

# Stochastic Geometry-based Analysis of Cache-Enabled Hybrid Satellite-Aerial-Terrestrial Networks with Non-Orthogonal Multiple Access

Xiaokai Zhang, Bangning Zhang, Kang An, Gan Zheng, *Fellow, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE* and Daoxing Guo

**Abstract**—Due to the emergence of non-terrestrial platforms with extensive coverage, flexible deployment, and reconfigurable characteristics, the hybrid satellite-aerial-terrestrial networks (HSATNs) can accommodate a great variety of wireless access services in different applications. To effectively reduce the transmission latency and facilitate the frequent update of files with improved spectrum efficiency, we investigate the performance of cache-enabled HSATN, where the user retrieves the required content files from the cache-enabled aerial relay or the satellite with the non-orthogonal multiple access (NOMA) scheme. If the required content files of the user are cached in the aerial relay, the cache-enabled relay would serve directly. Otherwise, the user would retrieve the content file from the satellite system, where the satellite system seeks opportunities for proactive content pushing to the relay during the user content delivery phase. Specifically, taking into account the uncertainty of the number and location of aerial relays, along with the channel fading of terrestrial users, the outage probability and hit probability of the considered network are, respectively, derived based on stochastic geometry. Numerical results unveil the effectiveness of the cache-enabled HSATN with the NOMA scheme and proclaim the influence of key factors on the system performance. The realistic, tractable, and expandable framework, as well as associated methodology, provide both useful guidance and a solid foundation for evolved networks with advanced configurations in the performance of cache-enabled HSATN.

**Index Terms**—Stochastic geometry, non-orthogonal multiple access, hybrid satellite-aerial-terrestrial networks, content caching.

## I. INTRODUCTION

**O**WING to the inherent advantages in terms of ubiquitous coverage, high throughput, and extreme resilience, hybrid satellite-aerial-terrestrial networks (HSATNs) have been

This work is supported by the Jiangsu Provincial Natural Science Foundation of China (No. BK20191328), the National Natural Science Foundation of China under Grant 61901502 and 62001517, the National Postdoctoral Program for Innovative Talents under Grant BX20200101, and the Research Project of NUDT under Grants ZK18-02-11 and 18-QNCXJ-029. (Corresponding author: D. Guo.)

Xiaokai Zhang, Bangning Zhang and Daoxing Guo are with the College of Communications and Engineering, Army Engineering University, Nanjing, 210007, China (e-mail: xiaokaizhang@foxmail.com, AEU\_zbn@163.com, xyzgfg@sina.com).

K. An is with the Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China. (ankang89@nudt.edu.cn)

G. Zheng is with Wolfson School of Mechanical, Electrical, and Manufacturing Engineering, Loughborough University, Loughborough, U.K. (g.zheng@lboro.ac.uk).

S. Chatzinotas is with Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg, Luxembourg (symeon.chatzinotas@uni.lu).

regarded as an emerging architecture by integrating the satellite systems, aerial platforms, and terrestrial infrastructures in the next-generation wireless networks for enhanced mobile broadband (EMBB) or three-dimensional integrated communications (3D-InteCom) [1]–[3]. The main motivation of HSATNs is that the obstacles and shadowing between the satellite and the terrestrial terminal may make line-of-sight (LoS) communication difficult to maintain [4]–[8]. The emergence of aerial platforms, such as unmanned aerial vehicles (UAVs), airships, and balloons, provides a promising solution for improving the scalability of hybrid networks, where the aerial layer acts as carriers for information acquisition, transmission, and processing. In the European FP7 ABSOLUTE program [8], the HSATN is pursued to satisfy the requirement of the network capacity through the combination of satellite, aerial/terrestrial communication links. Although the HSATN has advantages in terms of coverage and ease of rapid deployment, the transmission latency is an unavoidable problem for HSATNs due to the long distance between satellites and aerial or terrestrial terminals [8]. Besides, with an increasing number of devices, the scarcity of spectrum resources has gradually become an important issue [5], [9].

### A. Background and Motivation

With advances in the launch technologies as well as miniaturization of the satellite, the academic and industrial communities have put great efforts into developing new platforms offering more capacity to enable richer user cases and applications [10]. The demand for infotainment services has experienced an unprecedented increase over the last decade [7]. Therefore, the effective content distributions in the case of rural areas or underdeveloped countries certainly entail an important engineering effort in future satellite communication trends, especially collaborating with 5G infrastructure [11]. However, the current satellite Internet architecture is founded upon a host-centric communication model, such as Internet Protocol (IP)-based architectures [12]. The host-centric networks cannot guarantee the service continuity as well as proper handling of different content distributions services with high levels of quality of service (QoS)/quality of experience (QoE) [13] [14].

To address this problem, the advent of information-centric networking (ICN) architectures [15] has prompted the shift of traditional networking towards a content-oriented concept,

which can be directly implemented to the existing content delivery networks [16]. By defining information at the network layer, ICN favors the deployment of the in-network content caching mechanism, which facilitates the efficient and timely delivery of information to the users [17]. Content caching is one of the most important approaches in ICN architecture, where the popular contents are proactively cached on nodes with caching capabilities. Users can obtain the required content at the nearest node to reduce transmission latency. Moreover, the excessive propagation delay, as the inevitable long distance can also be effectively alleviated in HSATNs. Hence, through incorporating content caching enabled aerial relays into HSATN, both wide coverage area and lower transmission delay can be achieved.

As a novel paradigm of multiple access techniques, non-orthogonal multiple access (NOMA) transmits signals simultaneously in the same time/frequency resource block, which can be separated at receivers with successive interference cancellation (SIC) [18]–[21]. It encourages spectrum sharing among multiple users, instead of occupying orthogonal resource blocks as in orthogonal multiple access (OMA) [22]. Therefore, NOMA can significantly improve spectral efficiency, reduce transmission delay, and support massive connectivity. The incorporation of the NOMA scheme into HSATN is a promising direction to address some inherent problems.

### B. Related Work

A two-layered hybrid satellite-terrestrial relay network (HSTRN) is designed in [23]–[28] for the seamless integration of satellite with terrestrial components, which has been demonstrated as a simple yet effective architecture with improved reliability and coverage. In the foundation of autonomous working and mutual integration, HSTRNs integrate the advantages of satellite networks which could provide long transmission distance, wide coverage without the limitation of surroundings with terrestrial networks that have large network capacity and high rate. ITU-R S.2222 standard [8] is proposed to a cross-layer design based on HSTRN to provide multimedia services and ensure the user QoS requirements. However, the terrestrial relay can be blocked by obstacles and shadowing. As an extension, HSATN is further proposed, where the high and low altitude platforms (HAPs & LAPs) can provide broadband services complementing the terrestrial networks [29]. Compared with base stations or the relays in the terrestrial network, the aerial network has the features of low cost, easy deployment, and large coverage to offer wireless access services on a regional basis [30]. In recent years, we have witnessed a rapid proliferation of unmanned aerial vehicles (UAVs) in all areas due to the advance in their payload capacity and prolonged battery life. There exist some low cost-cost ways to provide a long time seamless coverage, for example, unmanned airships, high-altitude balloons, and solar-powered drones [31], [32]. One of the famous tests is Project Loon by Alphabet Inc. [33].

The paper in [3] presented a comprehensive review of recent research works concerning space-air-ground integrated

networks (SAGINs) from network design and resource allocation to performance analysis and optimization. The paper in [8] proved that HSATNs could provide better services than traditional infrastructures during the emergency, in which the HSATNs were surveyed and the key technologies were discussed from several aspects. A cooperative satellite-aerial-terrestrial system was considered in [34], considering the randomness of satellite and users and employing stochastic geometry, where the coverage probability in non-interference and interference scenarios was studied, and the outage performance was investigated. The paper in [4] investigated transmission for HSATNs from an energy-efficient perspective, where a multi-antenna UAV was employed as a relay to assist the satellite signal delivery. The authors in [5] were the first to study the cross-layer gateway selection problem for data delivery in the HSATN communication architecture with the practical inter-layer link capacity constraint, which achieved in such integrated network the best data delivery performance across multiple layers.

Wireless caching has been proved to be an efficient approach to enhance the system performance with respect to delay, QoS, and QoE for satellite communication [35], [36]. Particularly, the paper in [36] has justified the significant performance improvement of satellite-based overlay caching for terrestrial content delivery networks. The authors in [11] have analyzed content delivery over satellite integrated cognitive radio networks, which focused on throughput, energy efficiency, and quality as key performance metrics in the heterogeneous system. To alleviate the spectrum shortage and meet the requirements of improved spectral efficiency, the authors in [37] first proposed the cache-enabled relay in HSTRN, which confirmed the substantial performance improvement through caching capability. However, this paper only considers the uncertainty of the channel fading. As far as we know, there is currently no work to derive performance analysis while considering channel fading, the uncertainty of the number and location of relays in HSTRNs or HSATNs.

The NOMA scheme has been diffusely investigated in hybrid satellite-terrestrial networks (HSTN) from increasing the spectrum efficiency perspective [38]–[43], which proves the feasibility and effectiveness of the NOMA scheme in HSTN. The authors in [38] investigated the NOMA in the land mobile satellite system and gave the outage performance of the proposed system model. [39] and [40] implemented the NOMA scheme into cooperative spectrum sharing in HSTN for single relay and multiple relays, respectively. The paper in [41] and [42] introduced the NOMA scheme in HSTRN. The authors in [43] applied the NOMA-based scheme by integrating terrestrial networks and the satellite cooperatively to provide coverage for ground users while reusing the entire bandwidth. However, there is no prior work analyzing the performance of the NOMA from a stochastic geometry perspective in HSTNs.

The current proactive caching strategy assumes that the contents can be pushed to cache-enabled relays during the off-peak time in an error-free manner [44]. However, this content cache strategy is not applicable if caches need to be frequently updated or rapidly changing files, including latest

news, electronic commerce promotion with frequent pricing changes, newly released music videos, etc., especially via satellite transmission [35], [36]. Taking full advantage of the features of NOMA, additional files can be pushed to the aerial relays simultaneously, i.e. the satellite could seek opportunities for content pushing during the content delivery phase by the NOMA scheme. This strategy is important to efficiently use the limited resources reserved for content pushing, which facilitates the frequent update of the files cached at the aerial relays and improves the cache hit probability. Besides, the utilization of NOMA not only improves the reliability of content delivery but also ensures that more user requests can be served during the content delivery phase.

The deterministic and location-based models that have previously been applied to analyze hybrid satellite-aerial-terrestrial networks (HSATNs) are typically restricted to support simple network topology [34]. However, with the increase in the number of user terminals and the uncertain distribution of locations, the wireless network topology gradually shows the characteristics of randomization and dynamic network [18], [45]. To make the theoretical analysis as close to reality as possible, it is no longer possible to accurately characterize the characteristics of random networks based on determining the network topology. In recent years, the use of Stochastic Geometry to describe the randomness of the spatial location for nodes in wireless networks has been recognized by the industry [46], [47], which can characterize network interference and analyze average performance from a mathematical perspective. The biggest advantage is that when the nodes in the network obey a specific random distribution. Stochastic geometry plays a vital role in evaluating the system-level performance of wireless networks, where random spatial point patterns describe the nature of wireless systems statistically [48]. Therefore, we take the uncertainty of the channel, the number, and the location into consideration to evaluate the performance of the cache-enabled HSATN.

### C. Contribution and Organization

Motivated by the above observations, this paper investigates the stochastic geometry-based performance of cache-enabled HSTRN with the NOMA scheme. The main contributions of this paper are summarized as follows:

- 1) *Modeling the cache-enabled HSATN with NOMA:* A generic framework of cache-enabled HSATN is first modeled, where the user would retrieve the required content files. If the required content is not cached in the aerial relay, the user tends to retrieve the content file from the satellite, where the satellite simultaneously seeks opportunities for proactive content pushing to the relay during the user content delivery phase by applying the NOMA scheme. Due to the heterogeneity of different layers with various physical mechanisms, the propagation properties and mobility are considered for modeling and constructing the channels between each layer. The proposed framework can achieve enlarged coverage areas with increased spectrum efficiency and reduced latency.
- 2) *The performance metric:* For retrieving the content files from the satellite, the QoS requirement-based

NOMA power allocation strategy is first proposed, which guarantees that the retrieved content files of the user can decode the required files in the aerial relay of the user<sup>1</sup>. Specifically, the outage probability and hit probability of the user and cache-enabled aerial relays are derived for the proposed power allocation strategy. For retrieving the content files from the cache-enabled relay, the aerial relay serves users with transmission requirements by the NOMA scheme, where the co-channel interferences from other aerial relays are also taken into account. Particularly, a mathematical tractable expression for the outage probability of the user is derived in closed-form.

