Research Article

NOMA-Based Integrated Satellite-Terrestrial Networks with Wireless Caching

Qing Ye, Faxiang Zhao, and Weiyang Xu

1School of Big Data & Software Engineering, Chongqing University, 400044 Chongqing, China
2Research and Development Center of Transport Industry of Self-Driving Technology (China Merchants Chongqing Communications Technology Research and Design Institute Co., Ltd.), 400067 Chongqing, China
3School of Microelectronics and Communications, Chongqing University, 400044 Chongqing, China

Correspondence should be addressed to Weiyang Xu; weiyangxu@cqu.edu.cn

Received 18 March 2022; Revised 4 August 2022; Accepted 2 September 2022; Published 13 September 2022

1. Introduction

Future networks are expected to provide diverse services to cope with the ever-increasing traffic demands of various services. Nevertheless, limited by the network capacity and coverage, only depending on the terrestrial communication systems cannot provide wireless access with high data rate and reliability at every place on the earth, especially in environmentally harsh areas like oceans and mountains [1]. Hence, integrated satellite-terrestrial networks (ISTNs) are deemed to be new network architectures to accommodate diverse services and applications with different quality of service (QoS) requirements [2].

Meanwhile, as the latest member of the multiple access family, nonorthogonal multiple access (NOMA) is regarded as a promising technique in the next-generation wireless communication system [3]. In this paper, we focus on the power-domain NOMA, of which the key idea is to superimpose multiple signals in the power domain at the transmitter and employ successive interference cancellation (SIC) at the receiver [4]. With NOMA scheme, more users can be served simultaneously; hence, the spectrum efficiency is improved as a result. Therefore, the application of NOMA in ISTNs has attracted great attention in academia [5–8].

In future wireless networks, not only the spectral efficiency but also the transmission delay should be paid special attention to. Stemming from that, the idea of caching is introduced. Instead of retrieving information from a central server, users can ask for the cache-enabled server that has replicated popular information [9]. As a result, the response time required to fetch a content file can be reduced. The authors in [10] analyzed and optimized the outage performance of a multiple cache-enabled amplify-and-forward relay network. Besides, caching can also help to further enhance the spectral efficiency of NOMA [11]. Cache content placement optimization in NOMA networks was
investigated in [12, 13] which studied the corresponding coding mode. Furthermore, the authors in [14] proposed a NOMA-based multicast scheme that pushing and multicasting content objects can be accomplished at once, which boosts the spectrum efficiency. On this basis, two NOMA-assisted caching strategies were developed, namely, the push-then-deliver and the push-and-deliver [15]. However, these two strategies were studied individually, and the importance to combine them in practice was not investigated.

Recently, the significance of applying caching to the ISTN has become a consensus in the academic community. The authors in [16] compared the outage probability (OP) of a relay-assisted ISTN with two representative cache placement schemes. As for the NOMA-assisted caching schemes applied in ISTNs, [17] investigated the OP of the cache-enabled relays and the hit probability of users in the NOMA-based hybrid satellite-terrestrial content delivery network. Moreover, a satellite-aerial-terrestrial network with cache-enabled aerial relays was introduced in [18], where the NOMA scheme was implemented to deliver content and push other currently most popular content to cache-enabled aerial relays simultaneously.

The combination of NOMA and ISTNs can solve the problem of multiple nodes, but in addition to the requirements of massive nodes, future mobile communications also require low latency data transmission; so, cache is introduced to reduce latency. The combination of NOMA technology and cache technology can reduce the delay and increase spectral efficiency. However, the existing literature rarely considers how to make better use of the system spectrum resources and time resources in an environment with complex channel conditions, where the server needs to push files to users and place files to cache nodes at the same time. Therefore, this paper proposes a NOMA-based ISTN with wireless caching. Specifically, the active users request the active contents and are served by the satellite directly. In contrast, proactive users request proactive contents. However, due to heavy shadowing and masking effects, it is highly probable that proactive users lack direct links to the satellite and thus need help from cache-enabled relays. By exploiting the NOMA protocol, the content file placement of relays, and the files pushed to the active users from the satellite can be accomplished simultaneously.

The main contributions are summarized as follows:

(a) The overall communication includes the file-push-and-placement (FPAP) and file-push-and-delivery (FPAD) phases. Specifically, in the FPAP phase, active and proactive files are sent to the active users and cache-enabled relays from the satellite by using the NOMA scheme. In the FPAD phase, the satellite pushes new files to active users, while the relays deliver files to the proactive users by using the NOMA technique.

(b) The performance of the proposed system is thoroughly analyzed, with emphasis on users’ closed-form OP expressions. Then, the diversity gain is derived from the asymptotic behavior of OP. In addition, the hit probability of the relay node is studied. At last, the influence of key system parameters on outage performance is investigated.

(c) A comparison between the proposed scheme and the NOMA-based ISTN without caching is carried out, where the result demonstrates that the system performance is improved by our scheme.

2. System Model

The proposed two-tier heterogeneous cache-enabled NOMA-based ISTN is depicted in Figure 1. Due to the complexity and diversity of channel conditions, users with stable and reliable direct links with satellites are divided into the first tier and defined as active users. Users suffering from severe shadow effect without a direct link are divided into the second tier and defined as proactive users. It is worth noting that the satellite can directly serve the active users, but it needs to use the relay node with good channel conditions to assist the proactive users. Due to the long backhaul link of the satellite communication network, to avoid the high delay of proactive users, the cache-enabled relay nodes are used to help the communication of proactive users. Assuming that there is good channel state quality between the satellite and the cache-enabled relay node, the relay node is divided into the first tier. In addition, content files are divided into active files and proactive files according to different requirements, i.e., the active file is defined as the file currently requested by the user, and the proactive file belongs to the file that the user does not request at present but will request in the future. Therefore, when the satellite directly serves the first-tier users, the pushed files are active; when the satellite places the content file to the cache-enabled relay node, the pushed file is the proactive.

