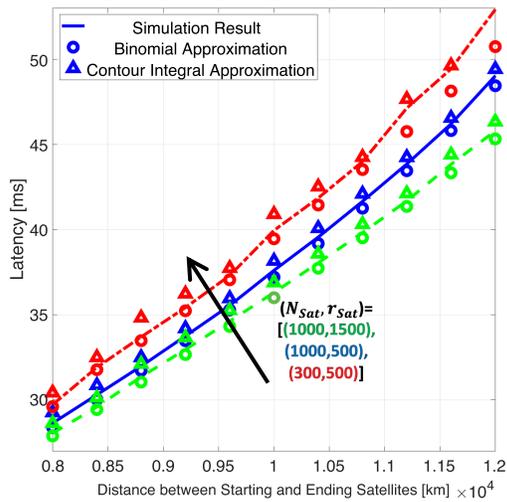


Fig. 4. Comparison of different approximations.

positions at each round, 2) recording the number of interrupt rounds, and 3) dividing the number of interrupt rounds by 10^6 as the probability of interruption. It can be seen that as long as the number of satellites obtained by any set of θ_i is smaller than the number of satellites in the actual constellation, type-II interruption does not occur. The opposite may not be accurate. For example, in the Kuiper constellation, when $\varepsilon = 0.01$ and number of hops is 8, the required number of satellites obtained is 5544, which exceeds the number of satellites of the Kuiper constellation 3236. However, the second type of error still does not occur.

The last discussion is about type-I interruption. For Oneweb constellation with parameter $\varepsilon = 0.1$, we get $n \leq 68.077$ from (21) the number of iterations reached 61. When $\varepsilon = 0.01$, the iterations do not start because the reliable angle $\theta_0 = 0.2026$ exceeded half of the maximum dome angle $\frac{\theta_{\max}}{2} = 0.1994$. Both situations lead to the type-I interruption, which further leads to the occurrence of type-II interruption. In addition, the algorithm has high efficiency for all constellations.

B. Comparison of Different Approximations

As shown in Fig. 4, the performances of the two estimation methods are compared, and the relationships between latency and constellation parameters are described. In Fig. 4, link tolerable probability of interruption $\varepsilon = 0.01$, $d_{\max} = 3000$ Km, the simulation result is the exact latency obtained by Monte Carlo simulation. Both approximation methods are accurate under different constellation altitudes, satellite numbers, and link distances. Under the existing groups of parameters, binomial approximation provides a tight lower bound for the latency. At the same time, the contour integral approximation gives a tight upper bound for the latency. The accuracy of the two approximations is reduced for scenarios where the number of satellites corresponding to the red dot and dash is insufficient. Especially, when the distance between the starting satellite and the ending satellite is large, binomial approximation has a relatively large gap with the actual results for the red line.

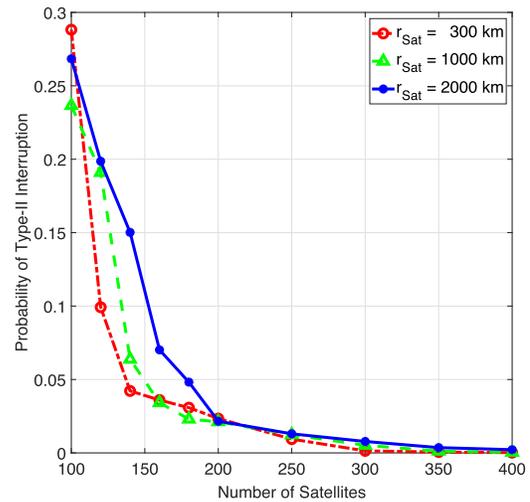


Fig. 5. Probability of type-II interruption under different parameters.

As the communication distance increases and the number of satellites is insufficient, the probability of link interruption increases. In this case, the introduction of the minimum deflection strategy brings larger latency.

Use the solid blue line (1000 satellites and 500 Km constellation altitude) in Fig. 4 as a baseline. When the communication distance is fixed, the latency is negatively correlated with the number of satellites and positively correlated with the constellation height. The decrease in the number of satellites lead to the locations of the found satellite deviating from the ideal optimal relay location, which increases latency. Although the increase of constellation height also causes the satellite location to deviate from the expected position, reducing the shortest inferior arc length has a more significant effect on the latency. A similar view can be found in proposition 1. In addition, the latency increases almost linearly with the increase of communication distance, and the constellation with larger latency has a larger slope of growth.