- 3) *System design guidelines:* We investigated the impact of the different parameters including the sequence number of the cache-enabled aerial relay, the transmission power of the satellite, the intensity of aerial relay and the content popularity parameter for retrieving the content files from the satellite, and the transmission power of aerial relay and the content popularity parameter for retrieving the cache-enabled relay. Both simulated and analytical results are provided to validate the effectiveness of the cache-enabled HSATN with NOMA and proclaim the influence of key factors on the performance.
- 4) *Comparison with OMA:* We investigate the superiority of the NOMA scheme in the HSATRN with the proper NOMA power allocation coefficient in terms of outage probability of the user and hit probability of the cache-enabled aerial relay.

The rest of this paper is organized as follows. Section II presents the stochastic cache-enabled in HSATNs with stochastic geometry and formulates the signal transmission inspired by the NOMA scheme. Section III and Section IV derive the key performance analysis for retrieving the content files from the satellite and the cache-enabled relay, respectively. Section V illustrates the simulation results. Finally, conclusions are drawn in Section VI.

## II. PROBLEM STATEMENT

### A. System Model

We consider a HSATN with heterogeneous preferences, in which multiple users request contents with the help of the satellite and multiple cache-enabled aerial relays, where all nodes are equipped with a single antenna and work in the half-duplex mode. If the request content is not cached in the aerial relay, the user tends to request the content from the satellite as shown in Fig. 1(a) on the next page, where the satellite can obtain the contents of the entire network from the network control center (NCC) through gateway station connecting to the core network [11], where the NOMA scheme is implemented for delivering content for the user and pushing other currently most popular content to cache-enabled aerial relays simultaneously. Otherwise, the users request the content from the cache-enabled aerial relay to retrieve the content with alleviated transmission delay, as shown in Fig. 1(b) on the next page, where the NOMA scheme is implemented in each cell

<sup>1</sup>The aerial relay retransmits the required content to the user in the next time slot, which avoids the transmission latency from the satellite in retransmission.

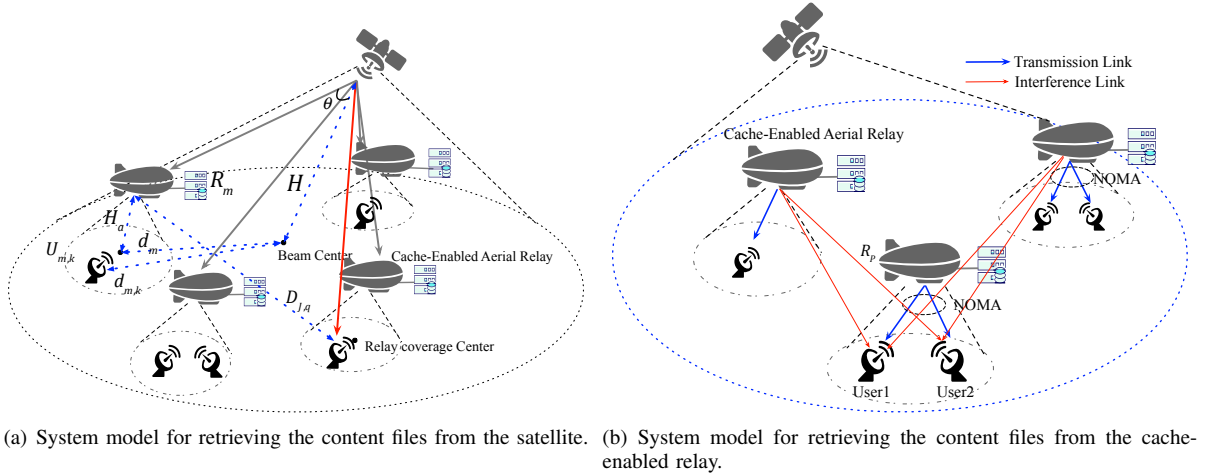


Fig. 1: System model

to increase spectrum efficiency. The users retrieve the content from the relay if the required files exist [37]. This paper mainly considers the downlink of the user content retrieving.

Without loss of generality, we consider that the distribution of cache-enabled aerial relays on a two-dimensional area, covering a circle with radius  $\mathcal{D}$ , is a homogeneous Poisson point process (HPPP) with an intensity of  $\lambda$  and denoted by  $\Phi_c$ . The height of the aerial relays is  $H_a$ . The  $m$ -th neighbor cache-enabled aerial relay from the beam center is denoted by  $R_m$ . Moreover, we consider the relay is the parent node of the cluster covering a circle whose radius is denoted by  $\mathcal{R}$ , where the spring users are uniformly distributed in the circle. We assume that there are  $K$  users associated with  $R_m$ , and the  $k$ -th user is denoted by  $U_{m,k}$ .

### B. Satellite-Link Propagation Model

The propagation factors of satellite links include the free-space path loss, the satellite antenna gain, the receiver antenna gain, the shadowing and channel fading as well as the other loss. The channel gain  $G$  of LoS is given by

$$G = \frac{G_R G_S(\theta) f_\varepsilon}{(4\pi H/\lambda_c)^2}, \quad (1)$$

where  $\varepsilon$  denotes the elevation angle from the receiver to the satellite,  $H$  is assumed to be the same as the satellite height for simplicity,  $\lambda_c$  denotes the wavelength, and  $f_\varepsilon$  denotes the other loss of the receiver.  $G_R$  is the receiving antenna gain. For simplicity of analysis, we consider the receiver antenna gains of the user and the aerial platform are fixed, where one of the typical scenarios is that the receiver antenna points in the direction of the satellite. According to ITU-R S.1528 [58], the LEO satellite antenna gain  $G_S$  [dBi] is given by

$$G_S(\theta) = \begin{cases} G_{s,\max} - 3\left(\frac{\theta}{\theta_b}\right)^2 & 0 \leq \theta < 1.5\theta_b \\ G_{s,\max} + L_s - 25 \log\left(\frac{\theta}{1.5\theta_b}\right) & 1.5\theta_b \leq \theta < Y \\ 0 & Y \leq \theta < 180^\circ \end{cases}, \quad (2)$$

where  $G_{s,\max}$  is the maximum gain at the on-board antenna boresight,  $L_N = -6.75\text{dB}$ ,  $a = 2.58$ ,  $b = 6.32$ ,  $Y =$

$b\theta_b 10^{0.04(G_{s,\max} + L_N)}$  by ITU-R S.1528,  $\theta_b$  is one-half of  $\theta_{3\text{dB}}$ , i.e.  $\theta_b = 0.5 \arctan(L/H)$ ,  $L$  is the half-power beamwidth,  $\theta$  represents the angular position from the receiver to beam center with respect to the satellite, which can be written as  $\theta = \arctan(d/H)$ , where  $d$  is the distance between the beam center and the receiver. For  $H \gg L$  and  $H \gg d$ ,  $\theta/\theta_b$  is approximately expressed as  $\theta/\theta_b \approx 2d/L$ . Therefore, the directional beam of the satellite is highly on the position of the aerial relay by  $d$ , which can be rewritten as  $G_s(d)^2$ .

The shadowing and channel fading of satellite link are considered to follow the Shadowed-Rician fading model [49], which is mathematically tractable and has been widely applied in various fixed and mobile satellite services for a variety of frequency bands. The probability density function (PDF) of  $|h|^2$  is shown as [40], [50]

$$f_{|h|^2}(x) = \alpha \exp(-\beta x) {}_1F_1(q, 1, \delta x), \quad (3)$$

where  $h$  is small scale fading of the satellite link,  ${}_1F_1(\cdot, \cdot, \cdot)$  denotes the confluent hypergeometric function and  $\alpha = \frac{(2cq)^q}{2c(2cq+\Omega)^q}$ ,  $\delta = \frac{\Omega}{2c(2cq+\Omega)}$ , and  $\beta = \frac{1}{2c}$ , with  $2c$  being the average power of the scatter component,  $\Omega$  being the average power of the line-of-sight (LOS) component,  $q$  being the Nakagami fading parameter. The corresponding cumulative distribution function (CDF) of  $|h|^2$  can be obtained as [51]

$$F_{|h|^2}(x) = \alpha \sum_{k=0}^{\infty} \frac{(q)_k \delta^k}{(k!)^2 \beta^{k+1}} \Gamma(k+1, \beta x). \quad (4)$$

We consider that the altitude of the aerial relays is sufficiently large without the shadowing and obstacles, such as Balloons<sup>3</sup>. Therefore, the channel gain between the satellite and the aerial relay  $R_m$  is  $G_m = G$ . For the user  $U_{m,k}$ , the line-of-sight satellite communication systems are vulnerable to

<sup>2</sup>Because the height of the LAP is invariant, we can hereby reduce the stochastic model within the aerial layer to two-dimensional. Compared to the distance from the beam center to aerial relay, the height of the LAP can be ignored.

<sup>3</sup>We consider that the Doppler Frequency Shift, caused by the mobility of the satellite, can be estimated and mitigated by the mature pre-compensation method for users [52]. Besides, since the movement of the aerial relay is slow when serving users, the Doppler shift is assumed to be well compensated, which has a negligible impact on the HSATN system [53].

be blocked by shadowing or obstacles. Therefore, the channel gain between the satellite and user is  $G_{m,k} = G|h_{m,k}|^2$  [39].

### C. Aerial-Terrestrial Link Propagation Model

The signal transmission of the aerial-terrestrial link is considered to experience the LoS propagation, however, it still may be blocked by buildings and bushes. Therefore, the radio signal attenuation probabilistically follows LoS propagation with a lower path-loss exponent and Nakagami- $m$  small-scale of fading.

The PDF of the channel power gain between  $P$ -th aerial relay  $R_p$  and the user  $q$  is given as [54]

$$f_{H_{P,q}}(x) = \frac{1}{\Gamma(m^A)(\theta_{P,q})^{m^A}} x^{m^A-1} e^{-\frac{x}{\theta_{P,q}}}, \quad (5)$$

where  $m^A$  is the fading severity, which is assumed as integer values in this paper,  $\theta_{P,q}$  is the average power,  $\Gamma(\cdot)$  denotes the complete gamma function<sup>4</sup>.

### D. Content Caching Model

We consider that the  $M$  files, required to be cached in the relay, are collected in a finite content catalog  $\mathcal{F} = \{f_1, \dots, f_m, \dots, f_M\}$ . According to the statistic of the popularity of the content, a large fraction of requests occurs for a relatively small fraction of the content, which implies that some contents are popular and requested with high frequency by different users [37]. We consider the popularity of the requested files obeys the *Zipf* distribution for different aerial relay. Hence, the popularity of  $f_m$  in  $i$ -th aerial relay is given by [56]

$$P_i(f_m) = \frac{m^{-\gamma_i}}{\sum_{l=1}^M l^{-\gamma_i}}, \quad (6)$$

where  $\gamma_i > 0$  denotes the shape parameter defining the content popularity skewness of  $i$ -th aerial relay. To be noticed, a large  $\gamma_i$  means the requests on the high popularity files, whereas a small  $\gamma_i$  is related to the requests with heavy-tailed popularity [37]. Notably, the popularity of the content is the probability of the requested file of relays<sup>5</sup>. Dissimilar from conventional assumptions, we consider that each file has the same length packets, while the information contained in the packets of different files is nonidentical. The reason is that packets belonging to dissimilar files have distinct priorities and QoS requirements, which requires the use of different transmission rates for each packet. Specifically, the required data rate for  $f_m$  is  $R^m$ .

<sup>4</sup>We consider that all channels are assumed to experience independent and identically distributed (i.i.d.) fading and follow quasi-static fading, i.e. the channel gains remain constant within each transmission block but vary independently between different blocks [37], [55].

<sup>5</sup>The popularity of the content files is statistically calculated. The interest of different users may be different, which can be adjusted by multiplying an interest factor. Therefore, we apply the system level popularity as the request probability of the users and aerial relays to proclaim the influence of key factors on the system performance.