Above all, select active users and cache-enabled relays as the first-tier nodes and proactive users as the second-tier nodes. As for the signal transmission, the satellite communicates with the nodes in the first tier directly, while with the nodes in the second tier with the help of relays. To facilitate the analysis, it is supposed that there is one satellite ($S$), one active user ($U_1$), one cache-enabled relay ($R$), and two proactive users ($U_2, U_3$) in the considered model. All nodes are equipped with a single antenna and operate in half-duplex mode. Moreover, $U_1$ and $R$ in the first tier, as well as $U_2$ and $U_3$ in the second tier, form NOMA clusters. It is worth noting that two-user NOMA-based downlink transmission has been proposed for long-term evolution (LTE) advanced [15].

As aforementioned, the overall communication consists of two phases. In the FPAP phase, by using the NOMA scheme, $S$ pushes the active files to $U_1$, and at the same time, pushes the proactive popular contents that will be acquired by $U_2$ to $R$. In the FPAD phase, $U_2$ and $U_3$ receive their corresponding files from $R$ using the NOMA protocol. Without loss of generality, we assume that a NOMA cluster consists of two users, where the user with good channel condition is denoted by the strong user, while the other one is the weak user. For ease of description, in the following analysis, we...
denote the FPAP and FPAD as the first and second phases, respectively.

2.1. Caching Model. In the light of [14], the content files are divided into two categories: active files which are requested by the active users currently and proactive files not requested by the proactive users now but will be requested soon. Correspondingly, the relay caches the proactive files in the first phase, and then it can provide service to the proactive users in the second phase. Besides, the active files are pushed to active users directly in the first phase.

The active files are collected in a finite content catalog denoted by \( F_a = \{f_{a_1}, \cdots, f_{a_N}\} \), and proactive files are collected in another one \( F_p = \{f_{p_1}, \cdots, f_{p_M}\} \), where \( N \) and \( M \) are the numbers of active and proactive files, respectively. Besides, the popularity of the requested files obeys a Zipf distribution, e.g., the popularity of \( f_{p_m} (1 \leq m \leq M) \) is given by [17]

\[
P(f_{p_m}) = \frac{m^{-\xi}}{\sum_{l=1}^{M} l^{-\xi}},
\]

where \( \xi > 0 \) denotes the shape parameter defining the content popularity skewness. The popularity of the file becomes more concentrated when \( \xi \) gets larger, while the cache utilization ratio depends on the number of files requested by users [10].

2.2. Signal Transmission

2.2.1. The First Phase (FPAP Phase). During this phase, \( S \) transmits the superimposed active signals to \( U_1 \) and \( R \). After channel propagation, the received signals at \( U_1 \) or \( R \) are

\[
y_i = \sqrt{Q_i} h_i \left( \sqrt{\omega_1 P_x x_1} + \sqrt{\omega_2 P_x x_2} \right) + n_i, i \in \{su, sr\},
\]

where \( h_i \) and \( \sqrt{Q_i} \) denotes the channel coefficient and composite fading distribution describing the links from \( S \) to \( U_1 \) or \( R \), respectively. The satellite transmit power is \( P_s \), and \( \omega_1 \) and \( \omega_2 \) are the power control coefficients with constraints \( \omega_1 + \omega_2 = 1 \) and \( \omega_1 > \omega_2 \). \( x_1 \) and \( x_2 \) are the information-bearing symbols intended for \( U_1 \) and \( R \), respectively. Without loss of generality, it is assumed that

![System model of the proposed cache-enabled NOMA-based ISTN.](image)
As for $U_1$, the signal reception consists of desired signal from satellite $S$ and interference from relay $R$; thus, its received signal is written as

$$y_{u_1}' = \sqrt{Q_{su}} h_{sr} \sqrt{\tau_1 P_x x_1} + h_{ru_1} \left( \sqrt{\tau_2 P_x x_2} + \sqrt{\tau_3 P_x x_3} \right) + n_3,$$

(8)

where $h_{ru_1}$ is the terrestrial channel link from $R$ to $U_1$, which suffers more severe fading than $h_{sr}$. $x_1'$ is the transmit signal from $S$ in the second phase, with $\mathbb{E}\{x_1^{2}\} = 1$. In addition, $\tau_1$ is the transmit power coefficient, and $n_3$ follows the same distribution as $n_1$. Hence, $U_1$ decodes $x_1'$ directly, and the corresponding SINR is expressed by

$$\lambda_{u_1,x_1'} = \frac{\tau_1 |h_{ru_1}|^2 \rho_r Q_{su}}{|h_{ru_1}|^2 \rho_r + 1}.$$  

(9)

### 2.3. Channel Model

For the satellite links, effects such as antenna gain, path loss, and link fading should be taken into account. We consider the GEO satellite; thus, the scaling parameter is given as

$$\sqrt{Q_i} = \frac{c \sqrt{G_i(\phi_i)^T T_i}}{4\pi f d_i \sqrt{K_B T B_i}}.$$  

(10)

where $c$ denotes the speed of light, $f$ is the carrier frequency, $d_i = 35786$ km is the distance between satellite $S$ and the first-tier user $i \in \{u_{1}, u_{2}\}$, $K_B = 1.38 \times 10^{-23} \text{J/K}$ is the Boltzmann constant, $T_i$ is the receiver noise temperature, and $B_i$ is the carrier bandwidth. Besides, $T_i$ denotes the antenna gain at user $i$, whereas $G_i(\phi_i)$ gives the satellite beam gain based on both the satellite beam pattern and position of user $i$ and is approximated by [19]

$$G_i(\phi_i) \approx G_{\max} \left( \frac{J_1(u_i)}{2u_i} + \frac{J_3(u_i)}{u_i^3} \right)^2.$$  

(11)

With $u_i = 2.07123(\sin \phi_i / \sin \phi_{2dB})$, $J_m(\cdot)$ represents the Bessel function of order $m$, while $G_{\max}$ denotes the maximal beam gain. $\phi_i$ is the angle between user $i$ and beam center with respect to the satellite.