Fig. 5 further explains the results in Fig. 4 through numerical results. In Fig. 5, the communication distance is 10000 Km and $d_{\max} = 3000$ Km. When number of satellites $N_{\text{Sat}} > 400$, type-II interruption rarely occurs. When $N_{\text{Sat}} < 200$, the probability of type-II interruption is significantly increased with the decrease of N_{Sat} and the increase in constellation height r_{Sat} . This suggests that when satellites are insufficient, the probability of type-II interruption is closely related to the number of satellites per unit sphere area. Furthermore, the influence of r_{Sat} on the probability of type-II interruption is not as significant as N_{Sat} , especially when $N_{\text{Sat}} < 200$.

C. Comparison of Different Strategies

Figs. 6 and 7 provide the results of latency changing with distance between starting and ending satellites for different strategies. In both figures, latitude is fixed as 500 Km and $d_{\max} = 3000$ Km. The number of satellites in Fig. 6 is sufficient (800 satellites) while the number of satellites in Fig. 7 is insufficient (100 satellites).

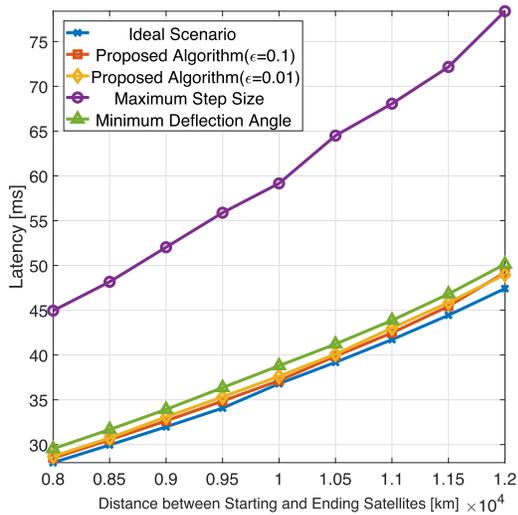


Fig. 6. Influence of communication distance on different strategies ($N_{\text{Sat}} = 800$).

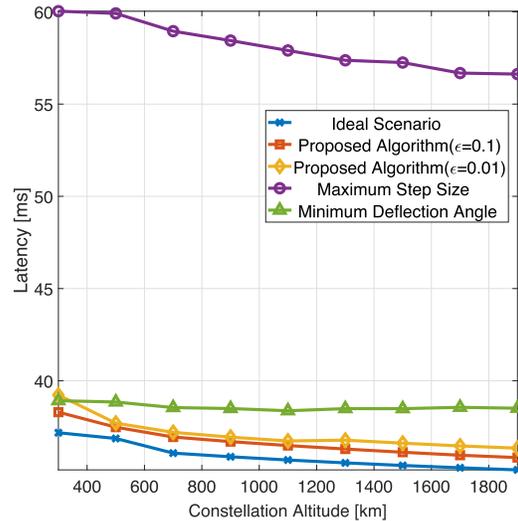


Fig. 8. Influence of constellation altitude on different strategies.

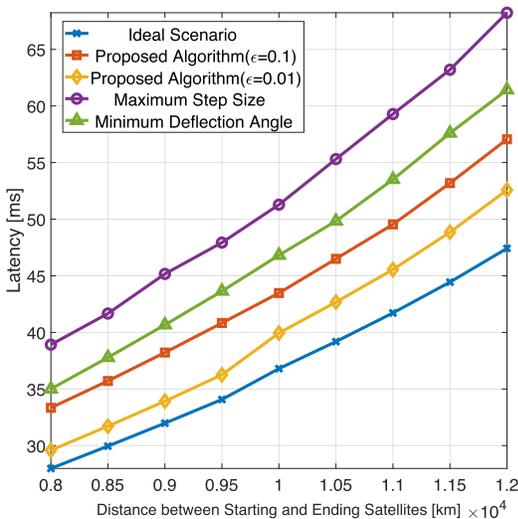


Fig. 7. Influence of communication distance on different strategies ($N_{\text{Sat}} = 100$).