## III. PERFORMANCE ANALYSIS FOR RETRIEVING CONTENT FROM SATELLITE

If the required content files are not cached in the aerial relay, the satellite would serve the user directly. To improve the spectrum efficiency and increase the frequency of cached file updates, the NOMA scheme is implemented to deliver the content for the user and push other currently most popular contents to cache-enabled aerial relays simultaneously, which can be seen as “add-on” transmission.

### A. NOMA Signal Transmission

By exploiting opportunities for content pushing during the content delivery phase, the satellite would serve the users directly and push the contents to the aerial relay simultaneously by the NOMA scheme. To be emphasized, the satellite server the user would not reuse the channel of aerial to the user to avoid interruption of the satellite link. Hence, there is no co-channel interference for retrieving content from the satellite. The  $M$  most popular files belonging to the same library can be pushed to all cache-enabled aerial relays.  $x_0$  and  $x_i$  denote the signal of the user  $U_{m,k}$  and the  $i$ -th content file, where  $\mathbb{E}[|x_0|^2] = \mathbb{E}[|x_i|^2] = 1$ . The received mixed-signal at  $z$  can be expressed as

$$y_z = \sqrt{\alpha_0 P_S} G_{S,z} x_0 + \sum_{i=1}^M \sqrt{\alpha_i P_S} G_{S,z} x_i + n_0, \quad (7)$$

where  $G_{S,z}$  is the power gain between satellite and node  $z$  the subscript  $z \in \{U_{m,k}, R_m\}$ ,  $P_S$  is the satellite transmitter power, and  $n_0$  is the AWGN with mean zero and variance  $\sigma_0^2$ .  $\alpha_0$  is the NOMA power allocation coefficient of the user,  $\alpha_i$  denotes the power allocation coefficient of  $f_i$ , where  $\sum_{i=0}^M \alpha_i = 1$ . The receivers implement SIC to separate the multiplexed signal.

According to the property of the NOMA scheme, the power allocation factors and the SIC decoding order would directly determine the performance of the users and relays [57]. We consider that the decoding order of the received signal is from  $f_0$  to  $f_M$  according to the importance and popularity.

### B. NOMA Power Allocation Strategy

If  $f_0$  could be successfully decoded on the  $m$ -th aerial relay, the  $m$ -th aerial relay can retransmit  $f_0$  to the user  $U_{m,k}$  with less delay even the frequent heavy shadowing happens between the satellite and the user  $U_{m,k}$ . Therefore, the following analysis is based on this strategy, which is named the QoS requirement strategy.

If the  $\rho G(d_m) < R^0$ , the file  $f_0$  cannot be decoded in the  $m$ -th aerial relay. Therefore, the satellite would allocate all power to the  $f_0$  to decrease the outage probability of the link between the satellite and the user  $U_{m,k}$ . If the  $\rho G(d_m) = R^0$ , all power is allocated to the  $f_0$  too. If the  $\rho G(d_m) > R^0$ , the remaining power could be allocated to other files to increase the spectrum efficiency.

Accordingly, the transmission rate of  $f_0$  in  $m$ -th aerial relay is given as

$$R_m^0 = \log \left( 1 + \frac{\alpha_0 G(d_m)}{\sum_{i=1}^M \alpha_i G(d_m) + \frac{1}{\rho}} \right). \quad (8)$$

Therefore, the power allocation coefficient of  $f_0$  in the satellite is

$$\alpha_0 = \min \left\{ 1, \frac{\varepsilon_0 [\rho G(d_m) + 1]}{\rho G(d_m) (\varepsilon_0 + 1)} \right\}. \quad (9)$$

Furthermore, the remain power coefficient  $\alpha^r$  of other files is given as

$$\alpha^r = \max \left\{ 0, \frac{\rho G(d_m) - \varepsilon_0}{\rho G(d_m) (\varepsilon_0 + 1)} \right\}. \quad (10)$$

To simplify the notation, the new power allocation power coefficient  $\tau_n$  denotes the ratio of  $f_n$  in  $P_r$  for  $n \in [1, \dots, M]$ , where  $\alpha_n = \tau_n \alpha^r$ . We consider  $\tau_n$  is fixed, which can be predefined by the priority or the popularity.

*Lemma 1:* The distance from  $m$ -th aerial relay and  $n$ -th aerial relay to the beam center are  $d_m$  and  $d_n$ , respectively. The joint PDF of  $d_m$  and  $d_n$  is given by

$$f_{d_m, d_n}(x, y) = 4y(\lambda\pi)^n e^{-\lambda\pi y^2} \frac{x^{2n-1}(y^2 - x^2)^{m-n-1}}{(m-n-1)!(n-1)!}. \quad (11)$$

*Proof:* See Appendix A.

According to the QoS requirement of the  $m$ -th aerial relay, we would investigate the performance of the aerial relay first.

### C. The Performance of the Relay

In the QoS requirement strategy,  $f_0$  is required to be successfully decoded in the  $m$ -th relay. Therefore, the outage and hit probability are highly related to the probabilities of  $\alpha_0$ , and their relation is summarized in *Theorem 1* below. The outage probability is defined as the probability that the instantaneous SINR  $\gamma_k$  falls below a predefined threshold  $\gamma_{th}$ , i.e.,

$$P_{out}(\gamma_{th}) = P(\gamma_k < \gamma_{th}), \quad (12)$$

where  $P(\cdot)$  is the probability of the corresponding event. Furthermore, the hit probability of  $n$ -th relay for the content category  $\mathcal{F}$  is defined as

$$P_n^{hit} = \sum_{i=1}^M P_n(f_i) (1 - P_n^i), \quad (13)$$

where  $P_n^i$  is the outage probability of  $f_i$ .

*Theorem 1:* If the power allocation coefficient  $\alpha_0$  makes sure  $f_0$  exactly decodable in the  $m$ -th aerial relay, the  $n$ -th aerial relay can successfully decode the  $f_0$  for  $\forall n < m$  and the  $p$ -th relay cannot decode the  $f_0$  for  $\forall p > m$ .

*Proof:* The transmission rate of  $f_0$  for  $k$ -th relay is rewritten as

$$R_k^0 = \log \left( 1 + \frac{\alpha_0}{1 - \alpha_0 + \frac{1}{G(d_k)\rho}} \right). \quad (14)$$

It can be observed in (14) that  $R_k^0$  decreases monotonically with  $d_k$ . Due to  $n < m < p$ , and therefore  $d_n < d_m < d_p$ , we have  $R_n^0 > R_m^0 > R_p^0$ . According to the assumption  $R_m^0 =$

$R^0$ , there exist  $R_n^0 > R^0$  and  $R_p^0 < R^0$ , i.e. the  $n$ -th relay can successfully decode  $f_0$  while the  $p$ -th relay cannot decode  $f_0$ . *Theorem 1* is proved.

If  $f_0$  cannot be decoded, the hit probability of each file for  $p$ -th relay ( $\forall p > m$ ) would be 0 since SIC can not achieve. Therefore, we mainly focus on the hit probability for  $m$ -th relay and  $n$ -th relay ( $\forall n < m$ ).

*Lemma 2:* The outage probability of  $m$ -th cache-enabled aerial relay of the  $f_0$  is given by

$$P_m^0 = e^{-\lambda\pi \left(G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^2} \sum_{k=0}^{m-1} \frac{(\lambda\pi)^k \left(G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^{2k}}{k!}, \quad (15)$$

and the outage probability of  $m$ -th cache-enable relay of  $f_i$  for ( $i \in [1, \dots, M]$ )

$$P_m^i = e^{-\lambda\pi \left(G^{-1}\left(\frac{\varepsilon_0}{\rho} + \frac{(1+\varepsilon_0)\eta_i}{\rho}\right)\right)^2} \times \sum_{k=0}^{m-1} \frac{(\lambda\pi)^k \left(G^{-1}\left(\frac{\varepsilon_0}{\rho} + \frac{(1+\varepsilon_0)\eta_i}{\rho}\right)\right)^{2k}}{k!}, \quad (16)$$

where  $\eta_i = \max \left\{ \frac{\varepsilon_1}{\omega_1}, \dots, \frac{\varepsilon_i}{\omega_i} \right\}$ , and  $\omega_i = \tau_i - \varepsilon_i \sum_{j=i+1}^M \tau_j$ .

*Proof:* See Appendix B.

*Lemma 3:* If  $\varepsilon_M \geq \varepsilon_0$ , the outage probability of  $n$ -th cache-enabled relay of the  $f_0$  is given by

$$P_n^0 = e^{-\lambda\pi \left(G_n^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^2} \sum_{k=0}^{n-1} \frac{(\lambda\pi)^k \left(G_n^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^{2k}}{k!}, \quad (17)$$

and the outage probability of  $n$ -th cache-enabled relay of the  $f_i$  for ( $i \in [1, \dots, M]$ ) is given by

$$P_n^i = P_m^0 + \frac{4(\lambda\pi)^n e^{-\lambda\pi y^2}}{(m-n-1)!(n-1)!} \sum_{k=0}^{t-m-1} (-1)^k \binom{t-m-1}{k} \times \sum_{l=1}^N \frac{\pi(\varphi_2 - \varphi_1)}{2N} \sqrt{1 - \omega_l^2} Z \left( \frac{\varphi_2 - \varphi_1}{2} \omega_l + \frac{\varphi_2 + \varphi_1}{2} \right), \quad (18)$$

where  $\varphi_1 = G^{-1}\left(\frac{(1+\varepsilon_0)\eta_i + \varepsilon_0}{\rho}\right)$ ,  $\varphi_2 = G^{-1}(\varepsilon_0/\rho)$ ,  $\mu_1 = G^{-1}\left(\frac{(1+\varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(y)}}\right)$ ,  $Z(y)$  is given by

$$Z(y) = \frac{e^{-\lambda\pi y^2} y^{2n-2m-2k-2}}{2n+2k} \left( y^{2n+2k} - \mu_1(y)^{2n+2k} \right). \quad (19)$$

$N$  is the Chebyshev-Gauss approximation parameter, and  $\omega_l = \cos\left(\frac{2l-1}{2N}\right)$ .

*Proof:* See Appendix C.

Based on the given outage probability  $P_n^i$  in *Lemma 3*, the corresponding cache hit probability for a user associated with  $R_k$  can be derived by (13).

### D. The Outage Probability of the User

The outage probability of the user for the QoS requirement strategy is determined by the NOMA power allocation coeffi-

cient. When  $\alpha_0 = 1$ , the outage probability of the user  $U_{m,k}$  is given as

$$\begin{aligned} P_{m,k}^{out} &= P\left(\rho G(d_{m,k}) |h_{m,k}|^2 < \varepsilon_0\right) \\ &= P\left(|h_{m,k}|^2 < \frac{\varepsilon_0}{\rho G(d_{m,k})}\right) \\ &= \mathbb{E}\left[F_{|h|^2}\left(\frac{\varepsilon_0}{G(d_{m,k})\rho}\right)\right]. \end{aligned} \quad (20)$$

where  $d_{m,k}$  is the distance between the  $U_{m,k}$  and the beam center,  $h_{m,k}$  denotes the fading between the satellite and  $U_{m,k}$ .