In (2), $h_i$ denotes the channel fading vector, which is usually modeled by composite fading distributions to describe more accurately the amplitude fluctuation of the signal envelope. Let the satellite link $\{ |h_i|, i \in \{u_{1}, u_{2}\} \}$ obey Shadowed-Rician fading, and the probability density function (PDF) is given by [20]

$$f_{|h_i|^2}(x) = \alpha_i e^{-\beta_i x} \mathcal{F}_1(m_i; 1; \delta, x),$$  

(12)

where $\alpha_i = (2b_i m_i/(2b_i m_i + \Omega_i))^{m_i/2b_i}$, $\beta_i = 1/2b_i$, and $\delta_i = 0.5\Omega_i/b_i/(2b_i m_i + \Omega_i)$, $2b_i$ is the average power of the multi-path, $\Omega_i$ is the line-of-sight (LoS) component, $m_i(m_i > 0)$ denotes the Nakagami-$m$ fading parameter, and $\mathcal{F}_1(a; b; c)$ represents the confluent hypergeometric function [21].
According to [21], the cumulative distribution function (CDF) of the satellite link gain is given by

\[
F_{|h_s|}(u) = \sum_{k=0}^{\infty} \frac{(m_j)^k \delta_{k}^k}{(k)!} P_{t}^k e^{\frac{k+1}{\Omega_j} y(k+1, \beta_j u)},
\]  

where \((\ast)_k\) is the Pochhammer symbol defined as \((x)_k = x(x+1)(x+k-1)\).

When it comes to the terrestrial link, which follows Nakagami-\(m\) fading, the PDF of the channel gain is represented by

\[
f_{|h_s|}(x) = \left( \frac{m_j}{\Omega_j} \right)^{m_j} x^{m_j-1} \Gamma(m_j) e^{-m_j/\Omega_j x},
\]

where \(m_j\) is the integer fading severity parameter, \(\Omega_j\) is the average power of each link, and \(\Gamma(\ast)\) is the Gamma function [21]. In addition, the CDF of the terrestrial link gain can be expressed as

\[
F_{|h_s|}(x) = \frac{1}{\Gamma(m_j)} \gamma(m_j, \frac{m_j x}{\Omega_j}),
\]

where \(\gamma(a, x) = \int_0^x e^{-t} t^{a-1} dt\) denotes the lower incomplete Gamma function [21].

### 2.4. Traditional Scheme without Caching

The model highly related to the proposed is the NOMA-based coordinated direct and relay transmission (CDRT) [22], which is depicted in Figure 2. In CDRT, a base station (BS) directly communicates with user equipment 1 (UE1) while communicating with user equipment 2 (UE2) via a relay. The communication consists of two phases, i.e., in the first phase, BS transmits the superposed signal \((\alpha x_1 + \alpha x_2)\) to UE1 and the relay directly with NOMA protocol. In the second phase, the relay retransmits the decoded signal \(x_2\) (required by UE2) to UE2, while the BS transmits a new data symbol to UE1. The main challenge of nonorthogonal CDRT can be solved by using the inherent property of NOMA that allows a receiver to obtain side information such as other UE’s data for interference cancellation. Stemming from that, it is reasonable to apply NOMA-based CDRT to achieve high spectrum resource utilization.

The main difference between proposed model and the traditional CDRT lies in the second phase. Specifically, in the proposed model, relay can serve multiple users simultaneously, while CDRT serves only one. To achieve the same time and spectrum utilization as the proposed scheme, the NOMA-based CDRT without caching, i.e., \(S\) serves \(U_1\) and \(R\) in the first slot, and \(R\) serves two users at the second slot, at the same time, \(S\) serves \(U_1\) with newly signal, can be described as follows. In the first slot, \(S\) transmits the superimposed signal to \(R\) and \(U_1\). The received signal can be

\[
y_1 = \sqrt{Q} h_1 \left( \sqrt{c_1 P_x x_1} + \sqrt{c_2 P_x x_2} + \sqrt{c_3 P_x x_3} \right) + n,
\]

where the power coefficients satisfy \(c_1 + c_2 + c_3 = 1, c_1 > c_2, c_2 > c_3\), and \(c_1 > c_2 + c_3\). Thus, the SINR of \(U_1\) to decode \(x_1\) is

\[
\lambda_{u_1, x_1} = \frac{c_1 |h_{su_1}|^2}{\rho |h_{sr}|^2} + 1.
\]

Besides, the SINR of decoding \(x_1, x_2,\) and \(x_3\) at \(R\) is

\[
\lambda_{r, x_1} = \frac{c_1 |h_{sr}|^2}{\rho |h_{sr}|^2} + 1,
\]

\[
\lambda_{r, x_2} = \frac{c_2 |h_{sr}|^2}{\rho |h_{sr}|^2} + 1,
\]

\[
\lambda_{r, x_3} = \frac{c_3 |h_{sr}|^2}{\rho |h_{sr}|^2} + 1.
\]

Meanwhile, the second time slot is the same as that of the proposed scheme; thus, the description is omitted for brevity.

### 3. Performance Analysis

#### 3.1. Outage Performance

The OP is defined as the probability that the instantaneous SINR \(\lambda\) is less than a predefined threshold \(\lambda_{th}\), i.e.,

\[
P_{out}(\lambda_{th}) = P(\lambda < \lambda_{th}) = F_\lambda(\lambda_{th}),
\]

where \(\lambda_{th} = 2^{R_{th}} - 1\) with \(R_{th}\) being the required data rate, and \(F_\lambda(\ast)\) is the CDF of \(\lambda\). We emphasize that the analytical derivation is based on the following assumptions. Firstly, the CSI and SIC are crucial to the outage performance, and there have been several studies concerning this issue [23]. The CSI can be obtained with channel estimation methods, which have been adopted in most the existing works [24, 25]. As the focus herein is the caching scheme, we assume that the perfect CSI is available, and the SIC is error-free. Secondly, both the terrestrial and satellite channels are independent and identically distributed (i.i.d.).
3.1.1. OP of $U_1$. In our model, $U_1$ receives signals within two phases, and then its OP needs to be discussed separately. In the first phase, the outage occurs when the decoding of $x_1$ is failed, i.e.,

$$p^{(1)}_{u_1,\text{out}} = P(\lambda_{u_1,x_1} < \lambda_{th_1}) = F_{|h_{u_1}|}(\phi_1),$$  \hspace{1cm} (20)

where $\phi_1 = \lambda_{th_1}/(\omega_1 - \omega_2 \lambda_{th_1})$ is the required SINR of decoding $x_1$, which satisfies $\lambda_{th_1} < \omega_1/\omega_2$.