In terms of latency, the optimal scenario, the proposed algorithm, the minimum deflection angle strategy, and the maximum step size strategy are sequentially ranked from small to large. When the number of satellites is sufficient, the latency of the maximum step size strategy is much larger than that of other methods. The minimum deflection angle strategy and the proposed algorithm's performances are close to the lower bound. When the number of satellites is insufficient, the proposed algorithm has a remarkable advantage over the minimum deflection angle strategy. For different tolerance rates, $\varepsilon = 0.01$ performs better with fewer satellites, while $\varepsilon = 0.1$ performs better when satellites are sufficient.

Fig. 8 considers the scenario where latency varies with constellation height. The number of satellites and is fixed as 800, the communication distance is fixed as 10000 Km and $d_{\text{max}} = 3000$ Km. Overall, the performance of the methods is similar to that in Fig. 6. The main difference is that for

the proposed algorithm and the maximum step size strategy, the latency decreases with the height of the constellation. The change of the minimum deflection angle strategy is not obvious.

V. FURTHER EXTENSIONS

Since the shortest routing problem on a three-dimensional (3-D) sphere is not an easy problem to deal with, we simplify the model for the convenience of analysis. Although our simple model has limitations when facing some practical issues, the model is fortunately extensible.

A. Expansion to Multitier Networks

Practically, LEO satellites may assist ground base stations [27] with global coverage or rely on ground gateways [17] to communicate. In addition, satellite systems at different altitudes (including those in synchronous orbits) also interact, such as satellites in the Kuiper constellation at three different altitudes. Therefore, cross-tier communication scenarios should be considered.

Hence, it is required to investigate routing in a spherical heterogeneous network consisting of ground stations, high altitude platforms (HAP), and multitier LEO satellites, where satellite communications start and end with ground stations. The theoretical analysis in this article is basically applicable to the above heterogeneous network, with the following three major changes. First, the values of some parameters such as maximum communication distance d_{max} vary with different types of the relay device. This means that global information will be harder to obtain and store for ground stations.

Second, as an essential parameter in analyzing the efficiency and reliability of the proposed algorithm, the expression and domain of the contact angle in a multitier network have minor modifications. Specifically, the contact angle will be replaced by the conditional contact angle, which is the contact angle of satellites distributed within

the reliable communication range of both the previous and next hop.

Finally, in the reliability analysis, the tier on which the relay device is located affects the probability of type-II interruption. Therefore, discrete Markov networks, state transition matrices, and absorption states are recommended for reliability analysis. Note that routine starts and ends on the ground stations, thus the first, middle, and last hops of the network need to be designed differently.

B. Latency of Computation and Search

Only transmission latency is considered as the objective function in this article. Computation and search latency should also be taken into account. As is mentioned, although the proposed algorithm provides a low computational complexity solution for finding the shortest latency routing on the closed sphere, its computational complexity still reaches $\mathcal{O}(N_{\min} \cdot N_{\text{Sat}})$. The latency corresponding to this computational complexity is still large for a real-time routing with a total transmission latency of tens of milliseconds. The algorithm complexity can be reduced to $\mathcal{O}(N_{\min})$ through any of the following two schemes since only steps (5)–(9) in algorithm 2 need to be executed for both of the schemes.

When ground stations are available, we can sacrifice storage space on the ground stations for less latency. A specific data structure called Two Line Elements (TLEs) can store the dynamic positions of the satellites, and the IDs of satellites around the target position can be quickly found by index when a routing task arrives. One possible disadvantage of this scheme is that when the source is not the ground equipment but the satellite, the source needs to spend extra latency to communicate with the ground equipment.

The second scheme applies to scenarios where ground stations are unavailable. The satellite transmits a signal to the target position [obtained in step (4)–(9) of algorithm 2], and the next-hop satellite within the beam forward this information in the above method and respond to the previous hop. Similarly, this scheme also includes extra search latency related to the contact angle, reliable angle, and beamwidth. When the satellite does not receive a response from the next hop, it assumes no satellite in the beam and continues to send messages to surrounding areas. In addition, when several satellites receive the messages from the previous hop and are busy, it requires short-distance communication to schedule a single satellite for routing.

C. Outage Probability and Buffering Latency

When considering power limits, the probability of interruption and latency are related not only to distance, but also to transmission signal power. Under the assumption that regenerative hops are used, a longer single-hop distance and a lower transmission power result in a larger probability of interruption and buffering latency. Under this circumstance, the signal-to-interference plus noise ratio (SINR) serves as a bridge between them. Since satellites are less dense than ground networks and the beam is highly directional, the interference caused by other satellites can be approximated

to a small constant. Assuming that the path loss of single-hop satellite-satellite channel follows the free-space fading model, the average SINR is a decreasing function of to the single-hop distance squared.