Conditioned on  $d_m$ , the PDF of the  $d_{m,k}$  is given by [45]

$$f_{d_{m,k}|d_m}(d) = \begin{cases} \frac{2\pi d}{\pi R^2} & d \leq R - d_m \\ \frac{2\pi d}{\pi R^2} - \Psi(d, d_m) & R - d_m < d \leq \sqrt{R^2 - d_m^2} \\ \Psi(d, d_m) & \sqrt{R^2 - d_m^2} < d < R + d_m \end{cases} \quad (21)$$

where  $\Psi(d, d_m)$  is defined as

$$\Psi(d, d_m) = \frac{2d \arccos \frac{d_m^2 + d^2 - R^2}{2dd_m}}{\pi R^2}. \quad (22)$$

According to the conditional probability of (21), the PDF of  $d_{m,k}$  is derived by obtaining the PDF of  $d_m$ . To avoid the complexity caused by segmented integration, we consider  $d = 0$ , i.e., the user is located at the beam center, and no aerial relay is located inside the circle  $\mathfrak{R}(o, \kappa\mathcal{R})$ , i.e., a circle with the beam center located at its origin and radius  $\kappa\mathcal{R}$  with  $\kappa > 1$ , and  $d_m \geq \kappa\mathcal{R}$  for all  $m \geq 1$ .  $\mathcal{A}$  represents the ring with inner radius  $\kappa\mathcal{R}$  and outer radius  $d$ . Therefore, the outage probability of the user  $U_{m,k}$  is given by

$$P_{m,k}^{out} = \int_{\kappa R}^{\infty} \int_{y-R}^{y+R} F_{|h|^2}\left(\frac{\varepsilon_0}{G(x)\rho}\right) \Psi(x, y) \tilde{f}_{d_m}(y) dx dy, \quad (23)$$

where  $\tilde{f}_{d_m}$  denotes the PDF of  $d_m$ . Since the cache-enabled aerial relays follow a HPPP, the complementary cumulative distribution function (CDF) of  $d_m$  is the probability that there are less than  $m$  relays closer than  $d$  in  $\mathcal{A}$ , which is given by

$$\begin{aligned} \tilde{F}_{d_m}(d) &= 1 - P(|\mathcal{A}| < m) \\ &= 1 - \sum_{i=1}^{m-1} \frac{[\lambda_c (\pi d^2 - \pi \delta^2 \mathcal{R}^2)]^i}{i!} e^{-\lambda_c (\pi d^2 - \pi \delta^2 \mathcal{R}^2)}, \end{aligned} \quad (24)$$

where  $|*|$  denotes the number of relays in the corresponding areas. Furthermore, the complementary PDF of  $d_m$  can be calculated as<sup>6</sup>

$$\tilde{f}_{d_m}(d) = 2\pi \lambda_c^m d e^{-\lambda_c (\pi d^2 - \pi \delta^2 \mathcal{R}^2)} \frac{(\pi d^2 - \pi \delta^2 \mathcal{R}^2)^{m-1}}{(m-1)!}. \quad (25)$$

Substituting (25) into (23), the outage probability of the user can be obtained for  $\alpha_0 = 1$ .

<sup>6</sup>The difference between (25) and (35) is due to the assumption that the user is located in the beam center to avoid the trivial piecewise integration.

When  $\alpha_0 < 1$ , the outage probability of the user  $U_{m,k}$  is given as

$$\begin{aligned} P_{m,k}^{out} &= P\left(\frac{\alpha_0}{1 - \alpha_0 + \frac{1}{\rho G(d_{m,k})|h_{m,k}|^2}} < \varepsilon_0\right) \\ &= P\left(\frac{\frac{\varepsilon_0[\rho G(d_m)+1]}{\rho G(d_m)(\varepsilon_0+1)}}{\frac{\rho G(d_m)-\varepsilon_0}{\rho G(d_m)(\varepsilon_0+1)} + \frac{1}{\rho G(d_{m,k})|h_{m,k}|^2}} < \varepsilon_0\right) \\ &= P\left(\frac{G(d_{m,k})|h_{m,k}|^2}{\varepsilon_0[G(d_m)-G(d_{m,k})|h_{m,k}|^2]+G(d_{m,k})|h_{m,k}|^2} < 1\right) \\ &= P\left(|h_{m,k}|^2 < \frac{G(d_m)}{G(d_{m,k})}\right) \\ &= \mathbb{E}\left[F_{|h|^2}\left(\frac{G(d_m)}{G(d_{m,k})}\right)\right]. \end{aligned} \quad (26)$$

Substituting (26) into the corresponding, the outage probability of the user for  $\alpha_0 < 1$  is given as

$$P_{m,k}^{out} = \int_{\kappa R}^{\infty} \int_{y-R}^{y+R} F_{|h|^2}\left(\frac{G(y)}{G(x)}\right) \Psi(x, y) \tilde{f}_{d_m}(y) dx dy. \quad (27)$$

**Theorem 2:** The outage probability of the user  $U_{m,k}$  is the same for NOMA and TDMA in the QoS requirement strategy. And the outage probability of the user  $U_{m,k}$  is irrelevant to the total transmission power of the satellite when the QoS requirement is satisfied.

*Proof:* For the situation  $\alpha_0 = 1$ , all power is allocated into  $f_0$ , therefore the NOMA scheme is TDMA scheme. For the situation  $\alpha_0 < 1$ , the outage probability of the TDMA is given by

$$\begin{aligned} P_{m,k}^{out} &= P\left(R^0 \frac{\log_2(1+\rho G(d_{m,k})|h_{m,k}|^2)}{\log_2(1+\rho G(d_m))} < R^0\right) \\ &= P\left(\frac{G(d_{m,k})|h_{m,k}|^2}{G(d_m)} < 1\right) \\ &= P\left(G(d_m) > G(d_{m,k})|h_{m,k}|^2\right) \\ &= P\left(|h_{m,k}|^2 < \frac{G(d_m)}{G(d_{m,k})}\right) \\ &= \mathbb{E}\left[F_{|h|^2}\left(\frac{G(d_m)}{G(d_{m,k})}\right)\right]. \end{aligned} \quad (28)$$

Therefore, the outage probability of the user  $U_{m,k}$  is the same for NOMA and TDMA. According to (26) and (28), the outage probability of the user  $U_{m,k}$  is irrelevant to the total transmission power of the satellite.

Hence, *Theorem 2* is proved.

#### IV. PERFORMANCE ANALYSIS FOR RETRIEVING CONTENT FROM AERIAL RELAY

In this section, we consider that the required content file is cached in the aerial relay. Hence, the aerial relay would serve the user to reduce the transmission latency, and the NOMA scheme is implemented to increase the spectrum efficiency.

##### A. NOMA Signal Transmission

To reduce the transmission latency of retrieving the required content, the user would request the content files from its aerial server. If the required content file is cached in the aerial relay,

the aerial relay would serve the user directly. We assume that each relay can find at least two users whose requests can be accommodated in its coverage area. By employing the NOMA scheme, the other aerial relay would cause interference to the users<sup>7</sup>. Without loss of generality, we consider that  $Z_P$  users are required to be served in one resource block by aerial relay  $R_P$ . To improve the spectrum efficiency, the frequency reuse is implemented among aerial to user links. Therefore, the received mixed signal can be expressed as

$$y_{P,q} = \underbrace{P_A H_{P,q} \sum_{l_P=1}^{Z_P} \alpha_{l_P}^A x_{l_P}}_{\text{Signals}} + \underbrace{\sum_{J \in \Phi_c \setminus P} P_J H_{J,q} \sum_{l_J=1}^{Z_J} \alpha_{l_J}^A x_{l_J}}_{\text{Interference}} + n_1, \quad (29)$$

where  $P_A$  is the transmitted power of  $P$ -th aerial relay,  $P_I$  is the transmitted power of other aerial relays, where the transmission power of other relays is the same to simplify the expression,  $\Phi_c$  is the HPPP set of aerial relay,  $\alpha_{l_X}^A$  is the NOMA power allocation for  $l$ -th user of  $X$ -th relay, where  $\sum_{l_x=1}^{Z_x} \alpha_{l_x}^A = 1$ ,  $H_{x,y}$  is the link between node  $x$  and node  $y$ , and  $n_1$  is the additive white Gaussian noise (AWGN) with mean zero and variance  $\sigma_0^2$ ,  $x_{l_X}$  denotes the signal for  $l$ -th user of  $X$ -th relay, where  $E[|x_{l_X}|^2] = 1$ , and  $E[\cdot]$  means the expectation operation.

### B. Outage Probability of The Users

To obtain tractable analytical results, we consider that the SIC decoding order is from  $x_{1_P}$  to  $x_{Z_P}$  and the NOMA power allocation coefficient is predefined. In such a general case, the transmission rate of user  $U_{P,m}$  is given by

$$R_{U_{P,m}} = \log \left( 1 + \frac{\rho_A H_{P,m} \alpha_m^A}{\sum_{l=m+1}^{Z_P} \rho_A H_{P,m} \alpha_l^A + \mathbf{I}_{\text{inter}} + 1} \right), \quad (30)$$

where  $\rho_A = \frac{P_A}{\sigma_1^2}$ ,  $\rho_I = \frac{P_I}{\sigma_1^2}$  and  $\mathbf{I}_{\text{inter}} = \sum_{J \in \Phi_c \setminus P} \rho_I H_{J,m}$ .

The successfully decoding probability of user  $U_{P,m}$  is given in (31) on the next page, where  $\varepsilon_{P,i}^A = 2^{R_{P,i}^A} - 1$ ,  $R_{P,i}^A$  is the required transmission rate of  $i$ -th user in  $P$ -th aerial relay coverage area, and  $\xi_{P,m} = \max \left\{ \frac{\varepsilon_{P,i}^A}{\alpha_i^A - \varepsilon_{P,1}^A \sum_{l=i+1}^{Z_P} \alpha_l^A}, \forall i \in (1, \dots, m) \right\}$ ,  $\mathcal{L}_{\mathbf{I}}(s)$  is the Laplace transform (LT) of  $\mathbf{I}$ , and  $\mathcal{L}_{\mathbf{I}}^{(i)}(s)$  is  $i$ -th derivative of  $\mathcal{L}_{\mathbf{I}}(s)$ . The radio signal attenuation follows LoS propagation with a lower path-loss exponent and Nakagami- $m$  small-scale fading for the aerial-terrestrial link, and the receive side signal-to-noise ratio (SNR)  $\rho H_{P,q}$  follows gamma distributions [54].

<sup>7</sup>Here we consider the general situation, which constitutes the worst case for the reception reliability of users. If no user requests the content for one aerial relay, there exists no interference. If the number of the user is large, the available method is that the users are divided into several clusters by implementing NOMA in different resource blocks.

TABLE I  
SIMULATION PARAMETERS

Parameters	Values
Center frequency	4 GHz
Satellite height	800 Km
Aerial relays distribution radius $\mathcal{D}$	200 Km
Aerial relay coverage radius $\mathcal{R}$	5 Km
Half-power beamwidth of satellite $L$	50 Km
Satellite antenna gain $G_s$	20 dBi
Cluster heads antenna gain $G_t$	6 dBi
PRs antenna gain $G_s$	5 dBi
Terrestrial noise temperature	300 K
Satellite noise temperature	350 K

According to (5),  $\theta_{P,q}$  is defined as  $\rho D_{P,q}^{-\alpha_L} / m_{P,q}$ . Therefore, the step (a) in (31) apply the CDF of gamma distribution. The step (b) in (31) is due to  $E_{\mathbf{I}}[\mathbf{I}^i \exp(-s\mathbf{I})] = (-1)^i \mathcal{L}_{\mathbf{I}}^{(i)}(s)$ , where  $\mathcal{L}_{\mathbf{I}}^{(i)}(s) = \left[ \frac{d^i}{ds^i} \mathcal{L}_{\mathbf{I}}(s) \right]_{s=s_0}$ .

To calculate  $\mathcal{L}_{\mathbf{I}}^{(i)}(s)$ , we utilize the Faa di Bruno formula [54], which is expressed as (32) in the next page, to deal with the higher order derivatives of the composite function by calculating the LT of the  $\mathbf{I}_{\text{inter}}$ .  $\mathcal{L}_{\mathbf{I}_{\text{inter}}}(s)$  is given as (33) on the next page, where  $D_{J,q}$  is distance between the other aerial relay and the user. The step (c) in (33) utilizes the probability generating functional (PGFL) of the HPPP. The  $i$ -th derivative is formulated as (34) on the next page, where  $\mathcal{N}_i = \left\{ (n_1, \dots, n_\tau, \dots, n_i) \mid n_\tau \in \mathbb{N}, \sum_{\tau=1}^i \tau n_\tau = i \right\}$ .

## V. NUMERICAL SIMULATIONS

In this section, representative simulation results are presented to confirm the effectiveness of the considered HSATN systems and evaluate the impact of key factors on the performance of the proposed framework. The aerial relay distribution radius  $\mathcal{D}$ , the height of the aerial relay, and the aerial relay coverage area radius  $\mathcal{R}$  are 200 kilometers (Km), 5Km and 5 Km, respectively, where the users are randomly distributed in the aerial relay coverage area. The key system parameters are provided in Table I.