In the second phase, the OP is expressed as

$$p^{(2)}_{u_1,\text{out}} = P\left(\lambda_{u_1,x_1'} < \lambda_{th_1}\right).$$  \hspace{1cm} (21)

According to (21), we can obtain $1 - p^{(2)}_{u_1,\text{out}}$ as follows:

$$1 - p^{(2)}_{u_1,\text{out}} = \frac{1}{\lambda_{th_1} - \lambda_{th_1}} \int_{j_1}^{j_2} f_{|h_{u_1}|}(x) dx = \frac{1}{\lambda_{th_1} - \lambda_{th_1}} \int_{j_1}^{j_2} F_{|h_{u_1}|}(\frac{\lambda_{th_1} + \lambda_{th_1} - 1}{\lambda_{th_1}}) f_{|h_{u_1}|}(x) dx,$$  \hspace{1cm} (22)

where $j_1$ and $j_2$ denote the first and second integrals, respectively. Afterwards, it is easy to find that $j_1 = 1$, and then $j_2$ is expanded as

$$j_2 = J_2 = \int_{0}^{\infty} \alpha_{m_1} \sum_{k=0}^{\infty} \frac{(m_{m_1})^k}{(k!)^2 \rho_{u_1}^k} \Gamma(k + 1, \frac{\rho_{u_1} \lambda_{th_1}}{\lambda_{th_1}} (\rho_{u_1} \lambda_{th_1} + 1)) \cdot \frac{m_{m_1}}{\Omega_{m_1}} x^{m_{m_1} - 1} e^{-\frac{\rho_{u_1} \lambda_{th_1}}{\lambda_{th_1}} x} dx.$$  \hspace{1cm} (23)

According to [21], one can arrive at

$$J_2 = E_1 - E_1 \left(\frac{m_{m_1}}{\Omega_{m_1}} \right)^m e^{-E_2} \int_{0}^{\infty} e^{-E_1 x - m_{m_1} \Omega_{m_1} x} x^m dx.$$  \hspace{1cm} (24)

Then, the closed-form expression of $p^{(2)}_{u_1,\text{out}}$ is shown as

$$p^{(2)}_{u_1,\text{out}} = E_1 - E_1 \left(\frac{m_{m_1}}{\Omega_{m_1}} \right)^m e^{-E_2} \int_{0}^{\infty} e^{-E_1 x - m_{m_1} \Omega_{m_1} x} x^m dx.$$  \hspace{1cm} (25)

where

$$E_1 = \alpha_{m_1} \sum_{k=0}^{\infty} \frac{(m_{m_1})^k}{k! \rho_{u_1}^k},$$  \hspace{1cm} (26)

$$E_2 = \frac{\beta_{m_1} \lambda_{th_1} \Omega_{m_1}}{\rho_{u_1} \lambda_{th_1}}.$$  \hspace{1cm} (27)

Hence, the derivation of the OP of $U_1$ concludes.

3.1.2. OP of $U_2$. As for $U_2$, outage occurs when the detection of $x_3$ is failed, or the detection of $x_2$ fails provided the successful detection of $x_3$. Therefore, the resultant OP can be derived by

$$p_{u_2,\text{out}} = P\left(f_{p_2}\right) \times P\left(\lambda_{u_2,x_3} < \lambda_{th_2}\right) + P\left(f_{p_2}\right) \times P\left(\lambda_{u_2,x_2} < \lambda_{th_2}\right).$$  \hspace{1cm} (28)

Besides, $\lambda_{th_2}$ and $\lambda_{th_2}$ are the corresponding SINRs of decoding $x_2$ and $x_3$, which satisfies $\lambda_{th_2} < r_1/r_2$.

3.1.3. OP of $U_3$. The OP of $U_3$ can be derived straightforward as

$$p_{u_3,\text{out}} = P\left(f_{p_3}\right) \times P\left(\lambda_{u_3,x_3} < \lambda_{th_3}\right) = P\left(f_{p_3}\right) \times F_{|h_{u_3}|}(\phi_2).$$  \hspace{1cm} (29)

3.2. Diversity Order. Due to their complex forms, it is difficult to have a deep understanding of the closed-form OP expressions derived above. On the other hand, the diversity order, defined as $d_i = \lim_{\rho_i \rightarrow \infty} \log P(\rho_i)/\log \rho_i$, is another key parameter to measure the system performance. To facilitate our analysis of diversity order, the asymptotic outage behaviour of OP at high SNR regimes should be investigated. First, the approximation of CDFs of terrestrial and satellite channels is addressed. For terrestrial channels, by expressing the exponential function in (15) in terms of the Maclaurin series, the PDF is approximated as
\[ f_{|h_j|^2}(x) \approx \frac{m_j^m x^{m_j-1}}{\Gamma(m_j)} j \in \{ru_1, ru_2, ru_3\}. \quad (30) \]

Thus, the corresponding CDF of (30) is obtained straightforward, shown as

\[ F_{|h_j|^2}(x) = \frac{1}{\Gamma(m_j + 1)} \left( \frac{m_j x}{\Omega_j} \right)^{m_j}. \quad (31) \]

As for satellite links, via using the series representation in [21], one can arrive at

\[ \gamma(k + 1, \beta_i, u) = \sum_{k=0}^{\infty} \frac{(-1)^n \beta_i^k}{n! (k + 1 + n)} , i \in \{su_1, sr\}. \quad (32) \]

Thus, the approximated CDF expression for the satellite links can be obtained as follows:

\[ F_{|h_j|^2}(u) = \alpha \sum_{k=0}^{\infty} \frac{(m_j)^k}{(k!)^2} \frac{\gamma(k + 1, \beta_i, u)}{\Gamma(m_j + 1) k + 1}. \quad (33) \]

And the final approximated result is

\[ F_{|h_j|^2}(u) = \alpha u, \quad (34) \]

With (31) and (34), diversity orders of users can be derived as follows.