Different from the qualitative analysis before, outage probability can be a quantitative substitute for type-II interruption and single-hop maximum reliable distance d_{\max} . The outage probability is defined as the probability of receiving SINR smaller than a predefined threshold $\mathbb{P}[\text{SINR} < \gamma]$. The maximum step size proposition may not be optimal because a long single-hop distance may lead to a high probability of communication failure [28]. Because of the randomness of fading, signal interruptions always occur, and the retransmission mechanism can be introduced [24].

Average achievable rate is regarded as an upper bound on the as the upper bound of the transmission rate and the lower bound of the buffering latency. It is defined as the ergodic capacity from the Shannon–Hartley theorem over a fading communication link [29], which is proportional to $\log_2(1 + \text{SINR})$. When the packet size is much larger than the maximum amount of data transmitted per millisecond under the average achievable rate, buffering latency is necessary to be taken into account. In order to decrease the buffering latency, a large data packet can be divided into parts and transmitted in multiple separate paths. The number of paths is determined by traffic and the average achievable rate of the relay satellites. According to proposition 1, the path corresponding to the inferior arc with a smaller central angle is selected preferentially.

D. Small Satellite Swarms and Storage-and-Forward Communication

In the case of an insufficient number of satellites swarms with large packet sizes [30], the accessibility of data transmission is restricted, and it is challenging to realize real-time communication. These networks are demonstrated as delay/disruption tolerant networks (DTN), in which satellites store information for an amount of time after receiving it [31]. The proposed algorithm can be extended to reduce the latency of networks with sufficient interactions. For example, with the accessibility of Earth-to-satellite links, the proposed algorithm applies to Earth observation satellite constellations.

Furthermore, small satellite swarms can help update the satellite's information (such as positions) around the relay satellite, which is beneficial for the proposed algorithm in this article that relies on information interaction. The strategy combining the proposed algorithm with store-and-forward communication is also extendable to small spacecraft swarms communicating for interstellar exploration [32].

VI. CONCLUSION

The latency minimization of multihop satellite links under the maximum distance constraints is studied. We propose a nearest neighbor search algorithm to determine the number of hops of multihop links and the position of

the relay satellite in each hop. Numerical results show that the algorithm achieves linear complexity and can complete iteration in finite steps. At the same time, the search area required by the algorithm only accounts for a tiny part of the whole sphere area. The latency performance of this algorithm is very close to the minimum latency in the ideal scenario. Take Starlink constellation for example, the algorithm only needs two iterations and searches 0.066% of the entire spherical area. The extra latency it needs to pay is no more than 1% of the total latency of the optimal case. Furthermore, two approximations are provided to estimate the maximum gap between the latency of the proposed algorithm and the lower bound of the latency in the ideal scenario. They provide tight upper and lower bounds for latency in most cases. Finally, the influence of system parameters on multihop link latency is studied.

APPENDIX A PROOF OF PROPOSITION 1

Among all circles passing x_{h_0} and x_{h_n} on the sphere, where the satellites are located, the circle centered at the origin has the largest radius. Therefore, the shortest inferior arc divided by these two points has the smallest central angle. Based on the fact that the smaller the central angle, the shorter the length of the arc, this inferior arc has the shortest length among all arcs passing through x_{h_0} and x_{h_n} .

For an arbitrary routing scheme, as shown in Fig. 1, we can always locate the corresponding relay satellite on the shortest inferior arc to achieve lower latency. The correspondence of satellite positions between the two schemes is shown in Fig. 1. In the scheme corresponding to the sky blue arrow, the distance of each hop is no longer than that of the scheme corresponding to the green arrow. Note that all subsequent concepts related to the central angle refer to the dome angle unless otherwise stated.