### A. Simulation Results for Retrieving the Content Files from the Satellite

We consider that the file category  $\mathcal{F} = \{f_1, f_2, f_3\}$ , and transmission rate requirement are  $\{0.75, 0.875, 1\}$  (bps/Hz), respectively<sup>8</sup>. The required rate for  $f_0$  of the user  $U_{m,k}$  is 0.6 bps/Hz.

As  $\alpha_0$  is determined by the channel gain of  $m$ -th aerial relay, the ratio  $\tau_i$  of the remaining power for  $\{f_1, f_2, f_3\}$  are  $\{0.5, 0.3, 0.2\}$ , respectively. Besides, the OMA scheme is provided for comparison, where the time slot allocation for  $f_0$  is determined by the QoS requirement of the user  $U_{m,k}$ , and the rest time slot allocation is proportionally allocated according to the rate requirements of each file<sup>9</sup>.

<sup>8</sup>Due to the existence of interference in NOMA scheme, a single time slot is only available for limited content files. If the number of content files is large, the content files can be divided into different groups for different time slots.

<sup>9</sup>In the TDMA, each user occupies orthogonal specific time blocks in practice system. For a clear comparison with NOMA, we consider the time blocks can be divided into several sub-time slots during the simulation.



$$\begin{aligned}
 P_{U_{P,m}} &= P \left( \frac{\rho_A H_{P,q} \alpha_1^A}{\sum_{l=2}^{Z_P} \rho_A H_{P,q} \alpha_l^A + \mathbf{I}_{\text{inter}} + 1} > \varepsilon_{P,1}^A, \dots, \frac{\rho_A H_{P,q} \alpha_m^A}{\sum_{l=m+1}^{Z_P} \rho_A H_{P,q} \alpha_l^A + \mathbf{I}_{\text{inter}} + 1} > \varepsilon_{P,m}^A \right) \\
 &= P \left( \rho_A H_{P,q} > \frac{\varepsilon_{P,1}^A (\mathbf{I}_{\text{inter}} + 1)}{\alpha_1^A - \varepsilon_{P,1}^A \sum_{l=2}^{Z_P} \alpha_l^A}, \dots, \rho_A H_{P,q} > \frac{\varepsilon_{P,m}^A (\mathbf{I}_{\text{inter}} + 1)}{\alpha_m^A - \varepsilon_{P,1}^A \sum_{l=m+1}^{Z_P} \alpha_l^A} \right) \\
 &= P(\rho_A H_{P,q} > \xi_{P,m} (\mathbf{I}_{\text{inter}} + 1)) \\
 &\stackrel{a}{=} E_{\mathbf{I}_{\text{inter}}} \left[ \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} (\mathbf{I}_{\text{inter}} + 1) \right) \sum_{k=0}^{m^A-1} \frac{1}{k!} \left( \frac{\xi_{P,m}}{\theta_{P,m}} (\mathbf{I}_{\text{inter}} + 1) \right)^k \right] \\
 &= \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} \right) E_{\mathbf{I}_{\text{inter}}} \left[ \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} \mathbf{I}_{\text{inter}} \right) \sum_{k=0}^{m^A-1} \frac{1}{k!} \left( \frac{\xi_{P,m}}{\theta_{P,m}} \right)^k \sum_{i=0}^k \binom{k}{i} \mathbf{I}_{\text{inter}}^i \right] \\
 &= \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} \right) \sum_{k=0}^{m^A-1} \frac{1}{k!} \left( \frac{\xi_{P,m}}{\theta_{P,m}} \right)^k \sum_{i=0}^k \binom{k}{i} E_{\mathbf{I}_{\text{inter}}} \left[ \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} \mathbf{I}_{\text{inter}} \right) \left( \frac{\xi_{P,m}}{\theta_{P,m}} \mathbf{I}_{\text{inter}} \right)^i \right] \\
 &\stackrel{b}{=} \exp \left( -\frac{\xi_{P,m}}{\theta_{P,m}} \right) \sum_{k=0}^{m^A-1} \left( \frac{\xi_{P,m}}{\theta_{P,m}} \right)^k \sum_{i=0}^k \binom{k}{i} \frac{(-1)^i}{k!} \mathcal{L}_{\mathbf{I}_{\text{inter}}}^{(i)} \left( \frac{\xi_{P,m}}{\theta_{P,m}} \right).
 \end{aligned} \tag{31}$$

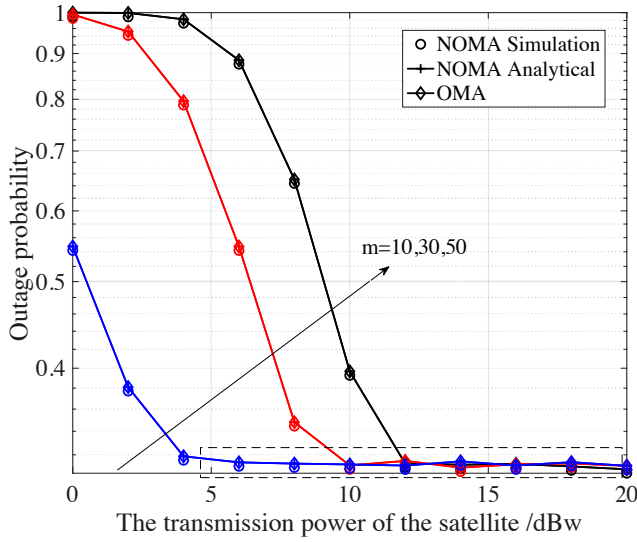
$$f^{(\kappa)}(g(s)) = \sum_{(n_1, \dots, n_\kappa) \in \mathcal{N}_\kappa} \frac{\kappa!}{\prod_{\tau=1}^{\kappa} (n_\tau!)} f^{(n_1 + \dots + n_\kappa)}(g(s)) \prod_{\tau=1}^{\kappa} \left( \frac{g^{(\tau)}(s)}{\tau!} \right)^{n_\tau}. \tag{32}$$

$$\begin{aligned}
 \mathcal{L}_{\mathbf{I}_{\text{inter}}}(s) &= E \left( \exp \left( -s \sum_{J \in \Phi_c \setminus P} \rho_I H_{J,q} \right) \right) = E_{\Phi_c \setminus P} \left( \exp \left( -s \sum_{J \in \Phi_c \setminus P} \rho_I D_{J,q}^{-\alpha^A} |h_{J,q}|^2 \right) \right) \\
 &\stackrel{c}{=} E_{\Phi_c \setminus P} \left[ \prod_{J \in \Phi_c \setminus P} E \left[ \frac{1}{\left( 1 + \frac{s \rho_I}{m^A} D_{J,q}^{-\alpha^A} \right)^{m^A}} \right] \right] = \exp \left[ -2\pi\lambda \int_{H_a}^{\infty} \frac{\sum_{n=1}^{m^A} \binom{m^A}{n} \left( \frac{s \rho_I}{m^A} \right)^n z^{-n\alpha^A}}{\left( 1 + \frac{s \rho_I}{m^A} D_{J,q}^{-\alpha^A} \right)^{m^A}} z dz \right] \\
 &= \exp \left( -2\pi\lambda \sum_{n=1}^{m^A} \binom{m^A}{n} \left( \frac{s \rho_I}{m^A} \right)^n \frac{H_a^{2-n\alpha^A}}{n\alpha^A-2} {}_2F_1 \left( m^A, n - \frac{1}{\alpha^A}, n - \frac{1}{\alpha^A} + 1, \frac{s \rho_I}{m^A H_a^{\alpha^A}} \right) \right).
 \end{aligned} \tag{33}$$

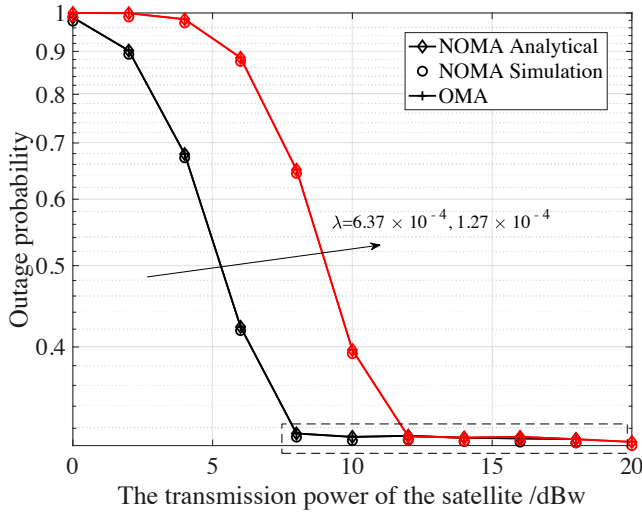
$$\begin{aligned}
 \mathcal{L}_{\mathbf{I}_{\text{inter}}}^{(i)}(s) &= \sum_{(n_1, \dots, n_i) \in \mathcal{N}_i} \frac{i!}{\prod_{\tau=1}^i (n_\tau!)} \mathcal{L}_{\mathbf{I}_{\text{inter}}}(s) \\
 &\times \prod_{\tau=1}^i \left( 2\pi\lambda (-1)^\tau \binom{m^A-1+i}{m^A-1} \left( \frac{\rho_I}{m^A} \right)^\tau \frac{H_a^{2-n\alpha^A}}{n\alpha^A-2} {}_2F_1 \left( m^A + \tau, \tau - \frac{2}{\alpha^A}, \tau - \frac{2}{\alpha^A} + 1, \frac{s \rho_I}{m^A H_a^{\alpha^A}} \right) \right)^{n_\tau}.
 \end{aligned} \tag{34}$$

Fig. 2 illustrates the  $f_0$  outage probability of the user  $U_{m,k}$  for the different orders of  $m$ . The shadowing and channel fading is considered as ILS, where the corresponding channel coefficient  $(c, q, \Omega)$  in (3) are given as (0.158, 19.4, 1.29), respectively [40], [49]–[51]. The analytical results match the simulation results, which verifies the accuracy of the outage probability analysis. When the power is insufficient for de-

coding the  $f_0$  in the aerial relay  $R_m$ , i.e.  $\alpha_0 = 1$ , the outage probability of the user is decreasing with the transmitted power increasing. In this situation, the outage probability of the user achieves superior performance with a shorter distance from the beam center. The main reason is that the antenna gain of the satellite is large for the user closing to the beam center, which decreases the outage probability of the user. In addition, the



**Fig. 2:** The outage probability of the user  $U_{m,k}$  versus different  $m$ , where  $\lambda = 1.27 \times 10^{-4}/km^2$ .

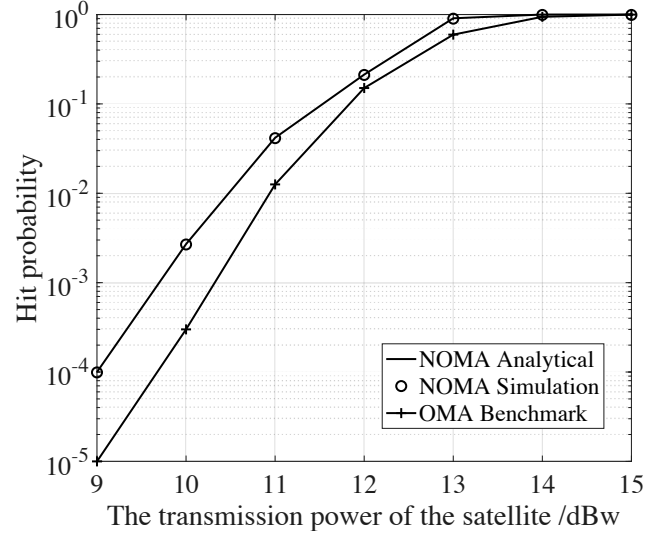


**Fig. 3:** The outage probability of the user  $U_{m,k}$  located in 50-th aerial relay coverage versus different intensity  $\lambda$ , where  $\lambda = 1.27 \times 10^{-4}/km^2$  and  $\lambda = 6.37 \times 10^{-4}/km^2$ , respectively, and the shadowing and channel fading is ILS.