3.2.1. Diversity Order of \( U_1 \). According to (20) and (34), the asymptotic OP of \( U_1 \) in the first phase is

\[ \hat{P}_{u_1, out}^{(1)} = \alpha u_1 \phi_1. \quad (35) \]

By assuming \( \rho_s \rightarrow \infty \), it is derived that the diversity order of \( U_1 \) is 1.

In the second phase, assuming \( \rho_s \rightarrow \infty \) and \( \rho_s \rightarrow \infty \), then \( \lambda_{u_1, x_1} \) is approximated as \( \tau_1 \frac{|h_{u_1}|^2 Q_{u_1}}{|h_{ru_1}|^2} \). Therefore, the asymptotic OP is calculated by

\[ \int_0^{\infty} f_{|h_{u_1}|^2}(x) f_{|h_{ru_1}|}(y) dy dx = \left( \frac{m_{ru_1}}{\Omega_{ru_1}} \right)^{m_{ru_1}} \frac{\alpha_{ru_1}}{\Gamma(m_{ru_1})} \]

\[ \times \int_0^{\infty} x^{m_{ru_1}-1} e^{-m_{ru_1} x^2} dx. \]

According to [21], the OP of \( U_1 \) in the second phase approximates to

\[ \hat{P}_{u_1, out}^{(2)} = \frac{\alpha_{ru_1} Q_{ru_1}}{\Gamma(m_{ru_1})} \left( \frac{m_{ru_1}}{\Omega_{ru_1}} \right)^{-1} m_{ru_1}. \quad (37) \]

Hence, the diversity order of \( U_1 \) in the second phase is 0.

3.2.2. Diversity Order of \( U_2 \). According to (27) and (31), the asymptotic OP of \( U_2 \) is shown as

\[ \hat{P}_{u_2, out} = \frac{p(f_p)}{\Gamma(m_{ru_1} + 1)} \left( \frac{m_{ru_1}}{\Omega_{ru_1}} \right)^{m_{ru_1}} \times \left( \phi_2^m - \phi_1^m \right) \times \left( \frac{m_{ru_1}}{\Omega_{ru_1}} \right) \times \frac{1}{\Gamma(m_{ru_1} + 1)}. \quad (38) \]

After straightforward mathematical manipulations, the diversity order is obtained as \( m_{ru_1} \).

3.2.3. Diversity Order of \( U_3 \). First, the asymptotic OP of \( U_3 \) is

\[ \hat{P}_{u_3, out} = p(f_p) \times \frac{1}{\Gamma(m_{ru_1} + 1)} \times \left( \frac{m_{ru_1}}{\Omega_{ru_1}} \phi_2 \right)^{m_{ru_1}}. \quad (39) \]

At last, one can obtain that the diversity order is \( m_{ru_1} \).

3.3. Hit Probability. When a user requests a certain file cached in \( R \), a cache hit event happens. Hence, an effective criterion to evaluate the performance of content pushing is the hit probability, which indicates the probability that, during the content delivery phase, a user finds its requested file in the associated cache-enabled \( R \) [15]. The hit probability can be expressed as

\[ P_{hit} = \sum_{n=1}^{N} P_{f_p} \times \left( 1 - P_{f_{r, out}}^{(r)} \right), \quad (40) \]

where \( P_{f_{r, out}}^{(r)} \) denotes the OP of \( R \) to decode file \( f_{p_r} \). In our considered system model, the OP of \( R \) to decode \( f_{p_2} \) is derived as

\[ P_{r, out}^{(2)} = F_{|h_{u_2}|^2} \left( \frac{\lambda_{h_2}}{(\omega_1 - \omega_2 \lambda_{h_2}) \rho_2 Q_{h_2}} \right) \times \left( 1 - F_{|h_{u_1}|^2} \left( \frac{\lambda_{h_1}}{(\omega_1 - \omega_2 \lambda_{h_1}) \rho_1 Q_{h_1}} \right) \right) \times F_{|h_{u_1}|^2}(\phi_4), \quad (41) \]

where \( \phi_4 = \lambda_{h_2}/(\omega_2 \rho_2 Q_{h_2}) \). By substituting (41) into (40), the hit probability of \( R \) is acquired.

3.4. Comparison to the Scheme without Caching. For comparison, this subsection gives the outage performance of users in the traditional CDRT configuration.
3.4.1. OP of $U_1$. In the first time slot, the OP of detecting $x_1$ is expressed as

\[ P_{u1,\text{out}}^{\text{no},1} = F \left| h_{u1} \right|^2 \left( \frac{\lambda_{th1}}{(s_1 - (s_2 + c_1)\lambda_{th1}) \rho_2 Q_{R_{u1}}} \right). \] (42)

While for the second slot, outage occurs in three cases. First, when $R$ fails to decode $x_1$ or $x_2$ given the success decoding of $x_1$; thus, the SINR of detecting $x_1'$ at $U_1$ is denoted by $\lambda_{u1,x1}' = \tau_1 |h_{u1}|^2 \rho_2 Q_{R_{u1}}$. Second, detection of $x_1$ and $x_2$ is successful at $R$, but decoding $x_3$ encounters with failure, and then the SINR is $\lambda_{u1,x1}' = \tau_1 |h_{u1}|^2 \rho_2 Q_{R_{u1}}/\tau_2 |h_{u1}|^2 \rho_3 + 1$. At last, all the signals are decoded correctly in the first slot; hence, $\lambda_{u1,x3}'$ is equal to $\lambda_{u1,x3}'. To sum up, the OP can be derived as

\[ P_{u1,\text{out}}^{\text{no},2} = (P_1 + (1 - P_1) \times P_2) \times P_{x1}' + ((1 - P_1) \times (1 - P_2) \times P_3) \times P_{x1}' + ((1 - P_1) \times (1 - P_2) \times (1 - P_3)) \times P_{x1}'. \] (43)