APPENDIX B PROOF OF PROPOSITION 2

Use (3a) and (3d) to construct the Lagrange function

$$\mathcal{L}(\theta_1^h, \theta_2^h, \dots, \theta_n^h) = \frac{1}{c} \sum_{i=1}^n 2r \sin\left(\frac{\theta_i^h}{2}\right) + \lambda \left(\sum_{i=1}^n \theta_i^h - \theta_{02n}^h \right) \quad (23)$$

take the partial derivative with respect to θ_i^h , we get

$$\frac{\partial \mathcal{L}(\theta_1^h, \theta_2^h, \dots, \theta_n^h)}{\partial \theta_i^h} = \frac{r}{c} \cos\left(\frac{\theta_i^h}{2}\right) + \lambda \quad (24)$$

set the result of the partial derivative to 0, the optimal θ_i^{h*} is

$$\theta_i^{h*} = 2 \arccos\left(-\frac{\lambda c}{r}\right) \quad (25)$$

which is not related to i . Finally, the proof can be completed by combining the constraint (3d).

APPENDIX C PROOF OF PROPOSITION 3

Assume that the satellites keep equal dome angles on the shortest inferior arc. The latency can be expressed as

$$T = \frac{2r}{c} \sum_{i=1}^n \sin\left(\frac{\theta_{02n}^h}{2n}\right) = \frac{2rn}{c} \sin\left(\frac{\theta_{02n}^h}{2n}\right) \quad (26)$$

take partial derivative with respect to n

$$\frac{\partial T}{\partial n} = \frac{2r}{c} \sin\left(\frac{\theta_{02n}^h}{2n}\right) - \frac{\theta_{02n}^h r}{cn} \cos\left(\frac{\theta_{02n}^h}{2n}\right) \quad (27)$$

since an inferior arc is chosen, $\theta_{02n}^h/(2n) < \pi/2$, when $n \neq 1$, we have $\cos(\theta_{02n}^h/2n) > 0$, and

$$\frac{\partial T}{\partial n} = \frac{1}{\cos(\theta_{02n}^h/2n)} \left(\tan\left(\frac{\theta_{02n}^h}{2n}\right) - \frac{\theta_{02n}^h}{2n} \right) \quad (28)$$

for the right-hand side of the equation, $\tan\left(\frac{\theta_{02n}^h}{2n}\right) > \frac{\theta_{02n}^h}{2n}$ when $\theta_{02n}^h/(2n) < \pi/2$. The above analysis shows that $\frac{\partial T}{\partial n} > 0$ is always satisfied. As n increases, the latency T increases, so we need to select the minimum number of hops that satisfies the constraints (3b) and (3c), the upper bound of θ_i^h is limited as θ_{\max} defined in (6), by solving

$$\theta_{02n}^h = \sum_{i=1}^{N_{\min}} \theta_i^h \leq N_{\min} \theta_{\max} \quad (29)$$

and based on the fact that N_{\min} is an integer, the final result is obtained.

APPENDIX D PROOF OF LEMMA 1

Start deriving the CDF of the contact angle distribution from the definition

$$\begin{aligned} F_{\theta_0}(\theta) &= 1 - \mathbb{P}[\theta_0 > \theta] \\ &= 1 - \mathbb{P}[\mathcal{N}(\mathcal{A}) = 0] \\ &\stackrel{(a)}{=} 1 - \left(1 - \frac{\mathcal{S}(\mathcal{A})}{4\pi r^2}\right)^{N_{\text{Sat}}} \\ &\stackrel{(b)}{=} 1 - \left(1 - \frac{2\pi r(r - r \cos \theta)}{4\pi r^2}\right)^{N_{\text{Sat}}} \\ &= 1 - \left(\frac{1 + \cos \theta}{2}\right)^{N_{\text{Sat}}} \end{aligned} \quad (30)$$

where $\mathcal{N}(\mathcal{A})$ counts the number of the satellites in the spherical cap \mathcal{A} shown in Fig. 2, $\mathcal{S}(\mathcal{A})$ is the area measure of spherical cap \mathcal{A} . According to step (a), for a homogeneous point process, the probability of having satellites on the spherical cap is equal to the ratio of the area of the spherical cap to the total surface area of the sphere with radius r . Step (b) comes from the area formula of a spherical cap, where $r - r \cos \theta$ is the height of the spherical cap. In addition, the domain of θ_0 should meet the constraints. It is easy to verify that for a constellation of hundreds of satellites, $F_{\theta_0}(\theta_{\max})$ is very close to 1 [8].