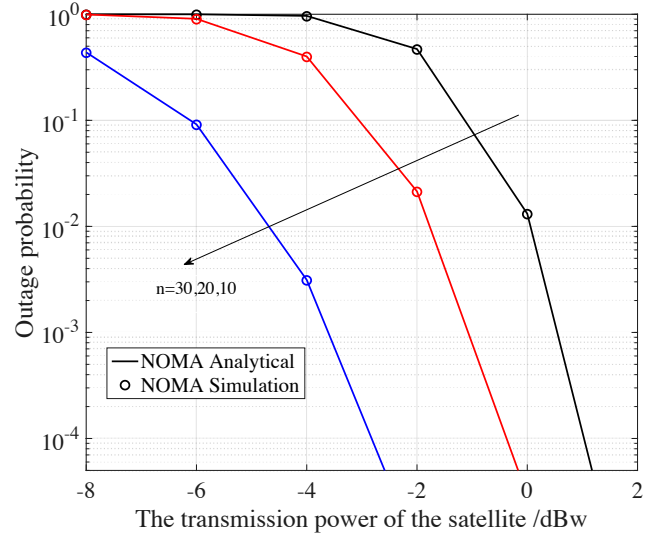
user's outage probability for  $f_0$  of the NOMA and TDMA is the same in the QoS requirement strategy, which proves the correctness of the *Theorem 2*.

Fig. 3 shows the outage probability of the user under the different intensity of the aerial relay. When the power is not sufficient for decoding the  $f_0$  in the aerial relay  $R_m$ , i.e.  $\alpha_0 = 1$ , the dense intensity would decrease the outage probability. When the QoS of the  $f_0$  is satisfied in  $R_m$ , the intensity of the aerial relay has only a limited impact on the outage probability of the user  $U_{m,k}$  as shown in the dotted rectangle area of Fig. 2 and Fig. 3.

Fig. 4 depicts the hit probability of 50-th aerial relay for the content category  $\mathcal{F}$ . Although the outage probability of the user  $U_{m,k}$  is the same for NOMA and TDMA, the hit



**Fig. 4:** The hit probability of the aerial relay  $R_m$ , where  $m = 50$ ,  $\lambda = 7.96 \times 10^{-4}/km^2$  and  $\gamma_m = 0.5$ .

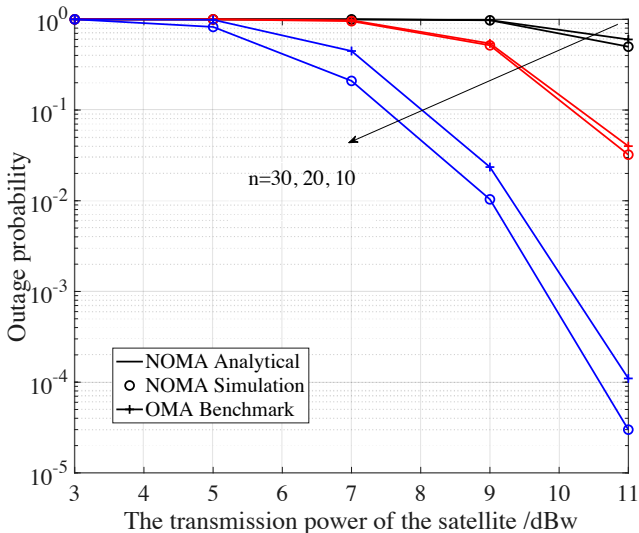


**Fig. 5:** The outage probability of  $f_0$  in the aerial relay  $R_n$  for different  $n$ , where  $m = 50$  and  $\lambda = 7.96 \times 10^{-4}/km^2$ .

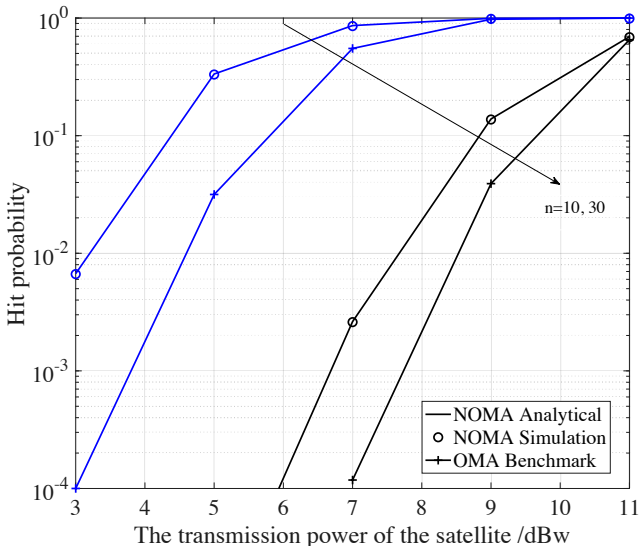
probability of content category  $\mathcal{F}$  in the NOMA scheme is superior to the OMA benchmark for  $R_m$ , which indicates the effectiveness of the NOMA scheme.

Fig. 5 shows the  $f_0$  outage probability of the aerial  $R_n$  within different order  $n$  when the  $f_0$  cannot be decoded in  $R_m$  under the transmitted power. As can be observed from the figure, increasing  $n$  will ascend the outage probability due to the larger distance from the beam center. When the  $f_0$  is decoded in  $R_m$ , the outage probability of the  $f_0$  in the aerial relay  $R_n$  equals zero. The outage probability of  $f_2$  in the aerial relay  $R_n$  for different order  $n$  is given in Fig. 6. It can be observed that the user achieves superior performance since the shorter distance from the beam center and results in higher beam gain.

Fig. 7 and Fig. 8 depict the hit probability for the content



**Fig. 6:** The outage probability of  $f_2$  in the aerial relay  $R_n$  for different  $n$ , where  $m = 50$  and  $\lambda = 7.96 \times 10^{-4}/km^2$ .

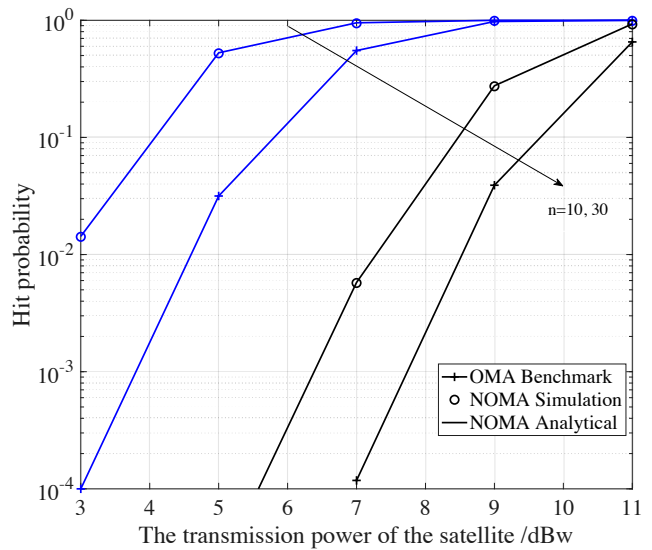


**Fig. 7:** The hit probability of the aerial relay  $R_n$  for the different aerial relay intensity and  $n$ , where  $\gamma_n = 0.5$ ,  $m = 50$  and  $\lambda = 7.96 \times 10^{-4}/km^2$ .

category  $\mathcal{F}$  of the aerial relay within the different content popularity parameter  $\gamma$  and order  $n$ . Since the expected hit probability of the OMA scheme is the same for different  $\gamma$ , we only plot one curve of TDMA for instance. A large  $\gamma$  means the requests on the high popularity files, whereas a small  $\gamma$  is related to the requests with heavy-tailed popularity. The NOMA scheme decodes the most popular content file at first. Therefore, the larger  $\gamma$  scenario obtains the better hit probability.

### B. Simulation Results for Retrieving the Content Files From the Cache-enabled Relay

In this subsection, we evaluate the outage probability of the users for retrieving the content from the relay. We consider



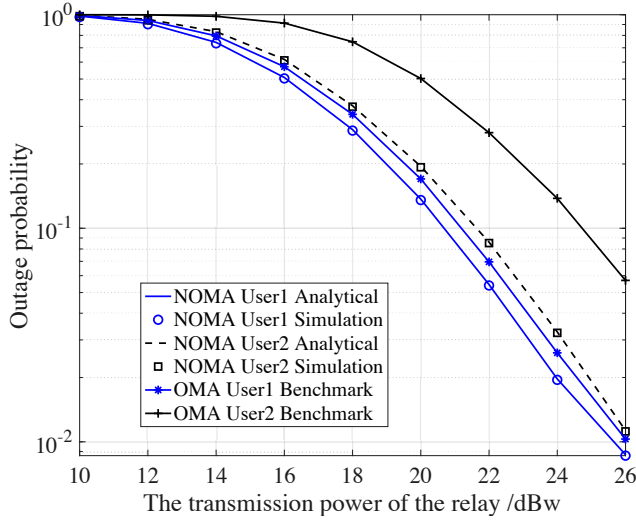
**Fig. 8:** The hit probability of the aerial relay  $R_n$  for the different aerial relay intensity and  $n$ , where  $\gamma_n = 5$ ,  $m = 50$  and  $\lambda = 7.96 \times 10^{-4}/km^2$ .

that the required content files are cached in the aerial relay. Therefore, the aerial relay serves the two users by the NOMA scheme simultaneously. The power allocation coefficient is 0.75 for far user 1 and 0.25 for near user 2. The required transmission rates are  $\{1.5, 2\}$  (bps/Hz) respectively. The fading severity  $m_A$  of Nakagami- $m$  is set as 3, and the path loss exponent  $\alpha^A$  is set as 2 for all aerial-terrestrial link. Besides, the TDMA scheme is provided for comparison, where the same transmission power as the NOMA scheme and the sub-time slot allocation is averagely allocated to two users. The transmission power of the interference relays is set as 10 W.

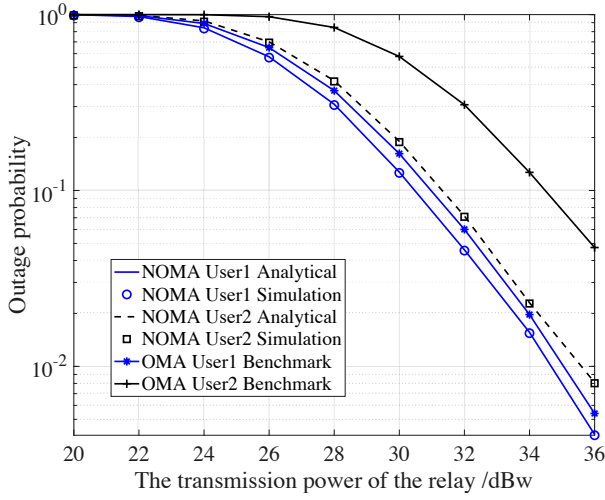
Fig. 9 and Fig. 10 depict the outage of the two users in 10-th aerial relay coverage area for the different dense intensity of the aerial relay. Both figures show that the derived outage probability results perfectly match the simulation results for two users, which confirms that our derivations are valid. The outage probability of the NOMA scheme for two users is superior to the OMA scheme, which implies the superiority of the NOMA scheme. In addition, the higher intensity of the aerial relays leads to higher outage probability, which is results from more interference from a larger number of aerial relays.

## VI. CONCLUSION

In this paper, to reduce the transmission latency and facilitate the frequent update of the files cached at the aerial relays, we investigated the performance of cache-enabled HSATN, where the user retrieved the required content files from the aerial relay or the satellite with the NOMA scheme. If the required content files of the user were cached in the aerial relay, the cache-enabled relay would serve the user directly. Otherwise, the user would retrieve the content file from the satellite, where the satellite seeks opportunities for proactive content pushing to the relay during the user content delivery phase. Specifically, taking into account the uncertainty of the



**Fig. 9:** The outage probability of the two users versus different transmission power of the cache-enabled relay, where  $\lambda = 1.27 \times 10^{-4} / km^2$ .



**Fig. 10:** The outage probability of the two users versus different transmission power of the cache-enabled relay, where  $\lambda = 1.27 \times 10^{-3} / km^2$ .

number and location of aerial relays, along with the channel fading of users, the outage probability and hit probability of the considered network were derived based on stochastic geometry. Finally, both simulation and analytical results were provided to validate the effectiveness of the cache-enabled HSATN with the NOMA scheme and proclaim the influence of key factors on the system performance.