where $P_1$, $P_2$, and $P_3$ denote the OP of decoding $x_1$, $x_2$, and $x_3$ at $R$, respectively. Besides, $P_{x1}'$, $P_{x1}'$, and $P_{x3}'$ indicate the OP of decoding $x_1'$ in three cases described above, respectively. Those outage probabilities can be calculated by

\[ P_1 = F \left| h_{u1} \right|^2 \left( \frac{\lambda_{th1}}{(s_1 - (s_2 + c_1)\lambda_{th1}) \rho_2 Q_{R_{u1}}} \right), \]
\[ P_2 = F \left| h_{u1} \right|^2 \left( \frac{\lambda_{th2}}{(s_2 - c_2 \lambda_{th1}) \rho_2 Q_{R_{u1}}} \right), \]
\[ P_3 = F \left| h_{u1} \right|^2 \left( \frac{\lambda_{th1}}{s_3 \rho_2 Q_{R_{u1}}} \right), \]
\[ P_{x1}' = F \left| h_{u1} \right|^2 \left( \frac{\lambda_{th1}}{s_2 \rho_2 Q_{R_{u1}}} \right), \]
\[ P_{x3}' = P_{u1,\text{out}}^{(2)}. \]

Moreover, by using the same approach in (25), $P_{x1}'$ is obtained as

\[ P_{x1}' = \Xi_1 - \Xi_1 \left( \frac{m_{ru1}}{m_{ru1}} \right)^{m_{ru1}} \frac{\exp(-\Xi_2 \tau_2 \rho_2 \sum_{m=0}^{\infty} (\Xi_2 \tau_2 \rho_2)^m)}{m!} \times \prod_{m=0}^{\infty} \left( \frac{1}{\rho_2 Q_{R_{u1}}} \right)^{m-n} \Xi_2 \tau_2 \rho_2 + \frac{m_{ru1}}{\Omega_{ru1}} \Gamma(n + m_{ru1}). \] (45)

With (43) and (44), $P_{u1,\text{out}}^{\text{no},2}$ is finally obtained.

### Table 1: The satellite link parameters (Q).

| Parameters                      | Value                  |
|---------------------------------|------------------------|
| Orbit ($d_j$)                  | GEO (36000 km)         |
| Carrier frequency $f$           | 2 GHz                  |
| Carrier bandwidth $B_i$         | 15 MHz                 |
| 3 dB angle of $\varphi_i$       | 0.3°                   |
| Angle between $U_1$ and satellite beam center $\varphi_{u1}$ | 0.3°                   |
| Angle between $R$ and satellite beam center $\varphi_{sr}$ | 0.8°                   |
| Antenna gain of user and relay $G_i$ | 4 dB                  |
| Maximal beam gain $G_{\text{max}}$ | 52.1 dB               |
| Receiver noise temperature $T$  | 300°C                  |

3.4.2. OP of $U_2$. As for $U_2$, the outage event occurs in situations as follows. First, the decoding of $x_1$ at $R$ fails. Second, the decoding of $x_2$ fails given the success decoding of $x_1$. Third, under the assumption of successfully detecting $x_1$ and $x_2$ at $R$, the decoding of $x_3$ at $R$ is failed. Finally, the detection at $R$ is successful, but the decoding of $x_3$ or $x_2$ fails at $U_2$. Therefore, the OP can be derived as

\[ P_{u2,\text{out}}^{\text{no}} = P_1 + (1 - P_1) \times P_2 + (1 - P_1) \times (1 - P_2) \times P_3 \times (P_5 + (1 - P_5) \times P_4), \] (46)

where $P_4$ and $P_5$ denote the OP of decoding $x_2$ and $x_3$ at $U_2$, respectively. The detailed expressions are

\[ P_4 = F \left| h_{u2} \right|^2 \left( \frac{\lambda_{th1}}{\tau_2 \rho_2} \right), \]
\[ P_5 = F \left| h_{u2} \right|^2 \left( \frac{\lambda_{th1}}{\tau_3 - \tau_2 \lambda_{th1}} \right) \rho_2 \Gamma. \] (47)

3.4.3. OP of $U_3$. As for $U_3$, an outage first occurs if the decoding of $x_1$ at $R$ fails. Then, the decoding of $x_2$ fails given the successful decoding of $x_1$. Moreover, the decoding of $x_3$ is failed when the detection of both $x_1$ and $x_2$ succeeds. At last, the detection at $R$ succeeds, but $U_3$ fails to detect $x_3$. Hence, the overall OP is

\[ P_{u3,\text{out}}^{\text{no}} = P_1 + (1 - P_1) \times P_2 + (1 - P_1) \times (1 - P_2) \times P_3 \times (1 - P_1) \times (1 - P_2) \times (1 - P_3) \times P_6, \] (48)

where $P_6$ denotes the OP of detecting $x_3$ at $U_3$, which is

\[ P_6 = F \left| h_{u3} \right|^2 \left( \frac{\lambda_{th1}}{(\tau_3 - \tau_2 \lambda_{th1}) \rho_2} \right). \] (49)
Figure 3: The OP of $U_1$ for the proposed scheme and the scheme without caching.