APPENDIX E
PROOF OF LEMMA 3

Since satellites' locations are assumed to be independent, the average probability interruption of each hop should be equal. For an n -hop link with tolerable probability of interruption ε , the tolerable probability of interruption of each hop is

$$\varepsilon_1 = 1 - (1 - \varepsilon)^{\frac{1}{n}}. \quad (31)$$

In the spherical cap determined by reliable angle, the probability of having a satellite should be greater than $1 - \varepsilon_1$. Since the reliable angle is the minimum angle that satisfies the above constraint, it can be obtained by the definition of the contact angle CDF

$$1 - \left(\frac{1 + \cos \theta_r(n)}{2} \right)^{N_{\text{Sat}}} = (1 - \varepsilon)^{\frac{1}{n}} \quad (32)$$

transpose and take the square root of N_{Sat} times on both sides

$$\frac{1 + \cos \theta_r(n)}{2} = \left(1 - (1 - \varepsilon)^{\frac{1}{n}} \right)^{\frac{1}{N_{\text{Sat}}}} \quad (33)$$

final conclusion can be reached through simple mathematical operations.

APPENDIX F
PROOF OF PROPOSITION 4

Since the reliable angle $\theta_r(n)$ is related to the number of hops, a θ_t unrelated to n is taken as the search radius to simplify the relationship. In this case, the minimum number of hops N_h is given as

$$N_t = \left\lceil \frac{\theta_{02n}^h}{\theta_{\max} - 2\theta_t} \right\rceil + 1. \quad (34)$$

$2\theta_t < \theta_{\max}$ ensures that N_t is positive. Substitute (34) into (32)

$$\left(\frac{1 + \cos \theta_t}{2} \right)^{N_{\text{Sat}}} \leq 1 - (1 - \varepsilon)^{1/\left(\left\lceil \frac{\theta_{02n}^h}{\theta_{\max} - 2\theta_t} \right\rceil + 1\right)} \quad (35)$$

take the logarithm of both sides, and divide by the $\ln\left(\frac{1+\cos\theta_t}{2}\right)$ of both sides to get the result. Note that (32) guarantees that the θ_t satisfying (15) must be greater than or equal to the reliable angle. A set of practical θ_t can be taken as

$$\left\{ \frac{1}{2} \left(\theta_{\max} - \frac{\theta_{02n}^h}{N_{\min} + k} \right), k = 0, 1, 2, \dots \right\}. \quad (36)$$

APPENDIX G
PROOF OF PROPOSITION 2

Assuming that the contact angles between the two relay positions and their nearest satellites are $\theta_0^{(1)}$ and $\theta_0^{(2)}$, respectively. In this case, these two satellites are uniformly distributed on circles $\mathcal{O}_1(\theta_0^{(1)})$ and $\mathcal{O}_2(\theta_0^{(2)})$ with radius $r \sin \theta_0^{(1)}$ and $r \sin \theta_0^{(2)}$, respectively. The average distance at contact angles $\theta_0^{(1)}$ and $\theta_0^{(2)}$ can be obtained by contour

integral around two circles with respect to single-hop distance d_1 . Therefore, the expectation of single-hop distance d_1 can be expressed as

$$\mathbb{E}[d_1] = \mathbb{E}_{\theta_0^{(1)}, \theta_0^{(2)}} \left[\oint_{\mathcal{O}_1(\theta_0^{(1)})} \oint_{\mathcal{O}_2(\theta_0^{(2)})} f_{d_1} d\mathcal{O}_2 d\mathcal{O}_1 \right] \quad (37)$$

where f_{d_1} is the PDF of d_1 , it is related to the contact angles $\theta_0^{(1)}, \theta_0^{(2)}$ and the positions on the corresponding $\mathcal{O}_1, \mathcal{O}_2$. The expression of f_{d_1} is hard to express in either rectangular or spherical coordinates. Let us split the problem in two. One of the satellites is fixed to the relay position, while the other is uniformly distributed on the circle. The uniform distribution of a satellite can be offset by changing the position of a relay position. This amount of change can be described by α . By symmetrically making the same change of the other relay position, the amount of change becomes $2\alpha - 1$.