#### APPENDIX A THE PROOF OF LEMMA 1

The number of cache-enabled relays is followed the HPPP within intensity  $\lambda$ , therefore the PDF of  $d_m$  is given by [45]

$$f_{d_m}(y) = \frac{2\lambda^m \pi^m y^{2m-1}}{(m-1)!} e^{-\lambda\pi y^2}. \quad (35)$$

Conditioned on  $d_n = x$ , the CDF of the  $d_m$  is given by

$$F_{d_m|d_n}(y) = P(d_m \leq y | d_n = x) = 1 - P(d_m > y | d_n = x) \quad (36)$$

The event  $(d_m > y | d_n = x)$  indicates that  $m$ -th cache-enabled relay is located in outer of the circle with radius  $y$  or inner of the circle with radius  $x$ . Without loss of general, we consider  $m > n$ . Therefore, there exist at most  $m - n - 1$  cache-enabled relay inside a ring with inner radius  $x$  and outer radius  $y$ , where the ring denotes by  $\mathfrak{R}(x, y)$ . Hence,  $F_{d_m|d_n}(y)$  is derived as

$$F_{d_m|d_n}(y) = 1 - \sum_{k=1}^{m-n-1} P(|\mathfrak{R}(x, y)| = k). \quad (37)$$

where  $|\mathfrak{R}(x, y)|$  represents the number of the cache-enabled located in ring  $\mathfrak{R}(x, y)$ . According to the property of HPPP,  $F_{d_m|d_n}(y)$  is given as

$$F_{d_m|d_n}(y) = 1 - \sum_{k=1}^{m-n-1} e^{-\lambda\pi(y^2-x^2)} \frac{\lambda^k \pi^k (y^2-x^2)^k}{k!} \quad (38)$$

We implement  $S_k = \frac{\lambda^k \pi^k (y^2-x^2)^k}{k!}$ . The PDF of  $f_{d_m|d_n}(y)$  can be obtained by the derivation of CDF  $F_{d_m|d_n}(y)$ . Based on [45], the conditional PDF for  $f_{d_m|d_n}(y)$  is derived as

$$\begin{aligned} f_{d_m|d_n}(y) &= 2y\lambda\pi e^{-\lambda\pi(y^2-x^2)} \left( \sum_{k=0}^{m-n-1} S_k - \sum_{k=1}^{m-n-1} S_{k-1} \right) \\ &= 2y(\lambda\pi)^{m-n} e^{-\lambda\pi(y^2-x^2)} \frac{x^{2m-1}(y^2-x^2)^{n-m-1}}{(m-n-1)!}. \end{aligned} \quad (39)$$

Therefore, the join PDF for  $d_m$  and  $d_n$  is obtained by Bayesian formula, which can be given by (11).

Therefore, the *lemma 1* is proved.

#### APPENDIX B THE PROOF OF LEMMA 2

According to (9), there exists a situation, where  $\alpha_0 = 1$ , the  $m$ -th relay still decodes the  $f_0$ . The PDF of  $d_m$  is given in (35). Therefore, the outage probability of  $f_0$  in  $m$ -th relay is given by

$$\begin{aligned} P_m^0 &= P\left(G(d_m) < \frac{\varepsilon_0}{\rho}\right) = P\left(d_m > G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right) \\ &= \frac{2(\lambda\pi)^m}{(m-1)!} \int_{G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)}^{\infty} y^{2m-1} e^{-\lambda\pi y^2} dy \\ &= e^{-\lambda\pi\left(G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^2} \sum_{k=0}^{m-1} \frac{(\lambda\pi)^k \left(G^{-1}\left(\frac{\varepsilon_0}{\rho}\right)\right)^{2k}}{k!}. \end{aligned} \quad (40)$$

To be noticed,  $G(d_m)\rho < \varepsilon_0$ , all power is allocated to the  $f_0$ . Hence,  $P\left(G(d_m) \leq \frac{\varepsilon_0}{\rho}\right)$  and  $P(\alpha_0 = 1)$  is the same. The outage probability of  $f_i$  in  $m$ -th relay is given by

$$P_m^i = P \left( \underbrace{\alpha_0 = 1, G(d_m) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\}}_{\text{event1}} \right) + P \left( \underbrace{\alpha_0 < 1, G(d_m) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\}}_{\text{event2}} \right), \quad (41)$$

where the  $G(d_m) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\}$  is always true for  $\alpha_0 = 1$  in *event 1*. Therefore  $P_{\text{event1}} = P(\alpha_0 = 1)$ . For the *event 2*, we have

$$\left\{ G(d_m) < \frac{\varepsilon_i}{\rho\zeta_i} \right\} = \left\{ G(d_m) < \frac{\varepsilon_i}{\frac{\rho G(d_m) - \varepsilon_0}{\rho\omega_i}} \right\} = \left\{ G(d_m) < \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\varepsilon_i}{\rho\omega_i} \right\}. \quad (42)$$

Hence, formula (41) can be rewritten as

$$P_m^i = P \left( G(d_m) < \frac{\varepsilon_0}{\rho} \right) + P \left( G(d_m) < \max \left\{ \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\varepsilon_k}{\rho\omega_k}, \forall k \in [1, \dots, i] \right\} \right) = P \left( d_m > G^{-1} \left( \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\eta_i}{\rho} \right) \right) = \frac{2(\lambda\pi)^m}{(m-1)!} \int_{G^{-1} \left( \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\eta_i}{\rho} \right)}^{\infty} y^{2m-1} e^{-\lambda\pi y^2} dy = e^{-\lambda\pi \left( G^{-1} \left( \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\eta_i}{\rho} \right) \right)^2} \times \sum_{k=0}^{m-1} \frac{(\lambda\pi)^k \left( G^{-1} \left( \frac{\varepsilon_0}{\rho} + \frac{(1 + \varepsilon_0)\eta_i}{\rho} \right) \right)^{2k}}{k!}, \quad (43)$$

where  $\eta_i = \max \left\{ \frac{\varepsilon_1}{\varsigma_1}, \dots, \frac{\varepsilon_i}{\varsigma_i} \right\}$ , and  $\varsigma_i = \tau_i - \varepsilon_i \sum_{j=i+1}^M \tau_j$ . Hence, *lemma 2* is proved.

### APPENDIX C

#### THE PROOF OF LEMMA 3

The premise of  $n$ -th relay decoding  $f_0$  is that  $m$ -th relay successfully decode the  $f_0$ . Therefore, the outage probability of  $n$ -th relay successfully decode  $f_0$  is given as

$$P_n^0 = P \left( G(d_m) < \frac{\varepsilon_0}{\rho}, G(d_n) < \frac{\varepsilon_0}{\rho} \right) = P \left( G(d_n) < \frac{\varepsilon_0}{\rho} \right) = P \left( d_n > G^{-1} \left( \frac{\varepsilon_0}{\rho} \right) \right) = e^{-\lambda\pi \left( G^{-1} \left( \frac{\varepsilon_0}{\rho} \right) \right)^2} \sum_{k=0}^{n-1} \frac{(\lambda\pi)^k \left( G^{-1} \left( \frac{\varepsilon_0}{\rho} \right) \right)^{2k}}{k!}. \quad (44)$$

The outage probability of  $n$ -th relay for  $f_i$  ( $\forall i \in [1, \dots, M]$ ) is given as

$$P_n^i = P \left( \underbrace{\alpha_0 = 1, G(d_n) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\}}_{\text{event3}} \right) + P \left( \underbrace{\alpha_0 < 1, G(d_n) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\}}_{\text{event4}} \right). \quad (45)$$

For *event 3*,  $\left\{ G(d_n) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (0, \dots, i) \right\} \right\}$  occurs, there must be  $\alpha_0 = 1$ . Therefore,  $P_{\text{event3}} = P(\alpha_0 = 1)$ . The probability of *event 4* can be rewritten as

$$P \left\{ G(d_m) > \frac{\varepsilon_0}{\rho}, G(d_n) < \max \left\{ \frac{\varepsilon_k}{\rho\zeta_k}, \forall k \in (1, \dots, i) \right\} \right\} = P \left\{ G(d_m) > \frac{\varepsilon_0}{\rho}, G(d_n) < \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G_m(d_m)}} \right\}. \quad (46)$$

Therefore, the probability of *event 4* is

$$P_4 = P \left( G(d_m) > \frac{\varepsilon_0}{\rho}, G(d_n) < \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(d_m)}} \right) = P \left( d_m < G^{-1} \left( \frac{\varepsilon_0}{\rho} \right), d_n > G^{-1} \left( \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(d_m)}} \right) \right). \quad (47)$$

If *event 4* always happens, there must exist

$$G^{-1} \left( \frac{\varepsilon_0}{\rho} \right) > G^{-1} \left( \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(d_m)}} \right). \quad (48)$$

Since  $G$  is monotonically decreasing,  $G^{-1}$  is monotonically increasing. We have

$$\frac{\varepsilon_0}{\rho} > \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(d_m)}}. \quad (49)$$

Furthermore, (49) is rewritten as

$$\frac{1}{G(d_m)} > \frac{\rho}{\varepsilon_0} \left[ 1 - \frac{(1 + \varepsilon_0)\eta_i}{\varepsilon_0} \right] = \frac{\rho}{\varepsilon_0} [1 - \Theta], \quad (50)$$

where  $\Theta$  can be formulated as

$$\Theta = \frac{(1 + \varepsilon_0) \max \left\{ \frac{\varepsilon_1}{\varsigma_1}, \dots, \frac{\varepsilon_M}{\varsigma_M} \right\}}{\varepsilon_0} \geq \frac{(1 + \varepsilon_0)\varepsilon_M}{\varepsilon_0\varsigma_M} = \frac{2^{R^0}\varepsilon_M}{\varepsilon_0\varsigma_M}. \quad (51)$$

The  $\varsigma_M < 1$  and  $2^{R^0} > 1$  hold. If we consider  $\varepsilon_M \geq \varepsilon_0$ ,  $\Theta \geq 0$  is obtained. Therefore, the relationship of (50) is always true. According to (47), the *event 4* holds the following constraint

$$d_m > G^{-1} \left( \frac{(1 + \varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(d_m)}} \right). \quad (52)$$

Further, we have

$$d_m > G^{-1} \left( \frac{(1 + \varepsilon_0)\eta_i + \varepsilon_0}{\rho} \right). \quad (53)$$

The joint PDF of  $d_m$  and  $d_n$  is obtained by *lemma*, the relationship and the range of  $d_m$  and  $d_n$  are calculated by above observations. Hence, the probability of *event 4* is given as

$$P_{event4} = \int_{\varphi_1}^{\varphi_2} \int_{\mu_1}^y f_{d_n, d_m}(x, y) dx dy, \quad (54)$$

where  $\varphi_1 = G^{-1}\left(\frac{(1+\varepsilon_0)\eta_i + \varepsilon_0}{\rho}\right)$ ,  $\varphi_2 = G^{-1}(\varepsilon_0/\rho)$ ,  $\mu_1 = G^{-1}\left(\frac{(1+\varepsilon_0)\eta_i}{\rho - \frac{\varepsilon_0}{G(y)}}\right)$ . The joint PDF  $f_{d_n, d_m}(x, y)$  can be expanded by binomial as

$$\begin{aligned} f_{d_n, d_m}(x, y) &= 4y(\lambda\pi)^n e^{-\lambda\pi y^2} \frac{x^{2n-1}(y^2 - x^2)^{m-n-1}}{(m-n-1)!(n-1)!} \\ &= \frac{4(\lambda\pi)^n e^{-\lambda\pi y^2}}{(m-n-1)!(n-1)!} \sum_{k=0}^{m-n-1} (-1)^k \\ &\quad \times \binom{m-n-1}{k} x^{2n+2k-1} y^{2m-2n-2k-2}. \end{aligned} \quad (55)$$

Substituting (55) to (54), the probability of *event 4* is rewritten as

$$\begin{aligned} P_4 &= \frac{4(\lambda\pi)^n e^{-\lambda\pi y^2}}{(m-n-1)!(n-1)!} \\ &\quad \times \sum_{k=0}^{t-m-1} (-1)^k \binom{t-m-1}{k} \int_{\varphi_1}^{\varphi_2} Z(y) dy. \end{aligned} \quad (56)$$

The integral of (56) is burdensome to get a closed-form solution. Hence, Chebyshev-Gauss quadrature is implemented to obtain an approximated solution. The approximated probability of *event 4* is given by

$$\begin{aligned} P_4 &= \frac{4(\lambda\pi)^n e^{-\lambda\pi y^2}}{(m-n-1)!(n-1)!} \sum_{k=0}^{m-n-1} (-1)^k \binom{m-n-1}{k} \\ &\quad \times \sum_{l=1}^N \frac{\pi(\varphi_2 - \varphi_1)}{2N} \sqrt{1 - \omega_l^2} Z\left(\frac{\varphi_2 - \varphi_1}{2}\omega_l + \frac{\varphi_2 + \varphi_1}{2}\right). \end{aligned} \quad (57)$$

where  $Z(y)$  is given as

$$Z(y) = \frac{e^{-\lambda\pi y^2} y^{2n-2m-2k-2}}{2n+2k} \left( y^{2n+2k} - \mu_1(y)^{2n+2k} \right). \quad (58)$$

$N$  is the Chebyshev-Gauss approximation parameter, which achieves a tradeoff between accuracy and complexity of the approximation. Specifically,  $N \rightarrow \infty$ , the approximation solution is the same as the precise result. Besides,  $\omega_l = \cos\left(\frac{2l-1}{2N}\right)$ . Finally, the outage probability of  $n$ -th relay for  $f_i$  ( $\forall i \in [1, \dots, M]$ ) is given as

$$P_n^i = P_m^0 + P_4. \quad (59)$$

Therefore, *Lemma 3* is proved.