Figure 4: The OP of $U_2$ and $U_3$ for the proposed scheme and the scheme without caching.
4. Numerical Results

In this part, numerical simulations are performed to show the performance of the proposed cache-enabled NOMA-based ISTNs. First, the satellite link parameters \( (Q_i) \) are shown in Table 1. In addition, we consider different Shadowed-Rician fading conditions for satellite links \( (h_i) \), i.e., frequent heavy shadowing (FHS), average shadowing (AS), and infrequent light shadowing (ILS) [17]. The sets of channel coefficients \( (b_i, m_i, \Omega_i) \) are \( (0.063, 0.739, 8.97 \times 10^{-4}) \), \( (0.126, 10.1, 0.835) \), and \( (0.158, 19.4, 1.29) \) for FHS, AS, and ILS, respectively. Moreover, \( h_{su_1} \) and \( h_{sr} \) are both modeled as ILS. Besides, \( m_{ru_1} = 2, \Omega_{ru_1} = 1, m_{ru_2} = 2, \Omega_{ru_2} = 1, m_{ru_3} = 1, \Omega_{ru_3} = 1 \). Furthermore, the transmit power of \( S \) is twice that of \( R \), namely, \( P_s = 2P_r \). A fixed power allocation scheme is applied that the power coefficients are set by \( \omega_1 = 0.6, \omega_2 = 0.4, r_1 = 1, r_2 = 0.3, r_3 = 0.7, \) and \( c_1 = 0.6, c_2 = 0.3, c_3 = 0.1 \). Moreover, the content popularity parameter is \( \xi = 0.7 \), and the number of files is \( M = 10 \). At last, the transmission rate requirement for \( x_1, x_2, \) and \( x_3 \) is 0.5 (bps/Hz), 0.8 (bps/Hz), and 0.8 (bps/Hz), respectively.

Figures 3 and 4 illustrate the outage performance of users for different schemes. First of all, the good match between the analytical and numerical results validates our derivations. As for the outage performance of \( U_1 \), due to the same power control and signal transmission, the OP is the same in the first phase for both the proposed and traditional schemes. While in the second phase, it is shown that the OP for the scheme without caching has a turning point at \( SNR = 10 \text{ dB} \). The reason behind this is that the decoding at \( R \) fails more frequently with small \( SNR \) (<10 dB), causing less interference to \( U_1 \); thus, the OP will drop when \( SNR \) increases. On the other hand, when the \( SNR \) becomes large, the decoding at \( R \) encounters much less failure, and then the OP will increase because the interference at \( U_1 \) is more severe. For comparison, it is clear that the OP of proactive users \( (U_2 \) and \( U_3 \) in our scheme is lower than the scheme without caching. This is because the proactive users are affected by two factors, namely, the channel links and files’ popularity, in the proposed scheme, while they are influenced by the channel links of \( h_{ru_1}, h_{ru_2}, \) and \( h_{ru_3} \) in the scheme without caching.

Table 2: Diversity order comparison.

| User | Analysis | Proposed scheme | Simulation | Analysis | Without caching | Simulation |
|------|----------|----------------|------------|----------|----------------|------------|
| \( U_1 \), first phase | 1 | 0.9999 | 1 | 1.0264 |
| \( U_1 \), second phase | 0 | 0 | 0 | 0 |
| \( U_2 \) | \( m_{ru_1} \) | 2.0406 | \( \min \{1, m_{ru_1}\} \) | 1.1606 |
| \( U_3 \) | \( m_{ru_3} \) | 1.0116 | \( \min \{1, m_{ru_3}\} \) | 0.9935 |

Figure 5: The impact of channel parameters on the OP of \( U_1 \) in two phases.
To evaluate the diversity order, we concentrate on OP values at SNR = 25 dB and SNR = 30 dB. Specifically, the comparison of the diversity order between the two schemes is shown in Table 2. Note that the diversity order of $U_1$ shows no difference in two schemes, which is not the case when considering $U_2$ and $U_3$. In particular, if $m_{ru_2} < 1$ and $m_{ru_3} < 1$, then the diversity orders achieved are the same for two schemes. While for larger $m_{ru_2}$ and $m_{ru_3}$ values, the proposed scheme is superior to the counterpart. This observation lies in that the signal transmission of proactive users is subject to the channel quality of $h_{ru_2}$ or $h_{ru_3}$ for the proposed scheme, while it is
Figure 5 illustrates the influence of channel parameters on the OP of $U_1$. In phase one, the OP is affected by $h_{su_1}$, which is set as ILS or AS. It is clearly observed that the OP reduces when $h_{su_1}$ changes from AS to ILS, since the channel condition becomes better. In contrast, the OP of $U_1$ in the second phase is decided by both $h_{su_1}$ and $h_{ru_1}$. The channel quality concerning $h_{ru_1}$ is tuned by modifying parameter $m_{ru_1}$. It is demonstrated that the OP almost keeps unchanged when $m_{ru_1}$ varies from 1 to 2, indicating that $h_{ru_1}$ has little impact on the OP of $U_1$. Furthermore, the impacts of $h_{ru_2}$ on $U_2$ and that of $h_{ru_3}$ on $U_3$ are depicted in Figure 6, where phase two is considered. As expected, the outage event occurs more frequently if the channel condition gets worse.

Figure 7 depicts the hit probability of $R$, where different content popularity parameters are included. The number of content files $M$ is set as 10, and $\xi$ varies from 5 to 1, and then to 0.1. A large $\xi$ indicates the request for files with high popularity, whereas a small one corresponds to the requests with heavy-tailed popularity. Hence, a higher hit probability is obtained with a larger $\xi$, which is consistent with the numerical results shown in Figure 7. It is also worth noting that enhancing the SNR value is beneficial to level up the hit probability.

In the simulations above, we assume that the perfect CSI is available to all nodes. However, in practice, the channel estimation can never be error-free; thus, it is necessary to investigate the impact of uncertainty in CSI on the outage performance. To this end, Figure 8 draws the outage performance of $U_1$ in phase one, where the CSI uncertainty is modeled as a Gaussian noise independent of the actual channel coefficients. Note that the independence between the actual CSI and estimation error holds if the MMSE channel estimation method is employed. In this figure, the case of zero variance indicates the perfect CSI acquisition. From this figure, the CSI uncertainty harms the outage probability, and the larger the uncertainty is, the higher the outage probability will be.

5. Conclusion

In this paper, a NOMA-based ISTN with wireless caching is proposed to reduce the transmission delay and improve spectrum efficiency. Specifically, the overall signal transmission includes two phases. In the first phase, the active user and cache-enabled relay are served by the satellite using NOMA protocol. In the second phase, the proactive users are served by the relay, and at the same time, the active user is served by the satellite. The exact-form and asymptotic OP, diversity order, and the hit probability are derived. Both simulation and analytical results are provided to validate the superior performance of the proposed scheme over the conventional one.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.