Since rotation does not affect the distribution of the satellite, let the spherical coordinate of the relay position be $(r, 0, 0)$. The coordinate of the fixed satellite is $(r, \theta^h, 0)$, where θ^h is the dome angle of the single hop. Assume the contact angles between the relay position and its nearest satellite is θ_0 , by equation

$$\frac{1}{\pi} \int_0^{\theta_{\max}} d(\theta_0, \varphi, \theta_h, 0) d\varphi = \alpha(\theta_0, \theta^h) 2r \sin\left(\frac{\theta_h}{2}\right) \quad (38)$$

where $d(\theta_0, \varphi, \theta_h, 0)$ is defined in (16), the amount of change $\alpha(\theta_0, \theta^h)$ can be written as

$$\alpha(\theta_0, \theta^h) = \frac{\sqrt{2}}{2\pi \sin\frac{\theta^h}{2}} \times \int_0^\pi \sqrt{1 - \cos\theta_0 \cos\theta_h - \sin\theta_0 \sin\theta^h \cos\varphi} d\varphi. \quad (39)$$

Note that for a small θ_0

$$\alpha(\theta_0, \theta^h) \approx \frac{\sqrt{2(1 - \cos\theta_h)}}{2 \sin\frac{\theta^h}{2}} = 1. \quad (40)$$

Take the expectation of $\alpha(\theta_0, \theta^h)$ with respect to θ_0

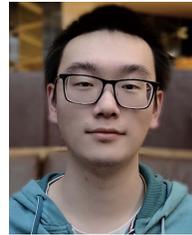
$$\bar{\alpha}(\theta^h) = \int_0^{\theta_{\max}} f_{\theta_0}(\theta) \alpha(\theta_0, \theta^h) d\theta \quad (41)$$

the result in (18) is derived. Since the propagation speed of the laser is constant, the ratio of latency is equivalent to the ratio of distance, the proof of theorem 2 is finished.

REFERENCES

- [1] I. Del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, 2019.
- [2] O. Kodheli et al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surv. Tut.*, vol. 23, no. 1, pp. 70–109, Jan.–Mar. 2021.
- [3] E. Yaacoub and M.-S. Alouini, "A key 6G challenge and opportunity-connecting the base of the pyramid: A survey on rural connectivity," *Proc. IEEE*, vol. 108, no. 4, pp. 533–582, Apr. 2020.