## REFERENCES

[1] S. Dang, O. Amin, B. Shihada, and M. S. Alouini, "What should 6G be?" *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, 2020.

[2] X. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, "Space/Aerial-assisted computing offloading for IoT applications: a learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, 2019.

[3] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-Air-Ground integrated network: a survey," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 4, pp. 2714–2741, 2018.

[4] Q. Huang, M. Lin, J. B. Wang, T. A. Tsiftsis, and J. Wang, "Energy efficient beamforming schemes for satellite-aerial-terrestrial networks," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3863–3875, 2020.

[5] Y. Shi, J. Liu, Z. M. Fadlullah, and N. Kato, "Cross-layer data delivery in satellite-aerial-terrestrial communication," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 138–143, 2018.

[6] N. Kato, Z. Md. Fadlullah, F. Tang, B. Mao, S. Tani, A. Okamura, and J. Liu, "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 140–147, 2019.

[7] T.X. Vu, Y. Poirier, S. Chatzinotas, N. Maturo, J. Grotz, F. Roelens, "Modeling and Implementation of 5G Edge Caching over Satellite", *International Journal of Satellite Communications*, vol. 38, no. 5, pp. 395–406, 2020.

[8] Y. Wang, Y. Xu, Y. Zhang, and P. Zhang, "Hybrid satellite-aerial-terrestrial networks in emergency scenarios: a survey," *China Commun.*, vol. 14, no. 7, pp. 204–216, 2017.

[9] P. K. Sharma, D. Deepthi, and D. I. Kim, "Outage probability of 3-D mobile UAV relaying for hybrid satellite-terrestrial networks," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 418–422, 2020.

[10] A. J. Roumeliotis, C. I. Kourgiorgas, and A. D. Panagopoulos, "Optimal dynamic capacity allocation for high throughput satellite communications systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 596–599, 2019.

[11] G. Gür and S. S. Kafilolu, "Layered content delivery over satellite integrated cognitive radio networks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 390–393, 2017.

[12] T. De Cola and A. Blanco, "ICN-based protocol architectures for next-generation backhauling over satellite," in *2017 IEEE International Conference on Communications (ICC)*, Paris, 2017, pp. 1–6.

[13] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: Research issues and challenges," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 2353–2371, 2018.

[14] G. Zhong, J. Yan, and L. Kuang, "QoE-driven social aware caching placement for terrestrial-satellite networks," *China Commun.*, vol. 15, no. 10, pp. 60–72, 2018.

[15] J. Li, K. Xue, J. Liu, Y. Zhang, and Y. Fang, "An ICN/SDN-based network architecture and efficient content retrieval for future satellite-terrestrial integrated networks," *IEEE Netw.*, vol. 34, no. 1, pp. 188–195, 2020.

[16] G. Gur, "Spectrum sharing and content-centric operation for 5G hybrid satellite networks: prospects and challenges for space-terrestrial system integration," *IEEE Veh. Technol. Mag.*, vol. 14, no. 4, pp. 38–48, Dec 2019.

[17] T. de Cola, A. Ginesi, G. Giambene, G. C. Polyzos, V. A. Siris, N. Fotiou, and Y. Thomas, "Network and protocol architectures for future satellite systems," *Found. Trends® Netw.*, vol. 12, no. 1-2, pp. 1–161, 2017.

[18] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, "NOMA assisted wireless caching: strategies and performance analysis," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4854–4876, 2018.

[19] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug 2016.

[20] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, 2014.

[21] V. K. Papanikolaou, G. K. Karagiannidis, N. A. Mitsiou, and P. D. Diamantoulakis, "Closed-form analysis for NOMA with randomly deployed users in generalized fading," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1253–1257, 2020.

[22] Z. Ding, R. Schober, P. Fan, and H. Vincent Poor, "OTFS-NOMA: an efficient approach for exploiting heterogenous user mobility profiles," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7950–7965, 2019.

[23] X. Liang, J. Jiao, B. Feng, S. Wu, B. Cao, and Q. Zhang, "Performance analysis of millimeter-wave hybrid satellite-terrestrial relay networks over rain fading channel," *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, IL, USA, 2018, pp. 1–5.

- [24] X. Liang, J. Jiao, S. Wu, and Q. Zhang, "Outage analysis of multirelay millimeter-wave networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 1046–1049, 2018.
- [25] J. Du, C. Jiang, H. Zhang, X. Wang, Y. Ren, and M. Debbah, "Secure satellite-terrestrial transmission over incumbent terrestrial networks via cooperative beamforming," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1367–1382, 2018.
- [26] B. Deng, C. Jiang, H. Yao, S. Guo, and S. Zhao, "The next generation heterogeneous satellite communication networks : integration of resource management and deep reinforcement learning," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 105–111, April 2020.
- [27] M. K. Arti and V. Jain, "Relay selection-based hybrid satellite-terrestrial communication systems," *IET Commun.*, vol. 11, no. 17, pp. 2566–2574, 2017.
- [28] P. K. Upadhyay and P. K. Sharma, "Max-Max user-relay selection scheme in multiuser and multirelay hybrid satellite-Terrestrial relay systems," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 268–271, 2016.
- [29] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, and S. Allsopp, "Designing and implementing future aerial communication networks," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 26–34, 2016.
- [30] S. Zhang, W. Quan, J. Li, W. Shi, P. Yang, and X. Shen, "Air-ground integrated vehicular network slicing with content pushing and caching," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2114–2127, 2018.
- [31] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: applications, challenges, and open problems," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
- [32] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 2, pp. 1123–1152, 2016.
- [33] A. Fotouhi et al., "Survey on UAV cellular communications: practical aspects, standardization advancements, regulation, and security challenges," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3417–3442, 2019.
- [34] G. Pan, J. Ye, Y. Zhang, and M.-S. Alouini, "Performance analysis and optimization of cooperative satellite-aerial-terrestrial systems," *IEEE Trans. Wireless Commun.*, to be published, 2020.
- [35] T. X. Vu, N. Maturo, S. Vuppala, S. Chatzinotas, J. Grotz, and N. Alagha, "Efficient 5G edge caching over satellite," in *36th International Communications Satellite Systems Conference (ICSSC 2018)*, Niagara Falls, ON, Canada, 2018, pp. 1–5.
- [36] C. Brinton, E. Aryafar, S. Corda, S. Russo, R. Reinoso, and M. Chiang, "An intelligent satellite multicast and caching overlay for CDNs to improve performance in video applications," in *Proc. AIAA*, Florence, Italy, Sep. 2013, pp. 1–9.
- [37] K. An, Y. Li, X. Yan, and T. Liang, "On the performance of cache-enabled hybrid satellite-terrestrial relay networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1506–1509, 2019.
- [38] S. A. Tegos, P. D. Diamantoulakis, J. Xia, L. Fan, and G. K. Karagiannidis, "Outage performance of uplink NOMA in land mobile satellite communications," *IEEE Wireless Commun. Lett.*, to be published, 2020.
- [39] X. Zhang, K. An, B. Zhang, Z. Chen, Y. Yan, and D. Guo, "Vickrey auction-based secondary relay selection in cognitive hybrid satellite-terrestrial overlay networks with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 628–632, May 2020.
- [40] X. Zhang, B. Zhang, K. An, Z. Chen, S. Xie, H. Wang, L. Wang, and D. Guo, "Outage performance of NOMA-based cognitive hybrid satellite-terrestrial overlay networks by amplify-and-forward protocols," *IEEE Access*, vol. 7, pp. 85 372–85 381, 2019.
- [41] X. Yan, H. Xiao, C. X. Wang, and K. An, "Outage performance of NOMA-based hybrid satellite-terrestrial relay networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 538–541, 2018.
- [42] X. Yan, H. Xiao, K. An, G. Zheng, and W. Tao, "Hybrid satellite terrestrial relay networks with cooperative non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 978–981, 2018.
- [43] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Non-Orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2253–2267, 2017.
- [44] A. Kalantari, M. Fittipaldi, S. Chatzinotas, T. X. Vu, and B. Ottersten, "Cache-assisted hybrid satellite-terrestrial backhauling for 5G cellular networks," in *2017 IEEE Global Communications Conference*, Singapore, 2017, pp. 1–6.
- [45] M. Haenggi, "On distances in uniformly random networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3584–3586, 2005.
- [46] N. Okati, T. Riihonen, D. Korpi, I. Angervuori, and R. Wichman, "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.
- [47] A. Talgat, M. A. Kishk, and M.-S. Alouini, "Stochastic geometry-based analysis of LEO satellite communication systems," *IEEE Commun. Lett.*, to be published, 2020.
- [48] J. Tang, G. Chen, J. P. Coon, and D. E. Simmons, "Distance distributions for Matérn cluster processes with application to network performance analysis," *2017 IEEE International Conference on Communications (ICC)*, Paris, 2017, pp. 1–6.
- [49] X. Zhang, B. Zhang, K. An, Z. Chen, and D. Guo, "Auction-based secondary relay selection on overlay spectrum sharing in hybrid satellite-terrestrial sensor networks," *Sensors (Switzerland)*, vol. 19, no. 22, 2019.
- [50] X. Yue, Y. Liu, Y. Yao, T. Li, X. Li, R. Liu, and A. Nallanathan, "Outage Behaviors of NOMA-Based Satellite Network over Shadowed-Rician Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6818–6821, 2020.
- [51] X. Zhang, D. Guo, K. An, Z. Chen, B. Zhao, Y. Ni, and B. Zhang, "Performance analysis of NOMA-based cooperative spectrum sharing in hybrid satellite-terrestrial networks," *IEEE Access*, vol. 7, pp. 172 321–172 329, 2019.
- [52] W. Wang, Y. Tong, L. Li, A. A. Lu, L. You, and X. Gao, "Near optimal timing and frequency offset estimation for 5G integrated LEO satellite communication system," *IEEE Access*, vol. 7, pp. 113 298–113 310, 2019.
- [53] T. Z. H. Ernest, A. S. Madhukumar, R. P. Sirigina, and A. K. Krishna, "NOMA-aided UAV communications over correlated rician shadowed fading channels," *IEEE Trans. Sig. Pro.*, vol. 68, pp. 3103–3116, 2020.
- [54] X. Wang, H. Zhang, K. J. Kim, Y. Tian, and A. Nallanathan, "Performance analysis of cooperative aerial base station-assisted networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5983–5999, 2019.
- [55] K. An, M. Lin, W. P. Zhu, Y. Huang, and G. Zheng, "Outage performance of cognitive hybrid satellite-terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397–9404, 2016.
- [56] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, "Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1189–1202, 2018.
- [57] Z. Ding, R. Schober, P. Fan, and H. Vincent Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, 2019.
- [58] X. Zhang et al., "On the performance of hybrid satellite-terrestrial content delivery networks with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, to be published, 2020.