References

[1] K. An, M. Lin, J. Ouyang, and W. Zhu, “Secure transmission in cognitive satellite terrestrial networks,” in *IEEE Journal on Selected Areas in Communications*, vol. 34no. 11, pp. 3025–3037, USA, November 2016.

[2] Z. Lin, H. Niu, K. An et al., “Refactoring RIS aided hybrid satellite-terrestrial relay networks: joint beamforming design and optimization,” in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34no. 4, pp. 3717–3724, USA, August 2022.

[3] Z. Ding, Y. Liu, J. Choi et al., “Application of non-orthogonal multiple access in LTE and 5G networks,” in *IEEE Communications Magazine*, vol. 55no. 2, pp. 185–191, USA, February 2017.

[4] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends,” in *IEEE Journal on Selected Areas in Communications*, vol. 35no. 10, pp. 2181–2195, USA, October 2017.

[5] F. Zhao, W. Xu, and W. Xiang, “Integrated satellite-terrestrial networks with coordinated C-NOMA and relay transmission,” in *IEEE Systems Journal, to be published*, pp. 1–11, USA, February 2022.

[6] Z. Lin, M. Lin, J. -B. Wang, T. de Cola, and J. Wang, “Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access,” in *IEEE Journal of Selected Topics in Signal Processing*, vol. 13no. 3, pp. 657–670, USA, June 2019.

[7] R. Liu, K. Guo, K. An, S. Zhu, and H. Shuai, “NOMA-based overlay cognitive satellite-UAV-terrestrial networks with multiple primary users,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2958864, 2022.

[8] K. An, T. Liang, G. Zheng, X. Yan, Y. Li, and S. Chatzinotas, “Performance limits of cognitive-uplink FSS and terrestrial FS for Ka-band,” in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55no. 5, pp. 2604–2611, USA, October 2019.

[9] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, “The role of caching in future communication systems and networks,” in *IEEE Journal on Selected Areas in Communications*, vol. 36no. 6, pp. 1111–1125, USA, June 2018.

[10] L. Fan, N. Zhao, X. Lei, Q. Chen, N. Yang, and G. K. Karagiannidis, “Outage probability and optimal cache placement for multiple amplify-and-forward relay networks,” in *IEEE Transactions on Vehicular Technology*, vol. 67no. 12, pp. 12373–12378, USA, December 2018.

[11] R. Liu, K. Guo, K. An, S. Zhu, and H. Shuai, “NOMA-based integrated satellite-terrestrial relay networks under spectrum sharing environment,” in *IEEE Wireless Communications Letters*, vol. 10no. 6, pp. 1266–1270, USA, June 2021.

[12] Y. Li, M. Jiang, Q. Zhang, and J. Qin, “Cache content placement optimization in non-orthogonal multiple access networks,” in *IEEE Transactions on Communications*, vol. 68no. 7, pp. 4580–4591, USA, July 2020.

[13] Y. Fu, Y. Liu, H. Wang, Z. Shi, and Y. Liu, “Mode selection between index coding and superposition coding in cache-based NOMA networks,” in *IEEE Communications Letters*, vol. 23no. 3, pp. 478–481, USA, March 2019.

[14] Z. Zhao, M. Xu, Y. Li, and M. Peng, “A Non-orthogonal multiple access-based multicast scheme in wireless content caching networks,” in *IEEE Journal on Selected Areas in Communications*, vol. 35no. 12, pp. 2723–2735, USA, December 2017.

[15] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, “NOMA assisted wireless caching: strategies and performance analysis,” in *IEEE Transactions on Communications*, vol. 66no. 10, pp. 4854–4876, USA, October 2018.

[16] K. An, Y. Li, X. Yan, and T. Liang, “On the performance of cache-enabled hybrid satellite-terrestrial relay networks,” in *IEEE Wireless Communications Letters*, vol. 8no. 5, pp. 1506–1509, USA, October 2019.

[17] X. Zhang, B. Zhang, K. An et al., “On the performance of hybrid satellite-terrestrial content delivery networks with non-orthogonal multiple access,” in *IEEE Wireless Communications Letters*, vol. 10no. 3, pp. 454–458, USA, March 2021.

[18] X. Zhang, B. Zhang, K. An et al., “NOMA-based proactive content caching in hybrid satellite-terrestrial networks,” in *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–6, USA, 2021.

[19] V. Singh and P. K. Upadhyay, “Exploiting cache-free/cachaided TWR-NOMA in cognitive hybrid satellite-terrestrial networks,” in *IEEE Transactions on Vehicular Technology*, vol. 71no. 2, pp. 1778–1793, USA, February 2022.

[20] X. Zhang, D. Guo, K. An et al., “Performance analysis of NOMA-based cooperative spectrum sharing in hybrid satellite-terrestrial networks,” in *IEEE Access*, vol. 7, pp. 172321–172329, USA, 2019.

[21] I. S. Gradsheyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th Academic, New York, NY, 2007.

[22] J. Kim and I. Lee, “Non-orthogonal multiple access in coordinated direct and relay transmission,” in *IEEE Communications Letters*, vol. 19no. 11, pp. 2037–2040, USA, November 2015.

[23] J. Zhao, X. Yue, S. Kang, and W. Tang, “Joint effects of imperfect CSI and SIC on noma based satellite-terrestrial systems,” in *IEEE Access*, vol. 9, pp. 12545–12554, USA, 2021.

[24] Q. Huang, M. Lin, W. -P. Zhu, S. Chatzinotas, and M. -S. Alouini, “Performance analysis of integrated satellite-terrestrial multi-antenna relay networks with multuser scheduling,” in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56no. 4, pp. 2718–2731, USA, August 2020.

[25] H. Shuai, K. Guo, K. An, Y. Huang, and S. Zhu, “Transmit antenna selection in noma-based integrated satellite-fap-terrestrial networks with imperfect CSI and SIC,” in *IEEE Wireless Communications Letters*, vol. 11no. 8, pp. 1565–1569, USA, August 2022.