- [4] S. C. Ekpo et al., "System-level multicriteria modelling of payload operational times for communication satellite missions in LEO," *Recent Prog. Space Technol. (Formerly Recent Patents Space Technol. (Discontinued))*, vol. 4, no. 1, pp. 67–77, 2014.
- [5] A. U. Chaudhry and H. Yanikomeroglu, "Free space optics for next-generation satellite networks," *IEEE Consum. Electron. Mag.*, vol. 10, no. 6, pp. 21–31, Nov. 2021.
- [6] F. Tang, H. Zhang, and L. T. Yang, "Multipath cooperative routing with efficient acknowledgement for LEO satellite networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 1, pp. 179–192, Jan. 2019.
- [7] F. He et al., "Delay-bounded and minimal transmission broadcast in LEO satellite networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–7.
- [8] A. Al-Hourani, "An analytic approach for modeling the coverage performance of dense satellite networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 897–901, Apr. 2021.
- [9] X. Sun and S. Cao, "A routing and wavelength assignment algorithm based on two types of LEO constellations in optical satellite networks," *J. Lightw. Technol.*, vol. 38, no. 8, pp. 2106–2113, Apr. 2020.
- [10] Y. Zhang, Y. Tang, and W. Wang, "Service deployment and service request optimization scheduling in MEC enabled LEO networks," in *Proc. IEEE Int. Conf. Comput. Commun. Netw. (ICCCN)*, Athens, Greece, 2021, pp. 1–6.
- [11] J. Li et al., "Temporal netgrid model-based dynamic routing in large-scale small satellite networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6009–6021, Jun. 2019.
- [12] Y. Lu et al., "Complexity of routing in store-and-forward LEO satellite networks," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 89–92, Jan. 2015.
- [13] Y. Lu et al., "Some analytical results for routing in LEO satellite networks," in *Proc. IEEE Int. Conf. Adv. Mechatronic Syst. (ICAMEchS)*, Shiga, Japan, 2019, pp. 359–362.
- [14] F. Wang, D. Jiang, and S. Qi, "An adaptive routing algorithm for integrated information networks," *China Commun.*, vol. 16, no. 7, pp. 195–206, 2019.
- [15] T. Pan et al., "OPSPF: Orbit prediction shortest path first routing for resilient LEO satellite networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, IEEE, 2019, pp. 1–6.
- [16] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [17] A. Talgat, M. A. Kishk, and M.-S. Alouini, "Stochastic geometry-based analysis of leo satellite communication systems," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2458–2462, Aug. 2021.
- [18] N. Okati et al., "Downlink coverage and rate analysis of low earth orbit satellite constellations using stochastic geometry," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5120–5134, Aug. 2020.
- [19] A. Talgat, M. A. Kishk, and M.-S. Alouini, "Nearest neighbor and contact distance distribution for binomial point process on spherical surfaces," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2659–2663, Dec. 2020.
- [20] K. Stamatiou and M. Haenggi, "The delay-optimal number of hops in poisson multi-hop networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TexasUSA, 2010, pp. 1733–1737.
- [21] H. S. Dhillon and G. Caire, "Wireless backhaul networks: Capacity bound, scalability analysis and design guidelines," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6043–6056, Nov. 2015.
- [22] M. J. Farooq, H. ElSawy, and M.-S. Alouini, "A stochastic geometry model for multi-hop highway vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2276–2291, Mar. 2015.
- [23] S. Sasaki, Y. Miyaji, and H. Uehara, "Energy budget formulation in progress-based nearest forwarding routing policy for energy-efficient wireless sensor networks," *IEICE Trans. Inf. Syst.*, vol. 100, no. 12, pp. 2808–2817, 2017.
- [24] M. Haenggi, "On routing in random rayleigh fading networks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1553–1562, Jul. 2005.
- [25] Y. Richter and I. Bergel, "Optimal and suboptimal routing based on partial CSI in random ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2815–2826, Apr. 2018.
- [26] F. A. Robert, M. Puteaux, and A. Najjar, "Small satellites market growth patterns and related technologies," *Handbook of Small Satellites: Technology, Design, Manufacture, Applications, Economics and Regulation*, pp. 1–41, 2020.
- [27] B. A. Homssi and A. Al-Hourani, "Modeling uplink coverage performance in hybrid satellite-terrestrial networks," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3239–3243, Oct. 2021.
- [28] M. Haenggi and D. Puccinelli, "Routing in ad hoc networks: A case for long hops," *IEEE Commun. Mag.*, vol. 43, no. 10, pp. 93–101, Oct. 2005.
- [29] R. Wang, M. A. Kishk, and M.-S. Alouini, "Ultra-dense LEO satellite-based communication systems: A novel modeling technique," *IEEE Commun. Mag.*, vol. 60, no. 4, pp. 25–31, Apr. 2022.
- [30] S. Nag, M. SanchezA. NetLi, and V. Ravindra, "Designing a disruption tolerant network for reactive spacecraft constellations," in *Proc. ASCEND*, 2020, paper no. 4009.
- [31] P. G. Madoery, J. A. Fraire, and J. M. Finochietto, "Analysis of communication strategies for earth observation satellite constellations," *IEEE Latin America Trans.*, vol. 14, no. 6, pp. 2777–2782, Jun. 2016.
- [32] K. L. Parkin, "The breakthrough starshot system model," *Acta Astronautica*, vol. 152, pp. 370–384, 2018.



Ruibo Wang received the B.S. degree in communication engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020, and the M.S. degree in electrical engineering from King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2022. He is currently working toward the Ph.D. degree with KAUST.

His current research interests include stochastic geometry, satellite communications, and

wireless networks.



Mustafa A. Kishk (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Giza, Egypt, in 2013 and 2015, respectively, and the Ph.D. degree in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2018.

He is currently an Assistant Professor with the Electronic Engineering Department, Maynooth University, Ireland. Before that, he was a Postdoctoral Research Fellow with the Communication Theory Laboratory, KAUST, Thuwal, Saudi Arabia. His current research interests include stochastic geometry, energy harvesting wireless networks, UAV-enabled communication systems, and satellite communications.



Mohamed-Slim Alouini (Fellow, IEEE) was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998.

He was a Faculty Member with the University of Minnesota then in the Texas A&M University at Qatar before joining in 2009 the KAUST, where he is now a Distinguished Professor of electrical and computer engineering. He is currently particularly interested in addressing the technical challenges associated with the uneven distribution, access to, and use of information and communication technologies in far-flung, rural, low-density populations, low-income, and/or hard-to-reach areas.

Prof. Alouini is a fellow of the OSA.